Toward Photon-Efficient Key Distribution Over Optical Channels

Yuval Kochman, *Member, IEEE*, Ligong Wang, *Member, IEEE*, and Gregory W. Wornell *Fellow, IEEE*

***Abstract*— This paper considers the distribution of a secret key over an optical (bosonic) channel in the regime of high photon efficiency, i.e., when the number of secret key bits generated per detected photon is high. While, in principle, the photon efficiency is unbounded, there is an inherent tradeoff between this efficiency and the key generation rate (with respect to the channel bandwidth). We derive asymptotic expressions for the optimal generation rates in the photon-efficient limit, and propose schemes that approach these limits up to certain approximations. The schemes are practical, in the sense that they use coherent or temporally entangled optical states and direct photodetection, all of which are reasonably easy to realize in practice, in conjunction with off-the-shelf classical codes.**

***Index Terms*— Information-theoretic security, key distribution, optical communication, wiretap channel.**

I. Introduction

INFORMATION-THEORETIC key distribution [1], [2] involves the generation of a sequence between the participating terminals, such that the mutual information between this sequence and any data obtained by other terminals is close to zero in an appropriate sense. Unlike secure communication through the wiretap channel [3], the sequence need not be known *a priori* to any of the terminals. Like the latter, however, the information-theoretic approach to key distribution hinges on knowledge of the channel through which an adversarial terminal listens to the communication, as opposed to computational approaches where the assumption is the inability of the adversary to perform certain computations in reasonable time. The computational hardness assumption may no longer be valid when future technology, e.g., quantum computers, becomes available, causing the computational approaches to fail. But the information-theoretic approach also has its drawback: the information obtained by the legitimate terminals cannot prove or disprove the channel assumption on which the key-distribution protocol is based, inhibiting security in a realistic setting.

The situation is much different when a quantum channel is employed [4], [5]. Loosely speaking, the "no-cloning" theorem [6] guarantees that information "stolen" by an eavesdropper will not reach the legitimate terminal, thus the situation where the adversary is stronger than initially assumed can be detected. In fact, even eavesdroppers that can actively transmit into the quantum channel can be detected, at the cost of key-rate loss, using measurements based on local randomness. We shall come back to these issues in the discussion at the end of the paper. For the main part of the paper, we rely on the existence of good detection methods to assume that the eavesdropper is passive, and that the complete statistical characterization of the eavesdropper's channel is known to the legitimate terminals.

Two-terminal quantum key distribution (QKD) protocols can be roughly divided into two classes. In "prepare and measure" protocols, one legitimate terminal (Alice) prepares quantum states that are sent via a quantum channel to the other terminal (Bob) and to the eavesdropper (Eve). In contrast, in entanglement-based protocols, a quantum source emits entangled states, which are observed by all terminals via quantum channels. These two classes are parallel to the "C" (channel) and "S" (source) models of [1]; in this work we shall use the C/S notation. In either approach, the quantum stage is followed by the use of a classical communication channel. This channel is assumed to be public, i.e., all information sent is received by Eve; however, it is assumed that Eve cannot transmit into this public channel. The performance of a QKD scheme is measured in terms of the size of the secret key normalized by the quantum-channel resources used. The classical channel is thus "free", although its use is limited by the assumption that Eve has full access to this channel.

A quantum channel most often encountered in practice is the optical channel, which is modeled in quantum mechanics as a bosonic channel. When used for communicating classical data at low average input power, it is asymptotically optimal to use a direct-detection receiver, which ignores the phase of the optical signal. This results in an equivalent classical channel where the output has a Poisson distribution whose mean is proportional to the channel's input [7]. Some of the first important works on this channel model are in [8]–[10]. The low-input-power regime can be thought of as a "photon-efficient regime". This is because, in the limit of low average photon number per channel use, the communication rate per photon is unbounded.

Manuscript received July 19, 2013; revised February 11, 2014; accepted May 30, 2014. Date of publication June 16, 2014, date of current version July 10, 2014. This work was supported in part by the DARPA InPho Program under Contract HR0011-10-C-0159 and in part by the Air Force Office of Scientific Research under Grant FA9550-11- 1-0183. Y. Kochman was supported by the Israel Science Foundation under Grant 956/12.

Y. Kochman was with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA. He is now with the School of Computer Science and Engineering, Hebrew University of Jerusalem, Jerusalem 91905, Israel (e-mail: yuvalko@cs.huji.ac.il).

L. Wang and G. W. Wornell are with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: wlg@mit.edu; gww@mit.edu).

Communicated by A. Holevo, Associate Editor for Quantum Information Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2014.2331060

In this work we consider QKD over the bosonic channel in the photon-efficient regime. We consider both C and S models, and show that in both, as happens in communication, the photon efficiency is unbounded and direct-detection receivers are asymptotically optimal. We further consider specific QKD protocols. We discuss the difficulty of finding code constructions that allow us to approach the theoretical performance limits, since in the photon-efficient regime they have to operate over highly-skewed sequences. We present protocols that overcome this difficulty: in the C model we use pulse-position modulation (PPM), while in the S model we parse the sequence of detections into frames. In both cases, coding over frames is an easier task than coding directly over the detection sequence.

The rest of the paper is organized as follows. We introduce our notation in Section II. In Section III we formally describe the problem setting. Then in Section IV we discuss, as a point of reference, photon-efficient communication. Sections V and VI include our main results for key distribution, regarding the C and S models, respectively. We conclude this paper in Section VII by discussing the gap between our results and fully quantum security proofs.

II. Notation

We use a font like $\mathbb{A}$ to denote a Hilbert space. Throughout this paper we shall focus on bosonic Hilbert spaces. We adopt Dirac's notation to use $|\psi\rangle$ to denote a unit vector in a Hilbert space, which can describe a pure quantum state, and use $\langle\psi|$ to denote the conjugate of $|\psi\rangle$. We follow most of the physics literature to slightly abuse our notation: we shall not make typographical distinctions between number states and coherent states. Hence $|n\rangle$, $n \in \mathbb{Z}_0^+$, (usually) denotes the number state that contains n photons; while $|\alpha\rangle$, $\alpha \in \mathbb{C}$, (almost everywhere) denotes a coherent state, whose exact characterization is given later. This abuse of notation will not cause confusion within the scope of this paper. We use a Greek letter like ρ to denote a density operator (i.e., a trace-one semidefinite operator) on a Hilbert space, which can describe a pure or mixed quantum state. Note that the density-operator description of a pure state $|\psi\rangle$ is $|\psi\rangle\langle\psi|$. When considering a system such as a beamsplitter, we reserve the letters $|\psi\rangle$ and ρ for input states, and $|\phi\rangle$ and σ for output states. Sometimes, to be explicit, we add a superscript to a state to indicate its Hilbert space so it looks like $|\psi\rangle^{\mathbb{A}}$ or $\sigma^{\mathbb{B}}$. We use the notation $\hat{a}$ to denote the annihilation operator on $\mathbb{A}$ (so $\hat{a}^\dagger$ is the creation operator on $\mathbb{A}$); similarly, $\hat{b}$ denotes the annihilation operator on $\mathbb{B}$, etc.

For a quantum state $\sigma^{\mathbb{AB}}$ on the Hilbert spaces $\mathbb{A}$ and $\mathbb{B}$, we use $H(\sigma^{\mathbb{A}})$, $H(\sigma^{\mathbb{A}}|\sigma^{\mathbb{B}})$, and $I(\sigma^{\mathbb{A}};\sigma^{\mathbb{B}})$ to denote the corresponding entropy, conditional entropy, and mutual information, respectively. These quantities are defined as follows (see [11] for more details):

$$H(\sigma^{\mathbb{A}}) \triangleq -\mathrm{tr}\left\{\sigma^{\mathbb{A}}\log\sigma^{\mathbb{A}}\right\} \tag{1}$$

$$H(\sigma^{\mathbb{A}}|\sigma^{\mathbb{B}}) \triangleq H(\sigma^{\mathbb{AB}}) - H(\sigma^{\mathbb{B}}) \tag{2}$$

$$I(\sigma^{\mathbb{A}};\sigma^{\mathbb{B}}) \triangleq H(\sigma^{\mathbb{A}}) + H(\sigma^{\mathbb{B}}) - H(\sigma^{\mathbb{AB}}). \tag{3}$$

For classical or mixed classical-quantum states, we simply replace the density operator by the classical random variable for the classical part in these expressions, so they look like, e.g., $H(X)$, $H(X|\sigma^{\mathbb{B}})$, and $I(\sigma^{\mathbb{A}};Y)$. Sometimes, to be more precise, we also write the mutual information as $I(\mathbb{A};\mathbb{B})|_\sigma$, indicating that it is the mutual information between space $\mathbb{A}$ and $\mathbb{B}$ evaluated for the joint state σ.

Throughout this paper, we use natural logarithms, and measure information in nats, though sometimes we do talk about "bits" and "binary representation".

We use the usual notation $O(\cdot)$ and $o(\cdot)$ to describe behaviors of functions of $\mathcal{E}$ in the limit where $\mathcal{E}$ approaches zero with other variables, if any, held fixed. Specifically, given a reference function $f(\cdot)$ (which might be the constant 1), a function denoted as $O(f(\mathcal{E}))$ satisfies

$$\overline{\lim_{\mathcal{E}\downarrow 0}}\left|\frac{O(f(\mathcal{E}))}{f(\mathcal{E})}\right| < \infty, \tag{4}$$

while a function denoted as $o(f(\mathcal{E}))$ satisfies

$$\lim_{\mathcal{E}\downarrow 0}\frac{o(f(\mathcal{E}))}{f(\mathcal{E})} = 0. \tag{5}$$

III. Problem Setting

In this section we describe our setups for optical communication and key distribution. To do so, we first recall some basic results in quantum optics.

A. Beamsplitting and Direct Detection

We briefly describe how *number (Fock) states* and *coherent states* evolve when passed through a beamsplitter, and what outcomes they induce when fed into a direct-detection receiver, i.e., a photon counter. We refer to [12] for more details. For some background in quantum physics and in quantum information theory, we refer to [11].

Let $\mathbb{A}$ and $\mathbb{V}$ be the two input spaces to a single-mode beamsplitter, and $\mathbb{B}$ and $\mathbb{E}$ be the two output spaces. Let the beamsplitter's transmissivity from $\mathbb{A}$ to $\mathbb{B}$ be $\eta \in [0,1]$. Then this beamsplitter is characterized in the Heisenberg picture by

$$\hat{b} = \sqrt{\eta}\,\hat{a} + \sqrt{1-\eta}\,\hat{v} \tag{6a}$$

$$\hat{e} = \sqrt{1-\eta}\,\hat{a} - \sqrt{\eta}\,\hat{v}. \tag{6b}$$

Throughout this paper we shall only consider situations where the second input space $\mathbb{V}$ (the "noise mode") is in its vacuum state $|0\rangle$.

Ideal direct detection (i.e., photon counting) measures an optical state in the number-state basis. For direct detection on $\mathbb{A}$, the observable is the Hermitian operator $\hat{a}^\dagger\hat{a}$. On state ρ, a photon counter gives outcome $n \in \mathbb{Z}_0^+$ with probability $\langle n|\rho|n\rangle$.

