On High-Efficiency Optical Communication and Key Distribution

Yuval Kochman
EECS Dept., MIT
Cambridge, MA 02139, USA
Email: yuvalko@mit.edu

Gregory W. Wornell
EECS Dept., MIT
Cambridge, MA 02139, USA
Email: gww@mit.edu

Abstract—**We investigate modulation and coding techniques that approach the fundamental limits of communication and key distribution over optical channels, in the regime of simultaneously high photon and bandwidth efficiencies. First, we develop a simple and robust system design for free-space optical communication that incorporates pulse-position modulation (PPM) over multiple spatial degrees of freedom in order to achieve high photon and spectral efficiency. Further, in the context of key distribution, we determine the optimal rate using a Poisson source of entangled photon pairs and photon detectors, and show how to approach it using PPM parsing of the detected photon stream.**

I. Introduction

Classical optical communication can be roughly divided into two regimes, according to the number of photons sent per channel use. If this number is high, homodyne or heterodyne detection may be used to effectively transform the channel into an equivalent additive white Gaussian noise (AWGN) channel. On the other hand, if the number is low, then photon detection, also known as direct detection, produces at the decoder a Poisson count, with mean equal to the energy sent - thus it is known as the Poisson regime.

In the Poisson regime, which is the focus of this work, the quantum nature of light is more evident, leading to somewhat results that defy common AWGN-based engineering intuition. The role of noise is played by spontaneous counts known as dark current. The capacity of the resulting classical channel under various constraints has been studied for the last half century, see e.g. [1]–[4].

Surprisingly, even without any noise (dark current), the channel capacity is finite. One may think of it as a result of the Poisson statistics effectively constraining the channel output to a finite alphabet; however, even if quantum-optimal measurements are used rather than photon detection, the capacity is still finite. In Section II we give a short account of the "clean" channel capacity in the low-energy limit, and present pulse-position modulation (PPM) as a way to aid the task of coding for that channel by introducing useful structure into the code.

Highly energy-efficient communication is inherently bandwidth-inefficient. Even in free-space communications where bandwidth may be abundant, it is still practically limited by the switching speed of the transmitter and receiver used. Thus, if both high transmission rates and high photon efficiency are required, one needs to resort to multiple modes; the most evident source for this is space. However, multiplexing the data over multiple parallel modes requires either very large spatial separation (thus large apertures), or the use of high-order orthogonal beams that are difficult to produce and degrade easily over poor atmospheric conditions. In Section III we propose to solve this problem using simple dense (thus non-orthogonal) Gaussian beams; by using spatial PPM modulation, interference between the beams is transformed into a simple noisy channel. We present the fundamental hardware tradeoffs for this solution.

In Section IV we turn to key distribution over optical channels, still with high photon efficiency. The basic principles of quantum mechanics help facilitate secure communications, see e.g. [5]. When applied to the optical channel, the "no-cloning" principle means that a photon detected at the destination could not be also detected by an eavesdropper. Furthermore, the terminals can detect active attacks, i.e., photons detected by the eavesdropper and then "replaced" by new ones. Popular key distribution algorithm are based on these principles, but typically use polarization, thus have limited photon efficiency. We consider an alternative where timing information is used

This work was supported in part by the DARPA InPho program under Contract No. HR0011-10-C-0159 and ARO Grant No. W911NF-10-1-0416, and by AFOSR under Grant No. FA9550-11-1-0183.

to generate the key. For this, we find the optimal key generation rate, and show how to construct a PPM-based coding scheme that is optimal in the limit of low photon count. Section V concludes the paper by discussing extensions to the work.

II. Background: Photon Efficiency Via PPM

We consider a discrete-time lossy bosonic channel, which serves as a good model for free-space optical communications. This channel is best described in terms of its effect on *coherent* states of light, which are the states generated by a classical laser. One of the important properties of a coherent state $|\bar{n}\rangle$, is that when fed to a photodetector it will produce a Poisson number of "clicks" with mean $\bar{n}$, see e.g. [6]. When the input of a bosonic channel is a coherent state $|\bar{n}\rangle$, the output is a coherent state $|\eta\bar{n}\rangle$, where $0 < \eta \leq 1$ is the transmissivity of the channel. We see, then, that the Poisson statistics of the photon count is maintained by the channel.

