# Optical Intensity-Modulated Direct Detection Channels: Signal Space and Lattice Codes

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

*Abstract*—Traditional approaches to constructing constellations for electrical channels cannot be applied directly to the optical intensity channel. This work presents a structured signal space model for optical intensity channels where the nonnegativity and average amplitude constraints are represented geometrically. Lattice codes satisfying channel constraints are defined and coding and shaping gain relative to a baseline are computed. An effective signal space dimension is defined to represent the precise impact of coding and shaping on bandwidth. Average optical power minimizing shaping regions are derived in some special cases. Example lattice codes are constructed and their performance on an idealized point-to-point wireless optical link is computed. Bandwidth-efficient schemes are shown to have promise for high data-rate applications, but require greater average optical power.

*Index Terms*—Free-space optical communications, lattice codes, optical intensity modulation, signal space, wireless infrared channel.

## I. INTRODUCTION

THE free-space, direct detection, optical intensity-modulated channel offers the modem designer interesting new challenges. Most practical wireless optical channels use light-emitting diodes as transmitters and photodiodes as detectors, as shown in Fig. 1. These devices modulate and detect solely the intensity of the carrier, not its phase, which implies that all transmitted signal intensities are nonnegative. Furthermore, biological safety considerations constrain the average radiated optical power, thereby constraining the average signal amplitude. Both multipath distortion in signal propagation and the limited response times of the optoelectronics create sharp constraints on the channel bandwidth. Conventional signal–space models and coded modulation techniques for electrical channels cannot be applied directly to this channel, since they do not take the signal amplitude constraints into consideration. Conventional transmission techniques optimized for broad-band optical channels (e.g., optical fibers) are not generally bandwidth efficient.

Historically, optical intensity channels have been modeled as Poisson counting channels. In the absence of background noise, the capacity of such channels is infinite [1], [2], and $M$-ary pulse-position modulation (PPM) can achieve arbitrarily small

Fig. 1. A simplified block diagram of an optical intensity direct detection communications system.

probability of error for any transmission rate [2], [3]. Under a peak optical power constraint, the capacity is finite and achieved with two-level modulation schemes [4], [5]. These schemes do not consider bandwidth efficiency; indeed, schemes based on photon counting in discrete intervals require an exponential increase in bandwidth as a function of the rate to achieve reliable communication [3]. In the more practical case of pulse-amplitude modulated (PAM) signals confined to discrete time intervals of length $T$ and with a given peak and average optical power, Shamai showed that the capacity-achieving input distribution is discrete with a finite number of levels increasing with $T$ [6]. For high-bandwidth cases ($T \to 0$) the binary level techniques found earlier are capacity achieving, however, lower bandwidth schemes require a larger number of levels.

There has been much work in the design of signal sets for use in optical intensity channels under a variety of optimality criteria; see, e.g., [7]–[11]. The most prominent modulation formats for wireless optical links are binary level PPM and on–off keying (OOK). For example, low-cost point-to-point wireless infrared infrared data association (IrDA) modems utilize 4-PPM modulation [12]. Spectrally efficient variations have been considered [13], [14], but these two-level schemes offer relatively limited improvement in bandwidth efficiency. The results of Shamai [6] show that nonbinary modulation is required to achieve capacity in bandwidth-limited channels. Subcarrier modulation [15] has been suggested as one possible multilevel scheme to achieve high bandwidth efficiency. Shiu and Kahn developed lattice codes for free-space optical intensity channels by constructing higher dimensional modulation schemes from a series of one-dimensional constituent OOK constellations [16].

In this paper, we present a signal–space model for the optical intensity channel using time-disjoint symbols, and define coding and shaping gain measures that are relevant in this framework. Unlike previous work, no assumptions are made about the underlying pulse shape, only that independent, time-disjoint symbols are employed. Previous work then appears as a special case of this work. A more accurate bandwidth measure is adopted which allows for the effect of shaping on

the bandwidth of the scheme to be represented as an effective dimension parameter. We suggest techniques for achieving coding and shaping gain using multidimensional lattices, and we compare the bandwidth efficiency of conventional PPM schemes with a new raised-quadrature amplitude modulation (QAM) scheme, and suggest that such schemes may be advantageous for high-rate short-distance optical communication.

The remainder of this paper is organized as follows. Section II gives background on the wireless optical channel and presents the communications model used in the remainder of the work. Section III presents a signal space model suited to the intensity-modulated channel and presents theorems regarding the set of signals satisfying the channel constraints. Lattice codes for the optical intensity channel are defined in Section IV and the gain is presented versus a PAM baseline. New expressions for the shaping and coding gain of these schemes are presented. Optimal shaping regions that minimize the average optical power are derived subject to certain conditions. The peak transmitted optical amplitude is taken as a design constraint and represented in the signal space. Section V presents example schemes and compares them versus a baseline. Design guidelines are given for the design of modulation schemes for this channel. Finally, in Section VI, the results of this work are summarized and general conclusions drawn.

## II. COMMUNICATIONS SYSTEM MODEL

### A. Channel Model

The *optical intensity* of a source is defined as the optical power emitted per solid angle. Wireless optical links transmit information by modulating the instantaneous optical intensity $I(t)$ in response to an input electrical current signal $x(t)$. This conversion can be modeled as $I(t) = gx(t)$, where $g$ is the optical gain of the device in units of $W/(A \cdot m^2)$. The photodiode detector is said to perform *direct-detection* of the incident optical intensity signal since it produces an output electrical photocurrent $y(t)$ proportional to the received optical intensity. The channel response from $I(t)$ to $y(t)$ in Fig. 1 is well approximated as

$$y(t) = rI(t) \otimes h(t) + n(t) \qquad (1)$$

where $\otimes$ denotes convolution and $r$ is the detector sensitivity in units of $A \cdot m^2/W$, $n(t)$ is the noise process, and $h(t)$ is the channel response [14], [18]–[20]. Substituting $I(t) = gx(t)$ into (1) gives

$$y(t) = rg \cdot x(t) \otimes h(t) + n(t)$$

where the product $rg$ is unitless. Without loss of generality, set $r = 1$, $g = 1$ to simplify analysis. In this manner, the free-space optical channel is represented by a baseband electrical model. Throughout the remainder of this paper, unless explicitly stated, all signals are electrical.

The additive noise $n(t)$ arises due to the high-intensity shot noise created as a result of ambient illumination. By the central limit theorem, this high-intensity shot noise is closely approximated as being Gaussian distributed. The noise is modeled as

additive, signal-independent, white, Gaussian with zero mean and variance $\sigma_n^2$ [14].

The channel response $h(t)$ has low-pass frequency response which introduces intersymbol interference. The low-pass channel response arises in two ways: i) front-end photodiode capacitance and ii) multipath distortion. The use of large photodiodes with high capacitances in free-space optical channels limits the bandwidth of the link. The achievable bandwidth in inexpensive systems, such as the point-to-point IrDA fast infrared (IR) standard [12] is on the order of 10–12 MHz which is approximately three orders of magnitude smaller than in wired fiber-optic systems. Multipath distortion gives rise to a linear, low-pass response which limits the bandwidth of some experimental links to approximately 10–50 MHz depending on room layout and link configuration [14], [18].

In this work, it is assumed that if a signaling scheme is "essentially band limited" to the frequency range $[-W_K, W_K]$ hertz, in the sense defined in Section IV-D, that the channel is nondistorting. Consequently, the received electrical signal $y(t)$ can be written as

$$y(t) = x(t) + n(t).$$

The physical characteristics of the optical intensity channel impose constraints on the amplitude of $I(t)$ which can equivalently be viewed as constraints on $x(t)$. Since the physical quantity modulated is a normalized power, this constrains all transmitted amplitudes to be nonnegative

$$(\forall t \in \mathbb{R}) \, x(t) \geq 0. \qquad (2)$$

The optical power transmitted is also limited by the biological impact that this radiation has on eye safety and thermal skin damage. Although limits are placed on both the average and peak optical power transmitted, in the case of most practical modulated optical sources, it is the average optical power constraint that dominates [17]. As a result, the average amplitude (i.e., average optical power)

$$P = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t) \, dt \qquad (3)$$

must be bounded. This is in marked contrast to conventional electrical channels in which the energy transmitted depends on the squared amplitude of transmitted signal.

Note that this channel model applies not only to free-space optical channels but also to fiber-optic links with negligible dispersion and signal independent, additive, white, Gaussian noise.

### B. Time-Disjoint Signaling

Let $M = \{1, 2, \ldots, M\}$ be a finite index set and let $X = \{x_m(t): m \in M\}$ be a set of optical intensity signals satisfying $(\forall m \in M) \, x_m(t) = 0$ for $t \notin [0, T)$ for some positive symbol period $T$. In the case where such time-disjoint symbols are sent independently, the optical intensity signal can be formed as

$$x(t) = \sum_{k=-\infty}^{+\infty} x_{A[k]}(t - kT) \qquad (4)$$

where $A[k]$ is an independent and identically distributed (i.i.d.) process over M. Since the symbols do not overlap in time, (2) is equivalent to

$$(\forall\, m \in \mathsf{M},\ t \in [0,\, T))\ x_m(t) \geq 0. \qquad (5)$$

The average optical power calculation in (3) can also simplified in the time-disjoint case as

$$P = \lim_{j \to \infty} \frac{1}{2(j+1)T} \sum_{k=-j}^{j} \int_0^T x_{A[k]}(t)\, dt$$

which, by the strong law of large numbers, gives

$$P = \sum_{m \in \mathsf{M}} \mathsf{Pr}(m) \left( \frac{1}{T} \int_0^T x_m(t)\, dt \right) \qquad (6)$$

with probability one, where $\mathsf{Pr}(m)$ is the probability of transmitting $x_m(t)$. Thus, the average optical power $P$ of a scheme is the expected value of the average amplitude of $x_m(t) \in \mathsf{X}$.

## III. Signal Space of Optical Intensity Signals

This section presents a signal space model which, unlike conventional models, represents the nonnegativity constraint and the average optical power cost of schemes directly. The properties of the signal space are then explored and related to the set of transmittable points and to the peak optical power of signal points.

### A. Signal Space Model

Let $\mathsf{N} = \{1, 2, \ldots, N\}$ be a finite index set, $N \leq M$, and let $\Phi = \{\phi_n(t): n \in \mathsf{N}\}$ be a set of real orthonormal functions time-limited to $t \in [0,\, T)$ such that $\mathsf{X} \subset \mathrm{span}(\Phi)$. Each $x_m(t) \in \mathsf{X}$ is represented by the vector $\boldsymbol{x}_m = (x_{m,1}, x_{m,2}, \ldots, x_{m,N})$ with respect to the basis set $\Phi$ and the signal constellation is defined as $\Omega = \{\boldsymbol{x}_m: m \in \mathsf{M}\}$.

