Optical wireless communications is realized by modulating the intensity of a light source and detecting intensity fluctuations at the receiver. This mode of operation, known as intensity-modulation and direct-detection (IM/DD), is simple to implement in practice. However, computing the channel capacity of the underlying channel is not straightforward due to the amplitude constraints that arise due to IM/DD operation. In particular, the transmit signal must be non-negative while the peak and average amplitudes are constrained due to practical and safety considerations. Though a closed-form for the capacity of IM/DD channels is not known, much work has been done to find capacity bounds and asymptotic capacity expressions. In this paper, a description of the IM/DD channel and its physical constraints is presented, followed by a review of recent progress pertaining to the capacity of IM/DD channels. Additionally, capacity achieving distributions are discussed along with simple constructions that approach capacity.

1. Introduction

Capacity is the ultimate “speed limit” of information transmission over any communications channel. It tells the system designer what communication rates are possible (i.e., reliable) and which are not. Although the capacity and methods to approach it are known for a wide host of channels (e.g., in particular the additive white Gaussian channel (AWGN)), there exist channels for which the capacity is ill understood.

One such channel is the IM/DD channel which models optical intensity communications. Understanding the capacity of the IM/DD channel became ever more important recently due to the renewed interest in optical wireless communications (OWC) in various forms such
as Light-Fidelity (LiFi), Visible-Light Communications (VLC), Free-Space Optics (FSO), Underwater Optical Communications (UOC), and non-line-of-sight optical communication [1–5]. Motivated by this, the IM/DD channel capacity has been investigated by several research teams recently, and significant advances have been achieved which have greatly contributed to our understanding of this channel.

In this work, we provide the reader with a review of these recent advances. We start with a description of optical wireless channels based on IM/DD. Then, we focus on a common IM/DD channel model which is the input-independent Gaussian channel model, and we review results on the capacity of this channel model, and most importantly for communication engineers, we review the form of distributions which approach capacity. Finally, we discuss the relation of the results with systems based on uni-polar orthogonal frequency-division multiplexing (OFDM), and discuss recent extensions to systems with multiple apertures, i.e., multiple-input multiple-output (MIMO) IM/DD channels.

2. Channel Model and Constraints

**Notation:** Throughout this paper, lower-case font denotes deterministic scalars and upper-case letters denote random variables. The notation $X \sim f_X$ indicates that $X$ is distributed according to a distribution $f_X$ (probability density or mass function), $\mathbb{E}_X[\cdot]$ denotes the expectation with respect to $X$, and $\mathbb{P}\{\cdot\}$ denotes the probability of an event. Calligraphic letters denote sets while the $Q$-function defined as $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$. For a random variable $X$, $H(X)$ denotes its entropy if $X$ is discrete and $h(X)$ denotes its differential entropy if $X$ is continuous. Additionally, for $X \sim f_X$ and $Y \sim f_Y$, $I(X;Y)$ denotes the mutual information between $X$ and $Y$, and $D(f_X \parallel f_Y)$ denotes the relative entropy (Kullback Leibler Divergence) between their distributions.

An optical wireless communication (OWC) system sends data by modulating the intensity of a light source (e.g., a light-emitting diode (LED), laser diode) in discrete time intervals of length $t$ as shown in Fig. 1. Let the sequence of modulating symbols transmitted in each $t$ interval be denoted as $\{x_i\}_{n=1}^\infty$, for some integer $n > 1$. The symbol $x_i$ physically represents the driving electrical current signal which is applied to the light source while transmitting the $i$th symbol which is fixed over a the symbol duration $t$. For both LEDs and lasers, the transmitted light intensity can be written as $\eta_{eo} x_i$ where $\eta_{eo}$ is the electrical-to-optical conversion efficiency of the light sources. Without loss of generality, $\eta_{eo} = 1$ in the balance of the paper.

Given that the light source is able to modulate only the output intensity or optical power of the source, rather than the amplitude and phase of the carrier as in radio systems, $\forall n, x_i \geq 0$. Additionally, for eye-safety and due to the limited dynamic range of the devices, a peak limitation is placed on the emissions from the source, that is $x_i \leq A$. Another unique feature of optical wireless channels is that for visible light communication (VLC) channels, users may dim or alter the average brightness of the source. An average optical power constraint constraint $\mathbb{E}_X [X] \leq E$ models a dimming level in VLC applications.

In the OWC channel, the transmitted signal is attenuated during propagation and it may be impacted an additive by background light. The attenuation is modelled as a linear factor $\alpha$ and arises due to beam divergence and is dominated by absorption and scattering in longer range
free-space optical links. Since this factor is simply a scale, to make the discussion more clear and without loss of generality, \( \alpha = 1 \) for the balance of the paper.

The overall channel model of an OWC system depends on the regime in which the system is operating. When the optical power at the receiver is very low, the discrete nature of the received photons must be considered. Notice that optical intensity or power is measure of photon flux. Thus, the intensity can also be considered as the rate of photon arrival. In this low power regime, the receiver is a photon counter which outputs a statistic depending on the number of received photons in each \( t \) interval. The received signal is composed of both the transmitted optical intensity signal as well as background light where background light is modelled as a source with a constant arrival rate of \( \lambda \) photons per second. Conditioned on the transmitted signal, the detected photon count \( y_i \) in interval \( i \) is conditionally Poisson, i.e.,

\[
P(y_i|x_i) = \frac{(x_i + \lambda)^y_i}{y_i!} e^{x_i + \lambda}, \quad x_i \in \mathbb{R}^+, y_i \in \mathbb{Z}^+
\]

where \( t = 1 \) for convenience and without loss of generality. Notice that in this discrete-time Poisson (DTP) channel the detection process for a given transmitted intensity is inherently stochastic and depends in general both on the transmitted signal as well as the background light. Such DTP links occur often in applications such as long range inter-satellite optical links operating at ranges of many thousands of kilometres [34,35].

**Definition 2.1 (DTP Channel).** Let \( X \) be a random variable representing the transmitted intensity distributed on \([0, A]\) with \( E[X] \leq E \). The received signal is corrupted by a background light of intensity \( \lambda \). In the discrete-time Poisson channel, the channel output, \( Y \), which represents a count of the number of photons, is conditionally Poisson with rate \( x + \lambda \).

When the intensity of the received signal increases, the distribution of the number of received photons in any interval approaches a Gaussian in distribution with mean and variance of \( x_i + \lambda \). In this regime, the channel is termed intensity-modulation with direct-detection (IM/DD). Ignoring the scaling constants as defined above, the received signal can be modelled as

\[
y_i = x_i + z_i,
\]

where \( z_i \) is Gaussian and has zero mean under the assumption that the bias to the mean due to the constant background light is removed. Notice that in general, the variance of the added noise depends on both the background light as well as the signal. This signal-dependent noise is a significant practical impairment when the signal power received at the detector is large with respect to the background light and is often termed the conditionally Gaussian IM/DD channel [6].