The classical-information capacity of a quantum channel is given by the Holevo rate [7]. For a bosonic channel under a mean photon-count constraint $\bar{n}$, this is given by $g(\eta\bar{n})$ nats per channel use, where: [8]

$$g(x) \triangleq (x+1)\log(x+1) - x\log(x). \tag{1}$$

Furthermore, this capacity is achieved by transmitting coherent states. We will assume throughout the work the use of these states. As a consequence, we may without loss of generality constrain the mean photon count at the channel output rather than input, or equivalently take $\eta = 1$. It is important to note, that while the Holevo capacity is achievable using a "classical" encoder, the decoder must still be "quantum", i.e., some general quantum measurement must be performed jointly over the channel output corresponding to multiple channel uses in order to approach the capacity.

In some applications, it is interesting to consider the *photon efficiency*, i.e., the information conveyed per transmitted photon. Interestingly, in the low-energy limit $\bar{n} \to 0$, while the capacity $g(\bar{n})$ approaches zero, the photon efficiency $g(\bar{n})/\bar{n}$ approaches infinity. In fact,

$$\frac{g(\bar{n})}{\bar{n}} = \log\frac{1}{\bar{n}} + 1 + o(1). \tag{2}$$

A simple scheme for high photon efficiency consists of the encoder using a binary code. We denote the probability of logical "1" in that code by p. The states $|\bar{n}/p\rangle$ and $|0\rangle$ are sent to represent "1" and "0", respectively. the receiver will declare logical "1" if the detector clicked at least once, "0" otherwise. Per pulse, a classical Z channel is created, where the probability to receive "0" given that "1" was transmitted is $\exp\{-\bar{n}/p\}$, according to the Poisson statistics. If a classical capacity-approaching code is used over this channel, a rate of

$$h_B\left(p\cdot\left(1-\exp\left\{-\frac{\bar{n}}{p}\right\}\right)\right) - p\cdot h_B\left(\exp\left\{-\frac{\bar{n}}{p}\right\}\right)$$

can be achieved, where $h_B(\cdot)$ is the binary entropy function. The exact analytical optimization of this rate over p is complicated, but in the high-efficiency regime the approximate optimum (which yields the best rate up to the approximation of interest) is given by:

$$p^*(\bar{n}) = \frac{\bar{n}}{2}\log\frac{1}{\bar{n}}. \tag{3}$$

Note that when a Z channel is used without constraints, the optimal input distribution satisfies $p > \exp(-1)$; however in our case, where the photon efficiency requirement translates into a constraint on the input, $p^*(\bar{n})$ can be arbitrarily small. The resulting efficiency is given by:

$$\frac{R_Z(\bar{n})}{\bar{n}} = \log\frac{1}{\bar{n}} - \log\log\frac{1}{\bar{n}} + \log 2 + o(1). \tag{4}$$

Comparing to the Holevo rate (2), we see that the efficiency loss of the Z-channel scheme with respect to the optimal performance grows as $\log\log 1/\bar{n}$ in the high-efficiency limit. This loss is inherent to any "classical" transmission scheme, even if general (non-binary) coherent states are considered, and the receiver is allowed to use feedback between measurements; see [9].

While the scheme described above is implementable, the task of coding is still difficult. Specifically, one needs mutual-information approaching codes for a Z channel with a highly skewed input. This can be overcome by yet another simplification. Every k binary symbols are grouped together. We now impose the constraint that any such super-symbol will include exactly one entry that is "1", i.e., they become PPM symbols. In the limit of large k, the channel input resembles a highly-skewed random binary input. Indeed, using a uniform prior over these PPM symbols, one may achieve efficiency

$$\frac{R_{PPM}(\bar{n})}{\bar{n}} = \frac{\log k}{\bar{n}k}\cdot\left(1-\exp\left\{-\bar{n}k\right\}\right). \tag{5}$$

With the optimum (to the approximation order) $k = 1/p^*(\bar{n})$ (3), this efficiency satisfies (4), i.e., the further efficiency loss incurred is $o(1)$. The channel for which one needs to code is now a large-alphabet erasure channel, much like a packet-erasure channel encountered

