

Ultimate Channel Capacity of Free-Space Optical Communications

Jeffrey H. Shapiro, Saikat Guha, and Baris I. Erkmen

*Research Laboratory of Electronics, Massachusetts Institute of Technology,
Cambridge, MA, 02139*

jhs@mit.edu, saikat@mit.edu, erkmen@mit.edu

The ultimate classical information capacity of multiple-spatial-mode, wideband, optical communications in vacuum between soft-aperture transmit and receive pupils is considered. The ultimate capacity is shown to be achieved by coherent-state encoding and joint measurements over entire codewords. This capacity is compared with the capacities realized with the same encoding and homodyne or heterodyne detection, which are single-channel-use measurements. Realistic background spectral radiance values are used to obtain tight bounds on the capacity of single-spatial-mode, narrowband 1.55- μm -wavelength free-space communications in the presence of background light. © 2005 Optical Society of America

OCIS codes: 060.4510, 060.1660

1. INTRODUCTION

Ubiquitous, reliable, high data-rate communication—carried by electromagnetic waves at microwave to optical frequencies—is an essential ingredient of our technological age. Information theory seeks to delineate the ultimate limits on reliable communication that arise from the presence of noise and other disturbances, and to establish means by which these limits can be approached in practical systems. The mathematical foundation for this assessment of limits is Shannon’s noisy channel coding theorem [1], which introduced the notion of channel capacity—the maximum mutual information between a channel’s input and output—as the highest rate at which error-free communication could be maintained. Textbook treatments of channel capacity [2, 3] study channel models—ranging from the binary symmetric channel’s digital abstraction to the additive white-Gaussian-noise channel’s idealization of thermal-noise-limited waveform transmission—for which classical physics is the underlying paradigm. Fundamentally, however, electromagnetic waves are quantum mechanical, i.e., they are Boson fields [4, 5]. Moreover, high-sensitivity photodetection systems have long been limited by noises of quantum mechanical origin [6]. Thus it would seem that determining the ultimate limits on optical communication would necessarily involve an explicitly quantum analysis, but such has not been the case. Nearly all work on the communication theory of optical channels—viz., that done for systems with laser transmitters and either coherent-detection or direct-detection receivers—uses semiclassical (shot-noise) models (see, e.g., [7, 8]). Here, electromagnetic waves are taken to be classical entities, and the fundamental noise is due to the random release of discrete charge carriers in the process of photodetection. Inasmuch as the quantitative results obtained from shot-noise analyses of such systems are known to coincide with those derived in rigorous quantum-mechanical treatments [9], it might be hoped that the semiclassical approach would suffice. But, Helstrom’s derivation [10] of the optimum quantum receiver for binary

coherent-state (laser light) signaling demonstrated that the lowest error probability, at constant average photon number, required a receiver that was neither coherent detection nor direct detection. That Dolinar [11] was able to show how Helstrom's optimum receiver could be realized with a photodetection feedback system which admits to a semiclassical analysis did not alleviate the need for a fully quantum-mechanical theory of optical communication, as Shapiro et al. [12] soon proved that even better binary-communication performance could be obtained by use of two-photon coherent state (now known as squeezed state) light, for which semiclassical photodetection theory did not apply.

In quantum mechanics, the state of a physical system together with the measurement that is made on that system determine the statistics of the outcome of that measurement, see, e.g., [13]. Thus in seeking the classical information capacity of a Bosonic channel, we must allow for optimization over *both* the transmitted quantum states *and* the receiver's quantum measurement. In particular, it is *not* appropriate to immediately restrict consideration to coherent-state transmitters and coherent-detection or direct-detection receivers. Imposing these structural constraints leads to Gaussian-noise (Shannon-type) capacity formulas for coherent (homodyne and heterodyne) detection [14] and a variety of Poisson-noise capacity results (depending on the power and bandwidth constraints that are enforced) for shot-noise-limited direct detection [15]–[18]. None of these results, however, can be regarded as specifying the ultimate limit on reliable communication at optical frequencies. What is needed for deducing the fundamental limits on optical communication is the analog of Shannon's noisy channel coding theorem—unfettered by unjustified structural constraints on the transmitter and receiver—that applies to transmission of classical information over a noisy quantum channel, viz., the Holevo-Schumacher-Westmoreland (HSW) theorem [19]–[21].

Recently [22], we have used the HSW theorem to find the classical capacity of the pure-loss Bosonic channel—in which signal photons may be lost in propagation, with the channel injecting the minimum quantum noise required to preserve the Heisenberg uncertainty principle—and showed that it is achieved by single-use encoding over coherent states with joint measurements over entire codewords. Single-use coherent-state encoding was also used to obtain a lower bound on the capacity of the thermal-noise channel—a lossy channel in which noise photons are injected from the external environment—which we have shown to be tight in the limit of low and high noise levels [23, 24]. Furthermore, if our conjecture concerning the minimum output entropy of this isotropic-Gaussian noise channel is correct [25], then single-use coherent-state encoding is capacity achieving.

Line-of-sight optical propagation between a finite-area transmitter exit pupil and a finite-area receiver entrance pupil constitutes a lossy Bosonic channel. The vacuum propagation case—whose fundamental propagation loss is due to diffraction—may approach the pure-loss (minimum noise-injection) channel. With atmospheric propagation there will be additional losses arising from absorption and scattering [26], as well as time-dependent fading resulting from atmospheric turbulence [27]. In addition, appreciable extraneous (background) light collection may occur in this case.

This paper addresses the ultimate limits on free-space optical communications by determining the capacity of the multiple spatial mode, wideband, pure-loss channel. Realistic background spectral radiance values are then used to obtain tight bounds on the capacity of single-spatial-mode, narrowband 1.55- μm -wavelength free-space communications in the presence of background light. In both cases—the wideband pure-loss channel, and the narrowband 1.55- μm -wavelength channel

with background light—we show that channel capacity is achieved with single-use coherent-state encoding plus joint measurements over entire codewords, and we quantify the capacity lost when coherent (homodyne or heterodyne) detection is used in lieu of optimum reception. We begin our development with a brief review of prior classical capacity results for lossy Bosonic channels.

2. LOSSY BOSONIC CHANNELS

Although we are interested in classical communication over multi-mode lossy Bosonic channels, it is convenient to begin with a treatment at the single-mode level. In this case the channel input is an electromagnetic-field mode with annihilation operator \hat{a} , and its output is another field mode with annihilation operator \hat{a}' . The descriptions of this channel when multiple temporal and/or spatial modes are employed can be built up from tensor-product constructions using the single-mode model. Neither the single-mode nor the multi-mode lossy channels constitute unitary evolutions, so they are governed by trace-preserving completely-positive (TPCP) maps [28] that relate their output density operators, $\hat{\rho}'$, to their input density operators, $\hat{\rho}$.