Obviously, when a number state $|n\rangle$, $n \in \mathbb{Z}_0^+$, is fed into an ideal photon counter, the outcome is n with probability one. But passing $|n\rangle$ through a beamsplitter is more complicated: if space $\mathbb{A}$ in (6) is in state $|n\rangle$, then the output state is an entangled state on $\mathbb{B}$ and $\mathbb{E}$:

$$|\phi\rangle^{\mathbb{BE}} = \sum_{i=0}^{n}\sqrt{\binom{n}{i}}\,\eta^{i/2}(1-\eta)^{(n-i)/2}|i\rangle^{\mathbb{B}}|n-i\rangle^{\mathbb{E}}. \tag{7}$$

This implies that performing direct detection on the output of this beamsplitter will yield a binomial distribution on the outcome: the probability of detecting m photons on space $\mathbb{B}$ is

$$\langle m|\sigma^{\mathbb{B}}|m\rangle = \binom{n}{m}\eta^m(1-\eta)^{n-m} \tag{8}$$

for $0 \le m \le n$, and is zero otherwise. It also implies that, if direct detection is performed both on $\mathbb{B}$ and on $\mathbb{E}$, then with probability one the sum of the two outcomes is equal to n.

A coherent state $|\alpha\rangle$, $\alpha \in \mathbb{C}$, can be written in the number-state basis as

$$|\alpha\rangle = e^{-|\alpha|^2/2}\sum_{n=0}^{\infty}\frac{\alpha^n}{\sqrt{n!}}|n\rangle. \tag{9}$$

Thus, when fed into a photon counter, the probability of n photons being observed in $|\alpha\rangle$ is

$$\langle n|\alpha\rangle\langle\alpha|n\rangle = |\langle n|\alpha\rangle|^2 = e^{-|\alpha|^2}\frac{|\alpha|^{2n}}{n!}. \tag{10}$$

Namely, the number of photons in $|\alpha\rangle$ has a Poisson distribution of mean $|\alpha|^2$.

Coherent states have the nice property that, when passed through a beamsplitter, the outcomes remain in coherent states. If $|\alpha\rangle$ is fed into the beamsplitter (6), the output state is

$$|\phi\rangle^{\mathbb{B}\mathbb{E}} = |\sqrt{\eta}\,\alpha\rangle^{\mathbb{B}} \otimes |\sqrt{1-\eta}\,\alpha\rangle^{\mathbb{E}}. \tag{11}$$

Therefore, if direct detection is performed both on $\mathbb{B}$ and on $\mathbb{E}$, the outcomes will be two *independent* Poisson random variables of means $\eta|\alpha|^2$ and $(1-\eta)|\alpha|^2$, respectively.

B. Optical Communication

A single-mode pure-loss optical (i.e., bosonic) channel can be described using the beamsplitter (6a), where we ignore the output space $\mathbb{E}$ and assume the noise space $\mathbb{V}$ to be in its vacuum state. In this formula, $\mathbb{A}$ is the input space controlled by the transmitter which, in consistency with the key-distribution part, we call Alice; $\mathbb{B}$ is the output space obtained by the receiver, Bob; and η is the transmissivity of the channel. Equivalently, the channel may be described in the Schrödinger picture as a completely-positive trace-preserving (CPTP) map from the input state $\rho^{\mathbb{A}}$ to the output state $\sigma^{\mathbb{B}}$:

$$\sigma^{\mathbb{B}} = \mathcal{C}(\rho^{\mathbb{A}}). \tag{12}$$

The explicit characterization of $\mathcal{C}$ is complicated and omitted.

We denote the blocklength of a channel code by k. Alice has a message of kR nats[1] to convey to Bob. In order to do this, she prepares a state ρ^k over $\mathbb{A}^k$, subject to an average-photon-number constraint $\mathcal{E}$ per channel use:

$$\mathrm{tr}\left\{\left(\sum_{i=1}^{k}\hat{a}_i^\dagger\hat{a}_i\right)\rho^k\right\} \le k\mathcal{E} \tag{13}$$

where $\hat{a}_i$ is the annihilation operator on the input space of the ith channel use. The channel is assumed to be memoryless, so the output is given by

$$\sigma^k = \mathcal{C}^{\otimes k}(\rho^k). \tag{14}$$

[1]We ignore the fact that the number of values that the message can take is not an integer.

Bob may perform any positive-operator valued measure (POVM) on σ^k to reconstruct the message. As usual, the capacity of the channel is defined as the supremum of rates for which there exist sequences of schemes with increasing blocklengths and with the error probabilities approaching zero.

We define the *photon efficiency* of transmission as the rate normalized by the expected number of photons that Bob receives per channel use:[2]

$$r(\eta,\mathcal{E}) \triangleq \frac{R(\eta,\mathcal{E})}{\eta\mathcal{E}}. \tag{15}$$

This quantity is upper-bounded by the channel's capacity divided by $\eta\mathcal{E}$.

C. Key Distribution Using an Optical Channel (Model C)

We next consider the problem where Alice and Bob use the channel of (6) to generate a secret key between them. The channel from Alice to Bob is still characterized by (6a) or by the CPTP (12), but we now assume that an eavesdropper, Eve, obtains the Hilbert space $\mathbb{E}$. Note that this is a worst-case assumption in the sense that Eve obtains the whole *ancilla* system of the channel. Also note that we assume Eve to be passive, so she cannot interfere with the communication; she can only try to distill useful information from her observations. This setting can be seen as a special case of the quantum version of "Model C" discussed in [1].

The aim of Alice and Bob is to use this channel, together with a two-way, public, but authentic classical channel, to generate a secret key. Let k denote the total number of uses of the optical channel. We impose the same average-photon-number constraint (13) on Alice's inputs. We assume the public channel is free so we can use it to transmit as many bits as needed, though all these bits will be known to Eve. By the end of a key-distribution protocol, Alice should be able to compute a bit string S_A and Bob should be able to compute S_B such that

- The probability that $S_A = S_B$ tends to one as k tends to infinity;
- The key S_A (or S_B) is almost uniformly distributed and independent of Eve's observations, in the sense that

$$\frac{H(S_A|\rho_{\mathrm{Eve}})}{\log|\mathcal{S}|}$$

tends to one as k tends to infinity, where ρ_{Eve} summarizes all of Eve's observations, and where $\mathcal{S}$ denotes the alphabet for S_A and S_B.

We define the *secret-key rate* of a scheme to be

$$R(\eta,\mathcal{E}) \triangleq \frac{\log|\mathcal{S}|}{k} \tag{16}$$

nats per use of the optical channel. The parameter $\mathcal{E}$ is the average photon number in (13).

A typical (and rather general) protocol to accomplish this task consists of the following steps:

[2]We adopt this definition rather than normalizing by transmitted photons, because this allows us to derive expressions which are less influenced by the transmissivity of the channel.

Step 1: Alice generates random variables $X_1, X_2, \ldots$ which are known to neither Bob nor Eve. She then prepares an optical state ρ^k on $\mathbb{A}^k$ based on $\mathbf{X}$ and sends the state into the channel, spread over k channel uses.

Step 2: Bob makes measurements on his output state to obtain a sequence $Y_1, Y_2, \ldots$.[3]

Step 3: (Information Reconciliation) Alice and Bob exchange messages $M_1, M_2, \ldots$ using the public channel. Then Alice computes her raw key K_A as a function of $(\mathbf{X}, \mathbf{M})$, and Bob computes his raw key K_B as a function of $(\mathbf{Y}, \mathbf{M})$. They try to ensure that $K_A = K_B$ with high probability, but Eve might have partial information about the raw key.

Step 4: (Privacy Amplification) Alice and Bob randomly pick one from a set of universal hashing functions. They apply the chosen function to their raw keys K_A and K_B to obtain the secret keys S_A and S_B, respectively.

Privacy amplification has been extensively studied in literature. Denote the quantum state that Eve obtained in Step 1 from the optical channel by $\sigma^{\mathbb{E}^k}$. It is shown in [13] that, provided $K_A = K_B$ with probability close to one, the privacy amplification step (i.e., Step 4) can be accomplished successfully with high probability, and the length of the secret key in nats, i.e., $\log |\mathcal{S}|$, can be made arbitrarily close to[4]

$$H(K_A|\mathbf{M}, \sigma^{\mathbb{E}^k}). \tag{17}$$

Hence, in this paper, we shall not discuss how to accomplish Step 4. As we shall see, in some cases Step 4 can be omitted. If not, then we shall concentrate on Steps 1 to 3, try to maximize (17), and compute the secret-key rate as

$$R(\eta, \mathcal{E}) = \frac{H(K_A|\mathbf{M}, \sigma^{\mathbb{E}^k})}{k}. \tag{18}$$

As mentioned previously, in Step 1, we impose the same average-photon-number constraint on Alice (13) as in the communications case. Consequently, we define the photon efficiency (of key distribution) $r(\eta, \mathcal{E})$ in the same way as in communications, namely, as in (15), except that now $R(\eta, \mathcal{E})$ is the secret-key rate.

D. Key Distribution Using a Photon Source (Model S)

In some key-distribution protocols, Alice and Bob make use of a random source, rather than Alice preparing states, to generate a secret key, as in the "Model S" discussed in [1]. In optical applications one can, for instance, generate a uniform stream of random, temporally-entangled photon pairs, which are very useful for key distribution; see [14].

An accurate model for such temporally-entangled photon sources divides the timeline into very fine temporal modes, where each temporal mode is in a pure, entangled state on its two output Hilbert spaces, with the number of photon pairs having a geometric (Bose-Einstein) distribution of a very small mean. Such a model, however, would be intractable for precise key-rate analyses. We hence choose a simplified model as follows. Let the timeline be divided into slots, where each slot can be thought of as one "use" of the source. Each slot contains many, e.g., a thousand, temporal modes. This results in the number of photon pairs in each slot having a Poisson distribution, whose mean $\mathcal{E}$ equals the total number of temporal modes times the mean photon number in each mode. We ignore the fine structures inside each slot and describe it with only two Hilbert spaces, $\mathbb{C}$ and $\mathbb{D}$. We also ignore the entanglement between the two spaces and simplify the optical state to a mixed one with classical correlation only. The optical state emitted by the source in every source use is thus given by

$$\rho^{\mathbb{C}\mathbb{D}} = \sum_{i=0}^{\infty} \frac{\mathcal{E}^i\, e^{-\mathcal{E}}}{i!} |i\rangle\langle i|^{\mathbb{C}} \otimes |i\rangle\langle i|^{\mathbb{D}}. \tag{19}$$

To justify the simplification we make, note the following.

- Discretization: In our schemes, Alice and Bob will never measure the arrival time of a photon with higher accuracy than the duration of one slot. In this case, it is easy to show that Eve cannot have any advantage by making finer measurements.
- Classical correlation: When Alice and Bob only make direct detection with the given time accuracy and Eve is listening passively via a beamsplitter, entanglement does not play any role. Note that this would not be the case if we were interested in a secrecy proof against a general (possibly active) Eve; see Section VII.

We assume that the source is collocated with Alice, who keeps $\mathbb{C}$; while the photons in $\mathbb{D}$ are sent to Bob through a lossy optical channel. To account for coupling losses, we can assume that Alice also only has access to a lossy version of $\mathbb{C}$. Specifically, $\rho^{\mathbb{C}}$ is passed through a beamsplitter, like the one in (6), of transmissivity η_A before it reaches Alice:

$$\hat{a} = \sqrt{\eta_A}\,\hat{c} + \sqrt{1-\eta_A}\,\hat{v} \tag{20a}$$

$$\hat{f} = \sqrt{1-\eta_A}\,\hat{c} - \sqrt{\eta_A}\,\hat{v}. \tag{20b}$$

But, except for Section VI-D, we shall ignore coupling losses and take $\eta_A = 1$. Similarly $\rho^{\mathbb{D}}$ is passed through a beamsplitter of transmissivity η_B before it reaches Bob:

$$\hat{b} = \sqrt{\eta_B}\,\hat{d} + \sqrt{1-\eta_B}\,\hat{u} \tag{21a}$$

$$\hat{e} = \sqrt{1-\eta_B}\,\hat{d} - \sqrt{\eta_B}\,\hat{u}. \tag{21b}$$

We assume $\eta_B < 1$ throughout. Both noise modes $\mathbb{V}$ and $\mathbb{U}$ are assumed to be in their vacuum states. Note that the two beamsplitters behave independently of each other.