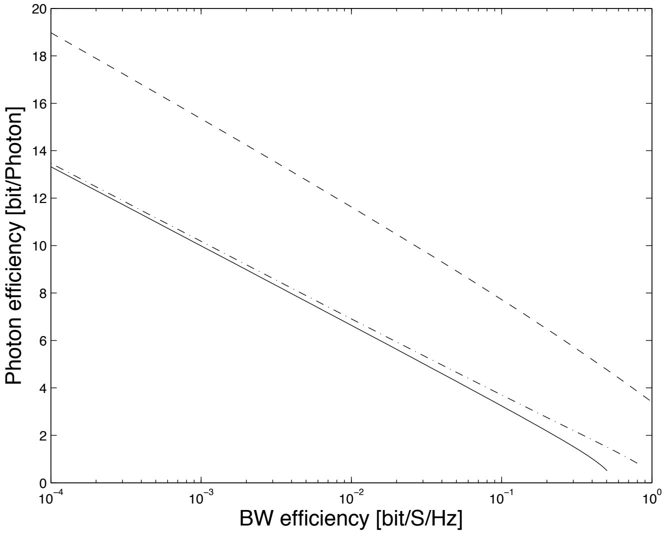

Fig. 1: Tradeoff between photon and bandwidth efficiencies. Dashed curve is the Holevo bound, dash-dotted is the Z-channel model, solid is PPM. The first satisfies the asymptotic expression (2), while the later two satisfy (4).

in internet applications, and good off-the-shelf codes are available.

Figure 1 depicts the tradeoff between photon and bandwidth efficiency, for the different rate expressions presented in this section. The BW efficiency is the rate, while the photon efficiency is the same rate divided by the average photon number $\bar{n}$. It can be appreciated that the tradeoff is inherent, and that, while the loss of "classical" operation (the Z channel model) is large (for fixed photon efficiency, it may be orders of magnitude in BW efficiency), the further loss of PPM is small. However, if one is concerned with this additional loss, it can be reduced by introducing an additional super-symbol to the PPM alphabet, where all entries are zero; see [10]

III. Spectral Efficiency via Spatial PPM

The Holevo-rate expression (2) represents a fundamental tradeoff: if high photon efficiency is required, the spectral efficiency $g(\bar{n})$, measured in nat per channel use, or in continuous time nat/S/Hz, must be low. In applications requiring both high photon- and spectral-efficiency, then, the premise of the problem must be changed. That is, degrees of freedom must be added in a domain other than time or frequency. One possibility is polarization, but we choose not to consider it as it cannot improve the spectral efficiency by more than a factor of two. We turn, then, to the *spatial domain*. Obviously, if we run m independent transmitter-receiver pairs in parallel, the spectral efficiency may be improved by a factor m without effecting the photon efficiency. For the system designer, such a solution implies additional hardware and space requirements; in the following, we quantify these costs and show how to reduce them using spatial modulation, rather than straightforward realization of multiple parallel time-modulated modes.

Let the required photon and spectral efficiencies be s nat/photon and R nat/S/Hz, respectively. Assuming PPM streams, the number of parallel modes needed is given by the solution m to

$$R_{PPM}\left(\frac{s}{m}\right) = \frac{R}{m}. \tag{6}$$

We are interested in the regime $m \gg 1$, and assume for simplicity that m is an integer. The number of degrees of freedom in an aperture is found by prolate spheroidal function analysis [11]. It turns out that the relevant geometry to the "spatial bandwidth" is encapsulated in the *Fresnel-number product* (see, e.g. [12])

$$\Gamma \triangleq \frac{A_t A_r}{(\lambda L)^2}, \tag{7}$$

where A_t and A_r are the transmit and receive apertures, respectively, λ is the carrier wavelength and L is the distance between transmitter and receiver. A large value of Γ translates to hardware cost, e.g., larger aperture. However, in order to support m modes, $\Gamma \geq m$ is required; that is, the total achievable rate is at most the rate per single mode times Γ.

A straightforward optimal implementation would include:

1) m parallel laser-detector pairs
2) Modulation/demodulation in order to carry the parallel lasers over an orthogonal set of modes, e.g. Hermite-Gaussian ones.

The implementation of a scheme which includes many light sources and modulation has a very high cost. However, we observe that in the high photon-efficiency regime the orthogonal approach may not be needed: with high probability, only a small portion of the modes is active, thus the potential interference is low. We use this to make the following modifications in the system.

1) Instead of using multiple parallel modes, over each of which a PPM-k constellation is sent, we can group k modes together and send the constellation over these modes at a single channel use. Consequently, we are assured that only one of these modes is "active" at any given instant. We can thus replace k sources by a single source and a switch which

directs its output according to the data. We coin this transmission method *spatial PPM*.

2) We abandon orthogonal modes. Instead, each mode is simply a laser beam, closely approximated by a Gaussian intensity profile, and different modes are separated by some translation. Since these modes are not orthogonal, some inter-mode interference is created; however, the use of spatial PPM, enables to translate this effect to noise within a single PPM signal.