The TPCP map $\mathcal{E}_\eta^N(\cdot)$ for the single-mode lossy channel can be derived from the commutator-preserving beam splitter relation

$$\hat{a}' = \sqrt{\eta}\hat{a} + \sqrt{1-\eta}\hat{b}, \quad (1)$$

in which the annihilation operator \hat{b} is associated with an environmental (noise) mode, and $0 < \eta < 1$ is the channel transmissivity. For the pure-loss channel, the \hat{b} mode is in its vacuum state; for the thermal-noise channel this mode is in a thermal state, viz., an isotropic-Gaussian mixture of coherent states with average photon number $N > 0$,

$$\hat{\rho}_b = \int d^2\beta \frac{\exp(-|\beta|^2/N)}{\pi N} |\beta\rangle\langle\beta|. \quad (2)$$

The classical capacity of the single-mode lossy channel is established by random coding arguments akin to those employed in classical information theory. A set of symbols $\{j\}$ is represented by a collection of input states $\{\hat{\rho}_j\}$ that are selected according to some prior distribution $\{p_j\}$. The output states $\{\hat{\rho}'_j\}$ are obtained by applying the channel's TPCP map $\mathcal{E}_\eta^N(\cdot)$ to these input symbols. The Holevo information associated with priors $\{p_j\}$ and states $\{\hat{\sigma}_j\}$ is given by,

$$\chi(p_j, \hat{\sigma}_j) = S\left(\sum_j p_j \hat{\sigma}_j\right) - \sum_j p_j S(\hat{\sigma}_j), \quad (3)$$

where $S(\hat{\sigma}) \equiv -\text{tr}(\hat{\sigma} \ln(\hat{\sigma}))$ is the von Neumann entropy. According to the Holevo-Schumacher-Westmoreland theorem [19]–[21], the capacity of this channel, in nats per use, is

$$C = \sup_n (C_n/n) = \sup_n \left\{ \max_{\{p_j, \hat{\rho}_j\}} [\chi(p_j, (\mathcal{E}_\eta^N)^{\otimes n}(\hat{\rho}_j))/n] \right\}, \quad (4)$$

where C_n is the capacity achieved when coding is performed over n -channel-use symbols and the supremum over n is necessitated by the fact that channel capacity may be superadditive.

We have previously shown that the capacity of the single-mode, pure-loss channel whose transmitter is constrained to use no more than \bar{N} photons on average is [22]

$$C = g(\eta\bar{N}) \text{ nats/use}, \quad (5)$$

where

$$g(x) \equiv (x + 1) \ln(x + 1) - x \ln(x) \quad (6)$$

is the Shannon entropy of the Bose-Einstein probability distribution. This capacity is achieved by single-use random coding over coherent states using an isotropic Gaussian distribution which saturates the transmitter's bound on average photon number. [Note that the optimality of single-use encoding means that the capacity of the single-mode pure-loss channel is *not* superadditive.] This capacity exceeds what is achievable with homodyne and heterodyne detection,

$$C_{\text{hom}} = \frac{1}{2} \ln(1 + 4\eta\bar{N}) \quad \text{and} \quad C_{\text{het}} = \ln(1 + \eta\bar{N}), \quad (7)$$

respectively, although heterodyne detection is asymptotically optimal as $\bar{N} \rightarrow \infty$. An analytical expression for the direct-detection capacity corresponding to this single-mode case is not known, but this capacity has been shown to satisfy [29],

$$C_{\text{dir}} \leq \frac{1}{2} \ln(\eta\bar{N}) + o(1) \quad \text{and} \quad \lim_{\bar{N} \rightarrow \infty} (C_{\text{dir}}) = \frac{1}{2} \ln(\eta\bar{N}), \quad (8)$$

and so is dominated by (5) for $\ln(\eta\bar{N}) > 1$.

For the pure-loss scalar channel in which the transmitter may use all frequencies $\omega \in [0, \infty)$ of a single electromagnetic polarization subject to an average power constraint P with all frequencies having the same channel transmissivity, we have shown that the resulting channel capacity is [22]

$$C_{\text{WB}} = \sqrt{\frac{\pi\eta P}{3\hbar}} \text{ nats/sec}, \quad (9)$$

which is $\pi/\sqrt{3}$ times higher than what can be achieved with homodyne or heterodyne detection. Once again, single-use encoding over a coherent-state ensemble is employed, with low frequencies being used preferentially because of the average power constraint. As yet, there is no corresponding wideband capacity result for direct detection, because existing results [15, 17] ignore the frequency dependence of photon energy by constraining photon flux rather than power.

For the thermal-noise channel, i.e., the lossy Bosonic channel with isotropic-Gaussian excess noise, we have obtained bounds on the channel capacity. For the sake of brevity, we will restrict our discussion to the single-mode case. A lower bound on the single-mode capacity for this channel is easily obtained. We assume coherent-state encoding over single channel uses with an isotropic Gaussian prior distribution. It then follows that

$$C \geq g(\eta\bar{N} + (1 - \eta)N) - g((1 - \eta)N). \quad (10)$$

We believe that this single-use coherent-state encoding with an isotropic Gaussian prior achieves channel capacity for the thermal-noise channel, i.e., we believe that the right-hand side of Eq. (10) gives the capacity of this channel [23, 24]. Because of the following upper bound on the single-mode channel capacity

$$C_n/n \leq \max_{\{p_j, \hat{\rho}_j\}} (S(\hat{\rho}')/n) - \min_{\hat{\rho}_j} (S(\hat{\rho}'_j)/n), \quad \text{where } \hat{\rho}' \equiv \sum_j p_j S(\hat{\rho}'_j) \quad (11)$$

$$= g(\eta\bar{N} + (1 - \eta)N) - \min_{\hat{\rho}_j} (S(\hat{\rho}'_j)/n), \quad (12)$$

the proof of our conjecture is intimately related to the problem of determining the minimum von Neumann entropy that can be realized at the output of the thermal-noise channel by choice of its input state [25]. So far, among many other things, we have shown that a coherent-state input leads to a *local* minimum in the output entropy, and we have shown that a coherent-state input minimizes the integer-order Rényi output entropies [30, 31]. A proof of our capacity conjecture would follow immediately from the latter result were a rigorous foundation available for the replica method of statistical mechanics. [The replica method has recently been applied to other problems in communication theory [32, 33], so establishing its rigorous basis would have additional import outside of statistical physics and Bosonic-channel communications.] Further support for our output entropy and capacity conjectures comes from the suite of lower bounds that we have obtained on the thermal-noise channel’s single-use output entropy [25]. These bounds provide fairly tight constraints on any possible gap between the channel’s minimum output entropy and the associated coherent-state upper bound on this quantity. Indeed these results imply that coherent-state encoding approaches the C_1 capacity at both low and high noise levels. We have also developed numerical results that favor an even stronger conjecture, viz., that the output states resulting from coherent-state inputs to the thermal-noise channel majorize the output states arising from all other inputs [25]. This majorization conjecture, if true, would immediately imply both the minimum output entropy and the capacity conjectures for the thermal-noise channel.