Since the source is collocated with Alice, we know the photons that are lost from $\mathbb{C}$ to $\mathbb{A}$ (in case $\eta_A < 1$) should *not* reach Eve; Eve only has access to the Hilbert space $\mathbb{E}$.

It is useful to describe the output states when $\rho^{\mathbb{D}}$ passes through the beamsplitter on Bob's side. We first write down $\rho^{\mathbb{D}}$ by taking partial trace of (19):

$$\rho^{\mathbb{D}} = \sum_{i=0}^{\infty} \frac{\mathcal{E}^i\, e^{-\mathcal{E}}}{i!} |i\rangle\langle i|, \tag{22}$$

which can be equivalently written as

$$\rho^{\mathbb{D}} = \frac{1}{2\pi} \int_0^{2\pi} \mathrm{d}\theta |\alpha(\theta)\rangle\langle\alpha(\theta)| \tag{23}$$

[3] We do not consider feedback from Bob to Alice during the first two steps. As in channel coding, feedback cannot increase the maximum key rate.

[4] To be precise, to achieve (17), Alice and Bob should repeat Steps 1 to 3 many times, and then do Step 4 on all the raw keys together.

where

$$\alpha(\theta) = \sqrt{\mathcal{E}}\, e^{i\theta}. \tag{24}$$

From (23) and (21) it is straightforward to obtain the output optical state on $\mathbb{BE}$:

$$\sigma^{\mathbb{BE}} = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\theta |\sqrt{\eta}\,\alpha(\theta)\rangle\langle\sqrt{\eta}\,\alpha(\theta)|^{\mathbb{B}} \otimes |\sqrt{1-\eta}\,\alpha(\theta)\rangle\langle\sqrt{1-\eta}\,\alpha(\theta)|^{\mathbb{E}}. \tag{25}$$

By taking partial traces of (25) we obtain Bob's and Eve's states:

$$\sigma^{\mathbb{B}} = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\theta |\sqrt{\eta}\,\alpha(\theta)\rangle\langle\sqrt{\eta}\,\alpha(\theta)| \tag{26}$$

$$= \sum_{i=0}^{\infty} \frac{(\eta\mathcal{E})^i\, e^{-\eta\mathcal{E}}}{i!} |i\rangle\langle i| \tag{27}$$

$$\sigma^{\mathbb{E}} = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\theta |\sqrt{1-\eta}\alpha(\theta)\rangle\langle\sqrt{1-\eta}\alpha(\theta)| \tag{28}$$

$$= \sum_{i=0}^{\infty} \frac{((1-\eta)\mathcal{E})^i\, e^{-(1-\eta)\mathcal{E}}}{i!} |i\rangle\langle i|. \tag{29}$$

The joint state $\sigma^{\mathbb{BE}}$ is not a tensor state, i.e., $\mathbb{B}$ and $\mathbb{E}$ are not independent. However, if direct detection—namely, projective measurement in the number-state basis—is performed on $\mathbb{B}$ (or on $\mathbb{E}$), the post-measurement state on $\mathbb{E}$ (or on $\mathbb{B}$) is independent of the measurement outcome; in particular, the photon numbers in $\mathbb{B}$ and in $\mathbb{E}$ are independent. Indeed, conditional on the measurement outcome on $\mathbb{B}$ being i, the post-measurement state on $\mathbb{E}$ is

$$\frac{\mathrm{tr}_{\mathbb{B}}\left\{|i\rangle\langle i|^{\mathbb{B}}\sigma^{\mathbb{BE}}\right\}}{\mathrm{tr}\left\{|i\rangle\langle i|^{\mathbb{B}}\sigma^{\mathbb{B}}\right\}} = \frac{\frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\theta |\langle i|\sqrt{\eta}\,\alpha(\theta)\rangle|^2 |\sqrt{1-\eta}\,\alpha(\theta)\rangle\langle\sqrt{1-\eta}\,\alpha(\theta)|}{\frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\theta |\langle i|\sqrt{\eta}\,\alpha(\theta)\rangle|^2} \tag{30}$$

$$= \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\theta |\sqrt{1-\eta}\,\alpha(\theta)\rangle\langle\sqrt{1-\eta}\,\alpha(\theta)| \tag{31}$$

$$= \sigma^{\mathbb{E}} \tag{32}$$

where (31) follows because

$$|\langle i|\sqrt{\eta}\,\alpha(\theta)\rangle|^2 = \frac{(\eta\mathcal{E})^i\, e^{-\eta\mathcal{E}}}{i!} \tag{33}$$

does not depend on θ.

We now describe a scheme (which is again rather general) for Alice and Bob to use this source k times to generate a secret key. In this scheme, Steps 3 and 4 are exactly the same as in Section III-C, but Steps 1 and 2 are now replaced by:

Step 1': Alice makes measurements on her state $\sigma^{\mathbb{A}^k}$ to obtain the sequence $X_1, X_2, \ldots$.

Step 2': Bob makes measurements on his state $\sigma^{\mathbb{B}^k}$ to obtain the sequence $Y_1, Y_2, \ldots$.

As in Section III-C, we shall concentrate on Steps 1', 2', and 3. The secret-key rate, denoted by $R(\eta_A, \eta_B, \mathcal{E})$, is again given by the right-hand side of (18), with unit "nats per source use". But the photon efficiency in this setting is defined as

$$r(\eta_A, \eta_B, \mathcal{E}) \triangleq \frac{R(\eta_A, \eta_B, \mathcal{E})}{\eta_A\eta_B\mathcal{E}}. \tag{34}$$

We choose this definition because $\eta_A\eta_B\mathcal{E}$ is the expected number of photon pairs in each source use that reach both Alice and Bob,[5] and because these photon pairs are those that contain correlated information that can be used to generate the secret key. When $\eta_A = 1$, we omit the subscript in η_B, and denote the secret-key rate and photon efficiency simply by $R(\eta, \mathcal{E})$ and $r(\eta, \mathcal{E})$, respectively. Obviously, they are again related by (15).

IV. Background: Photon-Efficient Communication Using Pulse-Position Modulation

Before we address key distribution, we give some results regarding communications over the bosonic channel described in Section III-B. These results serve as a point of reference, and the derivation provides tools later used in key distribution. See also [15]–[17].

The capacity of a quantum channel is characterized by the formula found by Holevo [18] and by Schumacher and Westmoreland [19]. For the pure-loss bosonic channel (6a) under constraint (13), this capacity is $g(\eta\mathcal{E})$ nats per channel use [20], where

$$g(x) \triangleq (x+1)\log(x+1) - x\log x, \qquad x > 0. \tag{35}$$

This immediately implies that the photon efficiency (15) satisfies:

$$r_{\mathrm{quantum}}(\eta, \mathcal{E}) = \frac{g(\eta\mathcal{E})}{\eta\mathcal{E}} = \log\frac{1}{\eta\mathcal{E}} + 1 + o(1). \tag{36}$$

Note that the efficiency is unbounded, that is,

$$\lim_{\mathcal{E}\downarrow 0} r_{\mathrm{quantum}}(\eta, \mathcal{E}) = \infty. \tag{37}$$

Hence, in terms of [21], [22], the capacity per unit cost $\sup_{\mathcal{E}} r(\eta, \mathcal{E})$ of the channel (12) is infinite.

The capacity $g(\eta\mathcal{E})$ is achievable by Alice using product (i.e., nonentangled), pure input states

$$|\psi^k\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_k\rangle. \tag{38}$$

Indeed, in this paper we limit our attention to such mode of operation, where the average-photon-number constraint (13) becomes

$$\frac{1}{k}\sum_{i=1}^{k} \langle\psi_i|\hat{a}_i^\dagger\hat{a}_i|\psi_i\rangle \le \mathcal{E}. \tag{39}$$

For the degenerate case $\eta = 1$, a simple capacity-achieving codebook consists only of number states, where the photon numbers' empirical distribution is independent and identically distributed (i.i.d.) geometric (i.e., Bose-Einstein).

[5] We interpret this quantity in a semi-classical way: each photon pair reaches Alice with probability η_A, and reaches Bob with probability η_B independently of whether it reaches Alice or not, hence the fraction of photon pairs that reach both Alice and Bob is $\eta_A\eta_B$. We do not know if there exists a physical observable, i.e., a Hermitian operator that corresponds to this value.

Bob's optimal measurement for this codebook is simply per-channel-use direct detection. We shall see in Section IV-A that, in the photon-efficient regime, this code construction can be further simplified and can be used also when $\eta < 1$, without sacrificing much photon efficiency.

For the general case where η may not be one, the capacity can be achieved if Alice's codebook consists of coherent states

$$|\psi^k\rangle = |\alpha_1\rangle \otimes |\alpha_2\rangle \otimes \cdots \otimes |\alpha_k\rangle, \tag{40}$$

and if Bob performs a general (not per-channel-use) POVM on the output state, which is

$$|\phi^k\rangle = |\sqrt{\eta}\,\alpha_1\rangle \otimes |\sqrt{\eta}\,\alpha_2\rangle \otimes \cdots \otimes |\sqrt{\eta}\,\alpha_k\rangle. \tag{41}$$

In this case, the average-photon-number constraint (13) becomes

$$\sum_{i=1}^{k} |\alpha_i|^2 \le k\mathcal{E}. \tag{42}$$

It is known that capacity-achieving codebooks of coherent states should have empirical distributions that resemble i.i.d. complex-Gaussian with mean zero and variance $\mathcal{E}$ [20]. The main problem with such a code is that Bob's POVM is almost impossible to implement using today's technology. Hence we are interested in "practical" schemes, in particular, in schemes where Bob uses per-channel-use direct detection while Alice sends coherent states. As we shall see in Section IV-B, this restriction induces a second-order-term loss in photon efficiency.

A. Alice Sends Binary Number States

Consider the case where the sequence of states sent by Alice consists only of the number states $|0\rangle$ and $|1\rangle$, and where Bob uses direct detection. Recalling (8), for input $|0\rangle$ Bob will always detect no photon, while for input $|1\rangle$ Bob detects one photon with probability η, and detects no photon otherwise. Thus the scheme induces a classical Z channel. The maximum achievable rate is, according to the classical channel coding theorem [23], the maximum mutual information over this channel.

Let

$$I_{\mathrm{Z}}(q, \mu) \triangleq H_2(q\mu) - q\,H_2(\mu) \tag{43}$$

be the mutual information over a Z channel with input probability $P_X(1) = q$ and transition probability $P_{Y|X}(1|1) = \mu$, where $H_2(\cdot)$ is the binary entropy function

$$H_2(x) \triangleq x \log \frac{1}{x} + (1-x) \log \frac{1}{1-x}, \qquad 0 < x < 1. \tag{44}$$

Due to the photon-number constraint, the input distribution must satisfy $q \le \mathcal{E}$.[6] It is easy to see that $I_{\mathrm{Z}}(q, \mu)$ is monotonically increasing in q for small enough q, and hence, in the regime of interest, we should choose $q = \mathcal{E}$, achieving rate $I_{\mathrm{Z}}(\mathcal{E}, \eta)$. The resulting photon efficiency can be readily shown to satisfy:

$$r_{\mathrm{num,Z}}(\eta, \mathcal{E}) = \frac{I_{\mathrm{Z}}(\mathcal{E}, \eta)}{\eta\mathcal{E}} = r_{\mathrm{quantum}}(\eta, \mathcal{E}) - \frac{H_2(\eta)}{\eta} + o(1), \tag{45}$$

reflecting a constant efficiency loss with respect to the optimum (36).