The nonnegativity constraint in (5) implies that the average amplitude value of the signals transmitted is nonnegative. It is possible to set the function

$$\phi_1(t) = \frac{1}{\sqrt{T}}\, \mathrm{rect}(t) \qquad (7)$$

where

$$\mathrm{rect}(t) = \begin{cases} 1, & 0 \leq t < T \\ 0, & \text{otherwise} \end{cases}$$

as a basis function for every intensity modulation scheme. Note that by assigning $\phi_1(t)$, the upper bound on the dimensionality of the signal space is increased by a single dimension to be $N \leq M + 1$. Due to the orthogonality of the other basis functions

$$\int_0^T \phi_n(t)\, dt = \begin{cases} \sqrt{T}, & n = 1 \\ 0, & 1 < n \leq N. \end{cases} \qquad (8)$$

The $\phi_1(t)$ basis function contains the average amplitude of each symbol, and, as a result, represents the average optical power of each symbol. In this manner, the average optical power requirement is represented in a single dimension. The average optical power of an intensity signaling set can then be computed from (6) as

$$P(\Omega,\, T) = \frac{1}{\sqrt{T}}\, P^{\mathsf{G}}(\Omega) \qquad (9)$$

where $P^{\mathsf{G}}(\Omega)$ is defined as

$$P^{\mathsf{G}}(\Omega) = \sum_{m \in \mathsf{M}} \mathsf{Pr}(m) x_{m,1}.$$

The term $P^{\mathsf{G}}$ can be interpreted as the component of $P$ which depends solely on the constellation geometry.

Note that unlike the case of the electrical channel where the energy cost of a scheme is completely contained in the geometry of the constellation, the average optical power of an intensity signaling scheme depends on the symbol period as well. This is due to the fact that $\phi_1(t)$ in (7) is set to have unit *electrical* energy since detection is done in the electrical domain. As a result, the average amplitude and hence average optical power must depend on $T$.

### B. Admissible Region

Not all linear combinations of the elements of $\Phi$ satisfy the nonnegativity constraint (5). Define the *admissible region* $\Upsilon$ of an optical intensity-modulation scheme as the set of all points satisfying the nonnegativity criterion. In terms of the signal space

$$\Upsilon = \left\{ \boldsymbol{v} \in \mathbb{R}^N : \mathsf{Min}(\boldsymbol{v}) \geq 0 \right\} \qquad (10)$$

where for $\boldsymbol{v} = (v_1, v_2, \ldots, v_N)$, $\mathsf{Min}: \mathbb{R}^N \to \mathbb{R}$ is defined as

$$\mathsf{Min}(\boldsymbol{v}) = \min_{t \in [0,\, T)} \sum_{n \in \mathsf{N}} v_n \phi_n(t).$$

The set is closed, contains the origin, and is convex. This claim can be justified since for any $\boldsymbol{b}_1, \boldsymbol{b}_2 \in \Upsilon$ and any $\alpha \in [0,\, 1]$, $\alpha \boldsymbol{b}_1 + (1 - \alpha)\boldsymbol{b}_2 \in \Upsilon$ since it describes a nonnegative signal.

It is instructive to characterize $\Upsilon$ in terms of its cross section for a given $\phi_1$ value or, equivalently, in terms of points of equal average optical power. Define the set

$$\Upsilon_r = \{(v_1, v_2, \ldots, v_N) \in \Upsilon : v_1 = r,\, r \in \mathbb{R},\, r \geq 0\}$$

as the set of all signal points with a fixed average optical power of $r/\sqrt{T}$. Each $\Upsilon_r$ forms an equivalence class or *shell* of transmittable symbols. This is analogous to spherical shells of equal energy in the conventional case [21]. As a result, the admissible region can be written in terms of this partition as

$$\Upsilon = \bigcup_{r \geq 0} \Upsilon_r. \qquad (11)$$

The set of signals represented in each $\Upsilon_r$ can further be analyzed in the absence of their common $\phi_1$ component by defining the *projection* map $\mathsf{Proj}: \mathbb{R}^N \to \mathbb{R}^N$ that maps $(x_1, x_2, \ldots, x_N)$ to $(0, x_2, x_3, \ldots, x_N)$. The important properties of $\Upsilon$ are summarized in Theorem 1.

*Theorem 1:* Let $\Upsilon$ denote the admissible region of points defined in (10).

1) For $u, v > 0$, $\Upsilon_u = (u/v)\Upsilon_v$.
2) $\Upsilon = \bigcup_{r \geq 0}(r\Upsilon_1)$.
3) $\Upsilon_1$ is closed, convex and bounded.

4) Let $\partial\Upsilon_1$ denote the set of boundary points of $\Upsilon_1$. Then
   i) $\mathbf{0} \notin \mathsf{Proj}(\partial\Upsilon_1)$ and
   ii) $\partial\Upsilon_1 = \{\boldsymbol{v} \in \Upsilon_1 \colon \mathsf{Min}(\boldsymbol{v}) = 0\}$.

5) $\Upsilon$ is the convex hull of a generalized $N$-cone with vertex at the origin, opening about the $\phi_1$-axis and limited to $\phi_1 \geq 0$.

   *Proof:*

*Property 1.* The set $(v/u)\Upsilon_u$ is a set of pulses with average optical power $v/\sqrt{T}$. Therefore, $(v/u)\Upsilon_u \subseteq \Upsilon_v$. Similarly, $(u/v)\Upsilon_v \subseteq \Upsilon_u$ which implies that $\Upsilon_v \subseteq (v/u)\Upsilon_u$.

*Property 2.* Follows directly from (11) and Property 1.

*Property 3.* The set $\Upsilon_1$ is closed by definition of $\Upsilon$ in (10). The convexity of $\Upsilon_1$ arises since for any $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \Upsilon_1$ and any $\alpha \in [0, 1]$ the average optical amplitude value of the signal represented by $\boldsymbol{x} = \alpha\boldsymbol{v}_1 + (1 - \alpha)\boldsymbol{v}_2$ is $1/\sqrt{T}$. Hence, $\boldsymbol{x} \in \Upsilon_1$ implies $\Upsilon_1$ is convex.

Recall that a set in $\mathbb{R}^N$ is said to be *bounded* if it is contained in an $N$-ball of finite radius. The region $\mathsf{Proj}(\Upsilon_1)$ is

$$\mathsf{Proj}(\Upsilon_1) = \left\{\boldsymbol{v} = (v_1, v_2, \ldots, v_N) \in \mathbb{R}^N \colon \right.$$
$$\left. v_1 = 0, \ \mathsf{Min}(\boldsymbol{v}) \geq -\frac{1}{\sqrt{T}}\right\}.$$

Every $\boldsymbol{v} \in \mathsf{Proj}(\Upsilon_1)$ represents a signal with zero average amplitude in $[0, T)$ and $\mathsf{Min}(\boldsymbol{v}) \in [-1/\sqrt{T}, 0]$. Furthermore, $\mathsf{Proj}(\Upsilon_1)$ is closed and contains the origin. Take some point $\boldsymbol{v} \in \mathsf{Proj}(\Upsilon_1)$ such that $\|\boldsymbol{v}\| = q$ for some $q > 0$. If no such point exists, then $\mathsf{Proj}(\Upsilon_1)$ is contained in a ball of radius $q$ since $\Upsilon_1$ convex. Otherwise, $k\boldsymbol{v} \in \mathsf{Proj}(\Upsilon_1)$ for $k \in [0, |1/(\sqrt{T}\mathsf{Min}(\boldsymbol{v}))|]$ is contained in an $N$-ball of radius greater than $|q/(\sqrt{T}\mathsf{Min}(\boldsymbol{v}))|$. The union of all such $N$-balls for all $\boldsymbol{v} \in \mathsf{Proj}(\Upsilon_1)$ contains $\mathsf{Proj}(\Upsilon_1)$ implying that $\mathsf{Proj}(\Upsilon_1)$ is bounded. Since the $\phi_1$ coordinate of all points in $\Upsilon_1$ is the same, $\mathsf{Proj}(\Upsilon_1)$ bounded implies $\Upsilon_1$ is bounded.

*Property 4.* Since $\Upsilon_1$ is closed, let $\partial\Upsilon_1$ be the set of boundary points of $\Upsilon_1$. Note that an $(N - 1)$-ball of some radius $\epsilon > 0$ of points in $\mathsf{Proj}(\Upsilon_1)$ exists about the origin. If this were not the case, it would imply that a signal point $\boldsymbol{x} \in \mathsf{Proj}(\Upsilon_1)$, $\boldsymbol{x} \neq \mathbf{0}$, were either nonnegative or nonpositive, which is impossible due to the construction of the signal space in (8). Therefore, $\mathbf{0} \notin \mathsf{Proj}(\partial\Upsilon_1)$.

From Property 3, for $\boldsymbol{v} \in \mathsf{Proj}(\Upsilon_1)$, $\boldsymbol{v} \neq \mathbf{0}$, $k\boldsymbol{v} \in \mathsf{Proj}(\Upsilon_1)$ for $k \in [0, |1/(\sqrt{T}\mathsf{Min}(\boldsymbol{v}))|]$. The boundary points of the set arise when $k$ is maximized. The set $\mathsf{Proj}(\partial\Upsilon_1)$ is then the set of these extremal points with minimum amplitude equal to $-1/\sqrt{T}$. Applying the inverse map, $\partial\Upsilon_1$ is then the set of points in $\Upsilon_1$ with minimum amplitude equal to zero.

*Property 5.* Recall that a generalized $N$-cone is a surface in $\mathbb{R}^N$ that can be parameterized as $\boldsymbol{C}(u, v) = \boldsymbol{c} + v\boldsymbol{C}'(u)$, where $\boldsymbol{c}$ is a fixed vector called the vertex of the cone and $\boldsymbol{C}'(u)$ is a curve in $\mathbb{R}^N$ [22]. Using Properties 1 and 2, $\partial\Upsilon$ can be parameterized in terms of the $\phi_1$ coordinate value as $\partial\Upsilon_r = r\partial\Upsilon_1$, for $r \geq 0$. Thus, $\partial\Upsilon$ is a generalized cone with vertex at the origin opening about the $\phi_1$-axis. Since $\Upsilon_r$ is the convex hull of $\partial\Upsilon_r$ (by trivial extension of [23, Theorem 8.1.3]) and using Property 2, $\Upsilon$ is then the convex hull of a generalized cone limited to the half-space $\phi_1 \geq 0$.   □



Fig. 2. Three-dimensional (3-D) admissible region $\Upsilon$ for the raised-QAM example in Section V.
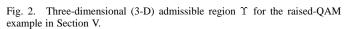
Fig. 2 illustrates a portion of the three-dimensional (3-D) admissible region for the raised-QAM basis functions described in Section V. Notice that the region forms a 3-D cone with circular cross sections in $\phi_1$, satisfying the properties of Theorem 1.

### C. Peak Optical Power Bounding Region

Although the communications model of Section II does not constrain the peak amplitude of $x(t)$, in any practical system device limitations limit the peak optical power transmitted.

The *peak optical power bounding region* $\Pi(p)$ is defined as the set of points in the signal space which correspond to signals which have amplitudes bounded from above by $p/\sqrt{T}$. Formally, for some $p \geq 0$

$$\Pi(p) = \left\{\boldsymbol{\pi} \in \mathbb{R}^N \colon \mathsf{Max}(\boldsymbol{\pi}) \leq \frac{p}{\sqrt{T}}\right\} \qquad (12)$$

where, for $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_N)$, $\mathsf{Max} \colon \mathbb{R}^N \to \mathbb{R}$ is defined as
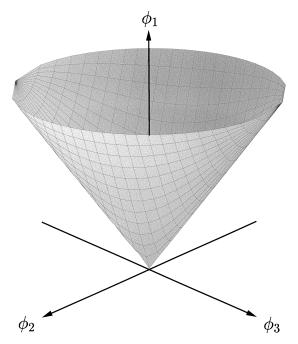
$$\mathsf{Max}(\boldsymbol{\pi}) = \max_{t \in [0, T)} \sum_{n \in \mathsf{N}} \pi_n \phi_n(t).$$

Not surprisingly, $\Upsilon$ and $\Pi(p)$ are closely related and their explicit relation is illustrated in the following.