3. MULTIPLE-SPATIAL-MODE, PURE-LOSS, FREE-SPACE CHANNEL

Although it serves a useful illustrative purpose, the wideband pure-loss channel with frequency-independent loss is not a realistic scenario. Thus we have also studied the far-field, scalar free-space channel in which line-of-sight propagation of a single polarization occurs over an L -m-long path from a circular transmitter pupil (area A_t) to a circular receiver pupil (area A_r) with the transmitter restricted to use frequencies $\{\omega : 0 \leq \omega \leq \omega_c \ll \omega_0 \equiv 2\pi cL/\sqrt{A_t A_r}\}$. This frequency range is the far-field power transfer regime, wherein there is only a single spatial mode that couples appreciable power from the transmitter pupil to the receiver pupil, and its transmissivity at frequency ω is $\eta(\omega) = (\omega/\omega_0)^2 \ll 1$. Figure 1 shows the geometry, the power allocations versus frequency for heterodyne, homodyne, and optimal reception, and their corresponding capacities versus normalized power, $P_0 \equiv 2\pi\hbar c^2 L^2/A_t A_r$, when only this dominant spatial mode is employed [22]. Because far-field, free-space transmissivity increases as ω^2 , high frequencies are used preferentially for this channel—unlike the case for frequency-independent loss—because the transmissivity advantage of high-frequency photons more than compensates for their higher energy consumption.

We have also explored the near-field behavior of the pure-loss free-space channel [23], by employing the full prolate-spheroidal wave function normal-mode decomposition associated with the propagation geometry shown in Fig. 1(a) [34, 35]. Near-field propagation at frequency $\omega = 2\pi c/\lambda$ prevails when $D_f = A_t A_r/(\lambda L)^2$, the product of the transmitter and receiver Fresnel numbers, is much greater than unity. In this case there are approximately D_f spatial modes with near-unity transmissivities, with all other modes affording insignificant power transfer from the transmitter pupil to the receiver pupil. In what follows we shall take another approach to the wideband capacity of the pure-loss free-space channel, by employing either the Hermite-Gaussian (HG) or Laguerre-Gaussian (LG) mode sets that are associated with the soft-aperture (Gaussian-attenuation pupil) version of the Fig. 1(a) propagation geometry. Two benefits will be derived from this approach.

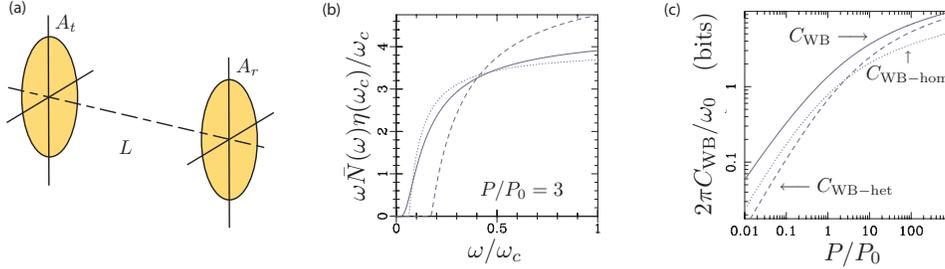


Fig. 1. Capacity results for the far-field, free-space, pure-loss channel: (a) propagation geometry; (b) capacity-achieving power allocations $\hbar\omega\bar{N}(\omega)$ versus frequency ω for heterodyne (dashed curve), homodyne (dotted curve), and optimal reception (solid curve), with ω_c and $\hbar\omega_c/\eta(\omega_c)$ being used to normalize the frequency and the power-spectra axes, respectively; and (c) wideband capacities of optimal, homodyne, and heterodyne reception versus transmitter power P , with $P_0 \equiv 2\pi\hbar c^2 L^2/A_t A_r$ used for the reference power.

First, closed-form expressions become available for the modal transmissivities, as opposed to the hard-aperture case [Fig. 1(a)], for which numerical evaluations or analytical approximations must be employed. Second, the LG modes have been the subject of a great deal of interest, in the quantum optics and quantum information communities [36], owing to their carrying orbital angular momentum. Thus it is germane to explore whether they confer any special advantage in regards to classical information transmission. As we shall see, in the next subsection, the modal transmissivities of the LG modes are isomorphic to those of the HG modes. Inasmuch as the latter do not convey orbital angular momentum, it is clear that such conveyance is not essential to capacity-achieving classical communication over the pure-loss free-space channel.

3.A. Propagation Model: Hermite-Gaussian and Laguerre-Gaussian Mode Sets

In lieu of the hard-aperture propagation geometry from Fig. 1(a), wherein the transmitter and receiver pupils are perfectly transmitting apertures within otherwise opaque planar screens, we now introduce the soft-aperture propagation geometry of Fig. 2. From the quantum version of scalar Fresnel diffraction theory [37], we know that it is sufficient, insofar as this propagation geometry is concerned, to identify a complete set of monochromatic spatial modes—for a single electromagnetic polarization of frequency $\omega = 2\pi c/\lambda = ck$ —that maintain their orthogonality when transmitted through this channel. The resulting two mode sets—i.e., the mode functions at the input and output of the Fig. 2 propagation geometry—constitute a singular-value decomposition (SVD) of the linear propagation kernel (spatial impulse response) associated with this geometry, which we will now develop.