**Definition 2.2 (Signal-Dependent IM/DD Channel).** Let \( X \) be a random variable distributed on \([0, A]\) with \( E[X] \leq E \), and \( Z(X) \) be Gaussian noise with zero mean and for \( X = x \) has variance \( \sigma_x^2 = \sigma_0^2 + \sigma_1^2 x \) for constant \( \sigma_0 \) and \( \sigma_1 \). A channel with input \( X \) and output \( Y = X + Z(x) \) is termed a signal-dependent IM/DD channel.

In most cases of practical interest for indoor and terrestrial OWC, the background light greatly dominates the intensity of the received signal intensity. In this case, the noise variance is well modelled as being independent of the transmitted signal giving rise to the following IM/DD channel definition.

**Definition 2.3 (IM/DD Channel).** Let \( X \) be a random variable distributed on \([0, A]\) with \( E[X] \leq E \), and \( Z \) be Gaussian noise with zero mean and constant variance \( \sigma^2 \). A channel with input \( X \) and output \( Y = X + Z \) is termed an IM/DD channel.
Notice that this channel is also appropriate in the case when front-end electronic noise of the receiver dominates background-induced shot noise. Given the great practical significance of the signal-independent IM/DD channel for visible light communication channels, this paper will focus on the channel in Def. 2.3. However, we also present some recent connections between the fundamental DTP channel in Def. 2.1 and IM/DD channels.

3. Channel Capacity

(a) What is Channel Capacity?

The information rate is the amount of information sent over a channel per unit time. The task of the communication system designer is to try to achieve the highest information rate. However, the rate can not be increased without bound due to the limited ability of the receiver to discern amongst different signals distorted by noise. What is this communications rate limit?

This limit is termed the channel capacity. As the name suggests, it is a property of the channel defined by inputs, outputs, and a channel law which relates the output to the input (\(X, Y\), and (2.2), respectively, in the IM/DD channel). It is not a property of the transmission scheme used to send bits over this channel but is rather a property of the channel itself. The channel capacity is the largest number of bits per transmission that can be sent ‘reliably’ over a channel using any transmission scheme that respects the imposed constraints [7].

To transmit information over the IM/DD channel using \(n\) symbols, the transmitter chooses a message \(M\) uniformly at random from amongst a set of messages \(\mathcal{M}_n = \{1, 2, \ldots, \ell_n\}\). Note that each message can be represented using \(\log_2(\ell_n)\) bits. The transmitter encodes the message \(M\) into a sequence \(\{x_i\}_{i=1}^n\) using an encoding function \(\phi_n: \mathcal{M} \mapsto [0, A]^n\) which satisfies all channel constraints, and sends \(\{x_i\}_{i=1}^n\) over the channel.

The receiver receives \(\{y_i\}_{i=1}^n\), and attempts to decode \(M\) using a decoding function \(\psi_n: \mathbb{R} \mapsto \mathcal{M}\). This operation results in a decoded message \(\hat{M}\). The error probability induced by this operation is \(p_n = \mathbb{P}(\hat{M} \neq M)\), and each transmitted message carries \(\log_2(\ell_n)\) bits.

The transmission rate \(R\) (in bits/transmission) is said to be an achievable rate if there exists a sequence of \(\mathcal{M}_n, \phi_n, \psi_n\) so that \(R = \frac{\log_2(\ell_n)}{n}\) and \(p_n \to 0\) as \(n \to \infty\) (which defines the meaning of ‘reliable’ communication). The maximum achievable rate \(R\) is the channel capacity \(C\). Since \(C\) depends on the channel constraints \(A\) and \(E\), we shall denote it \(C_p(A)\) where \(p = \frac{A}{E}\) is the peak-to-average optical power ratio.

The units of channel capacity can be express in bits/transmission or nats/transmission, where \(1\text{nats}=\frac{1}{\log(2)}\text{bits}\). We will use nats/transmission henceforth for convenience.

(b) Memoryless-Channel Capacity

The IM/DD channel belongs to the family of memoryless channels because noise is i.i.d., and hence, its capacity is well defined. In particular, the channel capacity can be written as [7]

\[
C_p(A) = \max_{f_X} I(X; Y),
\]

(3.1)

where \(f_X\) is the distribution of \(X\), and the maximization is over the set of feasible distributions (i.e., those that satisfy the channel amplitude constraints). For IM/DD channels, the non-negativity, peak and average amplitude constraints on \(X\) make solving this problem cumbersome.

This is in stark contrast with the AWGN channel with real-valued \(X\) with power constraint \(\mathbb{E}_X[X^2] \leq \rho\). For the AWGN channel the optimal input distribution is Gaussian and the capacity is given by \(\frac{1}{2}\log\left(1 + \frac{\rho}{P_E}\right)\) nats/transmission. The elegance and simplicity of this expression has inspired a wide range of studies, and led to significant developments in wireless communications. Alas, to date, no one has been able to derive such an elegant expression for IM/DD channels!

At this point, the best way to evaluate this capacity is to use a non-linear optimization framework to obtain the optimal input distribution. For example, the Blahut-Arimoto algorithm
can be used or numerical optimization methods can be applied directly. Though the optimal input distribution \( f_X \) and channel capacity can be found numerically, insights necessary for the design of communication systems are not apparent.

Recent research has led to capacity bounds and asymptotic capacity expressions that can help to circumvent this problem. Instead of calculating the capacity of the IM/DD channel, one can design of communication systems are not apparent.

**4. Capacity Upper Bounds**

Upper bounds on the capacity of IM/DD channels have been derived using two main approaches: (i) using a dual expression for the channel capacity [10] and (ii) using sphere-packing [11,12].

(a) Duality-Based Upper Bounds

Apart from (3.1), an alternate expression for the capacity of the IM/DD channel is [17]

\[
\mathcal{C}_\rho(A) = \inf_{f_Y} \sup_{f_X} \mathbb{E}_X \left[ D(f_Y|X \parallel f_Y) \right],
\]

where the infimum is over all distributions \( f_Y \) on the set of real numbers, the supremum is over all distributions \( f_X \) satisfying the constraints, and \( f_Y|X \) is the distribution of \( Y \) given \( X \) (channel law). Consequently, choosing \( f_Y \) freely leads to a capacity upper bound

\[
\mathcal{C}_\rho(A) \leq \sup_{f_X} \mathbb{E}_X \left[ D(f_Y|X \parallel f_Y) \right].
\]

The question is how to choose \( f_Y \) and how to upper bound this supremum. This has been tackled in [10] by choosing \( f_Y \) to be ‘similar’ to the distribution of a max-entropic \( X \), i.e., a distribution that maximizes \( h(X) \) (given in (5.1) and (5.2)). The intuition is that \( Y \) and \( X \) have ‘similar’ distributions when the noise variance is small. Thus, it is expected that the resulting bounds should be tight asymptotically, i.e., when \( A, E \gg \sigma \). The supremum and expectation in (4.2) are upper bounded in [10] using further analysis leading to the following bounds.