*Theorem 2:* The peak optical power bounding region $\Pi(p)$ can be written as $\Pi(p) = -\Upsilon + p\boldsymbol{\phi}_1$, where $\boldsymbol{\phi}_1$ is a unit vector in the $\phi_1$ direction.

   *Proof:* According to the definition in (10), $\Upsilon$ contains the set of all transmittable points, that is, the set of signals with nonnegative amplitudes. The set $-\Upsilon$ is the set of signals for which the maximum possible amplitude is zero. Since $\phi_1(t)$ is constant in a symbol period, the addition of $p\phi_1(t)$, $p \geq 0$, to each signal in $-\Upsilon$ yields the set of signals with maximum at most $p/\sqrt{T}$. The region $\Pi(p)$ is then given as

$$\Pi(p) = -\Upsilon + p\boldsymbol{\phi}_1. \qquad \qquad □$$

Since $\Pi(p)$ differs from $\Upsilon$ by an affine transform, by Property 5 of Theorem 1, $\Pi(p)$ is the convex hull of an $N$-dimensional generalized cone with vertex at $(p, 0, 0, \ldots, 0)$ and opening about the negative $\phi_1$-axis.

### D. Peak Optical Power per Symbol

The region $\Pi(p)$ is a set of points which satisfy the peak constraint, however, it does not reveal the maximum amplitudes of the signals within. Indeed, it would be useful to have some knowledge of which points have high optical peak values in the construction of modulation schemes. This section demonstrates how the peak value of every signal in the constellation can be determined from the geometry of $\Upsilon$.

As justified in Property 2 of Theorem 1, the region $\Upsilon$ can be completely characterized by looking at a single cross section. Therefore, determining the peak values of signals represented in $\Upsilon_1$ will give the peak values of all points in $\Upsilon_r$ through scaling. Additionally, the peak of only those points in $\partial\Upsilon_1$ need be considered since, as shown in the proof of Property 4 ii) of Theorem 1, $\partial\Upsilon_1$ arises by maximizing $k > 0$ so that $k\boldsymbol{v} \in \Upsilon_1$ for some $\boldsymbol{v} \in \Upsilon_1$. This maximization over $k$ implies that $\partial\Upsilon_1$ represents signal points of maximum amplitude in $\Upsilon_1$.

Take some $\boldsymbol{x} \in \mathsf{Proj}(\partial\Upsilon_1)$. As shown in Property 4 i) of Theorem 1, $\mathsf{Min}(\boldsymbol{x}) = -1/\sqrt{T}$. Note that the signal $-\boldsymbol{x}$ has $\mathsf{Min}(-\boldsymbol{x}) = \mathsf{Max}(\boldsymbol{x})$ and $\mathsf{Max}(-\boldsymbol{x}) = 1/\sqrt{T}$. Form the vector $\hat{\boldsymbol{x}} \in \mathsf{Proj}(\partial\Upsilon_1)$ such that

$$\hat{\boldsymbol{x}} = -k\boldsymbol{x} \tag{13}$$

for some unique $k > 0$. Now, since $\hat{\boldsymbol{x}} \in \mathsf{Proj}(\partial\Upsilon_1)$ implies that $\mathsf{Min}(\hat{\boldsymbol{x}}) = -1/\sqrt{T}$, however, $\mathsf{Min}(\hat{\boldsymbol{x}}) = -k\mathsf{Max}(\boldsymbol{x})$. Therefore,

$$\mathsf{Max}(\boldsymbol{x}) = \frac{k^{-1}}{\sqrt{T}}. \tag{14}$$

Using (13) the preceding formula can be simplified to

$$\mathsf{Max}(\boldsymbol{x}) = \frac{\|\boldsymbol{x}\|}{\|\hat{\boldsymbol{x}}\|} \cdot \frac{1}{\sqrt{T}} \tag{15}$$

which exists since $\|\hat{\boldsymbol{x}}\| > 0$ by Property 4 i) of Theorem 1. Note that (14) implies that $\mathsf{Max}(\hat{\boldsymbol{x}}) = k/\sqrt{T}$. Fig. 3 graphically illustrates the scenario in (15) for the 3-D PAM basis considered in Section V. In the figure it is possible to deduce that $\mathsf{Max}(\boldsymbol{x}_1) > \mathsf{Max}(\boldsymbol{x}_2)$ and that $\mathsf{Max}(\boldsymbol{x}_2) = \mathsf{Max}(\hat{\boldsymbol{x}}_2) = 1/\sqrt{T}$ by observing the relative magnitudes of the illustrated vectors. Since the region $\mathsf{Proj}(\Upsilon_1)$ is convex, the peak optical power values for all signals in $\mathsf{Proj}(\Upsilon_1)$ differ from those represented in (15) by a scaling factor in the interval $[0, 1]$.

Finally, the peak optical power of the constellation points in $\boldsymbol{v} \in \Upsilon_1$ can be represented as

$$\mathsf{Max}(\boldsymbol{v}) = \mathsf{Max}(\mathsf{Proj}(\boldsymbol{v})) + \frac{1}{\sqrt{T}}. \tag{16}$$

The *peak-to-average* optical power ratio (PAR), can be computed for all $\boldsymbol{v} \in \Upsilon_1$ using (15) and (16) is

$$\mathsf{PAR}(\Upsilon_1) = \max_{\boldsymbol{x} \in \mathsf{Proj}(\partial\Upsilon_1)} \frac{\|\boldsymbol{x}\|}{\|\hat{\boldsymbol{x}}\|} + 1 \tag{17}$$
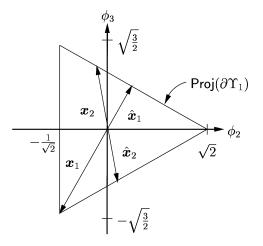


Fig. 3. Determination of peak amplitude value of elements in $\mathsf{Proj}(\partial\Upsilon_1)$ for 3-D PAM bases.

with $\hat{\boldsymbol{x}}$ as defined in (13). Again, maximization is done over points in $\mathsf{Proj}(\partial\Upsilon_1)$ since the signal with the largest peak value must be contained in this set.

The peak optical power of all other transmittable points in $\Upsilon$ can then be found by scaling the peak values found for $\boldsymbol{v} \in \Upsilon_1$. The PAR of points in $\Upsilon_r$ will be the same as (17) since both the average and peak optical power scale by a factor of $r$.

## IV. OPTICAL INTENSITY LATTICE CODES

The use of lattice codes over bandwidth-limited electrical channels has been explored extensively in the literature [21], [24]–[30]. Typically, optical channels are considered as being power limited rather than bandwidth limited. The case of signaling over bandwidth-limited optical channels, such as some wireless optical links, has not received much attention.

Early work in the development of signaling schemes for the optical intensity channel noted that unlike the conventional channel, where the electrical energy of the signals determines performance, the shape of the pulses used in transmission as well as the electrical energy determine the performance of an optical intensity scheme [8].

This section employs the signal space model of Section III to define lattice codes suited to the optical intensity channel. Figures of merit are defined and optical power gain with respect to a baseline are computed. Optimal shaping regions in the sense of minimum average optical power are derived and optical power gain is derived. The impact of a peak optical power constraint is investigated and gains computed.

### A. Definition of Lattice Codes

Lattice codes satisfying the constraints of the optical intensity channel can be defined for a given $\Phi$ using the regions defined in Section III-A. The *shaping region* $\Psi$ is defined as a closed set so that $\Upsilon \cap \Psi$ is bounded.

A finite $N$-dimensional lattice constellation is formed through the intersection of an $N$-dimensional lattice translate $\Lambda + \boldsymbol{t}$ and the bounded region $\Upsilon \cap \Psi$ to give

$$\Omega(\Lambda, \Upsilon, \Psi) = (\Lambda + \boldsymbol{t}) \cap \Upsilon \cap \Psi. \tag{18}$$

It is assumed in this work that the modulator selects symbols $\boldsymbol{x} \in \Omega$ independently and equiprobably from symbol period to symbol period. A modulation scheme is then described by the pair $(\Omega, \Phi)$, which defines the set $\mathsf{X}$ of signals transmitted.

In what follows, the performance of lattice codes so defined is related to the properties of $\Lambda$, $\Upsilon$ and $\Psi$.

### B. Constellation Figure of Merit, Gain

In conventional channels, the *constellation figure of merit* (CFM) is a popular measure of the energy efficiency of a signaling scheme [26]. An analogous measure for optical intensity channels which quantifies the *optical* power efficiency of the scheme is [11], [16]

$$\text{CFM}(\Omega, T) = \frac{d_{\min}(\Omega)}{P(\Omega, T)} = \sqrt{T} \cdot \frac{d_{\min}(\Omega)}{P^{\mathsf{G}}(\Omega)} \qquad (19)$$

where $d_{\min}(\Omega)$ is the minimum Euclidean distance between constellation points and $P(\Omega, T)$ is average optical power (9). The CFM in (19) is invariant to scaling of the constellation as in the case of the CFM for electrical channels [26]. The optical CFM is unaffected by $L$-fold Cartesian product of $\Omega$ so long as the symbol period also increases $L$-fold, that is, $\text{CFM}(\Omega^L, LT) = \text{CFM}(\Omega, T)$. In conventional channels the CFM is invariant under orthogonal transformations of $\Omega$ [31], whereas $\text{CFM}(\Omega, T)$ is invariant under a subset of orthogonal transformations which leave the $\phi_1$ coordinate unaffected. Additionally, the CFM in (19) is not unitless, since the average symbol amplitude depends on $T$ via (9) while $d_{\min}$ is independent of the symbol interval.

The optical power gain of one scheme versus another can be computed via the CFM. The probability of a symbol error can be approximated for a given $\Omega$ by the relation

$$P_e(P(\Omega, T)/\sigma_n) \approx \overline{N}(\Omega) \cdot Q\left(\frac{\text{CFM}(\Omega, T) \cdot P(\Omega, T)}{2\sigma_n}\right)$$

where $\overline{N}(\Omega)$ is the error coefficient related to the number of nearest neighbors to each constellation point and

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

Using the same analysis as in the case of electrical channels, the asymptotic optical power gain of $(\Omega, \Phi)$ over $(\Omega_\oplus, \Phi_\oplus)$, in the limit as $P_e(P(\Omega, T)/\sigma_n) \to 0$, can be shown to be

$$G(\Omega, T, T_\oplus) = \text{CFM}(\Omega, T)/\text{CFM}(\Omega_\oplus, T_\oplus)$$

which is independent of the error coefficients and noise variance [21].

### C. Baseline Constellation

A rectangular pulse, $M$-PAM constellation, $\Omega_\oplus$ is taken as the baseline. This constellation can be formed as in (18) where $\Lambda = d_{\min}\mathbb{Z}$, $\boldsymbol{t} = \boldsymbol{0}$, $\Upsilon = [0, \infty)$, and $\Psi = [0, (M-1)d_{\min}]$. The resulting modulation scheme can be written as $(\Omega_\oplus, \Phi_\oplus)$ where the single basis function of $\Phi_\oplus$ has symbol period $T_\oplus$. Since all constellation points are chosen equiprobably the baseline CFM is

$$\text{CFM}(\Omega_\oplus, T_\oplus) = \frac{2}{|\Omega_\oplus| - 1} \sqrt{T_\oplus}.$$

In $L$ dimensions, the constellation $\Omega_\oplus^L$ is formed through the $L$-fold Cartesian product of $\Omega_\oplus$ with itself and can be realized by transmitting a series of $L$ symbols of $\Omega_\oplus$ each of fixed symbol period $T_\oplus$. In this case, since

$$\text{CFM}(\Omega_\oplus^L, LT_\oplus) = \text{CFM}(\Omega_\oplus, T_\oplus)$$

and there is no asymptotic optical power gain.

