Quantum steganography via coherent- and Fock-state encoding in an optical medium

Bruno Avritzer*
Department of Physics and Astronomy, University of Southern California, Los Angeles, California 90089, USA

Todd A. Brun†
Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, USA

(Received 29 March 2023; accepted 6 February 2024; published 1 March 2024)

Steganography is an alternative to cryptography, where information is protected by secrecy (being disguised as innocent communication or noise) rather than being scrambled. In this work we develop schemes for steganographic communication using Fock and coherent states in optical channels based on disguising the communications as thermal noise, with a fidelity approaching 1. We derive bounds on their efficiency in terms of the communication rate and the required keyrate in the case of an all-powerful eavesdropper and provide explicit methods of encoding and error correction for the noiseless channel case.

DOI: 10.1103/PhysRevA.109.032401

I. INTRODUCTION

Covert communication is often associated with military or espionage applications, but some of the earliest applications were for personal use. In ancient Egypt, hieroglyphs were used by royal scribes to send the pharaoh's messages covertly. In Roman times, the Caesar cipher, where each letter in a message was "shifted" by a predetermined amount (akin to a secret key), was devised [1], this approach being refined into the Vigenere encoding around the 15th Century [2]. Over time, developments became more mathematically and technologically advanced. Famously, the use of the Enigma machine in World War II enabled the Germans to communicate covertly until the cipher was cracked by the Allied effort, giving them a strategic information advantage for the rest of the war [3,4].

The Allies were able to replicate, understand, and reproduce the technology of the Enigma machine to crack its cipher; thus the practical application of quantum computers and quantum information devices, which represent an additional level of technological sophistication, are evident. The question we would like to explore is how to leverage quantum computers and quantum information, a present and future technological advantage, to communicate covertly in a variety of situations.

The main scenario we will consider is one where the eavesdropper has access to the full contents of the messages being transmitted, but is made to believe they are innocuous, unlike in cryptography where the encrypted text is often nonsensical without a key and so arouses suspicion. This reflects a difference in the situation: In cryptography, the messages are often read secretly, while in steganography the eavesdropper is not secret and maintains everything out in the open. In many ways, this is advantageous (consider the wartime example). If the goal is not to arouse suspicion, it can be markedly more suspicious and dangerous to send messages of gibberish through wartime censors than to have a chat with your friend on the phone about normal topics, only this chat contains some hidden information. Other examples of steganography are an invisible watermark that can only be revealed with a procedure no one would think to do spontaneously [4] or a secret encoding of information in an audio file [5]. Some steganographic encodings are readable by anyone who thinks to look for them, relying on the concealment of the covering message, while others may require a shared secret key between the sender and receiver, just like many cryptosystems. Steganography can even be combined with cryptography as they are effectively independent measures, for example, by using any methods we describe in this paper to transmit an already cryptographically encoded message. This will, in general, require more shared secret key.

A large amount of work has been done on quantum cryptography [6–8]. The field of quantum steganography (and, broadly, covert communication) is smaller, but also includes a substantial body of relevant theoretical work. It was recently shown that over n uses of additive white Gaussian noise (AWGN), a number of bits proportional to $\sqrt{n}$ can be communicated covertly [9], which was later generalized in [10]. Furthermore, a number of methods have been devised for such communications using quantum systems, as in [11–13]. In this paper we will study the encoding of information in quantum states transmitted over an optical channel in such a way that it imitates thermal noise. This follows the broad approach of Brun and Shaw in [14,15] (for qubit channels), and has been studied for optical channels in [16]. Like the last work, our work follows the "secrecy" approach typical of steganography in which the message is protected by the fact that its existence is concealed. This is as opposed to the "security" approach of standard cryptography, as well as methods such as spread spectrum and chaotic communication

*avritzer@usc.edu
†tbrun@usc.edu

that are not generally secret at an information-theoretic standard [17,18]. Steganography, as studied in this paper, provides formal guarantees of secrecy based on metrics of fidelity and trace distance, unlike the aforementioned approaches, while also functioning in a narrow band. Compared to [16], our work is experimentally simpler, though its practical performance at scale remains to be demonstrated. It also does not require any assumptions about the ability of the eavesdropper to detect the noise beyond their expectation of a thermal state, which enables a potentially greater ability to communicate (as quantified by the communication rate and rate of secret key consumption). This current work only treats classical communication, but it shows the kind of methods that could be used in future quantum steganographic encodings for entanglement distribution or quantum communication, perhaps drawing on techniques similar to those in [19]. In this work, we prove secrecy by calculating the trace distance or fidelity between the "innocent" (thermal) state and the average state containing hidden information. This approach is sufficient to demonstrate the effectiveness of a steganographic method and is simpler than the proofs needed to show security in cryptography.

In Secs. II and III we develop the machinery required to understand the communication process. In Sec. IV we develop and analyze some simple encodings. In Secs. V and VI we do a more detailed analysis of their implementation and efficiency. Section VII summarizes the results and discusses future work.