**Theorem 4.1** ([10]). The capacity of the IM/DD channel in Def. 2.3 is upper bounded as \( \mathcal{C} \leq \mathcal{C}_{\rho}^{\text{lmw}}(A) \) where

\[
\mathcal{C}_{\rho}^{\text{lmw}}(A) = \left( 1 - 2Q \left( \frac{\delta + \frac{A}{\sigma}}{\sigma} \right) \right) \log \left( \frac{A + 2\delta}{\sqrt{2\pi} \sigma (\frac{A^2}{\sigma^2})} \right) - \frac{1}{2} + Q \left( \frac{\delta}{\sigma} \right) + \frac{\delta}{\sqrt{2\pi} \sigma} e^{-\frac{\delta^2}{2\sigma^2}},
\]

for \( \rho \leq 2 \), and

\[
\mathcal{C}_{\rho}^{\text{lmw}}(A) = \left( 1 - Q \left( \frac{\delta + \frac{A}{\sigma}}{\sigma} \right) \right) - Q \left( \frac{\delta + (A - \frac{A}{\sigma})}{\sigma} \right) \log \left( \frac{A(e^{\frac{\mu}{\sigma}} - e^{-\mu(1+\frac{\delta}{\sigma})})}{\sqrt{2\pi} \mu(1-2Q(\frac{\delta}{\sigma}))} \right) - \frac{1}{2} + Q \left( \frac{\delta}{\sigma} \right) + \frac{\delta}{\sqrt{2\pi} \sigma} e^{-\frac{\delta^2}{2\sigma^2}} + \frac{\sigma \mu}{A \sqrt{2\pi}} \left( e^{-\frac{\delta^2}{2\sigma^2}} - e^{-\frac{(\delta+\frac{A}{\sigma})^2}{2\sigma^2}} \right) + \frac{\mu}{\rho} \left( 1 - 2Q \left( \frac{\delta + \frac{A}{\sigma}}{\sigma} \right) \right),
\]

for \( \rho > 2 \), where \( \mu \) and \( \delta \) are positive free parameters.

Another upper bound can be derived using the dual expression in (4.2), by choosing \( f_Y \) to be a Gaussian distribution. The intuition is that when \( A \ll \sigma \), the output distribution \( f_Y \) is ‘similar’ to the distribution of noise, which is Gaussian. This leads to the following upper bound.
Theorem 4.2 ([10]). The capacity of the IM/DD channel in Def. 2.3 is upper bounded by $C_{\rho}(A) \leq C_{\rho}^0(A)$ where

$$C_{\rho}^0(A) = \frac{1}{2} \log \left( 1 + \frac{A}{\sigma^2} \left( A - \frac{A}{P} \right) \right), \quad (4.5)$$

where $\rho' = \max\{2, \rho\}$.

This upper bound was also derived in [12] using the following approach. First, we note that a random variable $X \in [0, A]$ with $\mathbb{E}[X] \leq E$ has a maximum variance of $\frac{A^2}{2P}(A - \frac{A}{P})$ where $\rho' = \max\{2, \rho\}$. This maximum variance is achievable if $X$ takes on values in $[0, A]$, with $P\{X = A\} = \frac{A}{P}$ and $P\{X = 0\} = 1 - \frac{A}{P}$. Then, $Y = X + Z$ has a maximum variance of $\frac{A^2}{2P}(A - \frac{A}{P}) + \sigma^2$. Consequently, $h(Y)$ can be no larger than the differential entropy of a Gaussian random variable with variance $\frac{A^2}{2P}(A - \frac{A}{P}) + \sigma^2$, since the Gaussian distribution maximizes differential entropy under a variance constraint [13]. This leads to the upper bound (4.5).

In addition to the results above, a bound derived by McKellips [14] for peak-constrained channels can be adapted to the IM/DD channel with $\rho = 2$. This bound, presented next, is also valid for any $\rho$, $C_{\rho}(A) \leq C_{2}(A)$ [10].

Theorem 4.3 ([14]). The capacity of the IM/DD channel in Def. 2.3 is upper bounded by $C_{\rho}(A) \leq C_{2}^{\rho}(A)$ where

$$C_{2}^{\rho}(A) = \log \left( 1 + \frac{A}{\sqrt{2\pi e}\sigma} \right). \quad (4.6)$$

This bound in this theorem was originally derived in [14] by maximizing $h(Y)$ when the support of $Y$ is parsed into two regions, $[0, A]$ and $(-\infty, 0) \cup (A, \infty)$. It was re-derived and refined in [15] using the duality approach (4.2).

Example 4.1. For an IM/DD channel with $A = 10$, $E = 4$, and $\sigma^2 = 1$, the bounds in Theorems 4.1, 4.2, and 4.3 evaluate to $C_{2}^{\text{law}}(A) = 1.3257$, $C_{\rho}^0(A) = 1.6094$, and $C_{2}^{\rho}(A) = 1.2296$ nats/transmission. Thus, the capacity $C_{\rho}(A)$ of this channel is no larger than 1.2296 nats/transmission.

While the duality-based bounds presented here rely on careful analysis, in the next section a geometric view of upper bounding the IM/DD capacity is presented.

(b) Sphere-Packing Upper Bounds

The channel requires a bound on the average optical power, i.e., $\mathbb{E}[X] \leq E$. By the law of large numbers, the transmitted sequence of symbols $\{x_i\}_{i=1}^n$ must satisfy $\frac{1}{n} \sum_{i=1}^n x_i \leq E$ almost surely as $n$ grows. Combining this with the constraint $X \in [0, A]$, we conclude that $\{x_i\}_{i=1}^n$ must lie within the intersection of an $n$-dimensional hypercube of side length $A$ and an $n$-simplex with side length $nE$. This is termed the permissible signal region.

Similarly, since the noise $Z$ satisfies $\mathbb{E}[Z^2] = \sigma^2$, then, for large $n$, the noise sequence $\{z_i\}_{i=1}^n$ must lie almost certainly near the surface of a ‘noise sphere’ in $n$-dimensions of radius $\sqrt{n\sigma^2}$ [18].

As a result, for large $n$, each transmit sequence $\{x_i\}_{i=1}^n$ will be mapped to a point on a sphere of radius $\sqrt{n\sigma^2}$ about $\{x_i\}_{i=1}^n$ at the receiver. Intuition suggests that, for reliable communication, spheres corresponding to different sequences $\{x_i\}_{i=1}^n$ must have negligible overlap at the receiver. How many such spheres can be fitted with their centers within the permissible signal region?

One can upper bound this number of spheres by dividing the volume of the $\sqrt{n\sigma^2}$-neighbourhood of the permissible region (the set of points in the $n$-dimensional space which are at at most a distance $\sqrt{n\sigma^2}$ outside the permissible region) by the volume of an individual
noise sphere. Then, taking the logarithm of this upper bound, dividing by \( n \), and taking the limit as \( n \to \infty \) leads to a capacity upper bound.\(^1\)

Finding the volume of the \( \sqrt{n} \sigma^2 \)-neighbourhood is not straightforward. To simplify the problem, let us consider the cube or the simplex separately, which still leads to an upper bound since dropping a constraint can not shrink the permissible region.