Let $u_i(\vec{x})$, for \vec{x} a 2D vector in the transmitter's exit-pupil plane, denote a frequency- ω field entering the transmitter pupil that is normalized to satisfy

$$\int d^2\vec{x} |u_i(\vec{x})|^2 = 1. \quad (13)$$

The resulting field that leaves the transmitter pupil is taken to be

$$u_T(\vec{x}) = \exp(-|\vec{x}|^2/r_T^2)u_i(\vec{x}), \quad (14)$$

which represents a soft-aperture (Gaussian-attenuation function) spatial truncation. After free-space Fresnel diffraction over an L -m-long path, $u_T(\vec{x})$ produces a field

$$u_R(\vec{x}') = \int d^2\vec{x} u_T(\vec{x}) \frac{\exp(ikL + ik|\vec{x} - \vec{x}'|^2/2L)}{i\lambda L}, \quad (15)$$

in the receiver's entrance-pupil plane, where \vec{x}' is a 2D vector in that plane. The receiver employs a soft-aperture (Gaussian-attenuation function) entrance pupil, so that the field immediately after this pupil is

$$u_o(\vec{x}') = \exp(-|\vec{x}'|^2/r_R^2)u_R(\vec{x}'). \quad (16)$$

Thus, the input-output ($u_i(\vec{x})$ -to- $u_o(\vec{x}')$) relation for the Fig. 2 channel is

$$u_o(\vec{x}') = \int d^2\vec{x} u_i(\vec{x})h(\vec{x}', \vec{x}), \quad (17)$$

where

$$h(\vec{x}', \vec{x}) \equiv \exp(-|\vec{x}'|^2/r_R^2) \frac{\exp(ikL + ik|\vec{x} - \vec{x}'|^2/2L)}{i\lambda L} \exp(-|\vec{x}|^2/r_T^2), \quad (18)$$

is the channel's spatial impulse response.

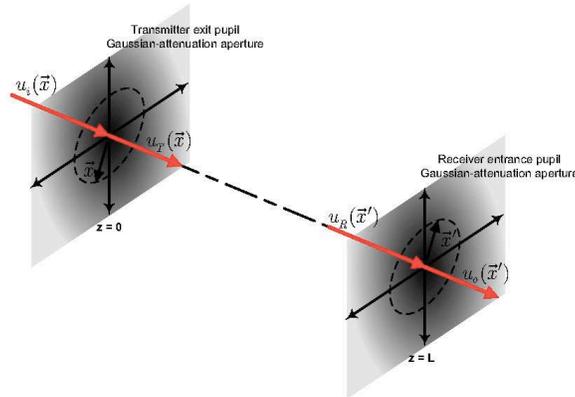


Fig. 2. Propagation geometry with soft apertures.

The singular-value (normal-mode) decomposition of $h(\vec{x}', \vec{x})$ is

$$h(\vec{x}', \vec{x}) = \sum_{m=1}^{\infty} \sqrt{\eta_m} \phi_m(\vec{x}') \Phi_m^*(\vec{x}), \quad (19)$$

where

$$1 \geq \eta_1 \geq \eta_2 \geq \eta_3 \geq \dots \geq 0, \quad (20)$$

are the modal transmissivities, $\{\Phi_m(\vec{x})\}$ is a complete orthonormal (CON) set of functions (input modes) on the transmitter's exit-pupil plane, and $\{\phi_m(\vec{x}')\}$ is a CON set of functions (output modes) on the receiver's entrance-pupil plane. Physically, this decomposition implies that $h(\vec{x}', \vec{x})$ can be separated into a countably-infinite set of parallel channels in which transmission of $u_i(\vec{x}) = \Phi_m(\vec{x})$ results in reception of $u_o(\vec{x}') = \sqrt{\eta_m} \phi_m(\vec{x}')$. Singular-value decompositions are unique

if their $\{\eta_m\}$ are distinct. When degeneracies exist—i.e., when there are multiple modes with the same η_m value—the SVD is not unique. In particular, a linear combination of input modes with the same η_m value produces $\sqrt{\eta_m}$ times that same linear combination of the associated output modes after propagation through $h(\vec{x}', \vec{x})$. As we shall soon see, owing to singular-value degeneracies, the HG and LG modes of the soft-aperture free-space channel are equivalent mode sets.

The spatial impulse response $h(\vec{x}', \vec{x})$ has both rectangular and cylindrical symmetries. The Hermite-Gaussian modes provide an SVD of this channel that has rectangular symmetry. With $\vec{x} = (x, y)$ in Cartesian coordinates, the HG input modes are as follows:

$$\begin{aligned} \Phi_{n,m}(x, y) &= \frac{\sqrt{2}(1+4D_f)^{1/4}}{r_T \sqrt{\pi n! m! 2^{n+m}}} H_n \left(\frac{\sqrt{2}(1+4D_f)^{1/4}}{r_T} x \right) H_m \left(\frac{\sqrt{2}(1+4D_f)^{1/4}}{r_T} y \right) \\ &\times \exp \left[- \left(\frac{(1+4D_f)^{1/2}}{r_T^2} + i \frac{k}{2L} \right) (x^2 + y^2) \right], \quad \text{for } n, m = 0, 1, 2, \dots, \end{aligned} \quad (21)$$

where $H_p(\cdot)$ is the p th Hermite polynomial and

$$D_f = \frac{kr_T^2}{4L} \frac{kr_R^2}{4L} \quad (22)$$

is the product of the transmitter-pupil and receiver-pupil Fresnel numbers for this soft-aperture configuration. The modal transmissivities for the HG modes are

$$\eta_{n,m} = \left(\frac{1+2D_f - \sqrt{1+4D_f}}{2D_f} \right)^{n+m+1}, \quad (23)$$

and the HG output modes are

$$\begin{aligned} \phi_{n,m}(x', y') &= \frac{\sqrt{2}(1+4D_f)^{1/4}}{i^{n+m+1} r_R \sqrt{\pi n! m! 2^{n+m}}} H_n \left(\frac{\sqrt{2}(1+4D_f)^{1/4}}{r_R} x' \right) \\ &\times H_m \left(\frac{\sqrt{2}(1+4D_f)^{1/4}}{r_R} y' \right) \exp \left[- \left(\frac{(1+4D_f)^{1/2}}{r_R^2} - i \frac{k}{2L} \right) (x'^2 + y'^2) \right], \\ &\text{for } n, m = 0, 1, 2, \dots, \end{aligned} \quad (24)$$

where $\vec{x}' = (x', y')$. Because channel capacity depends only on the modal transmissivities, it is worth noting that

$$D_f = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \eta_{n,m} = \int d^2 \vec{x}' \int d^2 \vec{x} |h(\vec{x}', \vec{x})|^2, \quad (25)$$

where the second equality is a consequence of (19) and the first equality can be obtained either by summing the series or evaluating the double integral. Far-field power transfer occurs when $D_f \ll 1$, in which case $\eta_{0,0} \approx D_f$ and all the other modal transmissivities are insignificantly small in comparison. Near-field power transfer occurs when $D_f \gg 1$, in which case there are many modes that couple appreciable power from the transmitter pupil to the receiver pupil. However, because the HG mode decomposition presumes soft-aperture pupils, the near-field modal transmissivities do not have the abrupt near-unity to near-zero transition that occurs for the hard-aperture singular values.