II. IMPORTANT MEASURES FOR STEGANOGRAPHY

In steganographic protocols, such as those we will propose in later sections, there is a key trade-off that must be taken into account: That of the effective communication rate achievable in the channel as opposed to the similarity of the targeted "innocent" state with the actual channel state. In some schemes, such as the Fock encoding we will discuss later, these can be quantified by just two measures: The communication rate [which is given by $R = 1 + p_{\text{err}} \ln p_{\text{err}} + (1 - p_{\text{err}}) \ln(1 - p_{\text{err}})$ for a binary channel with a probability p_{err} of mistaking one symbol for another] and the trace distance between the channel state ρ containing hidden information, represented by the density operator corresponding to the state of the channel over which information is being transmitted, and the "innocent" thermal state ρ_{th}. This trace distance is given by

$$D(\rho, \rho_{\text{th}}) = \tfrac{1}{2}||\rho - \rho_{\text{th}}|| \tag{1}$$

where the above norm is the trace norm. Alternatively, one can use the state fidelity

$$F(\rho, \rho_{\text{th}}) = [\text{Tr}(\sqrt{\sqrt{\rho}\rho_{\text{th}}\sqrt{\rho}})]^2 \tag{2}$$

as another measure of distance that functions similarly to the trace norm, although it is not a metric on the set of density matrices in the formal sense. The trace distance can be used to directly compute the minimum probability of mistaking ρ for ρ_{th} using a positive operator-valued measurement (POVM), and is therefore conceptually useful as a representation of secrecy. However, the fidelity is often easier to calculate and can be used to bound the trace distance [20] (it also can be interpreted as a probability of mistaking one state for another when at least one of the states is pure). In particular, if the fidelity is 1 (or approaching 1), the trace distance is 0 (or approaching 0) and the two states cannot be distinguished by any measurement.

In schemes where additional practical constraints are imposed to facilitate communication (for example, reducing the choice of possible states used in encoding to make distinguishing them easier, as in some of the coherent state protocols discussed below), we can assume that the sender and receiver draw on a preshared secret key, unknown to the eavesdropper. This key could be, for example, a secret string of random bits. This key usage can be quantified by a secret key rate K, the number of bits of secret key consumed per channel use. The ratio R/K or difference $R - K$ of the rates give additional measures of the usefulness of the scheme for steganographic communication.

These measures can help us evaluate the effectiveness of different potential protocols. On one extreme, if Alice and Bob simply send the "innocent" state at every time interval (i.e., with probability 1), the fidelity metric will have its highest possible value, which is 1. However, this encoding has no communication rate, $R = 0$. On the other hand, if Alice and Bob use a naive encoding that maps the input bits 0 and 1 to a fixed pair of orthogonal states, it will generally be impossible to have good fidelity with the thermal state, and Eve can easily detect that communication is happening. The goal of a good steganographic protocol is to encode the message (a string of input bits) into a sequence of states, such that after averaging over all possible messages (and also the preshared secret key, if any), the fidelity with a string of thermal states is close to 1, but Bob can also retrieve the encoded string with high probability.

The protocols we will consider deal with cases where the fidelity (or trace distance) is very close to 1, at least in an asymptotic sense; the communication rate is nonzero; and the system can be implemented physically via the transmission of physically realizable states. The first example we will discuss is a protocol using coherent states.

III. DISGUISING COHERENT STATES AS THERMAL NOISE IN A CHANNEL

A coherent state is a state of a quantum oscillator or field mode. It is defined, for some complex α, as

$$|\alpha\rangle = \sum_{n=0}^{\infty} \frac{\alpha^n e^{-|\alpha|^2/2}}{\sqrt{n!}} |n\rangle , \tag{3}$$

which is the result of acting with the displacement operator $D(\alpha) = e^{\alpha a^\dagger - \alpha^* a}$ on the $|0\rangle$ state of a harmonic oscillator. A thermal state is given for the same type of system by

$$\rho_{\text{th}} = \frac{1}{Z} \sum_{n=0}^{\infty} e^{-\frac{\hbar\omega(n+1/2)}{k_B T}} |n\rangle \langle n| , \tag{4}$$

where

$$Z = \sum_{n=0}^{\infty} e^{-\frac{\hbar\omega(n+1/2)}{k_B T}} = \frac{1}{2}\text{csch}\left(\frac{\hbar\omega}{2k_B T}\right) \tag{5}$$

is the partition function.