The intensity of a Gaussian beam of unit power and width σ is given by: (see e.g. [6])

$$I(\mathbf{z}) = \frac{2}{\pi\sigma^2} \exp\left\{-\frac{2\|\mathbf{z}\|^2}{\sigma^2}\right\}, \tag{8}$$

where $\mathbf{z}$ is the two-dimensional displacement from beam center, that is, if we place a detector which covers an infinitesimal area ds around location $\mathbf{z}$, the average number of detected photons will be $I(\mathbf{z}) \cdot ds$.[1] The beam width is related to the geometry of the system via:

$$\sigma^2 = \frac{\pi}{4} \cdot \frac{(\lambda L)^2}{A_t} = \frac{\pi}{4} \cdot \frac{A_r}{\Gamma}. \tag{9}$$

Due to the noise created by the non-orthogonality of beams, we can no longer hope to achieve the rate R using Fresnel-number product and number of detectors that equal m. Rather, we have higher Fresnel-number product (achievable by e.g. larger apertures) $\Gamma = \alpha^2 m$ and larger number of detectors $\beta^2 m$; our goal is to characterize the (α,β) pairs which support (s, R).

We make the following assumptions, that greatly simplify the analysis:

1) Each detector occupies a square area, and they fill a Cartesian grid (this reflects a small performance loss with respect to an array of hexagonal detectors).
2) The number of detectors is very large, thus we may ignore edge effects and treat each detector as a part of an infinite grid. The constellation size is also very large: $m, k \gg 1$.
3) If there is more than one detection, the receiver only makes use of the first one.

Proposition 1: Under the assumptions above, photon efficiency s and spectral efficiency R can be supported by Fresnel-number product $\alpha^2 m$ and number of detectors $\beta^2 m$, where m is found via (6) and (α,β) satisfy:

$$\log\beta \geq H(Z). \tag{10}$$

In this expression Z is a random variable over the integers, satisfying:

$$\begin{aligned}\Pr\{Z = z\} =& Q\left((2z-1)\frac{2}{\sqrt{\pi}} \cdot \frac{\alpha}{\beta}\right)\\ &-Q\left((2z+1)\frac{2}{\sqrt{\pi}} \cdot \frac{\alpha}{\beta}\right),\end{aligned}$$

where $Q(\cdot)$ is the complementary Gaussian cumulative distribution function.

Proof: The optimal transmitter directs a beam to the center of one of the detectors. Under the first assumption, the distribution of the detector having the first click is just as in quadratic amplitude modulation (QAM) with hard decision. The distance between adjacent QAM "constellation points" is found using (9):

$$d = \frac{2}{\sqrt{\pi}} \cdot \frac{\alpha}{\beta} \cdot \sigma. \tag{11}$$

The difference between the intended detector and the detector with the first click is independent between the axes, and in each axis is distributed as Z defined in the proposition. For a constellation of size k, the average number of photons is $k\bar{n}/m$. If the detectors corresponding to the constellation exactly overlap a receive array, the corresponding rate is the probability of detection multiplied by the QAM rate given that a detection has occurred:

$$R_G = \left(1 - \exp\left\{-\frac{k\bar{n}}{m}\right\}\right)\left(\log\beta^2 k - 2H(Z)\right).$$

We note that the detectors corresponding to a constellation may actually only occupy part of the aperture or span multiple ones, but this has no effect in the limit $k \gg 1$.[2] If the modes were orthogonal, the same aperture area would support a rate of

$$R = k \cdot R_{PPM}\left(\frac{\bar{n}}{m}\right),$$

where R_{PPM} is given by (5). Requiring $R_G = R$ in order to maintain the efficiencies translates to the condition (10). ■

We can analytically consider the pairs (α,β) on the boundary of the region (10) in two opposite limits. If

[1] Note that the standard deviation of the "location" of a click is $\sigma/2$ rather than the more intuitive σ; this is the common way used to describe Gaussian beams in Physics literature.

[2] Namely, if two constellations share the same aperture, we may keep a sufficient guard area between them, so that the interference between them has arbitrary small level, with vanishing cost as k grows. Similarly, the fact that a constellation may be split between apertures and thus the noise cannot carry the detection to any constellation point has a vanishing gain.