The HG modes' singular values have degeneracies, i.e., there are q HG modes whose modal transmissivities equal $\eta_{0,0}^q$, hence the HG-mode SVD of $h(\vec{x}', \vec{x})$ is not unique. The Laguerre-Gaussian modes provide an alternative SVD for this channel, one with cylindrical rather than rectangular symmetry. Using the polar coordinates $\vec{x} = (r, \theta)$ we have that the LG input modes are

$$\begin{aligned} \Phi_{p,\ell}(r, \theta) &= \sqrt{\frac{2p!}{\pi(|\ell|+p)!}} \frac{(1+4D_f)^{1/4}}{r_T} \left[\frac{\sqrt{2}(1+4D_f)^{1/4}}{r_T} r \right]^{|\ell|} \\ &\times L_p^{|\ell|} \left(\frac{2(1+4D_f)^{1/2}}{r_T^2} r^2 \right) \exp \left[- \left(\frac{(1+4D_f)^{1/2}}{r_T^2} + i \frac{k}{2L} \right) r^2 + i\ell\theta \right], \\ &\text{for } p = 0, 1, 2, \dots, \text{ and } \ell = 0 \pm 1, \pm 2, \dots, \end{aligned} \quad (26)$$

where $L_p^{|\ell|}(\cdot)$ is the p | ℓ |th generalized Laguerre polynomial. The corresponding modal transmissivities are given by

$$\eta_{p,\ell} = \left(\frac{1+2D_f - \sqrt{1+4D_f}}{2D_f} \right)^{(2p+|\ell|+1)}, \quad (27)$$

from which it can be seen that the HG modes with $n+m+1=q$ span the same eigenspace as the LG modes with $2p+|\ell|+1=q$, and hence are related by a unitary transformation. The LG output modes are

$$\begin{aligned} \phi_{p,\ell}(r', \theta') &= \sqrt{\frac{2p!}{\pi(|\ell|+p)!}} \frac{(1+4D_f)^{1/4}}{i^{2p+|\ell|+1} r_R} \left[\frac{\sqrt{2}(1+4D_f)^{1/4}}{r_R} r' \right]^{|\ell|} \\ &\times L_p^{|\ell|} \left(\frac{2(1+4D_f)^{1/2}}{r_R^2} r'^2 \right) \exp \left[- \left(\frac{(1+4D_f)^{1/2}}{r_R^2} - i \frac{k}{2L} \right) r'^2 + i\ell\theta' \right], \\ &\text{for } p = 0, 1, 2, \dots, \text{ and } \ell = 0 \pm 1, \pm 2, \dots, \end{aligned} \quad (28)$$

where $\vec{x}' = (r', \theta')$. As was the case for the HG modes, channel capacity when LG modes are employed depends only on the modal transmissivities.

A single frequency- ω photon in the LG mode $\Phi_{p,\ell}(r, \theta)$ carries orbital angular momentum $\hbar\ell$ directed along the propagation (z) axis, whereas that same photon in the HG mode $\Phi_{n,m}(x, y)$ carries no z -directed orbital angular momentum. The equivalence of the $\{\eta_{p,\ell}\}$ and the $\{\eta_{n,m}\}$ then implies that angular momentum does not play a role in determining the ultimate—channel capacity—limit on classical information transmission over the free-space channel shown in Fig. 2.

3.B. Wideband Capacities with Multiple Spatial Modes

Here we shall address the wideband capacities that can be achieved over the pure-loss, scalar free-space channel shown in Fig. 2 using either heterodyne detection, homodyne detection, or optimum (joint measurement over entire codewords) reception. We will allow the transmitter to use multiple spatial modes—from either the HG or LG mode sets—and all frequencies $\omega \in [0, \infty)$ subject to a constraint, P , on the average power in the field entering the transmitter's exit pupil. It follows from our prior work [22, 23] that the capacities we are seeking satisfy,

$$C(P) = \max_{\bar{N}_q(\omega)} \sum_{q=1}^{\infty} q \int_0^{\infty} \frac{d\omega}{2\pi} C_{\text{SM}}(\eta(\omega)^q, \bar{N}_q(\omega)), \quad (29)$$

where the maximization is subject to the average power constraint,

$$P = \sum_{q=1}^{\infty} q \int_0^{\infty} \frac{d\omega}{2\pi} \hbar\omega \bar{N}_q(\omega), \quad (30)$$

and

$$\eta(\omega)^q \equiv \left(\frac{1 + 2(\omega/\omega_0)^2 - \sqrt{1 + 4(\omega/\omega_0)^2}}{2(\omega/\omega_0)^2} \right)^q \quad (31)$$

is the modal transmissivity at frequency ω with q -fold degeneracy, with $\omega_0 = 4cL/r_t r_R$ being the frequency at which $D_f = 1$. In (29),

$$C_{\text{SM}}(\eta, \bar{N}) \equiv \begin{cases} g(\eta\bar{N}), & \text{for optimum reception} \\ \ln(1 + \eta\bar{N}), & \text{for heterodyne detection} \\ \frac{1}{2} \ln(1 + 4\eta\bar{N}), & \text{for homodyne detection} \end{cases} \quad (32)$$

are the relevant single-mode capacities as functions of the modal transmissivity, η , and the average photon number, \bar{N} , for that mode. Regardless of the frequency dependence of $\eta(\omega)$ the single-mode capacity formulas for heterodyne and homodyne detection imply that their wideband multiple-spatial-mode capacities bear the following relationship,

$$C_{\text{hom}}(P) = \frac{1}{2} C_{\text{het}}(4P). \quad (33)$$

Thus, only two maximizations need to be performed—both of which can be done via Lagrange multipliers—to obtain the wideband multiple-spatial-mode capacities for optimum reception, heterodyne detection, and homodyne detection.

The results we have obtained by performing the preceding maximizations are as follows. The optimum-reception capacity (in nats/sec) and its associated optimum modal-power spectra are given by

$$C(P) = \frac{P}{\hbar\omega_0\sigma} - \sum_{q=1}^{\infty} q \int_0^{\infty} \frac{d\omega}{2\pi} \ln[1 - \exp(-\omega/\omega_0\eta(\omega)^q\sigma)], \quad (34)$$

and

$$\hbar\omega \bar{N}_q(\omega) = \frac{\hbar\omega/\eta(\omega)^q}{\exp(\omega/\omega_0\eta(\omega)^q\sigma) - 1}, \quad (35)$$

respectively, where σ is a Lagrange multiplier chosen to enforce the average power constraint. The corresponding capacity and optimum modal-power spectra for heterodyne detection are

$$C_{\text{het}}(P) = \sum_{q=1}^{\infty} q \int \frac{d\omega}{2\pi} \ln(\beta\omega_0\eta(\omega)^q/\omega), \quad (36)$$

and

$$\hbar\omega \bar{N}_q(\omega) = \max[\hbar\omega_0(\beta - \omega/\omega_0\eta(\omega)^q), 0], \quad (37)$$

where β is another Lagrange multiplier, again chosen to enforce the average power constraint. We will omit the homodyne detection formulas, as they can be obtained via scaling the heterodyne formulas according to (33), and instead focus our attention on the behavior of the capacity and power spectrum results.