If we describe the thermal state of a mode in a channel in terms of the average number of photons transmitted, known as

$$\bar{n} = (e^{\frac{\hbar\omega}{k_B T}} - 1)^{-1}, \tag{6}$$

we can reformulate the expression for ρ_{th} in a simpler way:

$$\rho_{\text{th}} = \frac{1}{\bar{n}+1} \sum_{n=0}^{\infty} \left(\frac{\bar{n}}{\bar{n}+1}\right)^n |n\rangle \langle n| . \tag{7}$$

We want to represent this using coherent states over the phase space described by $|\alpha\rangle = |re^{i\theta}\rangle$, as in the Glauber P-Representation [21]

$$\begin{aligned}
\rho_{\text{th}} &= \frac{1}{N} \int d^2\alpha e^{-c|\alpha|^2} |\alpha\rangle \langle\alpha| \\
&= \frac{1}{N} \int_0^{\infty} r dr \int_0^{2\pi} d\theta e^{-cr^2} |re^{i\theta}\rangle \langle re^{i\theta}| \\
&= \frac{2\pi}{N} \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^{\infty} r^{2n+1} e^{-(c+1)r^2} |n\rangle \langle n| \\
&= \frac{\pi}{N} \sum_{n=0}^{\infty} \frac{1}{(c+1)^{n+1}} |n\rangle \langle n| , \qquad (8) \\
\Rightarrow \frac{\pi}{N(c+1)^{n+1}} &= \frac{1}{\bar{n}} \left(\frac{\bar{n}}{\bar{n}+1}\right)^{n+1}, \qquad (9) \\
\Rightarrow c = \frac{1}{\bar{n}}, N &= \pi\bar{n}, \qquad (10) \\
\Rightarrow \rho_{\text{th}} &= \frac{1}{\pi\bar{n}} \int_0^{\infty} dr \int_0^{2\pi} d\theta r e^{-\frac{r^2}{\bar{n}}} |re^{i\theta}\rangle \langle re^{i\theta}| . \qquad (11)
\end{aligned}$$

We can integrate over θ and consider this as a probability distribution over coherent states $|r\rangle$ with $p(r) = \frac{2}{\bar{n}} r e^{-\frac{r^2}{\bar{n}}}$, i.e., a Rayleigh distribution

$$\text{Rayleigh}\left(r; \sqrt{\frac{\bar{n}}{2}}\right) = \frac{2}{\bar{n}} r e^{-\frac{r^2}{\bar{n}}}. \tag{12}$$

The median of this distribution is given by $r_{1/2} = \sqrt{\bar{n}\ln 2}$, which is a convenient point of separation if we want to send binary messages.

Because the set of coherent states is overcomplete, the existence of a coherent-state representation is guaranteed. A related question is how well a set of coherent states with a set of M randomly chosen radii $\{r_j\}$ and uniformly random phases can approximate a thermal state. This gives a mixture

$$\rho_c = \frac{1}{M} \sum_{j=1}^{M} e^{-r_j^2} \sum_n \frac{r_j^{2n}}{n!} |n\rangle \langle n| . \tag{13}$$

The answer is remarkably simple:

$$\sqrt{F(\rho_{\text{th}}, \rho_c)} \geqslant 1 - \frac{\bar{n}}{2M} \Rightarrow ||\rho_{\text{th}} - \rho_c|| \leqslant \sqrt{\frac{4\bar{n}}{M} - \frac{\bar{n}^2}{M^2}}, \tag{14}$$

where F is the average fidelity $\sqrt{F(\rho_{\text{th}}, \rho_c)} = \text{Tr}(\sqrt{\rho_{\text{th}}^{1/2} \rho_c \rho_{\text{th}}^{1/2}})$. This is relevant for "Pairwise" protocols we will discuss later.

Another important bound is on the same kind of setup, but without averaging over θ. If we instead discretize the circle over θ, that is, we consider a set of states $|\alpha_{jk}\rangle = |r_j e^{\frac{2\pi i k}{L}}\rangle$ for sufficiently large L, we can do at least as well as the above result. A proof of both these bounds is contained in the Appendix.

IV. MAPPINGS FOR QUANTUM STEGANOGRAPHY

In this work we consider four main approaches to encoding information steganographically as states of light: the Fock state encoding, which requires no shared key; an encoding in coherent states without shared key; and two other encodings into coherent states that do require a shared key: The vertical angles encoding and the redefined Rayleigh distribution encoding.

A. Fock state methods

The scheme for Fock state methods is straightforward and has one clear advantage: Fock states are more easily distinguishable than coherent states, so the communication capacity is higher, although the problem of realizing an arbitrary Fock state in an experimental setting is also more challenging than for coherent states. The protocol we will describe requires only the preparation of multiphoton Fock states, and may be done using probabilistic operations such as photon addition. As such, it can, in principle, be done by using only single-photon sources, single-photon detectors, and beam splitters, although when $\bar{n}$ is high this may become experimentally difficult due to low observation probability. It should be noted that, in that regime, the coherent state methods described in this paper perform more competitively with the Fock state methods due to the greater ease of distinguishing between different coherent states at a higher amplitude. This method, however, requires an encoding system from binary digits to Fock states which depends on the value of $\bar{n}$, and is described in detail in Sec. VI and the Appendix. If we are dealing with Fock states in a noiseless channel, the problem is one of translating regular binary strings into binary strings with a certain number of 1s and 0s determined by $\bar{n}$. It is worth noting as an experimental consideration that, in cases where $\bar{n}$ is low, the weight of the encoded text will be low on average and a sophisticated encoding may not be necessary. In cases where $\bar{n}$ is very large, that may not be the case, and it would require us to define different Fock states as encoding the binary 1 or 0, according to the Boltzmann weights of such Fock states (for example, we might define all states below a certain value of n as belonging to 0, and the others to belong to 1). In Sec. VI we will consider a more intermediate case for communication of a message of length N bits and derive bounds on its efficiency.