For the scheme described above, the task of (classical) coding is difficult: one needs mutual-information-approaching codes for a Z channel with a highly skewed input. We can solve this problem by replacing the i.i.d. binary codebook by PPM: the input sequence consists of "frames" of length $\lceil 1/\mathcal{E} \rceil$, where each frame includes exactly one photon, whose position is uniformly chosen inside the frame. (If the blocklength is not divisible by $\lceil 1/\mathcal{E} \rceil$, then we ignore the remainder.) This scheme converts the channel to a $\lceil 1/\mathcal{E} \rceil$-ary erasure channel. By computing the capacity of this erasure channel, we easily see that the photon efficiency of the PPM scheme is:

$$r_{\mathrm{num,PPM}}(\eta, \mathcal{E}) = \log \frac{1}{\mathcal{E}} + o(1), \tag{46}$$

which again reflects only a constant loss compared to the optimal efficiency (36). The large-alphabet erasure channel is much like a packet-erasure channel encountered in internet applications, and good off-the-shelf codes are available.

B. Alice Sends Binary Coherent States

Generating the number state $|1\rangle$ is hard in practice. We hence turn to coherent states, which are a good model for light coming out of laser sources [7].

We consider a simple binary-coherent-state scheme. In this scheme, Alice first generates a classical binary codebook where the probability of 1 is q. She then maps 0 and 1 to the coherent states $|0\rangle$ and $|\mathcal{E}/q\rangle$, respectively. Note that doing this satisfies the average-power constraint (42). Bob uses direct detection that is not photon-number resolving (PNR), i.e., he views a measurement with no photon as a logical 0, and views any measurement with at least one photon as a logical 1. (Such a detector is easier to build than a PNR detector, which outputs the exact number of detected photons.) This results again in a classical Z channel, with

$$P_{Y|X}(1|1) = \mu_{\mathrm{coh}}(q, \mathcal{E}) \triangleq 1 - \exp\left(-\frac{\eta\mathcal{E}}{q}\right). \tag{47}$$

We can thus achieve $I_{\mathrm{Z}}(q, \mu_{\mathrm{coh}})$ nats per channel use, where q should be chosen to maximize $I_{\mathrm{Z}}(q, \mu_{\mathrm{coh}})$. The exact analytical optimization is complicated, but in the photon-efficient regime the approximate optimum (which yields the best rate up to the approximation of interest) is given by

$$q^*(\mathcal{E}) = \frac{\eta\mathcal{E}}{2} \log \frac{1}{\mathcal{E}}. \tag{48}$$

The resulting photon efficiency is given by:

$$r_{\mathrm{coh,Z}}(\eta, \mathcal{E}) = \frac{I_{\mathrm{Z}}\big(q^*(\mathcal{E}), \mu_{\mathrm{coh}}(q^*(\mathcal{E}), \mathcal{E})\big)}{\eta\mathcal{E}} \tag{49}$$

$$= \log \frac{1}{\eta\mathcal{E}} - \log\log \frac{1}{\mathcal{E}} + \log 2 - 1 + o(1). \tag{50}$$

Comparing to the quantum limit (36), we see that the efficiency loss of the coherent-state-and-direct-detection scheme with respect to the optimal performance grows as $\log\log 1/\mathcal{E}$ as $\mathcal{E}$ decreases in the photon-efficient regime. This loss is inherent to any "classical" transmission scheme, even if general

[6]The expected number of photons translates to a per-codeword constraint via a standard expurgation argument.

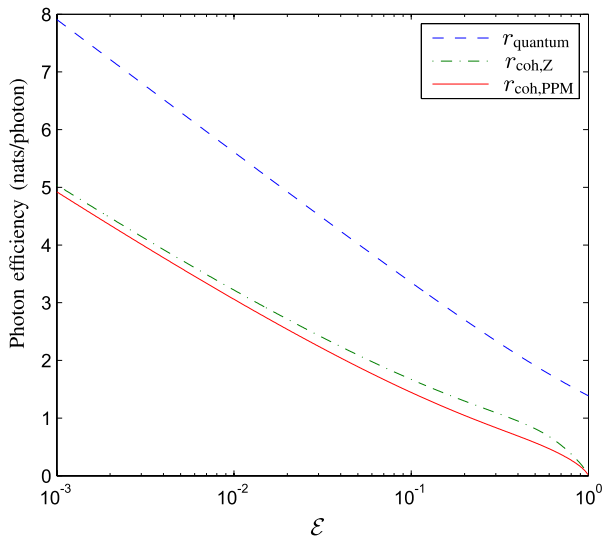

Fig. 1. Photon efficiency in the different cases discussed in Section IV. Efficiency in the quantum case r_{quantum} is computed from (36); efficiency for coherent-state inputs and Z-channel model $r_{\text{coh,Z}}$ from (49); and efficiency for coherent-state inputs and PPM $r_{\text{coh,PPM}}$ from (51). For all three we let the channel be lossless, i.e., we choose $\eta = 1$.

(non-binary) coherent states are sent [17], or if the receiver is allowed to use feedback between measurements [24].

Similarly to the case of Alice sending number states, we can alleviate the difficulty of coding by replacing the i.i.d. codebooks with PPM frames, an idea already exploited in [25], [26]. Indeed, using PPM frames of length b with the optimum (to the approximation order) choice of (48) and $b = \lceil 1/q^*(\mathcal{E}) \rceil$, this efficiency is

$$r_{\text{coh,PPM}}(\mathcal{E}) = \frac{\mu_{\text{coh}}\big(q^*(\mathcal{E}), \mathcal{E}\big) \log b}{\eta b \mathcal{E}}, \tag{51}$$

and has the same expression as on the right-hand side of (50), i.e., the further efficiency loss incurred by restricting to PPM is $o(1)$.

Figure 1 depicts the photon efficiency in the different cases discussed in this section. It can be appreciated that, while the loss of using coherent states with direct detection is large, the further loss of PPM is small. As we shall see, similar phenomena are also observed in key-distribution scenarios.

V. Key Distribution in Model C

In this section we study the key-distribution problem in Model C, which we set up in Section III-C.

To the best of our knowledge, the maximum secret-key rate, and hence also the maximum photon efficiency, in this setting are not yet known. However, in the photon-efficient regime we have the following asymptotic upper bound. (Later we show that this upper bound is tight within a constant term).

Proposition 1: The maximum photon efficiency for key distribution in Model C as described in Section III-C satisfies

$$r_{\max}(\mathcal{E}) \le \log \frac{1}{\eta \mathcal{E}} + 1 + o(1). \tag{52}$$

Proof: We use the fact that the maximum secret-key rate over a quantum channel cannot exceed the communication capacity of the same channel. This follows, e.g., from [27, Chapter I, Theorem 5.1]. Recalling (36), the proof is completed. ∎

As in the communication setting, we shall mostly focus on key-distribution schemes in which Bob only employs direct detection. As we shall see in Section V-A, if Alice can send number states—even only binary number states—the photon-efficiency loss of direct detection is at most a constant term in the photon-efficient regime. However, in Section V-B we show that if Alice can only send coherent states, then the loss in photon efficiency scales like $\log\log 1/\mathcal{E}$. These results are similar to their optical-communication counterparts. Also similar to the communication scenario is the fact that PPM is nearly optimal in terms of photon efficiency; in the context of key distribution, PPM allows us to greatly simplify the coding task in the information-reconciliation step.

A. Alice Sends Binary Number States

Consider the following key-distribution scheme.

Scheme C-1:

1) *Let $b \triangleq \lceil 1/\mathcal{E} \rceil$. We divide the whole block of k channel uses into frames each consisting of b consecutive uses (and ignore the remainder).*
2) *Alice generates a sequence of integers $\tilde{X}_1, \tilde{X}_2, \ldots$ i.i.d. uniformly in $\{1, \ldots, b\}$. These are the "pulse positions". Within the ith frame, $i \in \{1, 2, \ldots\}$, she sends the number state $|1\rangle$ in the $\tilde{X}_i$th channel use, and sends $|0\rangle$ in all other channel uses.*
3) *Bob makes direct detection on every channel output. Since Alice sends one photon per frame, Bob will either detect a single photon or no photon per frame. Let the set of frames where Bob had a detection be denoted as $\{i_1, i_2, \ldots\}$, and denote the detection positions inside these bins by $\tilde{Y}_{i_1}, \tilde{Y}_{i_2}, \ldots$. Bob tells Alice the values of $i_1, i_2, \ldots$ using the public channel.*
4) *Alice generates the secret key from $\tilde{X}_{i_1}, \tilde{X}_{i_2}, \ldots$, and Bob generates the secret key from $\tilde{Y}_{i_1}, \tilde{Y}_{i_2}, \ldots$, both by directly taking the binary representation of these integers.*

The average-photon-number constraint (13) is clearly satisfied. Scheme C-1 is rather simple in the sense that

- Alice's input states are either $|0\rangle$ or $|1\rangle$;
- Bob's detector can be non-PNR;
- The information-reconciliation step is uncoded, and only involves one-way communication from Bob to Alice;
- There is no privacy-amplification step.

As the next proposition shows, this simple scheme performs very well in the photon-efficient regime: it is at most a constant term away from optimum. Compared to the communication case (36), this proposition also shows that the loss in photon efficiency due to the secrecy requirement is at most a constant term.

Proposition 2: Scheme C-1 generates a secret key between Alice and Bob, and its photon efficiency is

$$r_{\text{C-1}}(\eta, \mathcal{E}) = \log \frac{1}{\mathcal{E}} + o(1) \tag{53}$$

for all $\eta \in (0, 1]$.

Proof: We first verify that Scheme C-1 indeed generates a secret key. To this end, first note that $\tilde{X}_{i_j} = \tilde{Y}_{i_j}$ for all $j \in \{1, 2, \ldots\}$. This is because Alice sends only one non-vacuum state in each frame, and because Bob cannot detect any photon in a channel use where Alice sends $|0\rangle$. Hence the keys obtained by Alice and by Bob are the same. Second, by the way Alice chooses $\tilde{\mathbf{X}}$, every $\tilde{X}_{i_j}$ (or, equivalently, $\tilde{Y}_{i_j}$) is uniformly distributed in $\{1, \ldots, b\}$, independently of $\tilde{X}_{i_{j'}}$ where $j' \neq j$. This shows that the key is uniformly distributed. It now remains to verify that the key is dependent neither on Eve's output states from the optical channel nor on the messages which Bob sends to Alice. It is independent of Eve's optical states because, in every selected frame, Bob detects the only photon that Alice transmits, so Eve's post-measurement state in this frame is the all-vacuum state. It is independent of Bob's messages because Bob only sends the labels of the selected frames to Alice, and because Alice chooses the pulse positions independently of the frame labels.

We next compute the photon efficiency achieved by Scheme C-1. Let $N(k)$ be the total number of frames selected by Bob within k channel uses. Since each frame is selected when Bob detects a photon in that frame, which happens with probability η, we have from the Law of Large Numbers that

$$\lim_{k\to\infty} \frac{N(k)}{k} = \lim_{k\to\infty} \frac{\eta\lfloor k/b \rfloor}{k} = \eta\mathcal{E} \tag{54}$$

with probability one. Each detected photon (or, equivalently, each selected frame) provides $\log b$ nats of secret key. So, as k tends to infinity, the achieved photon efficiency tends to

$$\lim_{k\to\infty} \frac{N(k)\log b}{k\eta\mathcal{E}} = \log b = \log\frac{1}{\mathcal{E}} + o(1). \tag{55}$$

■

B. Alice Sends Coherent States

We now restrict Alice to sending coherent states since, as discussed previously, generating the number state $|1\rangle$ is hard in practice. Under this restriction, Alice generates a sequence of complex numbers $\alpha_1, \alpha_2, \ldots, \alpha_k$ satisfying (42), prepares the coherent states $|\alpha_1\rangle, |\alpha_2\rangle, \ldots, |\alpha_k\rangle$, and sends them over the channel. As the next proposition shows, this restriction induces a loss of $\log\log 1/\mathcal{E}$ in the photon efficiency, even if the scheme employed is more sophisticated than Scheme C-1.