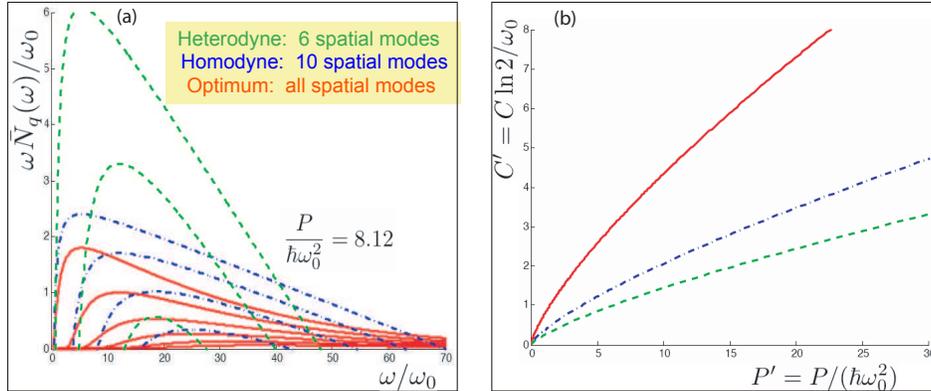


Fig. 3. (a) Capacity-achieving power spectra for wideband, multiple-spatial-mode communication over the scalar, pure-loss, free-space channel when $P = 8.12\hbar\omega_0^2$. Heterodyne detection (dashed curves) uses 6 spatial modes (from top to bottom, $1 \leq q \leq 3$), homodyne detection (dot-dashed curves) uses 10 spatial modes (from top to bottom, $1 \leq q \leq 4$) and optimum reception (solid curves) uses all spatial modes (although only top to bottom $1 \leq q \leq 6$ are shown). (b) Wideband, multiple-spatial mode capacities for the scalar, pure-loss, free-space channel that are realized with optimum reception (solid curve), homodyne detection (dot-dashed curve), and heterodyne detection (dashed curve). The capacities, in bits/sec, are normalized by $\omega_0 = 4cL/r_T r_R$, the frequency at which $D_f = 1$, and plotted versus the average transmitter power normalized by $\hbar\omega_0^2$.

The capacity-achieving power spectrum for optimal reception employs all spatial modes and all frequencies, whereas the capacity-achieving power spectra for heterodyne and homodyne detection are “water-filling” allocations, viz., they fill spatial-mode/frequency volumes above their appropriate noise-to-transmissivity-ratio contours until the average power constraint is met. That water-filling power allocation should be capacity achieving for these coherent detection cases is hardly a surprise, as water-filling power allocation has long been known to be optimal for additive Gaussian noise channels [38]. A consequence of water-filling power allocation is that heterodyne and homodyne detection only employ a finite number of spatial modes to achieve their respective capacities, whereas optimal-reception capacity needs all spatial modes. This behavior is illustrated in Fig. 3(a), where we have plotted the capacity-achieving power spectra for homodyne detection, heterodyne detection, and optimal reception when $P = 8.12\hbar\omega_0^2$. In this case, heterodyne detection uses $1 \leq q \leq 3$ (a total of 6 spatial modes) with non-zero power, and homodyne detection uses $1 \leq q \leq 4$ (a total of 10 spatial modes) with non-zero power. Optimum reception uses all spatial modes, but we have only plotted the spectra for $1 \leq q \leq 6$.

In Fig. 3(b) we have plotted the heterodyne detection, homodyne detection, and optimum reception capacities in bits/sec, normalized by ω_0 , versus the normalized power, $P/\hbar\omega_0^2$. Unlike the case seen in Fig. 1(c) for the wideband capacities of the single-spatial-mode, far-field pure-loss channel—in which heterodyne detection outperforms homodyne detection at high power levels—Fig. 3(b) shows that homodyne detection is consistently better than heterodyne detection for the multiple-spatial-mode scenario. This behavior has a simple physical explanation. Consider first the single-spatial mode wideband capacities. At low power levels, when capacity is power limited, homodyne detection outperforms heterodyne detection because

at every frequency it suffers less noise. On the other hand, at high enough power levels single-spatial mode communication becomes bandwidth limited. In this case heterodyne detection's factor-of-two bandwidth advantage over homodyne detection carries the day. Things are different when multiple spatial modes are available. In this case, increasing power never reaches bandwidth-limited operation; additional, lower transmissivity, spatial modes get employed as the power is increased so that the noise advantage of homodyne detection continues to give a higher channel capacity than does heterodyne detection.

Figure 3 shows that the wideband capacity realized with optimum reception, on the multiple-spatial-mode pure-loss channel, increasingly outstrips that of homodyne detection with increasing transmitter power. This advantage indicates that joint measurements over entire codewords—which are implicit in the Holevo information maximization procedure that leads to the optimum-reception capacity—afford performance that is unapproachable with homodyne detection, which is a single-use quantum measurement.

4. SINGLE SPATIAL-MODE COMMUNICATION AT 1.55 μm

The wideband results from the previous section set ultimate limits on free-space optical communications in the absence of atmospheric effects. In this final section, we consider a more constrained situation that will better approximate what can be accomplished over an atmospheric path. Specifically, we consider single-spatial-mode propagation at the fiber-compatible 1.55- μm -wavelength in the presence of atmospheric extinction and background light. The propagation geometry will be the soft-aperture configuration shown in Fig. 2, but the spatial-impulse response at the transmitter's center frequency ω will now be taken to be

$$h(\vec{x}', \vec{x}) = \exp(-|\vec{x}'|^2/r_R^2) \frac{\exp(-\alpha L/2 + ikL + ik|\vec{x} - \vec{x}'|^2/2L)}{i\lambda L} \exp(-|\vec{x}|^2/r_T^2), \quad (38)$$

where α is the atmospheric extinction (absorption plus scattering) coefficient. The transmitter will be assumed to be narrowband, i.e., its radian-frequency bandwidth, Ω , is small enough that the energy of all transmitted photons is approximately $\hbar\omega$. We shall assume that the transmitter uses the maximum-power-transfer HG input mode (which is identical to the maximum-power-transfer LG input mode) and that the receiver (whether it uses homodyne, heterodyne, or optimum reception) measures the corresponding output mode. Because of the combination of diffraction loss and extinction loss, the associated transmissivity for this atmospheric channel is

$$\eta_a = \left(\frac{1 + 2(\omega/\omega_0)^2 - \sqrt{1 + 4(\omega/\omega_0)^2}}{2(\omega/\omega_0)^2} \right) e^{-\alpha L}, \quad (39)$$

which, by our narrowband assumption, is constant over the frequencies used by the transmitter.