B. Distribution coherent state methods

Since Eq. (14) describes the average fidelity, it makes sense to examine different approaches to optimize this quantity, taking into account the nonorthogonality of the coherent states being measured. In all cases we will draw from distributions

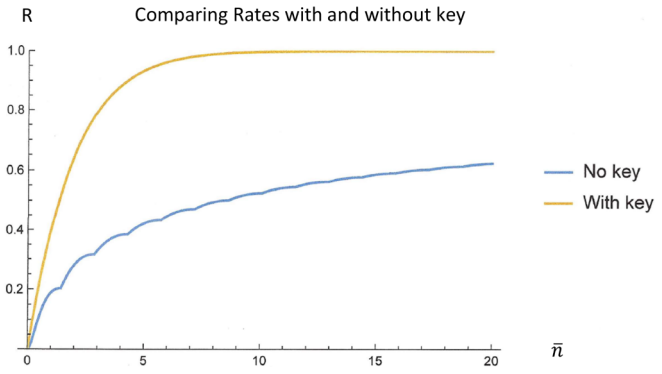

FIG. 1. A plot of the lower bound on the communication rate for the vertical angle (key) encoding scheme compared to the "Distribution" (no key) scheme.

defined by

$$\rho_0(r) = \frac{2}{\bar{n}} r e^{-\frac{r^2}{\bar{n}}}, \quad 0 < r < r_{\frac{1}{2}},$$
$$\rho_1(r) = \frac{2}{\bar{n}} r e^{-\frac{r^2}{\bar{n}}}, \quad r_{\frac{1}{2}} < r < \infty, \tag{15}$$

representing the transmission of 0 and 1 from the sender.

First, we consider the "distribution" case, which does not require a key, where Alice draws a coherent state randomly from ρ_0 or ρ_1 and Bob has to try and guess which distribution it came from. Bob's ability to do this is bounded by the trace distance between the states ρ_0 and ρ_1. This is difficult to evaluate, but can be evaluated in terms of the cumulative distribution function of the Poisson processes

$$Q_n = 2^{-(\bar{n}+1)} \sum_{k=0}^{n} \frac{\left(c r_{1/2}^2\right)^k}{k!} \tag{16}$$

for a process with $\lambda = (\bar{n}+1)\ln 2 = c r_{1/2}^2$ and $\tilde{Q}_n$ for $\lambda = \bar{n}\ln 2$, with $N_{\frac{1}{2}}$ the median of the Poisson process (a full derivation is included in the Appendix):

$$\frac{1}{2}||\rho_0 - \rho_1|| = \frac{1}{2(n+1)} \sum_{n=0}^{\infty} \left(\frac{\bar{n}}{\bar{n}+1}\right)^n |1 - 2Q_n|$$
$$= 2\left(1 + (2Q_{N_{\frac{1}{2}}} - 1)\left(\frac{\bar{n}}{\bar{n}+1}\right)^{N_{\frac{1}{2}}+1} - \tilde{Q}_{N_{\frac{1}{2}}}\right). \tag{17}$$

From this we can calculate the probability of error p_{err} and the communication rate R. Since the Poisson process is discrete, this curve has some kinks when N increases, as seen in Fig. 1. When $\bar{n}$ is large, the states ρ_0 and ρ_1 are almost orthogonal; but when $\bar{n}$ is $O(1)$ or smaller, the two states have significant overlap and are not perfectly distinguishable. In this case error correction may be necessary, and matching the thermal state may still require shared secret key, as discussed in Sec. VI. We consider two variations of this idea below.

C. Pairwise coherent state methods

A related method is to partition the thermal state as above and select a finite set of states from each half to send. As we show above, this finite set approximates the thermal state very well and can be distinguished more easily from each other. This method may be easier to implement practically, as only sampling from a subset of ρ_{th} is sufficient, but requires more secret key than the distribution methods, which will be quantified in Sec. VI.