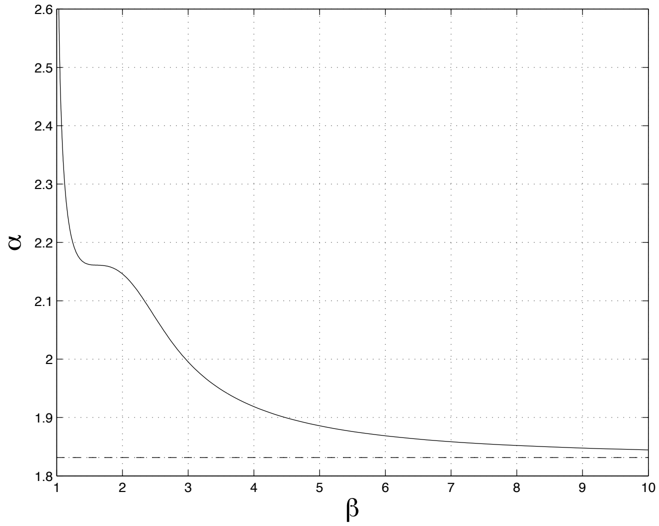

Fig. 2: Minimum Fresnel-number product redundancy α as a function of the detector number redundancy β. Dashed line depicts the asymptotic value α_0.

$\alpha \ll \beta$, Z is a high-resolution version of Gaussian noise, thus

$$H(Z) \cong \frac{1}{2}\log\frac{\pi\exp\{1\}\sigma^2}{2} - \log d = \frac{1}{2} + \log\frac{\pi\beta}{2\sqrt{2}\alpha}$$

thus (10) yields:

$$\alpha = \frac{\pi}{2\sqrt{2}}\exp\frac{1}{2} \triangleq \alpha_0 \cong 1.83.$$

This is a minimum value for α at any β, meaning that there must be a loss in the Fresnel-number product with respect to orthogonal modes, even if we use many small detectors. If, to the contrary, $\alpha \gg \beta$, then Z may have very low entropy. Thus in the limit of a very large aperture (almost orthogonal beams) we can have β close to one, representing approximately m detectors, as expected. Figure 2 depicts a numerical evaluation of the tradeoff (10). The rather strange curve is due to the shape of the Gaussian distribution. A working point that seems to be good for practical purposes is $(\alpha,\beta) = (2.15, 1.4)$. At this point, the radii of the transmit and receive apertures need to be multiplied by a factor 1.5 each, comparing to the ideal orthogonal system, and the number of detectors need to be doubled comparing to the same.

Finally, we note that there is a potential gain from considering more than one detection. While with orthogonal beams the extra detections cannot help, with Gaussian beams they can be used to reduce the noise. Let the number of detections be t, then we have t observations of the chosen location, with i.i.d. noises $Z_1, \ldots, Z_t$. Since t is known at the decoder, the rate in Proposition 1 is replaced by the mean mutual information given the number of detections t, where t has Poisson distribution with rate $\bar{n}$. In the high-resolution limit $\alpha \ll \beta$, the decoder averages the detections and the entropy is reduced by $E\{\log t\}$. Outside this limit we have a discrete additive channel with multiple looks. The exact capacity gain is an interesting information-theoretic problem; for $t = 2$, the noise entropy $H(Z)$ is replaced by the smaller

$$2H(Z) - H(\tilde{Z}),$$

where the distribution of $\tilde{Z}$ is that of the sum of two independent random variables, each one distributed as Z.

IV. Key Distribution: Parsing a Poisson Process

In quantum key distribution (QKD), Alice and Bob want to agree upon a message (“key”), while keeping the information that Eve has about the message close to zero. This is carried over in two stages.

1) Quantum stage. Following [13], we distinguish between two models. In model C, Alice prepares states, which are observed by Alice and Bob through a quantum channel. In model S, a source emits random states, which are observed by all three nodes.
2) Classical stage. In this stage, Alice and Bob exchange information over a classical channel. Eve is a passive listener that can decode all of this information.

Photon and spectral efficiencies are defined according to the key rate, and the photon and bandwidth resources consumed by the quantum stage. The classical stage is assumed to have no cost, still communication will be kept to a minimum since Eve has unlimited access.

Various QKD algorithms have been proposed. The most popular ones, BB84 and E91 (see e.g. [5], use polarization for key generation. However, such methods are limited to one bit per photon. For high photon efficiency, one may use time or frequency which may offer more degrees or freedom. We thus take an approach similar to the one used for communications in the previous sections.

For model C, Alice can choose random information and transmit it using a PPM communication scheme over a bosonic channel with efficiency η. In an implementation of model S, the source, co-located with Alice, emits entangled photon pairs according to a Poisson process with rate of γ pairs per second. One of the photons is sent locally to Alice and is always detected; the other is sent to Bob over the bosonic channel. In principle both models lead to similar results. We choose to analyze model S; at the end of this section we comment about model S.