Accompanying the atmospheric propagation loss is the collection of extraneous (background) light. For each temporal mode arriving at the receiver, this background light is an isotropic mixture of coherent states with average photon number [39]

$$N = \pi 10^6 \lambda^3 N_\lambda / \hbar\omega^2, \quad (40)$$

where N_λ is the background spectral radiance (in $\text{W}/\text{m}^2 \text{SR} \mu\text{m}$). A typical daytime value $N_\lambda \sim 10 \text{ W}/\text{m}^2 \text{SR} \mu\text{m}$ at 1.55 μm then leads to $N \sim 10^{-6}$; nighttime N_λ

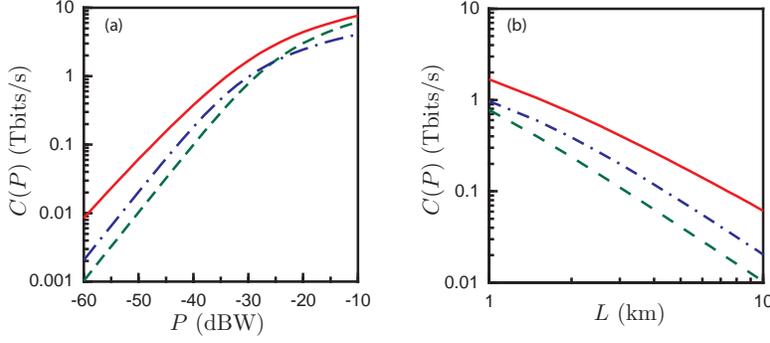


Fig. 4. Capacities of the single-spatial-mode, narrowband, 1.55- μm -wavelength channel for optimum reception (solid curves), homodyne detection (dot-dashed curves) and heterodyne detection (dashed curves) plotted in (a) versus average transmitter power and in (b) versus path length. Both (a) and (b) assume $\Omega/2\pi = 1$ THz, $\alpha = 0.5$ dB/km, $r_T = 1$ mm, and $r_R = 1$ cm; (a) assumes $L = 1$ km, and (b) assumes $P = 1$ mW.

values are several orders of magnitude lower [39]. Because the single-mode, thermal-noise channel \mathcal{E}_η^N is the concatenation of the pure-loss channel \mathcal{E}_η^0 with a classical-noise channel of average noise-photon number $(1 - \eta)N$ [25], it follows that the capacity of a thermal-noise channel can never exceed that of the pure-loss channel with the same transmissivity. As result, the optimal-reception capacity of a single temporal mode of our atmospheric channel satisfies

$$g(\eta_a \bar{N} + (1 - \eta_a)N) - g((1 - \eta_a)N) \leq C_{\text{SM}}(\eta_a, \bar{N}) \leq g(\eta_a \bar{N}), \quad (41)$$

and these single-mode-capacity upper and lower bounds are virtually coincident for $\eta_a \bar{N} \geq 10(1 - \eta_a)N$. For the narrowband, multiple-temporal-mode case, we then have

$$\frac{\Omega}{2\pi} [g(\eta_a 2\pi P / \hbar\omega\Omega + (1 - \eta_a)N) - g((1 - \eta_a)N)] \leq C(P) \leq \frac{\Omega}{2\pi} g(\eta_a 2\pi P / \hbar\omega\Omega). \quad (42)$$

Thus, as long as $\eta_a 2\pi P / \hbar\omega\Omega \geq 10(1 - \eta_a)N$, these bounds are exceedingly tight, and background light can be neglected in determining the capacity achieved with optimum reception. The impact of the background light on the multiple-temporal-mode channel capacities of homodyne and heterodyne detection is also negligible, because these capacities are given by

$$C(P) = \begin{cases} \frac{\Omega}{4\pi} \ln \left(1 + \frac{\eta_a 8\pi P / \hbar\omega\Omega}{1 + 2(1 - \eta_a)N} \right) & \text{for homodyne detection} \\ \frac{\Omega}{2\pi} \ln \left(1 + \frac{\eta_a 2\pi P / \hbar\omega\Omega}{1 + (1 - \eta_a)N} \right) & \text{for heterodyne detection,} \end{cases} \quad (43)$$

and $N \ll 1$.

In Fig. 4 we have plotted the single-spatial-mode, narrowband, 1.55- μm -wavelength capacities versus power [Fig. 4(a)] and path length [Fig. 4(b)]. Both of these plots assume $\Omega/2\pi = 1$ THz, $\alpha = 0.5$ dB/km, $r_T = 1$ mm, and $r_R = 1$ cm. Figure 4(a) assumes $L = 1$ km, and Fig. 4(b) assumes $P = 1$ mW. These numbers represent clear-weather propagation—neglecting atmospheric turbulence—with a

~ 1 mR transmitter-beam divergence, and transmitter powers commensurate with semiconductor-laser/power-amplifier technology. The curves show that optimum reception increasingly outstrips homodyne and heterodyne detection at lower transmitter powers and longer path lengths. The crossing of the heterodyne and homodyne capacity-versus-power curves is the same power-limited versus bandwidth-limited behavior we discussed in conjunction with single-spatial-mode, wideband, far-field communication, as is the asymptotic approach of the heterodyne capacity to the optimum-reception capacity with increasing transmitter power, cf. Fig. 1(c).

5. DISCUSSION

We have drawn upon the quantum description of Bosonic communication to establish ultimate limits on classical communications at optical frequencies. Although most of our results have addressed vacuum propagation, we showed how extinction and background noise can be incorporated into a single-spatial-mode capacity analysis. The principal propagation effect omitted from our development is the time-dependent fading incurred as a result of atmospheric turbulence. Some results are available on the classical capacity of the turbulent channel, see, e.g., [40], [41], but these studies have employed the semiclassical approach. Thus, a full quantum treatment is still needed.

6. ACKNOWLEDGMENTS

This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-00-1-0177.