Proposition 3: The maximum photon efficiency in Model C when Alice sends only coherent states and when Bob uses only direct detection satisfies

$$r_{\mathrm{coh}}(\eta, \mathcal{E}) \leq \log\frac{1}{\mathcal{E}} - \log\log\frac{1}{\mathcal{E}} + O(1) \tag{56}$$

for all $\eta \in (0, 1]$.

Proof: We note that, when Alice sends the coherent state $|\alpha\rangle$, Bob's measurement outcome Y has a Poisson distribution of mean $\eta|\alpha|^2$. We can bound the achievable secret-key rate as

$$R_{\mathrm{coh}}(\eta, \mathcal{E}) \leq \max_{\mathsf{E}[|X|^2]\leq\mathcal{E}} I(X; Y) \tag{57}$$

$$= \max_{\mathsf{E}[|X|^2]\leq\mathcal{E}} I(|X|^2; Y), \tag{58}$$

where (57) follows because the secret-key rate over a channel cannot be larger than the communication capacity of the channel (see [1]); and where (58) follows because $|X|^2$ is a deterministic function of X, and because $X \multimap\!\!-|X|^2\!\!-\!\!\multimap Y$ forms a Markov chain. Finally, the right-hand side of (58), which is the maximum mutual information over a Poisson channel under an average-photon-number constraint, is shown in [17] to satisfy

$$\max_{\mathsf{E}[|X|^2]\leq\mathcal{E}} I(|X|^2; Y) \leq \eta\mathcal{E}\left\{\log\frac{1}{\mathcal{E}} - \log\log\frac{1}{\mathcal{E}} + O(1)\right\}. \tag{59}$$

■

We do not specify the $O(1)$ term, as the derivation of (59) in [17] yields expressions that are rather involved. In the sequel we show that the bound (56) is tight within a constant term.

As in Section IV-B, to simplify the coding task for the information-reconciliation step, Alice and Bob can use a PPM-based scheme. We choose the PPM frame-length to be:

$$b \triangleq \left\lceil \frac{1}{\mathcal{E}\log 1/\mathcal{E}} \right\rceil. \tag{60}$$

This choice is optimal up to the order of approximation of interest. Note that b in (60) is half the frame-length chosen for the communication setting, where the latter is $\lceil 1/q^*(\mathcal{E})\rceil$ with $q^*(\mathcal{E})$ given in (48).

Scheme C-2:

1) *We divide the whole block of k channel uses into frames each consisting of b consecutive uses (and ignore the remainder).*
2) *Alice generates a sequence of integers $\tilde{X}_1, \tilde{X}_2, \ldots$ i.i.d. uniformly in $\{1, \ldots, b\}$. Within the ith frame, $i \in \{1, 2, \ldots\}$, she sends the coherent state $|\sqrt{b\mathcal{E}}\rangle$ in the $\tilde{X}_i$th channel use, and sends the vacuum state $|0\rangle$ in all other channel uses.*
3) *Bob makes direct detection on every channel-output. Since all channel input-states but one are in vacuum state, he will have detections in at most one output. Let the set of frames where Bob had a detection be denoted as $\{i_1, i_2, \ldots\}$, and denote the detection positions inside these bins by $\tilde{Y}_{i_1}, \tilde{Y}_{i_2}, \ldots$. He tells Alice the values of $i_1, i_2, \ldots$ using the public channel.*
4) *Alice generates the raw key K_A from $\tilde{X}_{i_1}, \tilde{X}_{i_2}, \ldots$, and Bob generates the raw key K_B from $\tilde{Y}_{i_1}, \tilde{Y}_{i_2}, \ldots$, both by directly taking the binary representation of these integers.*
5) *Alice and Bob perform privacy amplification on their raw keys to obtain the secret keys.*

The average-photon-number constraint (13) or (42) is clearly satisfied. Also note that, in this scheme,

- Alice's input states are binary: either $|0\rangle$ or $|\sqrt{b\mathcal{E}}\rangle$;
- Bob's detector can be non-PNR;
- The information-reconciliation step is uncoded, and only involves one-way communication from Bob to Alice.

In contrast to the restriction on Alice to sending only coherent states, which results in a loss of $\log\log 1/\mathcal{E}$ in photon efficiency, the further simplifications employed in Scheme C-2 induce at most a constant-term loss.

Proposition 4: Scheme C-2 achieves photon efficiency

$$r_{\text{C-2}}(\eta, \mathcal{E}) \geq \log \frac{1}{\mathcal{E}} - \log\log \frac{1}{\mathcal{E}} - (1-\eta) + o(1) \tag{61}$$

for all $\eta \in (0, 1]$.

The proof, which appears in Appendix A, is more involved than that of Scheme C-1, since in the case of coherent states, the raw key depends upon Eve's optical states (since, if Bob and Eve both see detections in some frame, then they must be in the same location). However, we bound the information leakage and show that it leads to at most a constant key-efficiency loss.

VI. Key Distribution in Model S

In this section we study the key-distribution problem in Model S, which we set up in Section III-D. Apart from Section VI-D, we shall focus on the case where $\eta_A = 1$. In this case, we omit the subscript of η_B to denote it simply as η.

Proposition 5: The maximum photon efficiency achievable in Model S satisfies

$$r_{\text{quantum}}(\eta, \mathcal{E}) \leq \log \frac{1}{\eta\mathcal{E}} + 1 + o(1). \tag{62}$$

Proof: We note that, without further constraints, the secret-key rate and hence the photon efficiency achievable in Model S cannot exceed those achievable in Model C. This is because any measurement Alice performs in Step 1' in Model S, which is described in Section III-D, can be simulated in Model C in the following way. Alice first generates random numbers that have the same statistics as the outcomes of the measurement that she would perform in Model S. Then, for each number, she generates the corresponding post-measurement state on $\mathbb{D}$ and sends it to Bob. Doing these will generate the same correlation between Alice, Bob, and Eve as the corresponding strategy in Model S would do. The claim now follows immediately from Proposition 1. ∎

Note: The above proof says that, when Alice and Bob can both use fully quantum devices, there is no advantage in Model S over Model C. However, as we later show, this need not be the case when Alice and Bob are restricted, e.g, to direct detection.

For practicality, for the rest of this section we restrict both Alice and Bob to using only direct detection on their quantum states. In fact, Alice and Bob will only use non-PNR direct detection. In contrast, we do not impose any constraint on Eve's measurement, thus our schemes are secure against a fully-quantum (though passive) Eve.

A. Direct Detection Combined With Optimal Binary Slepian-Wolf Codes

After Alice and Bob perform direct detection on their optical states, each of them has a binary sequence where 1 indicates photons are detected in the corresponding source use. Denote their sequences by **A** and **B**, respectively. Due to our source model, **A** and **B** are distributed i.i.d. in time, while each pair (A, B) has joint distribution according to a Z channel with

$$q \triangleq P_A(1) = 1 - e^{-\mathcal{E}} \tag{63a}$$

$$\mu \triangleq P_{B|A}(1|1) = \frac{1 - e^{-\eta\mathcal{E}}}{1 - e^{-\mathcal{E}}}. \tag{63b}$$

Bob can help Alice to know **B** by sending her a Slepian-Wolf code [28]. For the moment, we assume that Alice and Bob have an optimal Slepian-Wolf code for the joint distribution P_{AB} (Later we drop this assumption to find more realistic code constructions.) Then they can use the following key-distribution scheme.

Scheme S-3:

1) *Alice and Bob perform non-PNR direct detection to obtain binary sequences* **A** *and* **B**, *respectively.*
2) *Bob sends Alice an optimal Slepian-Wolf code so that Alice knows* **B** *with high probability. They use* **B** *as the raw key.*
3) *Alice and Bob perform privacy amplification on* **B** *to obtain the secret key.*

The key rate and photon efficiency of Scheme S-3 satisfy the following.

Proposition 6: Scheme S-3 achieves the key rate

$$R_{\text{S-1}}(\eta, \mathcal{E}) = I(A; B) \tag{64}$$

where the mutual information is computed on the joint distribution P_{AB} *given by* (63). *Furthermore, for all* $\eta \in (0, 1]$, *the photon efficiency of Scheme S-3 satisfies*

$$r_{\text{S-1}}(\eta, \mathcal{E}) = \log \frac{1}{\eta\mathcal{E}} + 1 - \frac{H_2(\eta)}{\eta} + o(1). \tag{65}$$

Proof: We first prove (64). Its converse part follows immediately from [27, Chapter I, Theorem 5.3], which states that the secret-key rate cannot exceed $I(A; B)$ even if Eve possesses no quantum state that is correlated to A and B. Its achievability part follows from [27, Chapter III, Theorem 2.2]: when we eliminate the "helper subalgebra", the theorem says that the forward key capacity (i.e., the maximum key rate achievable when Alice does not communicate to Bob) is lower-bounded by $I(A; B) - I(B; \mathbb{E})$ evaluated for the joint state consisting of Alice's and Bob's measurement outcomes and Eve's post-measurement state. As shown in Section III-D, Bob's measurement outcome is independent of Eve's post-measurement state, so $I(B; \mathbb{E}) = 0$.[7]

We next prove (65). Direct evaluation for the Z-channel mutual information (43) for the channel parameters q and μ of (63) gives:

$$I(A; B) = I_Z(q, \mu) \tag{66}$$

$$= H_2(e^{-\eta\mathcal{E}}) - \left(1 - e^{-\mathcal{E}}\right) H_2\left(\frac{1 - e^{-\eta\mathcal{E}}}{1 - e^{-\mathcal{E}}}\right) \tag{67}$$

$$= \eta\mathcal{E} \log \frac{1}{\eta\mathcal{E}} + \eta\mathcal{E} - \mathcal{E} H_2(\eta) + o(\mathcal{E}). \tag{68}$$

Substituting in (64) and dividing by $\eta\mathcal{E}$ yields (65). ∎

[7] In Section III-D we consider the case where Bob performs a complete projective measurement in the number-state basis, whereas here Bob's non-PNR detection only distinguishes between zero and positive photon numbers. But extending our claim for the former case to the latter is straightforward.

Hence the conceptually simple Scheme S-3, which only uses non-PNR direct detection both at Alice and at Bob, is at most a constant term away from the optimal quantum efficiency whose upper bound is given in (62) . Comparing this with (36) and (52) we see that the differences between the optimal photon efficiencies in communication, in Model C, and in Model S are at most constants. Interestingly, $r_{\text{S-1}}(\eta, \mathcal{E})$ is asymptotically the same as the photon efficiency in the communication scenario where Alice sends binary number states (45).

The problem with Scheme S-3 is, though, that the source distribution P_{AB} is highly skewed, which makes it difficult to find a good Slepian-Wolf code, much like the difficulty to obtain a channel code in the communication setting of Section IV. While in communication and in Model C Alice can use PPM to simplify code design, in Model S this is no longer possible, as the sequences **A** and **B** are governed by the source, over which neither Alice nor Bob have control. Nevertheless, Alice and Bob can use a PPM-like scheme by *parsing* the sequences into frames, as we next propose.