We formalize the model in discrete time. We use short time intervals τ such that $\bar{n} = \gamma\tau$. Alice and Bob use photon detection to obtain sequences A and B, respectively, representing the number of detections in each interval. Eve can make any measurement, or even perform an active attack, e.g., detect photons, then create new ones and send them to Bob with low additional delay. We formally prove security for the case where Eve obtains a sequence E in a similar manner to the sequences A and B. Proving that this implies quantum security is beyond the scope of this work, but we do note the following points.

- Continuous vs. discrete time. If Alice makes a measurement that is optimal in the sense of time-frequency uncertainty, i.e., uses narrow-band filters of width Δ_ω such that $\tau\Delta_\omega = 1/2$, then Eve cannot gain from measurements that have better resolution than τ.
- General measurement vs. photon detection. The photon-detection measurement by Alice causes the state that Eve may observe collapse to a number state. Consequently, Eve can not gain from any measurement other than detecting the presence of a photon in a time bin. In principle Eve could do that using a nondemolition measurement; that would detected by the same means as an active attack.
- Active attack. Such an attack may be detected by either interferometry, or alternating between measurement bases (high-reolution time and high-resolution frequency) similar to what is done in entanglement-based QKD. These methods do require that some of the photon pairs are dedicated to the task and not used for key generation; we ignore these photons, i.e., normalize efficiencies w.r.t. photons that are used for key generation.

Are model is thus defined as follows. A is an i.i.d. Poisson sequence with average count $\bar{n}$. Given that an element of A has value a, the corresponding element of B is binomial with parameters (a,η); the sequence E is the difference between A and B (this is a worst-case assumption: all losses are due to Eve). Alice and Bob need to agree upon a key K of rate R using public messages, which Eve also receives. As the blocklength increases, the probability that they agree of the same key should approach one, while the mutual information (per element) between Eve's data and the key must approach zero.

Theorem 1: The secret-key rate for the problem defined above is given by:

$$R_K = I(A;B)$$

Proof: Achievability. We describe the public communication stage. It consists of two steps. In the first, Bob sends Alice information such that with high probability Alice knows B. By the Slepian-Wolf (SW) Theorem [14], this can be done with rate $R_{SW} = H(B|A)$. By the assumed statistics, the sequence E os is independent of the sequence B. Thus, after the SW step Eve has information about B of (curly brackets denote the sequences, and the blocklength is l):

$$\begin{aligned} I_E = I(\{E\}, M; \{B\}) &= I(\{E\};\{B\}) + I(M;\{B\}|\{E\}) \\ &= I(M;\{B\}|\{E\}) \leq H(M) = l \cdot H(B|A). \end{aligned}$$

The second classical communication step will be secrecy amplification, producing from B a key of rate (see e.g. [15]):

$$H(B) - \frac{I_E}{l} \geq H(B) - H(B|A) = I(A;B) = R_K.$$

Converse. By the upper bound on key generation rate in [13] (see there remark 2 after Theorem 2, where we take V to be independent of (A,B,E)) to assert that the rate can not exceed $I(A;B)$ ■

In the limit of low average photon number $\bar{n}$ we can neglect the probability of two photon pairs within the same interval. Thus all sequences become binary: A is Bernoulli $(\bar{n})$; Given $A = 0$, $(B,E) = (0,0)$. Given $A = 1$, $(B,E) = (1,0)$ w.p. η, $(B,E) = (0,1)$ otherwise. The mutual information becomes:

$$I(A;B) = H_b(\eta\bar{n}) - \bar{n}H_b(\eta)$$

and the photon efficiency is given by:[3]

$$\frac{R_K(\bar{n})}{\eta\bar{n}} = \log\frac{1}{\bar{n}} + 1 - \frac{1-\eta}{\eta}\log\frac{1}{1-\eta} + o(1). \quad (12)$$

Favorably, this reflects no $\log\log 1/n$ loss; in fact, for $\eta = 1$ it coincides with (2). However, Slepian-Wolf coding as well as amplification need to be carried out

[3] Similar to the communication setting where we define photon efficiency w.r.t. the photons arriving at the receiver, here we define it w.r.t. pairs that arrive at both Alice and Bob

over heavily skewed binary sequences, which may be a very difficult task. To overcome this we use a PPM approach, similar to the one described in Section II. Since we cannot control the source, the actual transmission is not PPM, but we can still *parse* the sequences (A, B) into blocks of length k that typically contain at most one photon-pair.