One can find the volume of the \( \sqrt{n} \sigma^2 \)-neighbourhood of a cube (Fig. 2) using the Steiner-Minkowski theorem for polytopes \([21, \text{Proposition 12.3.6}]\). This theorem basically counts the volume of portions constituting the \( \sqrt{n} \sigma^2 \)-neighbourhood as depicted in Fig. 2b. This approach was used in \([12]\) to derive a capacity upper bound.

The volume of the \( \sqrt{n} \sigma^2 \)-neighbourhood of a simplex (Fig. 2c) is more difficult to bound. The Steiner-Minkowski theorem for polytopes requires calculating the dihedral angles of the simplex, which becomes difficult at higher dimensions. In \([11]\), the authors bounded these angles to obtain a capacity upper bound. Other bounds based on a similar approach were derived in \([22,23]\). Instead of using the Steiner-Minkowski theorem, a different approach for bounding the volume was developed in \([12]\) leading to a tighter bound.

By taking the minimum of the resulting sphere-packing bounds for a cube and for a simplex, one obtains a capacity upper bound for the IM/DD channel. This is given in the following theorem.

**Theorem 4.4** (\([12]\)). The capacity of the IM/DD channel in Def. 2.3 is upper bounded by \( C_p(A) \leq C_{p, \text{cma}}^{\text{lmw}}(A) \) where

\[
C_{p, \text{cma}}^{\text{lmw}}(A) = \min \left\{ \sup_{\alpha \in [0,1]} \alpha \log \left( \frac{A}{\sqrt{2\pi\sigma^2(1-\alpha)}} \right), \sup_{\alpha \in [0,1]} \alpha \log \left( \frac{\sqrt{A}}{\sqrt{2\pi\sigma^2} (1-\alpha)^{\frac{1}{2}}} \right) \right\}.
\]

**Example 4.2.** For an IM/DD channel with \( A = 10 \), \( E = 4 \), \( \sigma^2 = 1 \), the upper bound in Theorem 4.4 evaluates to \( C_{p, \text{cma}}^{\text{lmw}}(A) = 1.4104 \) nats/transmission. This is tighter than the bound \( C_p(A) \) in Example 4.1, but looser than the bound \( C_{p, \text{cma}}^{\text{lmw}}(A) \) and \( C_{p, \text{cma}}^{\text{cma}}(A) \) in the same Example. The main reason is that the sphere-packing bounds in Theorem 4.4 ignore one of the two constraints (A or E).

To provide a complete picture of the capacity, these upper bounds must be accompanied by capacity lower bounds. This is the topic of the next section.

### 5. Capacity Lower Bounds

A lower bound on capacity can be found by computing the mutual information for a given input distribution. Several capacity lower bounds have been developed for the IM/DD channel using different approaches. Some of the developed bounds are tight in some regimes of operations, allowing us to characterize capacity in these regimes. We present these lower bounds next.

We focus on lower bounds which have simple analytical expressions. These are namely lower bounds which use a continuous input distribution \( f_X \). Lower bounds based on discrete input distributions will be discussed in Sec. 6.

To derive a lower bound, let us assume that \( X \) is a continuous random variable. This might be restrictive, but will lead to a simple lower bound expression. Note that \( I(X;Y) = h(Y) - h(Y|X) \). Since \( h(Y|X) = h(Z) \) which does not depend on \( X \), one can maximize \( I(X;Y) \) by maximizing \( h(Y) \) with respect to \( f_X \). Instead of maximizing \( h(Y) \), we make the following observation: In the extreme case of zero noise, \( Y \) and \( X \) have the same distribution. Therefore, one expects that if noise is relatively small compared to the received signal, then \( Y \) has a distribution which is ‘similar’ to that of \( X \). In fact, the signal-to-noise ratio in VLC channels can often be high making

\(^1\)A formal proof of the sphere-packing approach for bounding capacity can be found in \([19, \text{Chapter 5}]\) and \([20, \text{Appendix B}]\).
A depiction of sphere-packing in a cube. With $n = 2$, the cube is a square, and the sphere is a circle.

(b) The $\lambda$-neighborhood of the square. It consists of the square itself, four quarter-disks, and four rectangles.

(c) A depiction of sphere-packing in a simplex. With $n = 2$, the simplex is a triangle, and the sphere is a circle.

(d) The $\lambda$-neighborhood of the triangle (simplex). It consists of the triangle itself, three portions of a disk, and three rectangles.

Figure 2: Spheres of radius $\lambda$ centered within a convex object will certainly be contained within the convex set whose boundary is at a distance $\lambda$ from the object (its $\lambda$-neighbourhood).

This observation practically relevant. This means that in this regime, a ‘good’ choice of $f_X$ is one which maximizes $h(X)$.

To find the maximum $h(X)$, one can use the maximum-entropy theorem [13, Theorem 12.1.1]. For a random variable $X \in [0, A]$ with $\mathbb{E}_X[X] \leq E$ and $E \geq \frac{A}{2}$, the maximum $h(X)$ is achieved when $X$ is uniformly distributed, i.e.,

$$f_X^u(x) = \frac{1}{A} \text{ for } x \in [0, A].$$

(5.1)

If $E < \frac{A}{2}$, then $h(X)$ is maximized when $X$ follows the truncated-exponential distribution

$$f_X^{te}(x) = \frac{\mu}{A(1 - e^{-\mu})} e^{-\frac{\mu x}{A}} \text{ for } x \in [0, A],$$

(5.2)

where $\mu$ is the unique solution of $\frac{E}{A} = \frac{1}{\mu} - \frac{e^{-\mu}}{1 - e^{-\mu}}$. Using the entropy-power inequality [13] and these distributions, the following simple capacity lower bounds were derived in [10].

**Theorem 5.1** ([10]). The capacity of the IM/DD channel in Def. 2.3 is lower bounded by $C \geq C_R^{\text{lmw}}(A)$ nats/transmission, where

$$C_R^{\text{lmw}}(A) = \begin{cases} \frac{1}{2} \log \left( 1 + \frac{A^2}{2\pi e \sigma^2} \right), & \rho \leq 2 \\ \frac{1}{2} \log \left( 1 + \frac{A^2 e^{2} e^{(1-e^{-\mu})^2}}{2\pi e \sigma^2 \mu^2} \right), & \rho > 2, \end{cases}$$

(5.3)

where $\mu$ is the unique solution of $\frac{1}{\rho} = \frac{1}{\mu} - \frac{e^{-\mu}}{1 - e^{-\mu}}$. 

The right hand side in (5.3) is an achievable rate which can be achieved using a uniform \( f_X \) (5.1) or a truncated-exponential \( f_X \) (5.2), respectively.

Another lower bound based on a continuous distribution was derived in [12] using a truncated-Gaussian input distribution. This distribution leads to an achievable rate which is higher than (5.3) for a wide range of \( A \), but is more complicated to evaluate. We restrict our attention to (5.3) here, which converges to capacity as \( A \to \infty \) as we shall see in Sec. 7.