The asymptotic optical power gain of $(\Omega, \Phi)$ over this baseline can then be computed as

$$\begin{aligned} G(\Omega, T, T_\oplus) &= \frac{\text{CFM}(\Omega, T)}{\text{CFM}(\Omega_\oplus, T_\oplus)} \\ &= \frac{|\Omega_\oplus| - 1}{2} \sqrt{\frac{T}{T_\oplus}} \, \frac{d_{\min}(\Omega)}{P^{\mathsf{G}}(\Omega)}. \end{aligned} \qquad (20)$$

### D. Spectral Considerations

In order to have a fair comparison between two signaling schemes, the spectral properties of each must be taken into account. Two schemes are compared on the basis of having equal *bandwidth efficiencies* $\eta = R/W$, where $R = \log_2(M)/T$ is the bit rate of the data source in bits per second and $W$ is a measure of the bandwidth support required by the scheme.

In systems employing time-disjoint symbols, the definition of the bandwidth of $(\Omega, \Phi)$ is nontrivial. In previous work, the first spectral null of the power spectral density was used as a bandwidth measure [13], [14], [16]. In this work, the bandwidth of schemes is measured using the *100 K%-fractional power bandwidth* $W_K$, defined as

$$\frac{\int_{-W_K}^{W_K} S_{\mathsf{X}}^c(f)\, df}{\int_{-\infty}^{+\infty} S_{\mathsf{X}}^c(f)\, df} = K \qquad (21)$$

where $S_{\mathsf{X}}^c(f)$ is the continuous portion of the power spectral density of $(\Omega, \Phi)$ and $K \in (0, 1)$ is fixed to some value, typically $0.99$ or $0.999$. This is a superior measure of signal bandwidth since it is defined as the extent of frequencies where the majority of the signal power is contained as opposed to arbitrarily denoting the position of a particular spectral feature as the bandwidth.

The channel model of Section II-A assumes that the frequency response of the channel is flat and that signals are limited to a bandwidth of $[-W_K, W_K]$. To a first approximation, if $K$ is chosen large enough, the energy outside of this band lies below the noise floor of the channel and neglecting it introduces little error. In this sense, $(\Omega, \Phi)$ is considered as being "essentially" band limited to the channel bandwidth. Subsection B of the Appendix contains the expressions for $S_{\mathsf{X}}^c(f)$ under the conditions of independent and equiprobable signaling and further justifies the use of this bandwidth measure.

As in the conventional case, to compare $(\Omega, \Phi)$ versus $(\Omega_\oplus, \Phi_\oplus)$ their bandwidth efficiencies must be equal. However, in the case of optical intensity schemes, the average optical power depends directly on the choice of $T$, as shown in (9). Changes in the symbol period will leave the constellation geometry unaffected because the basis functions are scaled to have unit electrical energy independent of $T$. Therefore, for the optical intensity-modulation scheme $(\Omega, \Phi)$, the geometry of the constellation does *not* completely represent the average optical power of the scheme. Thus, to make a fair comparison it

is necessary to fix both the bandwidth and the bit rate of $(\Omega, \Phi)$ and $(\Omega_\oplus, \Phi_\oplus)$. Equating the rate of two schemes gives

$$\log_2 |\Omega_\oplus| = \frac{T_\oplus}{T} \log_2 |\Omega|. \qquad (22)$$

Define $\kappa = 2W_K T$ for $(\Omega, \Phi)$ and define $\kappa_\oplus = 2W_{K\oplus} T_\oplus$, where $W_{K\oplus}$ is the fractional power bandwidth of the baseline scheme. Writing $W_K$ and $W_{K\oplus}$ in terms of $\kappa$ and $\kappa_\oplus$ and equating them defines

$$\nu = \frac{\kappa}{\kappa_\oplus} = \frac{T}{T_\oplus}. \qquad (23)$$

Combining the results of (22) and (23) gives

$$\log_2 |\Omega_\oplus| = \frac{1}{\nu} \log_2 |\Omega|. \qquad (24)$$

The term $\kappa$ can be viewed as the "essential" dimension of the set of signals time-limited in the range $[0, T)$ with fractional power bandwidth $W_K$. This definition is analogous to the use of the orthonormal family of prolate spheroidal wave functions as a basis for essentially band-limited functions [33]–[35]. These functions have the maximum energy in bandwidth $[-W, W]$ of all time-limited, unit energy functions in $[0, T)$ [33], [34]. The set of functions time-limited to $[0, T)$ and with $(1 - \epsilon)\%$-fractional energy bandwidth $W$ are approximately spanned by $2WT$ prolate spheroidal wave functions with an error which tends to zero as $\epsilon \to 0$.

Interpreting $\kappa$ the effective dimension of $\Omega$, the parameter $\nu$ in (23) can then be thought of as the effective number of dimensions of $(\Omega, \Phi)$ with respect to the baseline $(\Omega_\oplus, \Phi_\oplus)$. Equation (24) can then be interpreted as the *effective normalized rate* in units of bits per effective baseline dimension. This is analogous to the conventional expression of normalized bit rate in [26]. Since, in general, constellation shaping has an impact on the power spectrum of a scheme, $\nu$ must be determined for each choice of $\Omega$ and $\Phi$.

### E. Gain Versus a Baseline Constellation

The optical power gain in (20) can be simplified by substituting the effective dimension $\nu$ (23) and the effective normalized rate (24) to yield

$$G(\Omega, \nu) = \frac{|\Omega|^{1/\nu}}{2} \sqrt{\nu} \frac{d_{\min}(\Omega)}{P^G(\Omega)} \left(1 - |\Omega|^{-1/\nu}\right). \qquad (25)$$

Note that by specifying the spectral constraints as in Section IV-D, the gain is independent of the value of $T$ and $T_\oplus$, as in the conventional case. However, in this case, the gain depends on the effective dimension of the signal spaces through $\nu$ as opposed to the dimension of the Euclidean space $N$ as in the conventional case.

For large constellations, or more precisely, for a large effective normalized rates, the term $(1 - |\Omega|^{-1/\nu}) \approx 1$ and can be neglected.

### F. Continuous Approximation to Optical Power Gain

The *continuous approximation* [26] allows for the replacement of a discrete sum of a function evaluated at every $\boldsymbol{x} \in \Omega$ by a normalized integral of the function over the region $\Upsilon \cap \Psi$.

Specifically, for a function $f \colon \mathbb{R}^N \to \mathbb{R}$ which is Riemann-integrable over $\Upsilon \cap \Psi$

$$\sum_{\boldsymbol{x} \in \Omega} f(\boldsymbol{x}) \approx \frac{1}{V(\Lambda)} \int_{\Upsilon \cap \Psi} f(\boldsymbol{x}) \, dV(\boldsymbol{x}) \qquad (26)$$

where $V(\Lambda)$ is the fundamental volume of the lattice $\Lambda$ [21]. This approximation is good when $V(\Upsilon \cap \Psi) \gg V(\Lambda)$, where the notation $V(\cdot)$ evaluates to the volume of the region. In practical terms, this condition occurs when the scheme in question is operating at a high effective normalized rate.

Applying the continuous approximation, $P^G(\Omega)$ (9) is a function of the region $\Upsilon \cap \Psi$ and takes the form

$$P^G(\Upsilon \cap \Psi) \approx \int_{\boldsymbol{x} \in \Upsilon \cap \Psi} x_1 \frac{1}{V(\Upsilon \cap \Psi)} \, dV(\boldsymbol{x}). \qquad (27)$$

Similarly, $|\Omega|$ can be approximated as

$$|\Omega| \approx \frac{V(\Upsilon \cap \Psi)}{V(\Lambda)}.$$

Since the conditions of the continuous approximation assume that $|\Omega|$ is large, in (25) the term $(1 - |\Omega|^{-1/\nu}) \approx 1$.

Substituting these approximations, along with the fact that $d_{\min}(\Omega) = d_{\min}(\Lambda)$, into (25) yields

$$G(\Omega, \nu) \approx \gamma_c(\Lambda, \nu) \gamma_s(\Upsilon, \Psi, \nu) \qquad (28)$$

where the *coding gain* is given as

$$\gamma_c(\Lambda, \nu) = \frac{d_{\min}(\Lambda)}{V(\Lambda)^{1/\nu}} \qquad (29)$$

and the *shaping gain* is

$$\gamma_s(\Upsilon, \Psi, \nu) = \frac{\sqrt{\nu}}{2} \frac{V(\Upsilon \cap \Psi)^{1/\nu}}{P^G(\Upsilon \cap \Psi)}. \qquad (30)$$

### G. Coding Gain

In the electrical case, the coding gain is a normalized lattice-packing density known as *Hermite's parameter* $\gamma_n(\Lambda)$ [36], and is given as

$$\gamma_n(\Lambda) = \frac{d_{\min}^2}{V(\Lambda)^{2/N}}.$$

This electrical coding gain is a purely geometric property of the lattice selected. The coding gain for the optical intensity channel (29) can be written in terms of $\gamma_n(\Lambda)$ as

$$\gamma_c(\Lambda, \nu) = d_{\min} \cdot \left(\sqrt{\gamma_n(\Lambda)}/d_{\min}\right)^{N/\nu}. \qquad (31)$$

Through the effective dimension $\nu$, the optical coding gain depends on $\Phi$, $\Upsilon \cap \Psi$, and $\Lambda$. Thus, the densest lattice in $N$ dimensions, as measured by $\gamma_n(\Lambda)$, may not maximize the optical coding gain in (29).

In the case of transmission at high effective normalized rates, however, Subsection C of the Appendix demonstrates that the continuous approximation can be used to yield an estimate of the effective dimension $\nu^C$, independent of $\Lambda$. For a given $\Upsilon \cap \Psi$ and $\Phi$, substituting $\nu = \nu^C$ in (31) leaves $\gamma_n(\Lambda)$ as the only term dependent on the lattice chosen. As a result, the densest lattice in $N$ dimensions which maximizes $\gamma_n(\Lambda)$ also maximizes the optical coding gain $\gamma_c(\Lambda, \nu)$ at high effective normalized rates.

## H. Optimal Shaping Regions

Shaping is done to reduce the average optical power requirement of a scheme at a given rate. In the conventional case, the shaping gain depends solely on the geometry of the constellation. As is the case with $\gamma_c(\Lambda, \nu)$, the shaping gain for the optical intensity channel (30) is a function of the region $\Upsilon \cap \Psi$ as well as the effective dimension of the scheme. In general, the selection of $\Psi$ which maximizes $\gamma_s(\Upsilon, \Psi, \nu)$ is difficult to find since it depends on the specific basis functions selected.

In certain cases, $\nu$ can be assumed to be independent of $\Upsilon \cap \Psi$. If the $x(t) \in \mathsf{X}$ have $K\%$-fractional energy bandwidths which are all approximately the same and $K \to 1$, then every symbol occupies essentially the same bandwidth. In this case, $W_K$ and hence $\nu$, will be independent of $\Upsilon \cap \Psi$. Section V illustrates some practical situations when this approximation holds.