B. Simple Frame-Parsing

In a simple PPM-like scheme, Alice and Bob parse the source uses into frames, and only use the frames where each of them has exactly one detection to generate the key.

Scheme S-4:

1) *Alice and Bob perform non-PNR direct detection to obtain binary sequences* **A** *and* **B***, respectively.*
2) *Let* b *be as in* (60)*. We divide the whole block of* k *source uses into frames each consisting of* b *consecutive uses (and ignore the remainder).*
3) *Bob selects all the frames in which he detects at least one photon (*$B = 1$ *for at least one source use). Denote the labels of these frames by* $\{i_1, i_2, \ldots\}$*. He tells Alice the values of* $i_1, i_2, \ldots$ *using the public channel.*
4) *Alice selects the frames among* $i_1, i_2, \ldots$ *in which* $A = 1$ *for* exactly one *source use. Denote the labels of these frames by* $\{i_{j_1}, i_{j_2} \ldots\}$*, Alice's detection positions within these frames by* $\{Y_{i_{j_1}}, Y_{i_{j_2}}, \ldots\}$*, and Bob's (unique) detection positions within these frames by* $\{X_{i_{j_1}}, X_{i_{j_2}}, \ldots\}$*. She tells Bob the values of* $j_1, j_2, \ldots$ *using the public channel.*
5) *Alice and Bob generate the raw key by taking the binary representations of* $\{X_{i_{j_1}}, X_{i_{j_2}}, \ldots\}$ *and of* $\{Y_{i_{j_1}}, Y_{i_{j_2}}, \ldots\}$*, respectively.*
6) *Alice and Bob perform privacy amplification on the raw key to obtain the secret key.*

As in Schemes C-1 and C-2, the information-reconciliation step in Scheme S-4 is uncoded and hence very simple. The performance of Scheme S-4 is similar to that of Scheme C-2 where Alice sends coherent states, in the sense that it loses a $\log\log 1/\mathcal{E}$ term in photon efficiency compared to the optimum (62). Interestingly, here the loss does not come from the input states used, as they are identical to those in Scheme C-1, but rather from the parsing process.

Proposition 7: The photon efficiency of Scheme S-4 satisfies

$$r_{\text{S-2}}(\eta, \mathcal{E}) = \log \frac{1}{\mathcal{E}} - \log\log \frac{1}{\mathcal{E}} - 1 + o(1). \tag{69}$$

The scheme has some information leakage, since Eve can use her knowledge about the frames which were selected for key generation (obtained by listening to the public channel), in conjunction with the measurements she performs on the same frames. The proof, which appears in Appendix B, shows that this leakage is vanishing in the photon-efficient limit.

Note: If Alice uses PNR direct detection (which is technically more difficult than non-PNR), then Scheme S-4 can be simplified so that it does not contain an privacy-amplification step. Indeed, Alice can select those frames in which she detects *only one photon*. In this case, since Bob also detects photons (in fact, only one photon) in every such frame, we know that Eve's post-measurement states in these frames are all vacuum. Hence Eve has no information about $\tilde{X}$, and taking the binary representation of $\tilde{X}$ already gives Alice and Bob a secret key.

The information loss of Scheme S-4 compared to Scheme S-3 comes from two sources. First, the sequence $\{i_1, i_2, \ldots\}$ itself contains useful information that can be used to generate secret bits, but is not exploited in Scheme S-4. Second, frames in which Alice detects photons in two or more source uses are discarded. As it turns out, the first source of information loss is dominant in the photon-efficient regime; we next show how this loss can be recovered. (Loss from the second source can also be partially recovered, e.g., by varying the frame-lengths [29].)

C. Enhanced Frame-Parsing

Our idea of enhancing the frame-parsing scheme S-4 is to extract secret-key bits also from the sequence $\{i_1, i_2, \ldots\}$, which indicates the positions of frames selected by Bob. To this end, instead of sending this sequence uncoded, Bob uses a binary Slepian-Wolf code to send this information to Alice. Note that such a code is much easier to construct than the one in Scheme S-3, as the zeros (frames not selected by Bob) and ones (frames selected by Bob) are much more balanced than in the original binary sequence **B**; recall (60). Assuming that an optimal Slepian-Wolf can be found, we can completely recover the $\log\log 1/\mathcal{E}$ term and reduce the loss in photon efficiency to a constant term.

Scheme S-5:

1) *Alice and Bob use non-PNR direct detection to obtain binary sequences* **A** *and* **B***, respectively.*
2) *Let* b *be as in* (60)*. We divide the whole block of* k *source uses into frames each consisting of* b *consecutive uses (and ignore the remainder).*
3) *Let* $\tilde{B}_i$ *be the indicator that Bob detects at least one photon within the* i*th frame, and let* $\tilde{A}_i$ *be the same indicator for Alice. Bob sends a Slepian-Wolf code to Alice using the public channel, so that Alice can recover* $\tilde{\mathbf{B}}$ *based on the codeword together with* $\tilde{\mathbf{A}}$ *with high probability.*
4) *Corresponding to every* i *such that* $\tilde{B}_i = 1$*, Alice sends a binary symbol* C_i *to Bob:* $C_i = 1$ *if within the* i*th*

frame there is exactly *one source use where* $A = 1$*, and* $C_i = 0$ *otherwise. Note that since Alice knows* $\tilde{\mathbf{B}}$ *with high probability, she can send* C_i*s simply as a bitstream in an increasing order in* i *(and skip the* i*s for which* $\tilde{B}_i = 0$*).*

5) *Alice and Bob perform privacy amplification on* $\tilde{\mathbf{B}}$ *to obtain the first part of the secret key.*
6) *For every* i *such that* $\tilde{B}_i = C_i = 1$*, let* X_i *be the position where* $A = 1$*, and let* Y_i *be the (unique) position where* $B = 1$*. Alice and Bob generate the second part of the secret key by taking the binary representations of* X_i *and of* Y_i*, respectively, for all such* i*s, and by then performing privacy amplification.*

Proposition 8: Scheme S-5 achieves photon efficiency

$$r_{\text{S-3}}(\eta, \mathcal{E}) \geq \log \frac{1}{\mathcal{E}} - \frac{H_2(\eta)}{\eta} + o(1) \tag{70}$$

for all $\eta \in (0, 1]$.

The proof, which appears in Appendix C, evaluates the key rate that Step 5) adds over the rate of Scheme S-4. This part of the key consists of frame labels, thus it is obviously correlated with the messages sent over the public channel. However, we show that in the photon-efficient limit Eve must "lose synchronization" with the frame locations, thus the leakage is vanishing.

D. Extension to the Case $\eta_A < 1$

The results for the case where η_A in (20) is equal to one can be extended to the case where $\eta_A < 1$, though the expressions become considerably more cumbersome. We hence only give some heuristic explanations how our schemes should be modified, and how they perform. Note that for the following discussions the photon efficiency is defined in (34). Also recall that we assume that the source is co-located with Alice, such that the photons lost do not reach Eve.

Quantum Limit: Proposition 5 holds but with a different constant term. The same proof ideas apply.

Direct Detection: Scheme S-3 can be directly applied to the case where $\eta_A < 1$ without modification, and its photon efficiency is different from the right-hand side of (65) by a constant term, i.e., it is again at most a constant away from the quantum limit.

Simple Frame-Parsing: Scheme S-4 needs some modifications in order to work when $\eta_A < 1$. First, in Step 2) Bob should select only those frames in which there is *exactly* one source use where $B = 1$. This is because there can be frames in which Bob has more detections than Alice, due to the loss to Alice. Second, after Step 3) Bob needs to send Alice a b-ary Slepian-Wolf code on his detection positions inside the selected frames, so that Alice will know these positions with high probability. (This is a large-alphabet code for symmetric errors, and is relatively easy to construct.) This step is needed because, since both Alice and Bob only observe lossy versions of the source, their detection positions inside the selected frames might be different. Indeed, the two positions are equal if they come from the same source photon-pair, and are independent of each other if they come from two different source photon-pairs. Finally, for Step 5) (privacy amplification), Eve's side information needs to be examined more carefully compared to the case where $\eta_A = 1$. After these modifications, one can show that the photon efficiency is the same as the right-hand side of (69) up to the second term, i.e., the loss in photon efficiency scales like $\log \log 1/\mathcal{E}$.

Enhanced Frame-Parsing: If we incorporate the aforementioned modifications for Scheme S-4 to Scheme S-5, then Scheme S-5 also works for the case $\eta_A < 1$, and its photon efficiency is different from the right-hand side of (70) by a constant.

We finally note that, for all three cases in which we restrict Alice and Bob to using direct detection, we can also take *detector dark counts* into account. Statistically, a dark count at Alice can be treated as a source photon-pair that reaches Alice but not Bob; similarly for a dark count at Bob. For example, when the dark-count rates at Alice and at Bob are λ_A and λ_B counts per slot, respectively, we can model the system by replacing η_A, η_B, and $\mathcal{E}$ with η'_A, η'_B, and $\mathcal{E}'$ that are solved from

$$\eta'_A \mathcal{E}' = \eta_A \mathcal{E} + \lambda_A \tag{71a}$$

$$\eta'_B \mathcal{E}' = \eta_b \mathcal{E} + \lambda_B \tag{71b}$$

$$\eta'_A \eta'_B \mathcal{E}' = \eta_A \eta_b \mathcal{E} \tag{71c}$$

without introducing any new elements to the model. Note that this replacement of parameters yields the desired correlation between Alice's and Bob's photon counts, but does *not* yield the correct form for Eve's optical states after Alice's and Bob's measurements. However, as our proofs show, information in Eve's optical states does not affect the dominant terms in secret-key rate in the regime of interest. This observation combined with our results shows that dark counts only affect the constant term in photon efficiency, which is again similar to the previous results in optical communications [17].

VII. Discussion: Toward Secrecy With a General Adversary

In this work we have presented schemes that approach the optimal key rate in the photon-efficient limit, up to a constant efficiency loss. Moreover, these schemes are practical, both in the physical sense (utilizing realizable transmissions and measurements) and in the algorithmic sense (using simple protocols and off-the-shelf codes). However, throughout the work we have assumed that Eve is limited to passive eavesdropping through a beamsplitter channel. We now comment on the problems that may arise when this model does not hold, and point out ways to overcome them.

First, suppose that Eve is still passive, but is free to change the beamsplitter transmissivity η as a function of time, as long as it satisfies some average constraint $\bar{\eta}$. We now distinguish between two strategies that Eve can use:

1) Pre-scheduled transmissivity. Take, for example, Scheme C-2, and imagine that for each PPM frame, Eve uses $\eta = 0$ for half the block, and $\eta = 1$ for the other half. Then she knows that the key pertaining to this frame must correspond to the part where $\eta = 1$, gaining one bit per detected photon (thus reducing the

key efficiency by $\log 2$). This kind of attack can go undetected, provided that Eve randomizes the schedule. However, it is plausible that the efficiency loss is bounded by a constant for any schedule.

2) Measurement-dependent transmissivity. In principle, Eve can change η in a causal manner, based upon her measurement outcomes. However, we believe that the gain from using measurements can be shown to vanish in the photon-efficient limit, by the same techniques used to show that the information leakage is small.

It however still remains to be investigated whether our intuitions above are correct, i.e., whether Eve indeed cannot gain from changing the beamsplitter transmissivity.

If Eve is allowed to transmit as well, other types of attacks are possible. A very simple and efficient one is "intercept and resend": Eve uses direct detection on the channel meant for Bob, and then upon detection of a photon, transmits a substitute one to Bob. This way Eve can obtain information about Bob's sequence of detections, and if she uses much higher bandwidth than Bob, the delay will not be detected.