D. Vertical angles

In this approach, Alice and Bob choose α_0 and α_1 ahead of time to have opposite phases θ, i.e., $\alpha_0 = r_0 e^{i\theta}$ and $\alpha_1 = r_1 e^{i(\theta+\pi)} = -r_1 e^{i\theta}$, where θ can be chosen arbitrarily. These states correspond to the binary 0 or 1 and are drawn from the distributions ρ_0 and ρ_1, although since this specific protocol is more useful for the low-$\bar{n}$ case, ρ_0 will likely be sampled more often. Knowledge of the two possible states constitutes a secret key, and the θ correlation helps distinguish the distributions by minimizing the overlap

$$|\langle\alpha_0|\alpha_1\rangle| = e^{-\frac{|\alpha_0-\alpha_1|^2}{2}} = e^{-\frac{|r_0+r_1|^2}{2}}.$$

This is maximized for $r_0 = 0$, $r_1 = r_{\frac{1}{2}}$ which gives

$$|\langle\alpha_0|\alpha_1\rangle| \leqslant e^{-\frac{r_{1/2}^2}{2}} = 2^{-\frac{\bar{n}}{2}}, \tag{18}$$
$$\Rightarrow p_{\text{err}} = \tfrac{1}{2}(1 - ||\,|\alpha_0\rangle\langle\alpha_0| - |\alpha_1\rangle\langle\alpha_1|\,||)$$
$$= \tfrac{1}{2}(1 - \sqrt{1 - |\langle\alpha_0|\alpha_1\rangle|^2})$$
$$\leqslant \tfrac{1}{2}(1 - \sqrt{1 - 2^{-\bar{n}}}). \tag{19}$$

The associated communication rate is $R = 1 + p_{\text{err}}\ln(p_{\text{err}}) + (1 - p_{\text{err}})\ln(1 - p_{\text{err}})$ which can be seen in Fig. 1.

We can think of this setup as using a key since Alice and Bob must have prior information linking the two α values, in contrast to the aforementioned no-key case. Using this key allows a greater rate of communication since the overlap between the coherent states representing 0 and 1 is minimized. We can see this in Fig. 1.

E. Redefined Rayleigh distributions

The final approach involves simply drawing r_0 and r_1 from the corresponding distributions, communicating them over a channel and attempting to determine which distribution was sampled from based on the channel measurement. This approach is equivalent to determining the overlap of the distributions ρ_0 and ρ_1, where we are once again randomizing θ, but the two are not simultaneously diagonal in any basis and so the optimal measurement to distinguish them is not easy to find with no key. In this section we consider a specific type of measurement and attempt to optimize for a variable parameter m_c which will distinguish between the two states ρ_0 and ρ_1, although in Sec. VI we will also consider measurements that saturate the Helstrom bound [22]

$$p_{\text{err}} \geqslant \tfrac{1}{2} - \tfrac{1}{2}||(1-f)\rho_0 - f\rho_1|| \tag{20}$$

when discriminating between two states ρ_0 and ρ_1 occurring with probabilities $1 - f$ and f, respectively.

There is one notable difference between this approach and the previous one. For easy distinction, we can, without loss of generality, choose $\theta = 0$ for our analysis and redefine the

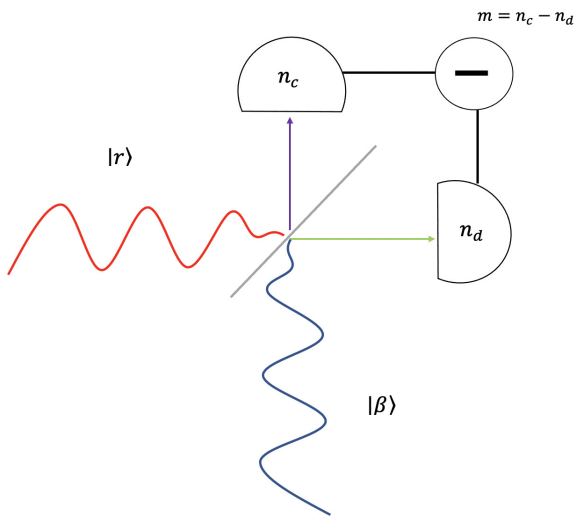

FIG. 2. In this setup, a beam splitter combines the coherent states denoted by $|r\rangle$ and $|\beta\rangle$, with n_c and n_d denoting detectors that measure the incidence of photons. The value of the homodyne measurement is given by $m = n_c - n_d$.

distribution $\tilde{\rho}_1$ as spanning $(-\infty, -r_{1/2}]$ to make it easier to distinguish from $\tilde{\rho}_0$, as this essentially rotates ρ_1 about the origin by π and creates the greatest possible distance between the means and medians of ρ_0 and ρ_1 by such an operation, while not fundamentally changing the nature of any calculations we will perform.

1. Setup

The coherent-state-based protocols we consider use a balanced homodyne measurement of a state $|r\rangle$. This state is coupled to an oscillator by means of the beam splitter shown in Fig. 2. If we have the operators a and $a^\dagger$, of which $|r\rangle$ is an eigenket of a, and likewise b and $b^\dagger$ for $|\beta\rangle$, once the states pass through the beam splitter the outcome is characterized by the new operators $c = \frac{a+b}{\sqrt{2}}$ and $d = \frac{a-b}{\sqrt{2}}$. As such, the desired observable is given by $m = n_c - n_d = c^\dagger c - d^\dagger d = a^\dagger b + b^\dagger a$.

2. Bounds

We can derive bounds using the generalized Markov inequality

$$P(X - \mu > \lambda) \leqslant \frac{M_n(X)}{\lambda^n} \tag{21}$$

for even n, where X is a random variable, μ is its mean, and $M_n(X)$ is its nth moment.