Under the assumption that $\nu$ is independent of $\Upsilon \cap \Psi$, the rate of a scheme is then dependent on the volume $V(\Upsilon \cap \Psi)$. The optimal choice of $\Psi$, in the sense of average optical power, is one which for a given volume, or rate, minimizes the average optical power $P^{\mathsf{G}}$. The optimum shaping region which maximizes shaping gain is the half-space

$$\Psi^*(r_{\max}) = \left\{ (\psi_1, \psi_2, \ldots, \psi_N) \in \mathbb{R}^N : \right.$$
$$\left. r_{\max} \geq 0, \ \psi_1 \in [0, r_{\max}] \right\} \quad (32)$$

for some fixed $r_{\max} \geq 0$ so that the desired volume is achieved. This assertion can be justified by noting that all points with equal components in the $\phi_1$ dimension, have the same average optical power. For a given volume and $\Upsilon$, the optimal shaping region can be formed by successively adjoining points of the smallest possible average optical power until the volume is achieved. Clearly, the region in (32) will result. This is much different than the case of the conventional channel where the $N$-sphere is the optimal shaping region in an average energy sense.

In practice, the set of signals transmitted are peak limited as well as average optical power limited. For a given $\Upsilon$, peak optical power $p/\sqrt{T}$ and volume, the optimum region which maximizes shaping gain is

$$\Psi^*(r_{\max}, p) = \Pi(p) \cap \Psi^*(r_{\max})$$

for $\Psi^*(r_{\max})$ (32) and $\Pi(p)$ as defined in (12). The form of this region can be justified in an identical manner as (32), except that here the points selected to form the given volume are taken from the set $\Upsilon \cap \Pi(p)$ which satisfy both the nonnegativity and the peak optical power constraints. Fig. 4 presents an example of such a region for raised-QAM defined in Section V.

## I. Shaping Gain

Suppose that $\Omega = \Lambda \cap \Upsilon \cap \Psi^*(r_{\max})$. The shaping gain in (30) can be simplified in this case by exploiting the symmetries of $\Upsilon$. By Property 1 of Theorem 1, the $\Upsilon_r$ are directly similar and scale linearly in $r$. As a result, the volume of each of the $\Upsilon_r$ must scale as $r^{N-1}$ for an $N$-dimensional signal space. Formally
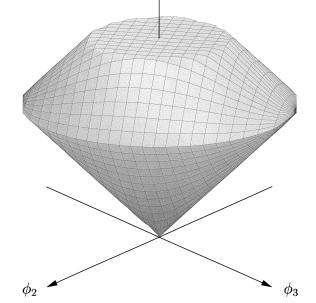
$$V(\Upsilon_r) = V(\Upsilon_1) r^{N-1}. \quad (33)$$



Fig. 4. The region $\Upsilon \cap \Psi^*(r_{\max}, p)$ in the case of raised-QAM for $r_{\max} = (3/4)p$.

The volume of $\Upsilon \cap \Psi^*(r_{\max})$ can then be computed simply as $\int_0^{r_{\max}} V(\Upsilon_r) \, dr$, which evaluates to

$$V(\Upsilon \cap \Psi^*(r_{\max})) = \frac{1}{N} V(\Upsilon_1) r_{\max}^N. \quad (34)$$

Exploiting the symmetry of the region in the $\phi_1$ dimension, the average optical power expression in (27) can be computed as an integral with respect to the $\phi_1$ only. Noting that $dV(\Upsilon \cap \Psi^*(r_{\max})) = V(\Upsilon_r) \, dr$ and substituting (33) and (34) into (27) gives

$$P^{\mathsf{G}}(\Upsilon \cap \Psi^*(r_{\max}))$$
$$= \int_0^{r_{\max}} r \, \frac{1}{V(\Upsilon \cap \Psi^*(r_{\max}))} \left( V(\Upsilon_1) r^{N-1} \right) dr$$

which simplifies to

$$P^{\mathsf{G}}(\Upsilon, \Psi^*(r_{\max})) = \frac{N}{N+1} r_{\max}. \quad (35)$$

The expression for the shaping gain is computed from $\nu^{\mathsf{C}}$, defined in Subsection C of the Appendix, and by substituting (34) and (35) into (30) to yield

$$\gamma_s(\Upsilon, \Psi^*(r_{\max}), \nu^{\mathsf{C}}) = \frac{\sqrt{\nu^{\mathsf{C}}}}{2} \left( \frac{(N+1)^{\nu^{\mathsf{C}}}}{N^{\nu^{\mathsf{C}}+1}} V(\Upsilon_1) r_{\max}^{N-\nu^{\mathsf{C}}} \right)^{1/\nu^{\mathsf{C}}}.$$

In contrast to the conventional electrical case, where shaping gain is invariant to scaling of the region, $\gamma_s(\Upsilon, \Psi^*(r_{\max}), \nu^{\mathsf{C}})$ depends on $r_{\max}$ since the dimension of the defined signal space is not equal to the effective dimension of $(\Omega, \Phi)$.

## J. Peak-Symmetric Signaling Schemes

A modulation scheme $(\Omega, \Phi)$ is termed *peak symmetric* if $\mathsf{Proj}(\Upsilon_1)$ is closed under inversion. From the point of view of signal amplitudes, using (13) and (15) of Section III-D, this condition implies that for $\boldsymbol{x}, \hat{\boldsymbol{x}} \in \mathsf{Proj}(\partial \Upsilon_1)$, $\boldsymbol{x} = -\hat{\boldsymbol{x}}$ and $\mathsf{Max}(\boldsymbol{x}) = \mathsf{Max}(-\boldsymbol{x}) = 1/\sqrt{T}$ and, hence, the term peak sym-

metric. Furthermore, for a peak-symmetric scheme using (16) for $\boldsymbol{v} \in \Upsilon_1$

$$\mathsf{Max}(\boldsymbol{v}) = \frac{2}{\sqrt{T}}. \tag{36}$$

As discussed in Section III-D, this maximum amplitude is achieved for $\boldsymbol{v} \in \partial \Upsilon_1$. Note that the 3-D PAM scheme in Fig. 3 is not peak symmetric. Section V-B presents examples of the $\mathsf{Proj}(\Upsilon_1)$ of 3-D peak-symmetric schemes.

Peak-symmetric schemes are desirable in the sense that the maximum amplitude value in $\Upsilon_1$ is achieved by all points in $\partial \Upsilon_1$. Maximization over all points in $\mathsf{Proj}(\partial \Upsilon_1)$ is not required in the calculation of the PAR (17) since it is satisfied at every point. Thus, peak-symmetric signaling schemes minimize the variation in the PAR over $\partial \Upsilon_1$, which may be beneficial in the design of transmit and receive electronics.

### K. Shaping Gain: Peak-Symmetric Schemes

Theorem 2 demonstrates that $\Pi(p)$ is the inversion of $\Upsilon$ with some constant shift in the $\phi_1$-axis. In the case of peak-symmetric schemes, since $\mathsf{Proj}(\partial \Upsilon_1)$ is closed under inversion, $\Pi(p)$ is a $\phi_1$-shifted *reflection* of $\Upsilon$ in the hyperplane $\phi_1 = 0$. As a result of this additional degree of symmetry, the cross sections of $\Upsilon$ and $\Pi(p)$ in the $\phi_1$-axis coincide for $\phi_1 = p/2$. In other words, the cross sections of $\Upsilon \cap \Psi^*(r_{\max}, p)$ for a given $\phi_1$ value are all directly similar to $\Upsilon_1$. Fig. 4 presents the region $\Upsilon \cap \Psi^*(r_{\max}, p)$ for a peak-symmetric, raised-QAM example defined in Section V.

Note that in the peak-symmetric case, for $r_{\max} < p/2$

$$\Upsilon \cap \Psi^*(r_{\max}, p) = \Upsilon \cap \Psi^*(r_{\max}).$$

The peak symmetry of the scheme requires that all points in $\partial \Upsilon_{r\max}$ have a maximum amplitude of $2r_{\max}/\sqrt{T}$ by (36). Thus, for $r_{\max} < p/2$, all points in $\Upsilon \cap \Psi^*(r_{\max})$ have a peak less than $p$.

The volume and $P^{\mathsf{G}}$ of the resulting region can be computed for $r_{\max} \in [p/2, p]$ as

$$V(\Upsilon \cap \Psi^*(r_{\max}, p)) = \frac{1}{N} V(\Upsilon_1) \left( 2 \left( \frac{p}{2} \right)^N - (p - r_{\max})^N \right) \tag{37}$$

and

$$
\begin{aligned}
&P^{\mathsf{G}}(\Upsilon \cap \Psi^*(r_{\max}, p)) \\
&= \frac{N}{N+1} \cdot \frac{(p/2)^N}{2(p/2)^N - (p - r_{\max})^N} \\
&\quad \cdot \left( p \frac{N+1}{N} - \left( \frac{p - r_{\max}}{p/2} \right)^N \left( \frac{1}{N} p + r_{\max} \right) \right).
\end{aligned} \tag{38}
$$

Substituting these expressions into (30) yields the shaping gain for these peak-constrained regions. Note also that peak optical power in the case of these regions is $p/\sqrt{T}$ and gives a PAR of

$$\mathsf{PAR}(\Upsilon \cap \Psi^*(r_{\max}, p)) = \frac{p}{P^{\mathsf{G}}(\Omega)}$$

which is independent of the symbol interval.

### L. Opportunistic Secondary Channels

In the case of optical intensity lattice codes defined in Section IV, an *opportunistic secondary channel* [26] exists which

can be exploited without a cost in average optical power. Using the shaping region $\Upsilon \cap \Psi^*(r_{\max})$, the $\Upsilon_{r\max}$ shell is generally unfilled. All lattice points in $\Upsilon_{r\max}$ are equivalent from an average optical power cost and the additional points in this shell can be selected without impact on the optical power of the scheme. This degree of freedom in selecting constellation points can be used to transmit additional data, introduce spectral shaping, or add a tone to the transmit spectrum for timing recovery purposes.

## V. EXAMPLES

This section presents examples of the signal space model defined in Section III-A and defines lattice codes using the techniques of Section IV. The gain versus the baseline is then computed. Design guidelines are presented based on the observed results.

### A. Definition of Example Schemes

As noted in Section IV and in [8], the performance of optical intensity modulation techniques depends not only on the electrical energy of the pulses (i.e., the geometry of the signal space), but also on the pulse shapes chosen to define the space. This section defines the basis functions used to form signals in the example schemes considered. Note that all symbols are limited to the interval $t \in [0, T)$ and $\phi_1(t)$ is specified as in (7).

QAM is a familiar modulation scheme in wireless communications. In the case of optical intensity channels, to satisfy the nonnegativity requirement, we define a *raised-QAM* scheme with $\phi_1(t)$ defined as before and

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

Adaptively biased QAM (AB-QAM) [11], [37] is a 3-D modulation scheme which is defined using the basis functions

$$
\begin{aligned}
\phi_2(t) &= \frac{1}{\sqrt{T}} \operatorname{rect}(t) - \frac{2}{\sqrt{T}} \operatorname{rect}(2t - T/2) \\
\phi_3(t) &= \frac{1}{\sqrt{T}} \operatorname{rect}(t) - \frac{2}{\sqrt{T}} \operatorname{rect}(2t - T).
\end{aligned}
$$

More generally, these functions are *Walsh functions*. This characterization is especially useful in light of the signal space definition in Section III-A, since the basis functions of AB-QAM are scaled and shifted versions of the first three Walsh functions [38].