References and Links

- [1] C. E. Shannon, "The mathematical theory of communication," *Bell System Tech. J.* **27**, 379–423, 623–656 (1948).
- [2] R. G. Gallager, *Information Theory and Reliable Communication* (Wiley, New York 1968) Chaps. 4, 5, 8.
- [3] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991) Chaps. 8, 10.
- [4] W. H. Louisell, *Quantum Statistical Properties of Radiation* (Wiley, New York, 1973) Sects. 4.3–4.4.
- [5] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge University Press, Cambridge, 1995) Sects. 10.1–10.3.
- [6] R. H. Kingston, *Detection of Optical and Infrared Radiation* (Springer-Verlag, Berlin, 1978).
- [7] R. M. Gagliardi and S. Karp, *Optical Communications* (Wiley, New York, 1976).
- [8] J. Gowar, *Optical Communication Systems* (Prentice Hall, Englewood Cliffs, 1984).
- [9] H. P. Yuen and J. H. Shapiro, "Optical communication with two-photon coherent states—Part III: quantum measurements realizable with photoemissive detectors," *IEEE Trans. Inform. Theory*, **26**, 78–92 (1980).

- [10] C. W. Helstrom, “Detection theory and quantum mechanics,” *Inform. Contr.* **10**, 254–291 (1964); C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976) Chaps. 4, 6.
- [11] S. J. Dolinar, “An optimum receiver for the binary coherent state quantum channel,” *M.I.T. Res. Lab. Electron. Quart. Prog. Rep.* **111**, 115-120 (1973).
- [12] J. H. Shapiro, H. P. Yuen, and J. A. Machado Mata, “Optical communication with two-photon coherent states—Part II: photoemissive detection and structured receiver performance,” *IEEE Trans. Inform. Theory*, **25**, 179–192 (1979).
- [13] W. H. Louisell, *Quantum Statistical Properties of Radiation* (Wiley, New York, 1973) Chaps. 1, 2.
- [14] J. P. Gordon, “Quantum effects in communications systems,” *Proc. IRE* **50**, 1898–1908 (1962).
- [15] M. H. A. Davis, “Capacity and cutoff rate for Poisson-type channels,” *IEEE Trans. Inform. Theory* **26**, 710–715 (1980).
- [16] J. R. Pierce, E. C. Posner, and E. R. Rodemich, “The capacity of the photon-counting channel,” *IEEE Trans. Inform. Theory* **27**, 61–77 (1981).
- [17] A. D. Wyner, “Capacity and error exponent for the direct detection photon channel—Parts I and II,” *IEEE Trans. Inform. Theory* **34**, 1449–1471 (1988).
- [18] S. Shamai (Shitz) and A. Lapidoth, “Bounds on the capacity of a spectrally-constrained Poisson channel,” *IEEE Trans. Inform. Theory* **39**, 19–29 (1993).
- [19] A. S. Holevo, “The capacity of a quantum channel with general signal states,” *IEEE Trans. Inform. Theory* **44**, 269–273 (1998).
- [20] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters, “Classical information capacity of a quantum channel,” *Phys. Rev. A* **54**, 1869–1876 (1996).
- [21] B. Schumacher and M. D. Westmoreland, “Sending classical information via noisy quantum channels,” *Phys. Rev. A* **56**, 131–138 (1997).
- [22] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, J. H. Shapiro, and H. P. Yuen, “Classical capacity of lossy Bosonic channels: the exact solution,” *Phys. Rev. Lett.* **92**, 027902 (2004).
- [23] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, J. H. Shapiro, B. J. Yen, and H. P. Yuen, “Classical capacity of free-space optical communication,” *Quantum Information and Computation* **4**, 489–499 (2004).
- [24] J. H. Shapiro, V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, and B. J. Yen, “Capacity of Bosonic communications,” in S. M. Barnett, E. Anderson, J. Jeffers, P. Öhberg, and O. Hirota, eds., *Proceedings of the Seventh International Conference on Quantum Communication, Measurement and Computing* (American Institute of Physics, New York, 2004), pp. 15–20.
- [25] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, and J. H. Shapiro, “Minimum output entropy of Bosonic channels: a conjecture,” *Phys. Rev. A* **70**, 032315 (2004).

- [26] S. Karp, R. M. Gagliardi, S. E. Moran, and L. B. Stotts, *Optical Channels: Fibers, Clouds, Water, and the Atmosphere* (Plenum, New York, 1988).
- [27] J. H. Shapiro, “Imaging and optical communication through atmospheric turbulence,” in J. W. Strohbehm, ed., *Laser Beam Propagation in the Atmosphere* (Springer-Verlag, Berlin, 1978) pp. 171–222.
- [28] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000) Sect. 8.2.4.
- [29] A. Lapidoth and S. M. Moser, “Bounds on the capacity of the discrete-time Poisson channel,” Proceedings of the 41st Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, Oct. 2003.
- [30] V. Giovannetti, S. Lloyd, L. Maccone, J. H. Shapiro, and B. J. Yen, “Minimum Rényi and Wehrl entropies at the output of Bosonic channels,” *Phys. Rev. A* **70**, 022328 (2004).
- [31] V. Giovannetti and S. Lloyd, “Additivity properties of a Gaussian channel,” *Phys. Rev. A* **69**, 062307 (2004).
- [32] T. Tanaka, “A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors,” *IEEE Trans. Inform. Theory* **48**, 2888–2910 (2002).
- [33] R. R. Miller, “Channel capacity and minimum probability of error in large dual antenna array systems with binary modulation,” *IEEE Trans. Signal Process.* **51**, 2821–2828 (2003).
- [34] D. Slepian, “Prolate spheroidal wave functions, Fourier analysis and uncertainty—IV: extensions to many dimensions; generalized prolate spheroidal functions,” *Bell. System Tech. J.* **43**, 3009–3057 (1964).
- [35] D. Slepian, “Analytic solution of two apodization problems,” *J. Opt. Soc. Am.* **55**, 1110–1115 (1965).
- [36] L. Allen, S. M. Barnett, and M. J. Padgett, eds., *Optical Angular Momentum* (Institute of Physics, Bristol, 2004).
- [37] H. P. Yuen and J. H. Shapiro, “Optical communication with two-photon coherent states—Part I: quantum state propagation and quantum noise reduction,” *IEEE Trans. Inform. Theory*, **24**, 657–668 (1978).
- [38] R. G. Gallager, *Information Theory and Reliable Communication* (Wiley, New York 1968) Chap. 8.
- [39] N. S. Kopeika and J. Bordogna, “Background noise in optical communication systems,” *Proc. IEEE* **58**, 1571–1577 (1970).
- [40] S. M. Haas and J. H. Shapiro, “Capacity of wireless optical communications,” *IEEE J. Sel. Areas in Commun.* **21**, 1346–1357 (2003).
- [41] B. Erkmen and J. H. Shapiro, “Performance analysis for near-field atmospheric optical communications,” Digest of 2004 IEEE Global Telecommunications Conference, November 2004, Dallas, TX.