Applying this to the distributions for r_1 and r_0, we have for a cutoff value $m_c \in [2\beta\bar{r}_1, 2\beta\bar{r}_0]$ that

$$\begin{aligned} P(m - \bar{m}_1 > \lambda) &\leqslant \frac{M_n(m_1)}{\lambda^n}, \\ P(m - \bar{m}_0 < \lambda) &\leqslant \frac{M_n(m_0)}{\lambda^n}, \end{aligned} \tag{22}$$

where here $\bar{m}_1$ and $\bar{m}_0$ refer to the mean expected m values for each distribution, $2\beta\bar{r}_1$ and $2\beta\bar{r}_0$, respectively.

There is an explicit formula:

$$M_n(X) = \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} E[X^k](E[X])^{n-k}. \tag{23}$$

We also have the explicit formula for moments of m, under the approximation $b \approx \beta$ (since only the leading order terms in β matter for sufficiently large real β):

$$\begin{aligned} E(m^k) &= \int p(r)\beta^k \langle r|(a + a^\dagger)^k|r\rangle \, dr \\ &= \frac{4\beta^k}{\bar{n}} \int dr (re^{-\frac{r^2}{\bar{n}}}) \mathcal{F}_k(r), \end{aligned} \tag{24}$$

where

$$\mathcal{F}_k(r) = \begin{cases} \sum_{j=0}^{k/2} \frac{k! 2^{3j-k/2}}{(2j)!(k/2-j)!} r^{2j}, & k \text{ even} \\ \sum_{j=0}^{(k-1)/2} \frac{k! 2^{3j-\frac{k-3}{2}}}{(2j+1)!(\frac{k-1}{2}-j)!} r^{2j+1}, & k \text{ odd.} \end{cases} \tag{25}$$

A derivation of the above is obtained by acting with $\langle r|D(\alpha)\rangle\, r = e^{2i\beta r - \frac{\beta^2}{2}}$ and equating the Taylor expansions of both sides to each order in β. This provides a means to evaluate the higher-moment Markov bound

$$\begin{aligned} p_{\text{err}} &= \frac{1}{2}[p(m - \bar{m} > m_c - \bar{m}|1) + p(m - \bar{m} < m_c - \bar{m}|0)] \\ &\leqslant \frac{1}{2}\left[\frac{M_n(m_1)}{(m_c - \bar{m}_1)^n} + \frac{M_n(m_0)}{(m_c - \bar{m}_0)^n}\right]. \end{aligned} \tag{26}$$

V. NUMERICAL SIMULATION AND PERFORMANCE

We expect that, for a coherent state, the distribution of m will be Gaussian [23]. It is straightforward to compute the mean and variance of m in the case where r is randomly sampled from the distributions without constraint, rather than being one of two possibilities:

$$\begin{aligned} \bar{m} &= 2\beta\bar{r}, \\ \bar{r}_0 &= \frac{4}{\bar{n}} \int_0^{\sqrt{\bar{n}\ln 2}} r^2 e^{-\frac{r^2}{\bar{n}}} dr \approx 0.516\sqrt{\bar{n}}, \\ \bar{r}_1 &= -\frac{4}{\bar{n}} \int_{-\infty}^{-\sqrt{\bar{n}\ln 2}} r^2 e^{-\frac{r^2}{\bar{n}}} dr \approx -1.256\sqrt{\bar{n}}, \\ \bar{r}_0^2 &= \frac{4}{\bar{n}} \int_0^{\sqrt{\bar{n}\ln 2}} r^3 e^{-\frac{r^2}{\bar{n}}} dr \approx 0.307\bar{n}, \\ \bar{r}_1^2 &= \frac{4}{\bar{n}} \int_{-\infty}^{-\sqrt{\bar{n}\ln 2}} r^3 e^{-\frac{r^2}{\bar{n}}} dr \approx 1.693\bar{n} \\ \Rightarrow \Delta r_0^2 &\approx .041\bar{n}\,; \quad \Delta r_1^2 \approx 0.131\bar{n}. \end{aligned} \tag{27}$$

In addition,

$$\begin{aligned} \text{Var}(m) &= E(m^2) - [E(m)]^2 \\ &= 4\beta^2\Delta r^2 + \beta^2 + \bar{r}^2, \\ \text{Var}(m_0) &= (0.164\bar{n} + 1)\beta^2 + 0.307\bar{n}, \\ \text{Var}(m_1) &= (0.524\bar{n} + 1)\beta^2 + 1.693\bar{n}. \end{aligned} \tag{28}$$

Sampling from a normal distribution with these parameters is simple, so we can empirically determine the optimal value