A 3-D PAM scheme can be constructed by transmitting three constituent one-dimensional symbols from $(\Omega_\oplus, \Phi_\oplus)$. This construction is analogous to the techniques used in conventional lattice coding literature [26] and is the case considered in earlier optical lattice coding work [16]. The basis functions for this 3-D scheme, according to the signal space model defined in Section III-A, are

$$
\begin{aligned}
\phi_2(t) &= \sqrt{\frac{2}{T}} \operatorname{rect}(t) - \frac{3}{\sqrt{2T}} \operatorname{rect}\left( \frac{3t - T}{2} \right) \\
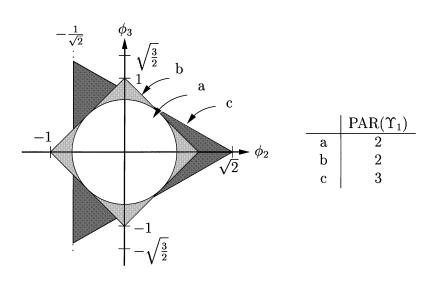\phi_3(t) &= \sqrt{\frac{3}{2T}} \operatorname{rect}(3t - T) - \sqrt{\frac{3}{2T}} \operatorname{rect}(3t - 2T).
\end{aligned}
$$

| | $\mathrm{PAR}(\Upsilon_1)$ |
|---|---|
| a | 2 |
| b | 2 |
| c | 3 |

Fig. 5. Example cross sections of 3-D admissible regions, $\mathsf{Proj}(\Upsilon_1)$, for (*a*) raised-QAM, (*b*) AB-QAM, (*c*) 3-D PAM scheme.

Note that the results derived for this case will differ from those in [16] due to the definition of bandwidth and the fact that unlike previous work, the precise impact of shaping on bandwidth is considered here.

A popular signaling scheme in optical communications is PPM. For $L$-PPM, the symbol interval is divided into a series of $L$ subintervals. A symbol is formed by transmitting an optical intensity in only one of the $L$ subintervals while the optical intensity is set to zero in the other subintervals. These schemes were originally conceived for the photon-counting channel and achieve high power efficiency at the expense of bandwidth efficiency [3]. Note that a PPM modulation scheme can be thought of as a coded version of the 3-D PAM scheme discussed earlier.

### B. Geometric Properties

Fig. 5 contains plots of the regions $\mathsf{Proj}(\Upsilon_1)$ of the example bases defined in Section V-A. Consistent with Property 3 of Theorem 1, the regions are all closed, convex, and bounded.

In the case of raised-QAM, every point in $\mathsf{Proj}(\Upsilon_1)$ is a sinusoid time limited to $[0, T)$ with amplitude determined by the squared distance from the origin. Since sinusoids of the same energy have the same amplitude, regardless of phase, all points equidistant from the origin have the same amplitude. As a result, a two–dimensional (2-D) disc naturally results as the region $\mathsf{Proj}(\Upsilon_1)$. In the case of 3-D PAM, $\mathsf{Proj}(\Upsilon_1)$ is an equilateral triangle with sides of length $\sqrt{6}$. It is easy to show that the signals corresponding to a 3-PPM scheme are represented by the vertices of the triangle. In this manner, PPM can be seen as a special case of the 3-D PAM scheme.

Note that the raised-QAM and AB-QAM scheme, (*a*) and (*b*) in Fig. 5, both represent peak-symmetric signaling schemes. While the $\mathsf{Proj}(\Upsilon_1)$ regions for these two modulation schemes are different, the peak-to-average amplitude value of the signals represented is 2 in both cases as given in (36). The 3-D PAM scheme is not peak symmetric. The largest peak values occur for the points at the vertices of the triangle ($\boldsymbol{x}_1$ in Fig. 3) to give a PAR of 3 for $\Upsilon_1$.
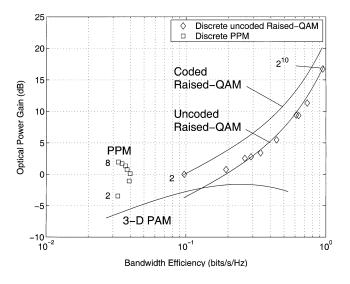


Fig. 6. Gain over baseline versus bandwidth efficiency. Note that points indicated with □ and ◇ represent discrete PPM and raised-QAM constellations, respectively, while the solid lines represent results using the continuous approximation.

### C. Gain of Example Lattice Codes

The optical power gain over the baseline versus the bandwidth efficiency is plotted for a variety of PPM and raised-QAM modulation schemes in Fig. 6.

Ten discrete raised-QAM constellations were formed as $\Omega = \mathbb{Z}^3 \cap \Upsilon \cap \Psi^*(r_{\max})$ by selecting the appropriate $r_{\max}$ to have each carrying from 1 to 10 bits per symbol. The power spectral density of each scheme was computed symbolically via (40) using a symbolic mathematics software package [39] and integrated numerically to determine $W_{0.99}$ and $W_{0.999}$ for a given $T$. These results were then combined to find the $\kappa$ for each $(\Omega, \Phi)$. The power spectral density of the baseline scheme is trivial to compute, and was integrated numerically to give $\kappa_\oplus = 20.572$ for $W_{0.99}$ and $\kappa_\oplus = 202.217$ for $W_{0.999}$. The effective dimension of the constellations considered are presented in Table I. The same procedure was repeated for discrete PPM

TABLE I
EFFECTIVE DIMENSION FOR OPTIMALLY SHAPED, DISCRETE RAISED-QAM CONSTELLATIONS

| $\lvert\Omega\rvert$ | $W_{0.99}$ Bandwidth $\nu$ | $W_{0.999}$ Bandwidth $\nu$ |
|---|---|---|
| cont. approx. | $\nu^{\mathsf{C}} = 1.006$ | $\nu^{\mathsf{C}} = 1.000$ |
| 2 | 1.000 | 1.000 |
| 4 | 1.004 | 1.000 |
| 8 | 1.107 | 1.108 |
| 16 | 1.335 | 1.365 |
| 32 | 1.428 | 1.458 |
| 64 | 1.319 | 1.338 |
| 128 | 1.114 | 1.119 |
| 256 | 1.224 | 1.245 |
| 512 | 1.194 | 1.189 |
| 1024 | 1.024 | 1.033 |

TABLE II
EFFECTIVE DIMENSION FOR DISCRETE PPM CONSTELLATIONS

| $\lvert\Omega\rvert$ | $W_{0.99}$ Bandwidth $\nu$ | $W_{0.999}$ Bandwidth $\nu$ |
|---|---|---|
| 2 | 2.990 | 2.994 |
| 3 | 3.939 | 4.002 |
| 4 | 4.893 | 4.993 |
| 5 | 5.985 | 5.993 |
| 6 | 6.810 | 7.002 |
| 7 | 7.855 | 7.984 |
| 8 | 8.907 | 8.982 |

constellations of size 2 through 8. Table II presents the $\nu$ values computed for these constellations.

In both the raised-QAM and PPM examples, the effective dimension in Tables I and II are essentially independent of the value of $K$ in $W_K$. This suggests that $\nu$ is not sensitive to the choice of fraction of total power used to compute bandwidth for values of $K$ near 1. Note also that $\nu$ for the PPM constellations increases as $\lvert\Omega\rvert$ since each signal point is orthogonal to all others. In the case of raised-QAM, the effective dimension remains approximately constant as $\Omega$ increases.

The gain of the raised-QAM and PPM examples was computed via (25) using the $\nu$ derived from the $W_{0.99}$ and plotted in Fig. 6. Note that the raised-QAM schemes provide large optical power gain over the baseline scheme while operating at high bandwidth efficiencies, while the PPM schemes provide small gain at low bandwidth efficiencies.

Care must be taken when using the optical power gain as a figure of merit since it depends on the baseline scheme chosen. Since the raised-QAM and PPM examples operate at different bandwidth efficiencies, a direct comparison of their performance is not possible using this measure. Indeed, this plot suggests that PPM and raised-QAM are suited for operation under highly different channel conditions. Section V-D presents a comparison technique based on an idealized point-to-point link which illustrates the conditions under which PPM or raised-QAM are appropriate.

Direct computation of the power spectral density to find $\nu$ is a time-consuming process. In order to verify the asymptotic accuracy of the continuous approximation to the optical power gain

(28), the parameter $\nu^{\mathsf{C}}$ was computed for the raised-QAM examples. The continuous approximation to the power spectral density was computed symbolically using (41) and integrated numerically to get an estimate of $\nu^{\mathsf{C}}$, which is presented in Table I. The results are plotted in Fig. 6 and show that the continuous approximation to the optical power gain approaches the discrete case for large constellations.

The continuous approximation of gain for the 3-D PAM case was computed in an identical fashion to give a $\nu^{\mathsf{C}} = 3.629$ and is also plotted in Fig. 6. The baseline scheme is more power efficient than the 3-D PAM scheme in spite of the fact that the 3-D PAM constellation arises as a shaped version of the baseline using $\Psi^*(r_{\max})$. This is a consequence of the fact that the $\nu$ depends on $\Upsilon \cap \Psi$. The effective dimension for the shaped case is larger than the baseline value of 3 which eliminates any shaping gain. In the case of 3-D PAM, the approximation that the bandwidth of each symbol is approximately constant no longer holds, and as a result, the interpretation of $\Psi^*(r_{\max})$ as optimal in average optical power at a given rate is not true.

A 24-dimensional example was constructed by specifying symbols consisting of blocks of eight consecutive symbols of the raised-QAM scheme. The resulting $\Upsilon$ was intersected with the Leech lattice, $\Lambda_{24}$, to form the constellation. The optical power gain was calculated using the continuous approximation. As in the 3-D case, the effective dimension was approximated by integrating over the 24-dimensional region symbolically to give $\nu^{\mathsf{C}} = 8.268$. The optical gain was then plotted in Fig. 6 for comparison. The use of $\Lambda_{24}$ over $\mathbb{Z}^{24}$ gives a coding gain of approximately 3 dB in optical power. This is less than the 6-dB electrical coding gain that arises in conventional channels, since, as alluded to in (31), the optical coding gain depends on the square root of Hermite's parameter. Qualitatively, an electrical coding gain corresponds to a reduction in the mean square value of the signal while an optical coding gain corresponds to the reduction in the mean value of the signal.

### D. Idealized Point-to-Point Link

The gain over a baseline is highly dependent on the baseline scheme that is chosen. In order to have a more concrete comparison between PPM and raised-QAM schemes, the optical power efficiency was measured by the distance that each scheme could transmit over an eye-safe, point-to-point wireless optical channel operating at a given symbol error rate and data rate. The receiver and transmitter are assumed to be aligned and a distance $D$ centimeters apart. The average transmitter intensity is limited to $\overline{I} = 104$ mW/sr, which is the eye-safety limit of a commercially available wireless infrared transceiver [40]. The detector sensitivity is taken to be $r = 25\ \mu\text{A} \cdot \text{m}^2/\text{mW}$ over an active area of 1 cm$^2$ and the channel noise standard deviation $\sigma_n = 11.5 \times 10^{-12}\sqrt{\text{W/Hz}}$, both of which have been reported for a similar experimental link [41]. The symbol error rate in all cases was set to $10^{-8}$, which corresponds to the IrDA fast IR specification [12]. Assuming operation in the far-field case, the transmission distance under these constraints is

$$D = \sqrt{\frac{r\overline{I} \cdot \text{CFM}(\Omega, \nu)}{2k\sigma_n}}$$

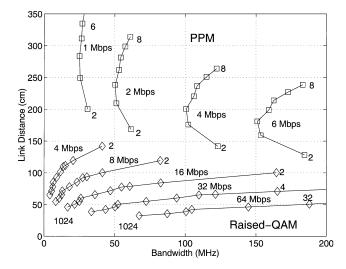where $k \approx 5.6120$, which is set by the symbol error rate.