**Example 5.1.** For an IM/DD channel with \( A = 10 \), \( E = 4 \), and \( \sigma^2 = 1 \), the lower bound in Theorem 5.1 evaluates to \( C^\text{low}_{\rho}(A) = 0.9111 \) nats/transmission. Combining this with the Example 4.1, we conclude that the capacity of this channel is between 0.9111 and 1.2296 nats/transmission.

### 6. Approaching Capacity

The lower bounds above are based on a continuous input distribution. This section focuses on discrete input distributions.

**(a) Discreteness of Optimal Input Distribution**

In [24], it was shown that the capacity achieving input distribution for the IM/DD channel is discrete, with a finite number of mass points. Thus, the optimal \( X \) is distributed according to

\[
\begin{align*}
\hat{f}_X(a_k) &= p_k, \quad k = 0, 1, \ldots, K, \\
\end{align*}
\]

where \((a_0, a_1, \ldots, a_K)\) is the support with \( a_k \in [0, A] \quad \forall k = 0, 1, \ldots, K \), and \((p_0, p_1, \ldots, p_K)\) are the associated probabilities.

Although this is attractive from an implementation perspective, finding the optimal discrete distribution can be difficult. The optimization problem to be solved is

\[
\begin{align*}
\max_{K, a_k, p_k} & \quad I(X; Y) |_{X \sim \hat{f}_X} \\
\text{subject to} & \quad K \geq 2; \quad \sum_{k=0}^{K} p_k = 1; \quad \sum_{k=0}^{K} p_k a_k \leq E \\
& \quad a_k \in [0, A] \quad \forall k = 0, 1, \ldots, K,
\end{align*}
\]

where

\[
I(X; Y) = - \int_{-\infty}^{\infty} \left( \sum_{k=0}^{K} \frac{p_k}{\sqrt{2\pi}\sigma} e^{-\frac{(a_k - a)^2}{2\sigma^2}} \right) \log \left( \sum_{k=0}^{K} \frac{p_k}{\sqrt{2\pi}\sigma} e^{-\frac{(a_k - a)^2}{2\sigma^2}} \right) dy - \frac{1}{2} \log(2\pi e\sigma^2). \tag{6.3}
\]

This is a nonlinear optimization problem, which can be solved numerically. If \( K \) and \( a_k \) are fixed, then the elegant Blahut-Arimoto algorithm [8,9] can be used to find \( p_k \). However, how can we determine the optimal \( K \) and \( a_k \)? To tackle this problem, [24] provides an algorithm and conditions which enable finding the optimal distribution. First \( K \) is initialized to 1 (two mass points). Then, the optimization (6.2) is solved for \( a_k \) and \( p_k \) numerically using standard solvers (numerical integration and maximization). The obtained \( a_k \) and \( p_k \) are then tested using the conditions in [24, Corollary 3]. If the conditions are satisfied, then the obtained distribution is optimal. Otherwise, \( K \) is incremented and the process is repeated. The result is a discrete input distribution which is capacity achieving.

**Example 6.1.** For an IM/DD channel with \( A = 10 \), \( E = 4 \), and \( \sigma^2 = 1 \), the solution of the optimization in (6.2) is \( K = 4 \) (5 mass points), \((a_0, \ldots, a_4) = (0, 2.7752, 5.0525, 7.3046, 10)\), and \((p_0, \ldots, p_4) = (0.3476, 0.1934, 0.1517, 0.1395, 0.1678)\) as shown in Fig. 3. The capacity is \( C_p(A) = 1.1853 \) nats/transmission.
The calculated capacity $C_ρ(A) = 1.1853$ nats/transmission is between the bounds in Examples 4.1 and 5.1, given by $C_{lmw}^1(A) = 1.3257$ and $C_{lmw}^0(A) = 0.9111$ nats/transmission. However, neither of the bounds are tight for this example. Nevertheless, we shall see later that this is not the case asymptotically in the limit of large $A, E$. Before we delve deeper into asymptotic capacity, let us ask the following question: Is there a way to approach the capacity of the channel without solving problem (6.2)?

The answer is yes, as demonstrated in [16], and is discussed next.

(b) Capacity Approaching Distributions

Rather than solving (6.2), let us fix the discrete input distribution to have the support $(a_0, \ldots, a_K) = (0, \ell, 2\ell, \ldots, A-\ell, A)$, where $\ell = A^{\frac{1}{K+1}}$. This way, we do not have to find the optimal support, but the optimal probabilities $p_k$ and number of mass points $K+1$. This simplifies the problem slightly.

For a further simplification, instead of maximizing $I(X;Y)$ under this input distribution, let us maximize $H(X)$ and find the max-entropic distribution with the given support. The intuition is that the max-entropic distribution approaches the maximum $h(Y)$ when $\sigma^2 \to 0$. Thus, one needs to solve the simpler problem

$$\max_{p_k} H(X)$$

subject to $\sum_{k=0}^{K} p_k = 1, \sum_{k=0}^{K} p_k \frac{kA}{K} \leq E$.  

(6.4)

The problem can be solved using using the Lagrangian, leading to the following solution

$$p_k = \begin{cases} 
\frac{\frac{1}{K+1}}{\sum_{j=0}^{K} \frac{1}{j+1}} & \rho \leq 2 \\
\frac{\frac{1}{K+1}}{\sum_{j=0}^{K} \frac{1}{j+1}} \left(1 - \frac{\frac{1}{K+1}}{\sum_{j=0}^{K} \frac{1}{j+1}}\right) & \rho > 2, 
\end{cases} \quad (6.5)$$

where $t_0$ is the unique positive root of $\sum_{k=0}^{K} \frac{1}{k+1} (1 - \frac{kA}{K}) t^k$.

Using this solution, one can initialize $K = 1$ (i.e., two mass points), calculate $p_k$, and evaluate the achievable rate $I(X;Y)$. Then, $K$ is incremented, and the process is repeated until the achievable rate stops increasing. This leads to the best achievable rate using a max-entropic discrete input distribution with equally spaced mass points, which we shall denote $C_{fh}^0(A)$. This distribution deviates from the optimal input distribution, but achieves close-to-optimal performance. The following example illustrates this point.
Example 6.2. For an IM/DD channel with $A = 10$, $E = 4$, and $\sigma^2 = 1$, the best rate that can be achieved using a max-entropic discrete input distribution with equally spaced mass points is $C_{\text{fh}}(A) = 1.1724$ nats/transmission, achieved with $K = 4$ (5 mass points), $(a_0, \ldots, a_4) = (0, 2.5, 5, 7.5, 10)$, and $(p_0, \ldots, p_4) = (0.2884, 0.2353, 0.1920, 0.1566, 0.1278)$ as shown in Fig. 3.

The achievable rate of the proposed distribution is generally close to capacity for any $A$, $E$.

Before comparing the capacity bounds with the achievable rates of the discrete input distributions above in Section 7, we make the following connection to the underlying DTP channel.