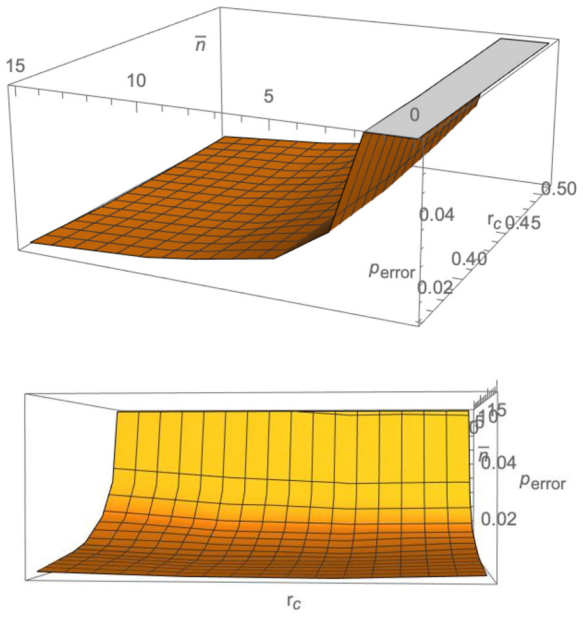

FIG. 3. Two views of the plot of p_{err} as a function of $r_c = -\frac{m_c}{2\beta\sqrt{\bar{n}}}$ and $\bar{n}$. The optimal value of r_c is between 0.4 and 0.5.

of m_c by simulating the transmission homodyne measurement procedure directly. The results are displayed in Fig. 3.

The above refers to the case where r is randomly sampled from the distributions without constraint, rather than being one of two possibilities. In the second case, where the key designates one of two specific states to be distinguished between by homodyne measurement for each transmitted bit (one state $|r_0\rangle$ from ρ_0 and one $|r_1\rangle$ from ρ_1), the optimal value calculation for $m_c \approx \beta(r_0 + r_1)$ is more straightforward: It derives from optimizing the accumulated probability

$$\int_{-\infty}^{m_c} \frac{e^{-\frac{(m-2\beta r_1)^2}{2(r_1^2+\beta^2)}}}{\sqrt{r_1^2+\beta^2}} - \frac{e^{-\frac{(m-2\beta r_0)^2}{2(r_0^2+\beta^2)}}}{\sqrt{r_0^2+\beta^2}} dm \tag{29}$$

at m_c, and gives the result

$$p_{\text{err}} = \frac{1}{4}\left[2 + \text{erf}\left(\frac{m_c - 2\beta r_0}{\sqrt{2(\beta^2 + r_0^2)}}\right) - \text{erf}\left(\frac{m_c - 2\beta r_1}{\sqrt{2(\beta^2 + r_1^2)}}\right)\right], \tag{30}$$

which does not exceed $1/2$. We will compare this to the Helstrom bound once we derive a bound for Fock state communication in Sec. VI, so they can all be seen side-by-side.

VI. PARTICULAR ENCODING METHODS

A. Fock state methods

Continuing from the discussion in Sec. IV, if we denote any Fock states $|n\rangle$ with $n \geqslant 1$ as the binary 1, we must have a quantity of $n_z = \frac{N}{\bar{n}+1}$ 0s and $N - n_z = \frac{N\bar{n}}{\bar{n}+1}$ 1s in the encoded string, which comes directly from the Fock state representation of ρ_{th}. Thus the problem is one of encoding from the set of all binary strings of length N to the set of binary strings with such a ratio, of which there are $\binom{N}{n_z}$. It is straightforward to calculate the channel capacity using such an encoding, as the number of bits we can encode is simply $Nh(\frac{1}{\bar{n}+1})$: This provides evidence for the simplicity of the cases of extreme $\bar{n}$ as noted in Sec. IV.

We can approach this rate using a "by value" encoding. Think of an N-bit string as representing an integer w, and encode this as the wth smallest bit string (by value) that satisfies the above criteria for the number of 0s and 1s. We can use a theorem called the "Christmas Stocking Theorem" [24] to efficiently generate the encoding, going from either the least or most significant bit. This theorem states that

$$\sum_{i=0}^{k-1} \binom{n+i}{i} = \binom{k+n}{k-1}, \tag{31}$$

which provides a straightforward way of counting down digits. More details and examples are contained in the Appendix.

B. Coherent state methods

We want to emulate the statistics of the thermal state using coherent states sampled from our distributions ρ_1 and ρ_0. If we have an equal number of 0s and 1s in the message, this is straightforward: We can sample from the distributions and simply transmit the result. There is a catch, however: Since the coherent states are not orthogonal, there is a probability of mistaking ρ_1 for ρ_0 at the time of measurement, which produces something similar to a Pauli X error. We can protect against this by using error correction and encoded keywords; however, this requires secret key, since the encoded messages will no longer appear to be sampled from ρ_{th}.