In fact, all QKD protocols face this problem. For example, in the BB84 protocol [4], the key is generated using the polarization of a photon; Eve can make a measurement, then transmit to Bob a photon with the same polarization. The solution for BB84 is that Alice and Bob measure in either of two *mutually unbiased* bases, according to local randomness. Only if they happened to measure in the same basis, the measurement results are used, inflicting a rate loss of factor 2. By sacrificing rate, they can now *a posteriori* find out whether they used the same basis, and compare the correlation of the polarizations to the expected statistics, thus authenticating the received photons.

Extending this idea to schemes based on photon arrival times involves an extension of the concept of mutually unbiased bases to continuous variables; see [30]. Specifically, in Model C the modulation and measurements can be performed either in the time or in the frequency domain with the help of dispersive optics; see [31]. Alternatively, in Model S, one can use interferometry to verify that the photons received by Alice and Bob are indeed entangled; see [32].

APPENDIX

A. Proof of Proposition 4

By the same argument as in the proof of Proposition 2, we know that the raw keys generated by Alice and Bob (before privacy amplification) are the same, and are independent of Bob's messages in the information-reconciliation step. It is, however, dependent on Eve's optical states. We thus need to determine how much secret key can be distilled from the raw key.

The quantum states in different frames are mutually independent, so we need only to analyze one frame that is selected by Bob. We note that, when Alice sends the coherent state $|\sqrt{b\mathcal{E}}\rangle$, Eve's output state is $|\sqrt{(1-\eta)b\mathcal{E}}\rangle$, and is independent of Bob's measurement outcome conditional on Alice's input. Thus, using (17), we know that the number of secret nats we can obtain in each selected frame can be arbitrarily close to $H(\tilde{X}|\rho^{\mathbb{E}^b})$, where $\tilde{X}$ is uniformly distributed over $\{1,\dots,b\}$, and where $\rho^{\mathbb{E}^b}$ is a b-mode bosonic state described as follows: conditional on $\tilde{X}=i$, $i\in\{1,\dots,b\}$, $\rho^{\mathbb{E}^b}$ has the coherent state $|\sqrt{(1-\eta)b\mathcal{E}}\rangle$ in the ith mode and has the vacuum state $|0\rangle$ in all other modes. Note that the total number of photons in $\rho^{\mathbb{E}^b}$ is $(1-\eta)b\mathcal{E}$, so

$$H(\rho^{\mathbb{E}^b}) \le b\big\{\big(1+(1-\eta)\mathcal{E}\big)\log\big(1+(1-\eta)\mathcal{E}\big) - (1-\eta)\mathcal{E}\log\big((1-\eta)\mathcal{E}\big)\big\} \tag{72}$$

$$= \left\lceil \frac{1}{\mathcal{E}\log 1/\mathcal{E}} \right\rceil \left\{ (1-\eta)\mathcal{E}\log\frac{1}{\mathcal{E}} + O(\mathcal{E}) \right\} \tag{73}$$

$$= (1-\eta) + o(1). \tag{74}$$

Here, (72) follows from the well-known fact that the maximum entropy of a b-mode bosonic state with a certain average photon number is achieved by the state consisting of b i.i.d. thermal states [33]. Now the number of secret nats per selected frame satisfies

$$H(\tilde{X}|\rho^{\mathbb{E}^b}) = H(\tilde{X}) - I(\tilde{X};\rho^{\mathbb{E}^b}) \tag{75}$$

$$\ge H(\tilde{X}) - H(\rho^{\mathbb{E}^b}) \tag{76}$$

$$= \log b - H(\rho^{\mathbb{E}^b}) \tag{77}$$

$$= \log\frac{1}{\mathcal{E}} - \log\log\frac{1}{\mathcal{E}} - (1-\eta) + o(1). \tag{78}$$

We next consider the number of frames per k channel uses that will be selected by Bob, which we denote by $N(k)$. When Alice sends $|\sqrt{b\mathcal{E}}\rangle$, Bob's output has a Poisson distribution of mean $\eta b\mathcal{E}$, so the probability that Bob detects at least one photon is $1-e^{-\eta b\mathcal{E}}$. Hence, by the Law of Large Numbers,

$$\lim_{k\to\infty}\frac{N(k)}{k} = \lim_{k\to\infty}\frac{(1-e^{-\eta b\mathcal{E}})\lceil k/b\rceil}{k} = \frac{1-e^{-\eta b\mathcal{E}}}{b} \tag{79}$$

with probability one. Using

$$e^{-x} \le 1 - x + \frac{x^2}{2}, \qquad x\ge 0, \tag{80}$$

the right-hand side of (79) can be lower-bounded as

$$\frac{1-e^{-\eta b\mathcal{E}}}{b} \ge \eta\mathcal{E}\left(1-\frac{\eta b\mathcal{E}}{2}\right). \tag{81}$$

The photon efficiency of the proposed scheme can now be lower-bounded as

$$r_{\text{C-2}} = \frac{1}{\eta\mathcal{E}}\cdot\frac{1-e^{-\eta b\mathcal{E}}}{b}\cdot H(\tilde{X}|\rho^{\mathbb{E}^b}) \tag{82}$$

$$\ge \left(1-\frac{\eta b\mathcal{E}}{2}\right)\left\{\log\frac{1}{\mathcal{E}} - \log\log\frac{1}{\mathcal{E}} - (1-\eta) + o(1)\right\} \tag{83}$$

$$= \left(1-\frac{\eta\left\lceil\frac{1}{\mathcal{E}\log 1/\mathcal{E}}\right\rceil\mathcal{E}}{2}\right) \cdot\left\{\log\frac{1}{\mathcal{E}} - \log\log\frac{1}{\mathcal{E}} - (1-\eta) + o(1)\right\} \tag{84}$$

$$= \log\frac{1}{\mathcal{E}} - \log\log\frac{1}{\mathcal{E}} - (1-\eta) + o(1), \tag{85}$$

which is as claimed.

B. Proof of Proposition 7

We first observe that, in every selected frame, the detection positions of Alice and Bob must be the same. This is because, due to (63), $B = 1$ can happen only if $A = 1$, and because by our choice each selected frame contains only one source use where $A = 1$. We thus know that Alice's and Bob's raw keys are the same with probability one.

To obtain the secret-key rate, we need to compute the entropy of the raw key conditional on Eve's observations. Note that the quantum states inside different frames are mutually independent. We consider one frame that is selected by Alice and Bob. Denote the detection position in the frame by $\tilde{X}$. It is clear that $\tilde{X}$ is uniformly distributed over $\{1, \ldots, b\}$ and is independent of the label of this frame. All Eve's information about $\tilde{X}$ is in her optical state from the b source uses that form this frame: if the source use where $A = B = 1$ contains more than one photons, then Eve could also detect a photon in this source use, hence knowing Alice's and Bob's detection position. But, as we next show, this information leakage is small. To this end, we first note that in source uses where $A = B = 0$, Eve's optical state is vacuum. Indeed, according to our source model, the number of photons in Alice's state equals the sum of the numbers of photons in Bob's and Eve's states with probability one. Therefore, when both Alice and Bob make direct detections on a source use and observe no photon, Eve's post-measurement state in the same source use becomes the vacuum state. In the (unique) source use where $A = B = 1$, Eve's post-measurement state is the same as her state *without* the condition $A = B = 1$ given in (29). This is because $A = B = 1$ means nothing but that Bob's photon number is positive, but Eve's post-measurement state is independent of Bob's photon number, as shown in Section III-D. Denote Eve's state over the whole frame by $\sigma^{\mathbb{E}^b}$. We now know that it consists of $b - 1$ vacuum states and one state of the form (29) whose position inside the frame is random. The expected number of photons in $\sigma^{\mathbb{E}^b}$ is $(1-\eta)\mathcal{E}$, so the entropy of $\sigma^{\mathbb{E}^b}$ is upper-bounded by [33]

$$H(\sigma^{\mathbb{E}^b}) \le b \cdot g\left(\frac{(1-\eta)\mathcal{E}}{b}\right) = o(1). \tag{86}$$

Thus the amount of secret information extractable from one selected frame is lower-bounded by

$$\begin{aligned} H(\tilde{X}|\sigma^{\mathbb{E}^b}) &= H(\tilde{X}) - I(\tilde{X}; \sigma^{\mathbb{E}^b}) && (87)\\ &\ge H(\tilde{X}) - H(\sigma^{\mathbb{E}^b}) && (88)\\ &\ge \log\left\lceil \frac{1}{\mathcal{E}\log 1/\mathcal{E}} \right\rceil + o(1) && (89)\\ &= \log\frac{1}{\mathcal{E}} - \log\log\frac{1}{\mathcal{E}} + o(1). && (90) \end{aligned}$$

It now remains to compute the probability that a specific frame will be selected by Alice and Bob. A simple lower bound on the probability of a frame being selected is the following: suppose both Bob and Eve make PNR direct detections on their states, then a frame is selected by Alice and Bob if (but not only if) Bob detects exactly one photon in the frame while Eve detects no photon. Bob's photon number has a Poisson distribution of mean $\eta b\mathcal{E}$, while Eve's photon number has a Poisson distribution of mean $b(1-\eta)\mathcal{E}$, and the two photon numbers are independent. Hence the probability a frame being selected is lower-bounded by

$$\left(\eta b\mathcal{E}\, e^{-\eta b\mathcal{E}}\right) \cdot \left(e^{-b(1-\eta)\mathcal{E}}\right) = \eta b\mathcal{E} - \eta b^2\mathcal{E}^2 + o\left(\frac{1}{\log 1/\mathcal{E}}\right). \tag{91}$$

Multiplying (91) with $H(\tilde{X}|\sigma^{\mathbb{E}^b})$ gives us the secret-key nats per frame, where we count both selected and unselected frames. Simple normalization then yields the photon efficiency

$$\begin{aligned} r_{\text{S-2}}(\eta, \mathcal{E}) &\ge \frac{1}{\eta b\mathcal{E}} \cdot \left(\eta b\mathcal{E} - \eta b^2\mathcal{E}^2 + o\left(\frac{1}{\log 1/\mathcal{E}}\right)\right) \\ &\quad \cdot \left(\log\frac{1}{\mathcal{E}} - \log\log\frac{1}{\mathcal{E}} + o(1)\right) && (92)\\ &= \log\frac{1}{\mathcal{E}} - \log\log\frac{1}{\mathcal{E}} - 1 + o(1). && (93) \end{aligned}$$

C. Proof of Proposition 8

The second part of the secret key, which is generated in Step 5) in Scheme S-5, is exactly the (whole) secret key generated by Scheme S-4, and hence contributes to the total photon efficiency by the right-hand side of (69). It is clear that this is independent of the first part of the key, as the latter only contains information of the frame labels. We thus only need to evaluate the contribution to the photon efficiency from the first part of the key which is generated in Step 5).

Consider a block of ℓ length-b frames. To compute the length of the first part of the key that can be obtained from these frames, we first consider the information leakage due to Bob's message to Alice. Note that $(\tilde{A}^\ell, \tilde{B}^\ell)$ is distributed i.i.d. in time, where each pair (A, B) has joint distribution according to a Z channel with

$$\tilde{q} \triangleq P_A(1) = 1 - e^{-b\mathcal{E}} \tag{94a}$$

$$\tilde{\mu} \triangleq P_{B|A}(1|1) = \frac{1 - e^{-\eta b\mathcal{E}}}{1 - e^{-b\mathcal{E}}}. \tag{94b}$$

The optimal Slepian-Wolf code for Bob to convey $\tilde{B}^\ell$ to Alice should contain, asymptotically, $H(\tilde{B}|\tilde{A})$ nats per frame [28]. Let M_B be the message which Bob sends to Alice, then

$$H(M_B) = \ell H(\tilde{B}|\tilde{A}) + \ell\epsilon \tag{95}$$

where ϵ tends to zero as ℓ tends to infinity.