Fig. 7. Idealized point-to-point link length versus bandwidth for the discrete uncoded PPM ($\square$) and raised-QAM ($\diamond$) constellations of Fig. 6 with SER $\approx 10^{-8}$.

TABLE III
EFFECTIVE DIMENSION FOR RAISED-QAM CONSTELLATIONS
SHAPED BY $\Psi^*(r_{\max}, p)$

| $r_{\max}/p$ | $W_{0.99}$ Bandwidth $\nu^{\mathsf{C}}$ | $W_{0.999}$ Bandwidth $\nu^{\mathsf{C}}$ |
|---|---|---|
| 0.50 | 1.006 | 1.000 |
| 0.55 | 1.006 | 1.000 |
| 0.60 | 1.005 | 1.000 |
| 0.65 | 1.005 | 1.000 |
| 0.70 | 1.005 | 1.000 |
| 0.75 | 1.005 | 1.000 |
| 0.80 | 1.005 | 1.000 |
| 0.85 | 1.004 | 1.000 |
| 0.90 | 1.004 | 1.000 |
| 0.95 | 1.004 | 1.000 |
| 1.00 | 1.004 | 1.000 |

Fig. 7 presents the link distance versus bandwidth for the discrete uncoded examples presented in Section V-C with a fixed symbol error rate and for a variety of data rates. For a given con-stellation, the data rate is varied by varying $T$. In both schemes, an increase in the rate for the same constellation causes a reduction in transmission distance. This is because, through (9), $P$ is inversely proportional to $\sqrt{T}$. An increase in the symbol rate for a given constellation then increases the optical power required to achieve the given symbol error rate and, hence, reduces transmission range. As discussed in Section III-A, this is due to the unit energy normalization of basis functions in the signal space.

PPM schemes provide long-range transmission at the price of bandwidth while raised-QAM schemes provide high data rates at the expense of short link range. Longer range links are limited by the amount of power which can be collected at the receiver. As a result, the power-efficient PPM scheme must be employed at the cost of bandwidth expansion or equivalently rate loss. As the link distance becomes smaller, the amount of power collected increases. In this case, bandwidth-efficient raised-QAM techniques can be employed to increase the link data rate. There is, thus, a rate versus link distance tradeoff in the design of point-to-point wireless optical links.

Fig. 7 provides a design guide for the construction of modulation schemes for a point-to-point link. Suppose that an industry standard IrDA fast IR link using 4-PPM and operating at at 4 Mb/s is taken as the operating point. In the bandwidth measure defined, the optoelectronics of this link limit the bandwidth to 100 MHz. This link is able to support a 4-Mb/s data rate at the specified $10^{-8}$ symbol error rate. If the same physical link is required to transmit at 125 cm, a 2-raised-QAM scheme can be used to achieve rates of 8 Mb/s. If the distance is reduced to 75 cm, rates of up to 16 Mb/s are possible using 4-raised-QAM. Data rates of 32 Mb/s over 60 cm and 64 Mb/s over 40 cm are also possible using 32-raised-QAM and 256-raised-QAM constellations, respectively. Thus, for a given set of optoelectronics, which set the channel bandwidth, the data rate can be optimized

by the proper selection of modulation scheme. Power-efficient schemes operate at low data rates over long-range links and bandwidth-efficient schemes offer high data rates over shorter distances.

Note that in Fig. 6 the gain of raised-QAM schemes increases as $|\Omega|$ while in Fig. 7 the link distance decreases with $|\Omega|$. This is due to the fact that the optical power requirement of the baseline scheme increases more rapidly than raised-QAM at high bandwidth efficiencies. The use of link distance is a more practical measure of optical power and permits the direct comparison of schemes.

*E. Peak Optical Power*

In the preceding examples, the peak optical power of the schemes in question was not discussed. In this subsection, the impact of shaping with $\Psi^*(r_{\max}, p)$ on fractional power bandwidth and on average optical power are investigated.

Table III presents the $\nu^{\mathsf{C}}$ values for raised-QAM constellations shaped with $\Psi^*(r_{\max}, p)$. For a given $r_{\max}$ and $p$, scaling of $\Psi^*(r_{\max}, p)$ does not alter $\nu^{\mathsf{C}}$ and so the set of regions can be parameterized by $k = r_{\max}/p$. In this case, since the symbols are all nearly band limited in the sense of Section IV-H, the effective dimension is approximated as being nearly independent of $\Upsilon \cap \Psi$. Thus, for the raised-QAM example, $\Psi^*(r_{\max})$ and $\Psi^*(r_{\max}, p)$ are optimal in the average optical power sense.

At a given rate, the shaping region $\Psi^*(r_{\max}, p)$ provides a reduction in the peak optical power of a scheme at the cost of increasing the average optical power over the case using $\Psi^*(r_{\max})$. Using the raised-QAM bases, consider forming two constellations $\Omega_1 = \Lambda \cap \Upsilon \cap \Psi^*(r_{\max})$ and $\Omega_2 = \Lambda \cap \Upsilon \cap \Psi^*(r'_{\max}, p)$, for some fixed $r_{\max}$ and $k = r'_{\max}/p$ for $k \in [0.5, 1]$. Under the assumption that $\nu$ is unaffected by $\Psi$, fixing

$$V(\Upsilon \cap \Psi^*(r_{\max})) = V(\Upsilon \cap \Psi^*(r'_{\max}, p))$$

defined in (34) and (37), fixes the rates of the schemes to be equal. The peak constraint causes an excess average optical power penalty, $P(\Upsilon \cap \Psi^*(kp, p))/P(\Upsilon \cap \Psi^*(r_{\max}))$, which can be computed via (35) and (38). The normalized peak optical power of $\Omega_2$ with respect to $\Omega_1$ is defined as $p/(2r_{\max})$, where the peak optical power of $\Omega_1$ is defined in (36).
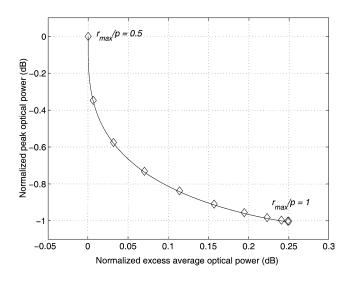
Fig. 8. Peak optical power versus excess average optical power (at same rate). Solid line represents a case where $\nu$ is assumed independent of $\Psi$, and points $\diamond$ indicate values for which $\nu^C$, in Table III, was explicitly computed.

Fig. 8 presents the tradeoff between the peak and average optical power at a given rate for the uncoded raised-QAM example. The figure illustrates the independence of $\nu$ and $\Psi$ and shows that the peak value of the constellation can be reduced by approximately 1 dB over the case of $\Psi^*(r_{\max})$ at a cost of less than 0.25-dB increase in average optical power. This tradeoff is important to note since a scheme with high peak optical amplitudes is more vulnerable to channel nonlinearities and requires more complex modulation circuitry.

## VI. CONCLUDING REMARKS

Modem design for the free-space optical intensity channel is significantly different than for the conventional electrical channel. Whereas in electrical channels the constraint is typically on the mean-square value of the transmitted signal, the optical-intensity channel imposes the constraint that all signals are nonnegative and that the average signal amplitude is limited. In this work, we impose a further constraint on the bandwidth of the channel since wireless optical channels are typically bandwidth constrained due to the multipath distortion and optoelectronic capacitance. The practical constraint of peak optical power is also addressed in this work and the tradeoff between average and peak optical power is quantified.

This work has shown that popular PPM schemes provide a means to trade off optical power efficiency for bandwidth efficiency. In point-to-point links, this translates to a tradeoff between transmission distance and data rate. Bandwidth-efficient schemes, such as raised-QAM introduced in this paper, provide much higher data rates at the expense of a greater required optical power. In comparison to PPM point-to-point links, modems based on raised-QAM provide far higher data rates at lower transmission distances.

This paper has further demonstrated that coding alone, although necessary to approach capacity, provides relatively limited optical power gain. The optical coding gain is shown to be proportional to the square root of the electrical coding gain. In

the design of modems for such a channel, physical improvements to improve the optical power efficiency should first be exploited before complex coding schemes are considered.

There exist a variety of physical techniques which can be employed to improve the optical power efficiency of the free-space optical channel. Optical concentrators, such as mirrors and lenses, can be used to increase the receive power at the price of higher implementation cost [42], [43]. Multiple spatially separated light emitters can be used to form a distributed source, thereby increasing the amount of optical power transmitted. At longer wavelengths (in the 1.3- and 1.5-$\mu$m range) the human eye is nearly opaque. As a result, an order of magnitude increase in the optical power transmitted can be realized at the price of costlier optoelectronics [14].

Through the use of such techniques it is possible to engineer an optical channel which for a given distance offers a significantly improved optical power at the receiver. Bandwidth-efficient raised-QAM-type modulation can then be applied in this new channel to provide improved data rates over the given transmission distance. Thus, these physical techniques increase the range of transmission distances in which high-rate, bandwidth-efficient modulation is appropriate.

## APPENDIX

This appendix will briefly outline how the power spectral density is calculated for the examples in Section V and present an approximation of the power spectral density based on the continuous approximation (26).

### A. Definitions

The power spectral density of $(\Omega, \Phi)$, $S_X(f)$, can be written as the sum of two terms, the discrete spectrum $S_X^d(f)$ and the continuous spectrum $S_X^c(f)$. The fractional power bandwidth measure in (21) is defined using only the continuous portion of the power spectral density. Discrete spectral components are typically undesirable since they do not carry any information but require electrical energy to be transmitted. In the optical channel model, the discrete spectral component at $f = 0$ represents the average optical power of $(\Omega, \Phi)$ while all other discrete components of the spectrum represent zero average optical power. These discrete components, except at $f = 0$, can be eliminated through the prudent construction of $\Omega$. Thus, the frequency extent of $(\Omega, \Phi)$ can then be considered as being set by $S_X^c(f)$.

It should be noted that $E = \int_{-\infty}^{\infty} S_X(f)\, df$ is *not* equal to the optical power cost of the scheme. The power spectrum is the distribution of electrical energy in the received photocurrent $y(t)$ in Fig. 1, while the average optical power is the average photocurrent amplitude. As discussed in Section II-A, the optical channel can be modeled as a baseband electrical system with constraints on the amplitude of signals transmitted. As a result, the use of the bandwidth of the electrical photocurrent signal is appropriate.