In the most naive approach, Alice and Bob first search for blocks in which each of them has exactly one detection, and then use the indices within the block to form the key. Since the location of the blocks is independent of the key, the key rate per such block is just $\log k$, and there is no need for SW coding or amplification. The probability that a block is usable is given by $\eta k \bar{n} \exp\{-k\bar{n}\}$, thus the efficiency is given by:

$$\begin{aligned}\frac{R_{K-PPM}(\bar{n})}{\eta\bar{n}} &= \max_k \left[(1-\exp\{-k\bar{n}\})\log k\right] \\ &= \log\frac{1}{\bar{n}} - \log\log\frac{1}{\bar{n}} - 1 + o(1), \quad (13)\end{aligned}$$

where the maximizer (to the approximation needed) is $k = 2/p^*(\bar{n})$ where p^* was defined in (3).

We see that PPM parsing inflicts the $\log\log(1/n)$ loss, similar to the communication efficiency (4). This happens because Alice and Bob ignore part of the common randomness they obtain via the channel, namely the number of detections per block and the location of detections beyond the first one. As the analysis that follows shows, the information that needs to be used in order to recover most of the loss is whether Bob had any detection in a block or not. The following two-stage strategy efficiently extracts this information.

1) Bob sends Alice a SW code describing the indices of blocks in which he has at least one detection. The first part of the key is generated by applying secrecy amplification to this sequence.
2) Alice replies by stating in which of the blocks where Bob has a detection, she has exactly one detection (thus Bob must have exactly one as well). The second part of the key is generated from the indices of the detection within these blocks, as with the original PPM parsing.

As the second part of the key is statistically independent of the first part and of all the information sent over the public channel, we can find the key rate by evaluating the additional rate in the first part of the key.

For calculating the additional key rate, let the number of detections per block that Alice and Bob have be N_A and N_B, respectively. Let I_B be an indicator of the event $N_B > 0$. By arguments similar to the proof of Theorem 1, a rate of $\Delta R = I(N_A; I_B)$ may be achieved. We use as a lower bound a Z channel with $\Pr\{N_A = 1\} = k\bar{n}\exp\{-k\bar{n}\}$ and $\Pr\{I_B = 1 | N_A = 1\} = \eta$ as in the original channel (thus we collapse all $N_A > 1$ to $N_A = 0$, reducing the mutual information). We have, then, additional photon efficiency of:

$$\begin{aligned}\frac{\Delta R}{\eta k\bar{n}} &\geq \frac{1}{\eta k\bar{n}} H_b\left(\eta k\bar{n}\exp\{-k\bar{n}\}\right) - \frac{\exp\{-k\bar{n}\}}{\eta} H_b(\eta) \\ &= \log\log\frac{1}{\bar{n}} + 1 - \frac{1-\eta}{\eta}\log\frac{1}{1-\eta} + o(1), \quad (14)\end{aligned}$$

where the calculation was carried out with $k = 2/p^*(n)$ as in (13). Comparing (12)-(14), we have:

$$\frac{R_{K-PPM}(\bar{n})}{\eta\bar{n}} + \frac{\Delta R}{\eta k\bar{n}} \geq \frac{R_K(\bar{n})}{\eta\bar{n}} - 1 + o(1),$$

showing that we have eliminated the $\log\log 1/\bar{n}$ loss. A constant efficiency loss of one nat per photon remains; in order to also close this gap, while remaining within the parsing framework, one would have to extract information from multiple detections within the same block, complicating the algorithm considerably.

Extracting the additional rate ΔR does require SW coding, as in the original Z-channel model. However, the sequences are far less skewed, making the coding task easier. In fact, the probability of "1" in the sequence A behaves as

$$\frac{1}{\log\frac{1}{\bar{n}}},$$

cf. $\bar{n}$ originally. At reasonable values of $\bar{n}$, the sequence A may be almost balanced.

We conclude this section by commenting on the implementation of model C, where entanglement is not used. For that case, quantum security holds for similar arguments to those presented above for model S, except that interferometry cannot be used; rather, Alice will must randomly use PPM either in time or frequency, and Bob will measure in a random basis, similar to both of them measuring in random basis for model S. Further, for extracting the additional rate ΔR, Alice needs to have side information, specifying in which PPM blocks at least one photon was sent to the channel; this may be obtained in principle by a nondemolition quantum measurement, but it is not known in practice how to implement such a measurement. An alternative that is a hybrid between models S and C is to have a source of entangled pairs with variable Poisson rate. According to randomness generated by Alice it will be switched to generation rate $\bar{n}/k$ for one bin in every PPM block, and will be kept to zero otherwise.