(c) Connection to DTP Channels

Although this review centres on the signal-independent IM/DD channel with Gaussian noise (see Def. 2.3), the relationship to the underlying DTP channel in Def. 2.1 is less well explored. It is well known that the the capacity achieving distribution for the DTP channel under peak and average constraints is discrete. In the case of large background light, i.e., large $\lambda$, the channel law of the DTP channel with high intensity background light will approach the Gaussian statistics of the IM/DD channel in Def. 2.3. The following theorem has been established for DTP channels with large background light.

Theorem 6.1 ([35]). The capacity-achieving distribution for the DTP channel in Def. 2.1 with $\rho > 2$ and $\lambda$ sufficiently large is binary with support at points $(0, A)$ and associated probabilities

$$
\left(\frac{\rho - 1}{\rho}, \frac{1}{\rho}\right).
$$

(6.6)

Notice that the distribution (6.6) is identical to the maxentropic source distribution given in (6.5) for $K = 1$ and it satisfies both peak and average amplitude constraints with equality. In the case of high $\lambda$ the channel law of the DTP channel approaches that of the Gaussian IM/DD channel and the results on capacity-achieving distributions coincide.

7. Comparison and Asymptotic Capacity

For comparison, we fix the peak-to-average-ratio $\rho$, and we vary $A/\sigma$, which we call the signal-to-noise ratio (SNR). We start with a comparison for $\rho = 2$. This case represents an IM/DD channel with a peak constraint only, since the average constraint becomes redundant when $\rho = 2$ [10]. Fig. 4 shows the upper bounds and achievable rates. By comparing the bounds with the capacity $C_{\rho}(A)$ evaluated by solving problem 6.2 numerically, we can see that the upper bound $C_{\text{cma}}(A)$ is asymptotically tight at low SNR, and the bound $C_{\text{lmw}}(A)$ is asymptotically tight at high SNR. The upper bound $C_{\text{w}}(A)$ is also fairly tight at high SNR. The upper bound $C_{\text{cma}}(A)$ is tight at high SNR when $\rho = 2$, and nearly tight overall. Moreover, we see that a max-entropic equally-spaced discrete input distribution achieves a rate $C_{\text{fh}}(A)$ which is close to capacity for any SNR. The lower bound $C_{\text{lmw}}(A)$, which has the advantage of a closed-form expression, is only tight at high SNR. Fig. 5 shows a similar comparison with $\rho = 4$. In this case, the bound $C_{\text{lmw}}(A)$ is slightly tighter than $C_{\text{fh}}(A)$.

The convergence of upper and lower bounds at high and low SNR has been also proved, leading to the asymptotic capacity of the IM/DD channel. This is described next.

Theorem 7.1 (High-SNR capacity [10]). For a fixed $\rho = \frac{A}{E}$, the capacity of the IM/DD channel in Def. 2.3 satisfies

$$
\lim_{A \to \infty} \left( C_{\rho}(A) - \frac{1}{2} \log \left( \frac{A^2}{2\pi e \sigma^2} \right) \right) = 0,
$$

(7.1)
Figure 4: Capacity bounds and achievable rates for an IM/DD channel with $\rho = 2$.

Figure 5: Capacity bounds and achievable rates for an IM/DD channel with $\rho = 4$.

for $\rho \leq 2$; and

$$\lim_{A \searrow 0} \left( C_\rho(A) - \frac{1}{2} \log \left( \frac{A^2 e^{2 \rho} (1 - e^{-\rho})^2}{2\pi e \sigma^2 \mu^2} \right) \right) = 0, \quad (7.2)$$

for $\rho > 2$, where $\mu$ is the unique solution of $\frac{1}{\rho} = \frac{1}{\rho} - \frac{e^{-\rho}}{1-e^{-\rho}}$.

To simplify this asymptotic capacity expression, [12] showed that the high-SNR capacity can be well approximated by $\min \left\{ \frac{1}{2} \log \left( \frac{A^2}{2\pi e \sigma^2} \right), \frac{1}{2} \log \left( \frac{e A^2}{2\pi e \sigma^2} \right) \right\}$, within a gap $\approx 0.1$ nats/transmission. On the other hand, at low SNR, the following statement holds.

**Theorem 7.2** (Low-SNR capacity [10]). For a fixed $\rho = \frac{A}{e}$, the capacity of the IM/DD channel in Def. 2.3 satisfies

$$\lim_{A \searrow 0} \frac{C_\rho(A)}{A^2 (1 - \frac{1}{\rho'}) \frac{A^2}{2\pi e}} = 1, \quad (7.3)$$

where $\rho' = \max\{2, \rho\}$. 
Finally, we note that similar bounds and asymptotic capacity results have been established for the case when $A = \infty$, i.e., when the channel only has an average intensity constraint. These bounds can be found in [10,11]. The low-SNR slope in this case was bounded in [10]. However, the low-SNR capacity remains an open problem.

8. Discussion and Extensions

(a) On Uni-Polar OFDM Schemes

Orthogonal Frequency-Division Multiplexing (OFDM) is a transmission scheme which enjoys simple construction. However, it is not directly applicable in IM/DD channels due to the non-negativity constraint of $x_i$. To circumvent this, several uni-polar OFDM schemes have been proposed in the literature [30,31]. Perhaps the most famous is DC-offset Optical OFDM (DCO-OFDM) [32], which constructs an real-valued OFDM signal by forcing frequency-domain symbols to be Hermitian-symmetric, and then applies a DC-offset, and clips the resulting signal at 0.

This construction shifts the encoding/decoding from the time domain to the frequency domain, where the frequency-domain channel for each subcarrier is a complex Gaussian channel. Namely, for a given subcarrier $k$, the transmitter encodes information into a sequence $\{s_k[\ell]\}_{\ell=1}^{m}$ where $s_k[\ell] \in \mathbb{C}$. Then, the operation in (8.1) is applied to $\{s_k[\ell]\}_{\ell=1}^{m-1}$ to generate $\{x_i[\ell]\}_{i=1}^{m-1}$, which is then transmitted. The receiver receives $\{y_i[\ell]\}_{i=1}^{m-1}$ where $y_i[\ell] = x_i[\ell] + z_i[\ell]$, and applies a Fourier transform to obtain

$$r_k[\ell] = s_k[\ell] + q_k[\ell] + p_k[\ell], \quad (8.2)$$

where $q_k[\ell]$ is complex Gaussian noise (the Fourier transform of $z_i$) and $p_k[\ell]$ is clipping distortion. The receiver collects $\{r_k[\ell]\}_{\ell=1}^{m}$ from which it decodes $\{s_k[\ell]\}_{\ell=1}^{m}$. If we assume that the clipping distortion is Gaussian, and that $s_k[\ell]$ is subjected to a power constraint $\mathbb{E}|s_k[\ell]|^2 \leq p_k$, then we know exactly how to achieve the capacity of the channel in (8.2); we construct the codewords $\{s_k[\ell]\}_{\ell=1}^{m}$ as i.i.d. complex Gaussian with zero mean and variance $p_k$.