### B. Calculation of Power Spectral Density

For digital modulation schemes, where the signal transmitted can be described as in (4) and the correlation from symbol to

symbol can be described by a Markovian model, the power spectral density can be shown to be [44]

$$S_X(f) = S_X^d(f) + S_X^c(f)$$

where

$$S_X^d(f) = \frac{1}{T^2} \sum_{n=-\infty}^{\infty} \left| \sum_{i \in M} \Pr(i) x_i^F \left( \frac{n}{T} \right) \right|^2 \delta \left( f - \frac{n}{T} \right)$$

$$S_X^c(f) = \frac{1}{T} \sum_{k \in M} \sum_{l \in M} \Pr(k) x_k^F(f) x_l^{F^*}(f)$$
$$\cdot \left\{ \sum_{m=-\infty}^{\infty} \left( a_{kl}^{(m)} - \Pr(l) \right) \exp(-j2\pi f m T) \right\}.$$

$x_i^F(f)$ is the Fourier transform of signal $x_i(t) \in X$, $\Pr(i)$ is the steady-state probability of transmitting symbol $x_i(t)$, and $a_{kl}^{(m)}$ is the $m$-step conditional probability of transmitting symbol $x_l(t)$ given the current symbol is $x_k(t)$. The power spectral density depends on two factors: the pulse shapes, through the $x_i^F(f)$, and on the correlation between symbols. Consider the case when all of the $x_i(t)$ are strictly band limited to $B$ hertz. It is clear that the resulting $S_X^c(f)$ must necessarily be band limited to $B$ hertz, independent of coding or shaping. This is the condition under which classical lattice coding results are derived for the electrical channel [26].

Under the conditions of independent and equally likely signaling, as in Section IV, the term $a_{ij}$ can be simplified as

$$a_{ij}^{(m)} = \begin{cases} 1/M, & m \neq 0 \\ \delta_{ij}, & m = 0 \end{cases}$$

where $\delta_{ij}$ is the Kronecker delta function. The power spectral density in (39) can be simplified to yield

$$S_X(f) = \frac{1}{M^2 T^2} \sum_{n=-\infty}^{\infty} \left| \sum_{i \in M} x_i^F \left( \frac{n}{T} \right) \right|^2 \delta \left( f - \frac{n}{T} \right)$$
$$+ \frac{1}{T} \left[ \sum_{i \in M} \frac{1}{M} \left| x_i^F(f) \right|^2 - \left| \sum_{i \in M} \frac{1}{M} x_i^F(f) \right|^2 \right]. \quad (40)$$

### C. Continuous Approximation of Power Spectral Density

The summation in the expression of $S_X^c(f)$ can be simplified using the continuous approximation (26). The resulting continuous approximation to power spectral density $S_X^c(f)^C$ can be written in terms of the Fourier transforms of the basis functions, $\phi_i^F(f)$, $i \in N$, as

$$S_X^c(f)^C = \frac{1}{T} \sum_{n \in N} J_n \left| \phi_n^F(f) \right|^2$$
$$+ \frac{1}{T} \sum_{\substack{n, l \in N \\ n \neq l}} K_{nl} \text{Re} \left\{ \phi_n^F(f) \phi_l^{F^*}(f) \right\} \quad (41)$$

where $*$ denotes conjugation and

$$J_n = \frac{1}{V(\Upsilon \cap \Psi)} \int_{\Upsilon \cap \Psi} x_n^2 \, dV(\boldsymbol{x})$$
$$- \left( \frac{1}{V(\Upsilon \cap \Psi)} \int_{\Upsilon \cap \Psi} x_n \, dV(\boldsymbol{x}) \right)^2$$
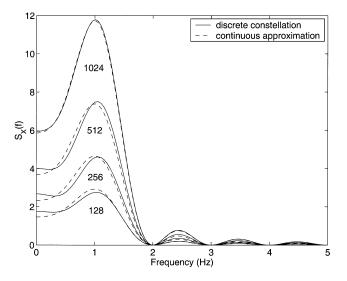
Fig. 9. Plot of power spectral density of uncoded, discrete raised-QAM constellations versus the continuous approximation for the power spectral density in those cases ($T = 1$ in all cases).

$$K_{nl} = \frac{1}{V(\Upsilon \cap \Psi)} \int_{\Upsilon \cap \Psi} x_n x_l \, dV(\boldsymbol{x})$$
$$- \frac{1}{V^2(\Upsilon \cap \Psi)} \int_{\Upsilon \cap \Psi} x_n \, dV(\boldsymbol{x}) \int_{\Upsilon \cap \Psi} x_l \, dV(\boldsymbol{x}).$$

Thus, to calculate the continuous approximation for the power spectral density, the first- and all second-order moments of the $N$-dimensional random vector uniformly distributed over $\Upsilon \cap \Psi$ must be determined.

Fig. 9 shows $S_X^c(f)^C$ plotted on the same axis as the power spectral density calculated via (40) for various sizes of discrete, optimally shaped, uncoded, raised-QAM examples presented in Section V. Notice that $S_X^c(f)^C$ approaches the true power spectral density (40) for high rates.

The effective dimension $\nu$ in (23) can be approximated by estimating the fractional power bandwidth of a scheme via numerical integration of (41) to yield $\nu^C$. As verified in Fig. 9, the accuracy of this approximation improves as the rate of the scheme increases.

### REFERENCES

[1] J. P. Gordon, "Quantum effects in communication systems," *Proc. IRE*, vol. 50, pp. 1898–1908, Sept. 1962.

[2] J. R. Pierce, "Optical channels: Practical limits with photon counting," *IEEE Trans. Commun.*, vol. COM-26, pp. 1819–1821, Dec. 1978.

[3] R. J. McEliece, "Practical codes for photon communication," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 393–398, July 1981.

[4] M. H. A. Davis, "Capacity and cutoff rate for Poisson-type channels," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 710–715, Nov. 1980.

[5] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel—Part I and II," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1449–1471, Nov. 1988.

[6] 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.

[7] B. Reiffen and H. Sherman, "An optimum demodulator for Poisson processes: Photon source detectors," *Proc. IEEE*, vol. 51, pp. 1316–1320, Oct. 1963.

[8] K. Abend, "Optimum photon detection," *IEEE Trans. Inform. Theory*, vol. IT–12, pp. 64–65, Jan. 1966.

[9] R. M. Gagliardi and S. Karp, "$M$-ary Poisson detection and optical communications," *IEEE Trans. Commun. Technol.*, vol. COM-17, pp. 208–216, Apr. 1969.

[10] S. Karp, E. L. O'Neill, and R. M. Gagliardi, "Communication theory for the free-space optical channel," *Proc. IEEE*, vol. 58, pp. 1611–1626, Oct. 1970.

[11] S. Hranilovic, "Modulation and constrained coding techniques for wireless infrared communication channels," M.A.Sc. thesis, Dept. Elec. Comp. Eng., Univ. Toronto, Toronto, ON, Canada, 1999.

[12] Infrared Data Association. (1998) Infrared Data Association serial infrared physical layer specification. Version 1.3. [Online]. Available: www.irda.org

[13] H. Park and J. R. Barry, "Modulation analysis for wireless infrared communications," in *Proc. IEEE Int. Conf. Communications*, 1995, pp. 1182–1186.

[14] J. M. Kahn and J. R. Barry, "Wireless infrared communications," *Proc. IEEE*, vol. 85, pp. 263–298, Feb. 1997.

[15] J. B. Carruthers and J. M. Kahn, "Multiple-subcarrier modulation for nondirected wireless infrared communication," *IEEE J. Select. Areas Commun.*, vol. 14, pp. 538–546, Apr. 1996.

[16] D. Shiu and J. M. Kahn, "Shaping and nonequiprobable signalling for intensity-modulated signals," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2661–2668, Nov. 1999.

[17] International Electrotechnical Commission, "Safety of laser products—Part 1: Equipment classification, requirements and user's guide," Group safety publication, reference number 825-1, 1993.

[18] J. M. Kahn, W. J. Krause, and J. B. Carruthers, "Experimental characterization of nondirected indoor infrared channels," *IEEE Trans. Commun.*, vol. 43, pp. 1613–1623, Feb./Mar./Apr. 1995.

[19] J. B. Carruthers and J. M. Kahn, "Modeling of nondirected wireless infrared channels," *IEEE Trans. Commun.*, vol. 45, pp. 1260–1268, Oct. 1997.

[20] F. R. Gfeller and U. Bapst, "Wireless in-house communication via diffuse infrared radiation," *Proc. IEEE*, vol. 67, pp. 1474–1486, Nov. 1979.

[21] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signalling for Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 39, pp. 913–929, May 1993.

[22] A. Gray, *Modern Differential Geometry of Curves and Surfaces*. Boca Raton, FL: CRC, 1993.

[23] M. J. Panik, *Fundamentals of Convex Analysis: Duality, Separation, Representation and Resolution*. Boston, MA: Kluwer, 1993.

[24] G. D. Forney, Jr., R. G. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Qureshi, "Efficient modulation for band-limited channels," *IEEE J. Select. Areas Commun.*, vol. SAC-2, pp. 632–647, Sept. 1984.

[25] G. D. Forney, Jr., "Coset codes—Part I: Introduction and geometrical classification," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1123–1151, Sept. 1988.

[26] G. D. Forney, Jr. and L.-F. Wei, "Multidimensional constellations—Part I: Introduction, figures of merit, and generalized cross constellations," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 877–892, Aug. 1989.

[27] G. R. Lang and F. M. Longstaff, "A Leech lattice modem," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 968–973, Aug. 1989.

[28] A. R. Calderbank and L. H. Ozarow, "Nonequiprobable signaling on the Gaussian channel," *IEEE Trans. Inform. Theory*, vol. 36, pp. 726–740, July 1990.

[29] F. R. Kschischang and S. Pasupathy, "Optimal shaping properties of the truncated polydisc," *IEEE Trans. Inform. Theory*, vol. 40, pp. 892–903, May 1994.

[30] A. K. Khandani and P. Kabal, "Shaping multidimensional signal spaces—Part I: Optimum shaping, shell mapping," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1799–1808, Nov. 1993.

[31] G. D. Forney, Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2384–2415, Oct. 1998.

[32] E. Biglieri and G. Caire, "Power spectrum of block-coded modulation," *IEEE Trans. Commun.*, vol. 42, pp. 1580–1585, Feb./Mar./Apr. 1994.

[33] B. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty—I," *Bell Syst. Tech. J.*, pp. 43–63, Jan. 1961.

[34] H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty—II," *Bell Syst. Tech. J.*, pp. 65–84, Jan. 1962.

[35] ——, "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.*, pp. 1295–1336, July 1962.

[36] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, 2nd ed. New York: Springer-Verlag, 1993.

[37] S. Hranilovic and D. A. Johns, "A multilevel modulation scheme for high-speed wireless infrared communications," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. VI, 1999, pp. 338–341.

[38] H. F. Harmuth, *Transmission of Information by Orthogonal Functions*, 2nd ed. New York: Springer-Verlag, 1972.

[39] Waterloo Maple Inc., *Maple V Programming Guide: Release 5*. New York: Springer-Verlag, 1997.

[40] Hewlett-Packard Corp. (1997) Compliance of infrared communication products to IEC 825-1 and CENELEC EN 60825-1. Application Note 1118. [Online]. Available: www.semiconductor.agilent.com/ir/app_index.html.

[41] K. L. Sterckx, J. M. Elmirghani, and R. A. Cryan, "On the use of adaptive threshold detection in optical wireless communication systems," in *Proc. IEEE GLOBECOM*, vol. 2, 2000, pp. 1242–1246.

[42] J. R. Barry, *Wireless Infrared Communications*. Boston, MA: Kluwer, 1994.

[43] R. Otte, L. P. de Jong, and A. H. M. van Roermund, *Low-Power Wireless Infrared Communications*. Boston, MA: Kluwer, 1999.

[44] S. G. Wilson, *Digital Modulation and Coding*. Upper Saddle River, NJ: Prentice-Hall, 1996.