We next bound the information leakage due to the message which Alice sends to Bob. A simple upper bound is: for each frame where $\tilde{B} = 1$, Alice needs to send Bob at most one bit. From (94) we can obtain

$$P_{\tilde{B}}(1) = 1 - e^{-\eta b\mathcal{E}}. \tag{96}$$

Let M_A be the message which Bob sends to Alice for ℓ frames, then

$$H(M_A) \le \ell\left(1 - e^{-\eta b\mathcal{E}}\right) + \ell\epsilon. \tag{97}$$

We finally consider Eve's quantum state from the optical channel. Denote this state over ℓ frames by $\rho^{\mathbb{E}^{b\ell}}$. Since,

as shown in Section III-D, it is independent of Bob's measurement (direct detection) outcomes, and since $\tilde{B}$ is a function of Bob's measurement outcomes, we know that

$$I\left(\tilde{B}^{\ell};\rho^{\mathbb{E}^{b\ell}}\right)=0. \tag{98}$$

We now use (96), (97) and (98) to bound the length of the first part of the key for ℓ frames which, according to (17), is given by

$$\begin{aligned} H(\tilde{B}^{\ell}|M_A,M_B,\rho^{\mathbb{E}^{b\ell}}) &= H(\tilde{B}^{\ell})-\underbrace{I(\tilde{B}^{\ell};\rho^{\mathbb{E}^{b\ell}})}_{=0}-\underbrace{I(M_A,M_B;\tilde{B}^{b\ell}|\rho^{\mathbb{E}^{b\ell}})}_{\leq H(M_A)+H(M_B)} && (99)\\ &\geq \underbrace{H(\tilde{B}^{\ell})}_{=\ell H(\tilde{B})}-\underbrace{H(M_A)}_{\leq \ell(1-e^{-\eta b\mathcal{E}})+\ell\epsilon}-\underbrace{H(M_B)}_{=\ell H(\tilde{B}|\tilde{A})+\ell\epsilon} && (100)\\ &\geq \ell H(\tilde{B})-\ell\left(1-e^{-\eta b\mathcal{E}}\right)-\ell H(\tilde{B}|\tilde{A})-2\ell\epsilon && (101)\\ &= \ell I(\tilde{A};\tilde{B})+\ell\eta b\mathcal{E}-2\ell\epsilon+o(\mathcal{E}). && (102)\end{aligned}$$

Hence, for large enough ℓ, the length of the first part of the key *per frame* is given by

$$I(\tilde{A};\tilde{B})+\ell\eta b\mathcal{E}+o(\mathcal{E}). \tag{103}$$

We next evaluate $I(\tilde{A};\tilde{B})$. Comparing the parameters (94) to (63), we see that $I(\tilde{A};\tilde{B})$ is the same as $I(A;B)$ (68), replacing $\mathcal{E}$ with $b\mathcal{E}$, where, recalling (60),

$$b\mathcal{E}=\frac{1}{\log 1/\mathcal{E}}+o(\mathcal{E}). \tag{104}$$

Thus,

$$\begin{aligned} I(\tilde{A};\tilde{B}) &= H_2(e^{-\eta b\mathcal{E}})-\left(1-e^{-b\mathcal{E}}\right)H_2\left(\frac{1-e^{-\eta b\mathcal{E}}}{1-e^{-b\mathcal{E}}}\right) && (105)\\ &= \eta b\mathcal{E}\log\log\frac{1}{\mathcal{E}}+\eta b\mathcal{E}-b\mathcal{E}H_2(\eta)\\ &\quad +o\left(\frac{1}{\log 1/\mathcal{E}}\right). && (106)\end{aligned}$$

We can now compute the photon efficiency coming from the first part of the secret key in Scheme S-5 by dividing (103) by $\eta b\mathcal{E}$ (the average number of photons Bob detects per frame), and by using (106). This photon efficiency is at least

$$\log\log\frac{1}{\mathcal{E}}+1-\frac{H_2(\eta)}{\eta}+o(1). \tag{107}$$

Adding (107) to the right-hand side of (69), i.e., to the photon efficiency coming from the second part of the secret key, we conclude that

$$r_{\text{S-3}}(\eta,\mathcal{E})\geq\log\frac{1}{\mathcal{E}}-\frac{H_2(\eta)}{\eta}+o(1). \tag{108}$$

ACKNOWLEDGMENTS

The authors would like to thank Nivedita Chandrasekaran and Jeffrey Shapiro for helpful discussions, and the anonimous reviewers for useful comments.

REFERENCES

[1] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography. I. Secret sharing," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1121–1132, Jul. 1993.

[2] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 733–742, May 1993.

[3] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, 1975.

[4] C. H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," in *Proc. IEEE Int. Conf. Comput., Syst. Signal Process.*, Bangalore, India, Dec. 1984, pp. 175–179.

[5] A. K. Ekert, "Quantum cryptography based on Bell's theorem," *Phys. Rev. Lett.*, vol. 67, pp. 661–663, Aug. 1991.

[6] W. K. Wootters and W. H. Zurek, "A single quantum cannot be cloned," *Nature*, vol. 3, pp. 802–803, Oct. 1982.

[7] J. H. Shapiro, "The quantum theory of optical communications," *IEEE J. Sel. Topics Quantum Electron.*, vol. 15, no. 6, pp. 1547–1569, Nov./Dec. 2009.

[8] I. Bar-David, "Communication under the Poisson regime," *IEEE Trans. Inf. Theory*, vol. 15, no. 1, pp. 31–37, Jan. 1969.

[9] Y. M. Kabanov, "The capacity of a channel of the Poisson type," *Theory Probab. Appl.*, vol. 23, no. 1, pp. 143–147, 1978.

[10] M. Davis, "Capacity and cutoff rate for Poisson-type channels," *IEEE Trans. Inf. Theory*, vol. 26, no. 6, pp. 710–715, Nov. 1980.

[11] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[12] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics*. Cambridge, U.K.: Cambridge Univ. Press, 1995.

[13] R. Renner and R. König, "Universally composable privacy amplification against quantum adversaries," in *Theory of Cryptography*, J. Kilian, Ed. New York, NY, USA: Springer-Verlag, 2005, pp. 407–425.

[14] T. Zhong, F. N. C. Wong, A. Restelli, and J. C. Bienfang, "Efficient single-spatial-mode periodically-poled $KTiOPO_4$ waveguide source for high-dimensional entanglement-based quantum key distribution," *Opt. Exp.*, vol. 20, pp. 26868–26877, Nov. 2012.

[15] Y. Kochman and G. W. Wornell, "On high-efficiency optical communication and key distribution," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2012.

[16] B. Erkmen, B. Moision, S. J. Dolinar, K. M. Birnbaum, and D. Divsalar, "Approaching the ultimate limits of communication efficiency with a photon-counting detector," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2012.

[17] L. Wang and G. W. Wornell, "A refined analysis of the Poisson channel in the high-photon-efficiency regime," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4299–4311, Jul. 2014.

[18] A. S. Holevo, "The capacity of the quantum channel with general signal states," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 269–273, Jan. 1998.

[19] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels," *Phys. Rev. A*, vol. 56, no. 1, pp. 131–138, 1997.

[20] 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.*, vol. 92, no. 2, pp. 027902-1–027902-4, 2004.

[21] R. G. Gallager, "Energy limited channels: Coding, multiaccess, and spread spectrum," Laboratory for Information and Decision Systems, MIT, Cambridge, MA, USA, Tech. Rep. 1714, Nov. 1987.

[22] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, Sep. 1990.

[23] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, Jul./Oct. 1948.

[24] H. W. Chung, S. Guha, and L. Zheng, "On capacity of optical channels with coherent detection," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, St. Petersburg, Russia, Jul./Aug. 2011, pp. 284–288.

[25] J. Pierce, "Optical channels: Practical limits with photon counting," *IEEE Trans. Commun.*, vol. 26, no. 12, pp. 1819–1821, Dec. 1978.

[26] J. L. Massey, "Capacity, cutoff rate, and coding for a direct-detection optical channel," *IEEE Trans. Commun.*, vol. 29, no. 11, pp. 1615–1621, Nov. 1981.

[27] R. Wilmink, "Quantum broadcast channels and cryptographic applications for separable states," Ph.D. dissertation, Univ. Bielefeld, Bielefeld, Germany, 2003.

[28] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.

[29] H. Zhou and G. W. Wornell, "Adaptive pulse-position modulation for high-dimensional quantum key distribution," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 359–363.
[30] S. Weigert and M. Wilkinson, "Mutually unbiased bases for continuous variables," *Phys. Rev. A*, vol. 78, no. 2, p. 020303(R), 2008.
[31] J. Mower, Z. Zhang, P. Desjardins, C. Lee, J. H. Shapiro, and D. Englund, "High-dimensional quantum key distribution using dispersive optics," *Phys. Rev. A*, vol. 87, no. 6, pp. 062322-1–062322-6, 2012.
[32] J. Mower, F. N. C. Wong, J. H. Shapiro, and D. Englund, "Sending classical information via noisy quantum channels," *Phys. Rev. A*, vol. 56, no. 1, pp. 131–138, 1997.
[33] A. S. Holevo, M. Sohma, and O. Hirota, "Capacity of quantum Gaussian channels," *Phys. Rev. A*, vol. 59, no. 3, pp. 1820–1828, 1999.

Yuval Kochman (S'06–M'09) received his B.Sc. (cum laude), M.Sc. (cum laude) and Ph.D. degrees from Tel Aviv University in 1993, 2003 and 2010, respectively, all in electrical engineering. During 2009-2011, he was a Postdoctoral Associate at the Signals, Informtion and Algorithms Laboratory at the Massachusetts Institute of Technology (MIT), Cambridge. Since 2012, he has been with the School of Computer Science and Engineering at the Hebrew University of Jerusalem. Outside academia, he has worked in the areas of radar and digital communications. His research interests include information theory, communications and signal processing.

Ligong Wang (S'08–M'12) received the B.E. degree in electronic engineering from Tsinghua University, Beijing, China, in 2004, and the M.Sc. and Dr.Sc. degrees in electrical engineering from ETH Zurich, Switzerland, in 2006 and 2011, respectively. He is currently a Postdoctoral Associate at the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, Cambridge, MA, USA. His research interests include classical and quantum information theory, and optical communication.

Gregory W. Wornell (S'83–M'91–SM'00–F'04) received the B.A.Sc. degree in electrical engineering from the University of British Columbia, Vancouver, BC, Canada, and the S.M. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 1985, 1987, and 1991, respectively.

Since 1991, he has been on the faculty at MIT, where he is the Sumitomo Professor of Engineering in the department of Electrical Engineering and Computer Science (EECS). He leads the Signals, Information, and Algorithms Laboratory in the Research Laboratory of Electronics, and co-chairs the EECS department graduate program. He has held visiting appointments at the former AT&T Bell Laboratories, Murray Hill, NJ, USA, the University of California, Berkeley, CA, USA, and Hewlett-Packard Laboratories, Palo Alto, CA, USA. His research interests and publications span the areas of signal processing, digital communication, and information theory, and include algorithms and architectures for wireless networks, sensing and imaging systems, digitally enhanced analog circuits and systems, multimedia applications, and aspects of computational biology and neuroscience.

Dr. Wornell has been involved in the Information Theory and Signal Processing Societies of the IEEE in a variety of capacities, and maintains a number of close industrial relationships and activities. He has won a number of awards for both his research and teaching.