While this is an attractive scheme in practice, it has some limitations which prevent it from achieving the IM/DD channel capacity. One of the limitations is the clipping distortion $p_k[\ell]$. This is introduced due to clipping, and would not be present if $x_i$ is chosen to follow any of the distributions discussed in Sections 5 and 6. To limit the clipping distortion, one has to choose $p_k$ to be small so that the transformed signal $\frac{1}{\sqrt{m}} \sum_{k=0}^{m-1} s_k e^{j2\pi ik}$ has a small amplitude with high probability. This results in a decrease in the achievable rate.

Even if we neglect this clipping distortion, this construction has other limitations. By studying (8.1), one can find that $X_i$ follows a clipped-Gaussian distribution [31], and that...
where $h(\cdot|\cdot)$ is the conditional differential entropy, and $\tilde{X}_i$ is a random variable distributed on $[0, A]$ following the optimal input distribution (solution of (6.2)). Here, the first inequality follows since conditioning ‘strictly’ reduces entropy when variables are dependent, which is the case when $x_i$ is constructed using DCO-OFDM. The second inequality follows since the capacity achieving distribution for an IM/DD channel is discrete and unique [24], i.e., does not coincide with a clipped-Gaussian distribution. From this, we conclude that the achievable rate of DCO-OFDM, even with the clipping distortion neglected, is strictly less than capacity.

A similar discussion applies to other unipolar-OFDM constructions. Reference [30] studies the achievable rates of several unipolar-OFDM schemes, and clearly shows a gap between these achievable rates and capacity bounds from [11, 33]. Whether this is a limitation of unipolar-OFDM in general or a limitation of existing constructions of unipolar-OFDM remains to be seen. Nevertheless, [30] shows that some multi-layer uni-polar OFDM schemes achieve rates which are fairly close to capacity at high SNR for the average-constrained IM/DD channel ($A = \infty$).

(b) MIMO IM/DD Channels

The results in Sections 4-6 which consider a Single-Input Single-Output (SISO) IM/DD channel, have been recently extended to IM/DD systems with multiple transmitters (multiple channel inputs) and/or multiple receivers (multiple channel outputs). This can be classified into Single-Input Multiple-Output channels (SIMO), Multiple-Input Single-Output (MISO), and Multiple-Input Multiple-Output (MIMO) channels.

(i) SIMO IM/DD Channels

In the SIMO case, the transmitter has one aperture, and the receiver has $n_r$ apertures. The receiver receives a vector $Y \in \mathbb{R}^{n_r}$ with $Y = hX + Z$, where $h = (h_1, \ldots, h_{n_r})^T$ is the vector of channel ‘attenuation’ from the transmitter to the receivers’ apertures, and $Z$ is a vector of independent Gaussian noises with zero mean variance $\sigma^2$. One can easily express the capacity of this case as a function of the capacity of a SISO IM/DD channel, and show that maximum-ratio combining is optimal. If we define $\tilde{Y} = U^T Y$ where $U$ is an orthonormal matrix with $h_{\|h\|}$ as its first column, then $\tilde{Y}$ will be given by $\tilde{Y} = (\|h\|X + \tilde{Z}_1, \tilde{Z}_2, \ldots, \tilde{Z}_{n_r})^T$, where $\tilde{Z}$ has the same distribution as $Z$. The channel from $X$ to $Y$ has the same capacity as the channel from $X$ to $\tilde{Y}$ since the multiplication by $U$ is an invertible transformation. Hence, the capacity of the channel is equal to $C_{p}^{\text{sim}}(\|h\|A)$. The bounds presented in Sections 4-6 can be directly applied.

(ii) MIMO IM/DD Channels

In the MIMO channel, the transmitter has $n_t$ apertures and the receiver has $n_r$ apertures. We can model the channel as one whose input is $X \in [0, A]^{n_t}$ where $n_t$ is the number of transmitter
apertures, and output

\[ Y = HX + Z, \]  

(8.4)

where \( Z \) is a vector of independent Gaussian noises with zero mean variance \( \sigma^2 \), and \( H = (h_1, \ldots, h_{nt}) \) where \( h_t \) is the vector of channel ‘attenuation’ from transmitter aperture \( t \) to the receivers’ apertures. The MISO channel is a special case of the MIMO channel with \( nt = 1 \). The average constraint is expressed as a total average constraint \( \sum_{t=1}^{nt} \mathbb{E}[X_t] \leq E \).

For the MISO and MIMO cases, we face another difficulty. Now, the SISO codebooks whose codewords are scalar sequences \( \{x_i\}_{i=1}^{n_t} \) do not directly apply. Instead, we need to generate vector codebooks with codewords \( \{x_i\}_{i=1}^{n_t} \) where \( x_i \in [0, A]^{n_t} \). The optimal codebook is one whose symbols are independent and identically distributed according to a distribution which maximizes \( I(X; Y) \), with respect to \( f_X \) subject to peak and average constraints. This problem has been investigated in [25–29].

In [25], capacity bounds were derived for two cases: \( nt \leq n_r \) and \( n_t > n_r \). For \( nt \leq n_r \), it was shown that if \( \rho \leq \frac{2}{n_t} \) (peak constraint is dominant) and \( H \) has full column rank, then the following characterizes the asymptotic capacity at high SNR

\[ \frac{1}{2} \log \left| \frac{A^2}{2 \sigma^2} H^T H \right|. \]  

(8.5)

This high-SNR capacity is asymptotically achievable as \( A \rightarrow \infty \) with a fixed \( \rho \leq \frac{2}{n_t} \), using a QR-decomposition scheme. Namely, the transmitter sends \( n_t \) independently coded streams from its apertures, where the symbols \( X_t \) are uniformly distributed on \([0, A] \). To resolve crosstalk, the receiver multiplies \( Y \) by \( Q \) where \( H = QR \) is the QR-decomposition of \( H \). \( Q \) is an orthogonal matrix, and \( R \) is upper triangular. This way, the receiver obtains an equivalent channel with upper triangular structure given by \( \tilde{Y} = RX + \tilde{Z} \). The receiver then starts by decoding \( X_{n_t} \) without crosstalk, then it removes the contribution of \( X_{n_t} \) from \( \tilde{Y} \) to decode \( X_{n_t-1} \) without crosstalk, and so on.

If \( nt \leq n_r \), \( H \) has full column rank, and \( \rho > \frac{2}{n_t} \) (average constraint is dominant), then the asymptotic capacity at high SNR (\( A \rightarrow \infty \)) is within a small gap of

\[ \frac{1}{2} \log \left| \frac{e}{2 \sigma^2} \min \left\{ \frac{A^2}{\rho^2 n_t}, \frac{A^2}{e^2} \right\} H^T H \right|. \]  

(8.6)

This is achievable using a similar QR-decomposition scheme with a truncated-Gaussian input distribution for each transmitted stream.