V. Future Directions

In Section III we presented spatial PPM as a simple way to achieve spectral efficiency without sacrificing photon efficiency. In Section IV we suggested PPM parsing of a Poisson source of entangled photon pairs, for key distribution. A natural goal would be to connect both: in order to achieve high spectral efficiency in key distribution, one would need multiple spatial modes.

The channel model we have analyzed is highly idealized. In particular, the following two effects have a significant impact on performance and coding.

1) Coupling and detection losses. These are additional to the path loss η. In general they can be included in η, except that in the key distribution scenario we have assumed that Alice has perfect access to the source. When Alice also has losses, the optimal key rate is still $I(A;B)$. This rate has now an additional constant asymptotic loss w.r.t. (12). Furthermore, a variant of the PPM parsing algorithm can be still used, and the asymptotic gap from the optimal efficiency is still one nat per photon. However, the second part of the key cannot be simply Bob's indices, since they may differ from the ones that Alice has in cases where the source generated multiple pairs, and each terminal received a photon from a different one; SW coding is now needed for this stage as well.
2) Dark current. These are detections that are independent of transmission. Although they typically arrive at a very slow rate, they will dominate performance in the limit of high photon efficiency (slow rate of true detections). In the context of spatial PPM, detections may appear in a detector independent of the intended one (instead or in addition to the true one); the code over the PPM symbols should account now for two kinds of errors: small (due to the Gaussian beams) and large (due to the dark current). In the key distribution setting, the effect of dark current is similar to that of both users having efficiency smaller than 1, discussed above.

Finally, although the PPM approach is a useful tool in simplifying the task of coding, we have still not considered specific codes. In the SW setting, low-density parity check (LDPC) codes may be useful, see e.g. [16].

Acknowledgement

The authors would like to thank Ligong Wang and Jeffrey H. Shapiro for helpful discussions.

References

[1] K. Abend. Optimum photon detection (corresp.). *IEEE Trans. Info. Theory*, 12(1):64–65, jan 1966.

[2] I. Bar-David. Communication under the Poisson regime. *IEEE Trans. Info. Theory*, 15(1):31–37, jan 1969.

[3] S. Shamai and A. Lapidoth. Bounds on the capacity of a spectrally constrained Poisson channel. *IEEE Trans. Info. Theory*, 39(1):19–29, jan 1993.

[4] A. Lapidoth and S.M. Moser. On the capacity of the discrete-time Poisson channel. *IEEE Trans. Info. Theory*, 55(1):303–322, jan. 2009.

[5] M. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Inormation*. Cambridge University Press, 2002.

[6] L. Mandel and E. Wolf. *Optical Coherence and Quantum Optics*. Cambridge University Press, Cambridge, UK, 1995.

[7] A. S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Trans. Info. Theory*, 44(1):269–273, jan 1998.

[8] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, J. H. Shapiro, and H. P. Yuen. Classical capacity of the lossy bosonic channel: The exact solution. *Phys. Rev. Lett.*, 92, Jan 2004.

[9] H. W. Chung, S. Guha, and L. Zheng. On capacity of optical channels with coherent detection. In *Proceedings of ISIT-11, St. Petersburg, Russia*, pages 284–288, 2011.

[10] S. Guha, J. H. Shapiro, and Z. Dutton. Addressing the ultimate limits of photon-efficiency vs. spectral-efficiency tradeoffs for the multiple-spatial-mode free-space optical communication channel. In *Proc. Updating Quantum Cryptography and Communications (UQCC) 2010, Tokyo*, 2010.

[11] D. Slepian. Prolate spheroidal wave functions, Fourier analysis and uncertainty-IV: Extension to many dimensions; generalized prolate spheroidal functions. *Bell Syst. Techn. J.*, 43, 1962.

[12] H. P. Yuen and J. H. Shapiro. Optical communication with two-photon coherent states, part I: Quantum state propagation and quantum noise reduction,.

[13] R. Ahlswede and I. Csiszár. Common randomness in information theory and cryptography. I. secret sharing. *IEEE Trans. Info. Theory*, 39(4):1121 –1132, jul 1993.

[14] D. Slepian and J.K. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Info. Theory*, IT-19:471–480, July 1973.

[15] U. Maurer and S. Wolf. Information-theoretic key agreement: From weak to strong secrecy for free. In *Lecture Notes in Computer Science*, pages 351–368. Springer-Verlag, 2000.

[16] T.P. Coleman, A.H. Lee, M. Medard, and M. Effros. Low-complexity approaches to Slepian-Wolf near-lossless distributed data compression. *IEEE Trans. Info. Theory*, 52(8):3546–3561, aug. 2006.