The asymptotic capacity expressions in (8.5) and (8.6) reveal a multiplexing gain equal to \( nt \), when \( H \) has rank \( n_t \) and \( n_r \leq n_t \). If the rank of \( H \) is less than \( n_r \) which in turn is less than \( n_t \), then we can use the same transmission scheme to achieve a multiplexing gain equal to the rank of \( H \) after deactivating some transmit apertures (switching them off, or on a constant intensity).

If \( n_t > n_r \), then [25] provides capacity bounds and scaling laws. In this case, transmission schemes can use a pre-coding matrix \( G \in \mathbb{R}^{n_r \times n_t} \) to reduce the effective MIMO channel to an \( n_r \times n_t \) channel, over which the above results can be applied. This case was studied more extensively in [29], which develops minimum energy signalling schemes, capacity lower bounds, upper bounds, and asymptotic capacity expressions.

The capacity of the MIMO IM/DD channel was characterized in the low-SNR regime \( (A \rightarrow 0) \) in [25], where it was shown to be equal asymptotically to (for both \( nt \leq n_r \) and \( n_t > n_r \))

\[ \max_{a_i, \sum_{i=1}^{nt} a_i \leq \frac{A^2}{2 \sigma^2}} \sum_{i=1}^{nt} \sum_{j=1}^{n_r} h_i^T h_j \min\{a_i, a_j\} (1 - \max\{a_i, a_j\}). \]  

(8.7)

This is achievable using a maximally-correlated \( n_t \)-ary binary input distribution. In this distribution, the transmitted vector symbols can take values in \([0, A]^{n_t} \) such that if \( X_i = A \) and \( E_i \leq E_j \), then \( X_j = A \). For instance, if \( n_t = 2 \) and \( E_1 \geq E_2 \), then \( X \) can take only three values, \((0, 0)\), \((A, 0)\) and \((A, A)\).
(iii) MISO IM/DD Channels

In the MISO channel, the transmitter has \( n_t \) apertures, and the receiver has one aperture. The input-output relation can be written as \( Y = h^T X + Z \) where \( Z \) is Gaussian with zero mean and variance \( \sigma^2 \), \( h = (h_1, \ldots, h_{n_t})^T \) and \( h_i \) is the channel attenuation from transmitter aperture \( i \) to the receiver aperture. We can assume without loss of generality that \( h_1 \geq h_2 \geq \cdots \geq h_{n_t} \).

Transmission schemes for the MISO channel can be extracted from [25] as a special case when \( n_t = 1 \). However, this paper only provides achievable rates and high-SNR capacity scaling laws. In [27], properties of capacity achieving input distributions for the MISO case were derived. In general, [27] shows that one can restrict the optimization \( \max_{f_X} I(X, Y) \) to distributions satisfying \( X_i > 0 \Rightarrow X_j = A \forall j < i \). In other words, the transmitter modulates LEDs at two levels: (i) modulating the number of fully-ON LEDs (\( X_j = A \)) from 0 to \( n_t - 1 \) so that fully-ON LEDs are those with the strongest channels, and (ii) modulating the intensity of the next LED (next in terms of channel strength) using some probability distribution. Note that a similar behaviour has been identified in [26] for the MIMO channel at low SNR.

Using this result capacity lower bounds were derived in [27], in addition to capacity upper bounds. It shows that when the peak constraint is dominant, i.e. \( \rho \leq \frac{2}{n_t} \), then the MISO channel capacity is equal to \( C_\rho (\sum_{i=1}^{n_t} h_i A) \). Asymptotic capacity results in Section 7 can be used in this case. Namely, the high SNR (\( \frac{A}{\sigma^2} \to \infty \)) capacity is

\[
\frac{1}{2} \log \left( \frac{(\sum_{i=1}^{n_t} h_i)^2 A^2}{2\pi e \sigma^2} \right).
\]

The above asymptotic capacity also applies if the average constraint is dominant, i.e., \( \rho > \frac{2}{n_t} \), and the peak-to-average ratio \( \rho \) satisfies \( \rho \leq \rho_{th} \triangleq \left( \frac{1}{2} + \frac{1}{\sum_{i=1}^{n_t} h_i} \sum_{k=1}^{n_t} \lambda (k-1) \right)^{-1} \).

On the other hand, if the average constraint is dominant and \( \rho > \rho_{th} \), then the high-SNR capacity is

\[
\frac{1}{2} \log \left( \frac{(\sum_{i=1}^{n_t} h_i)^2 A^2}{2\pi e \sigma^2} \right) + \nu,
\]

where

\[
\nu = \sup_{\lambda \in \left( \max\{0, \frac{1}{n_t} + \frac{1}{\lambda_{th}}, \min\{\frac{1}{n_t}, \frac{1}{\lambda_{th}}\}\} \right)} \left( 1 - \log \frac{\mu(\lambda)}{1 - e^{-\mu(\lambda)}} + \log\left( \frac{\mu(\lambda)}{1 - e^{-\mu(\lambda)}} \right) - \sum_{i=1}^{n_t} \frac{h_i a^i}{\sum_{k=1}^{n_t} h_k a^k} \right),
\]

\( \mu(\lambda) \) is the solution of \( \frac{1}{\rho} - \frac{1}{1-e^{-\rho}} = \lambda \), \( p = (p_1, \ldots, p_{n_t}) \), \( p_i = \frac{h_i a^i}{\sum_{k=1}^{n_t} h_k a^k} \), and \( a \) is the unique positive solution of \( \sum_{i=1}^{n_t} \lambda_i h_i a^i = \frac{1}{\rho} - \lambda + 1 \).

The low-SNR capacity of the MISO case has been given in [26] and [27], and coincides with (8.7) with \( n_t = 1 \).

9. Conclusion

Optical wireless channels utilize a largely under utilized huge portion of the spectrum which is available to enable future ultra high speed networks. In spite of ongoing work to develop OWC technologies, there exist practical and theoretical open problems to solve. In particular, this review has highlighted a string of results over the previous decade which have endeavoured to characterized the fundamental channel capacity of optical wireless IM/DD channels. Many bounds have been derived which are tight in different regimes. Capacity-approaching and in some cases capacity-achieving distributions for IM/DD channels have been derived. In spite of this intense effort, the unique amplitude characteristics of IM/DD optical wireless channels have resulted in a large number of open areas of research.
Current challenges exist in extending these fundamental results to practical VLC and optical wireless implementations. The design of tractable and efficient modulators and encoders is essential to the future progress of OWC technologies. In particular, VLC luminaries require energy efficient approaches which can leverage multiple emitters and receivers. In addition, a multi-disciplinary approach is required to understand the capacity of IM/DD channels in the presence of modulator nonlinearities such as those in high-illumiance LEDs. Many theoretical and practical challenges exist in these scenarios which are ripe for ongoing investigation.

Authors’ Contributions. Chaaban drafted the primary technical sections of the paper. Hranilovic edited and added additional technical content. Both authors designed the paper structure, edited the paper, jointly drafted the final manuscript and have read and approved this manuscript.

References