For example, if we use a simple three-bit Hamming code, there are only two codewords (000 and 111) and eight possible, equally likely bit-strings we would expect to see if sampling from ρ_{th}. We can make these strings appear random again by doing a bitwise XOR with a random three-bit string, selected by generating a random number between 0 and 7. This scrambled codeword will still protect against a bit-flip error, but it requires that Bob also knows the random number that was chosen. So such a scheme requires Alice and Bob to share a secret key in advance. If we wish to cut down on the amount of secret keys used, we could use a shared seed for a pseudorandom number generation protocol, such as the rabbit cipher, which is thought to be cryptographically secure [25]. However, this would reduce the secrecy below the information-theoretic level we are assuming up to this point.

1. Distribution method

We can quantify the results of the coherent state encoding. Supposing that, instead of splitting the coherent distribution in half radially, we split it so that a fraction of the density f is on the right and denotes a binary 1, and a fraction $1 - f$ on the left denotes the binary 0, we can find the optimal communication rate in terms of f. The results in Fig. 4 show that, asymptotically, no secret key is needed for sufficiently large $\bar{n}$, when we use an f value closer to 0. Interestingly, there is a trade-off between the optimal communication rate in terms of minimizing p_{err} and making the states as easy as

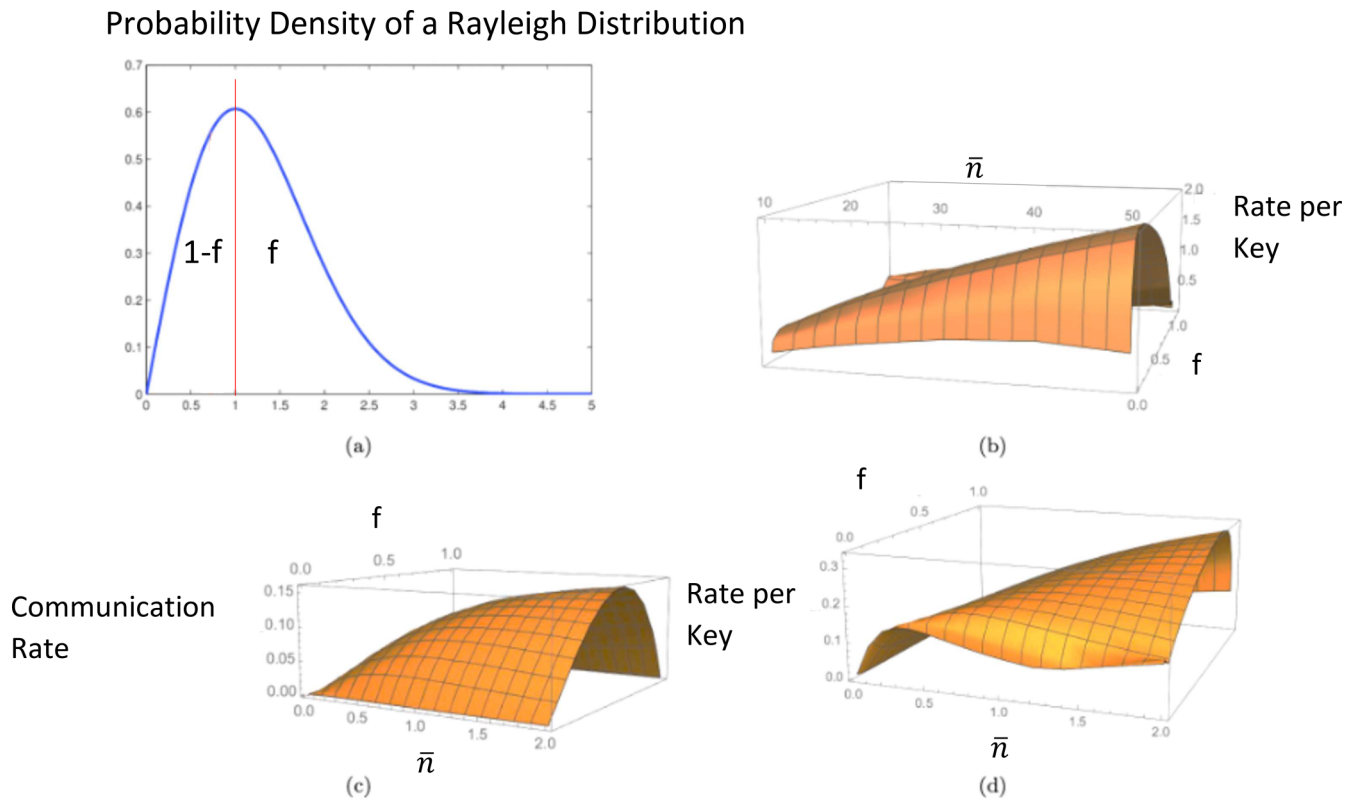

FIG. 4. The division of (a) the thermal state in terms of r (adapted from [26]) and the associated communication rates for various values of $\bar{n}$ and f. In descending order, (b) the plots display the optimized value of the communication rate per bit of secret key and (c) both the communication rate and (d) rate per bit of key when the quantity to be optimized is simply the communication rate.

possible to disguise. The key rate K required here is given by

$$\begin{aligned} K &= h(f) - h[(1-f)p(0|0) + fp(0|1)] \\ &\quad + fh[p(0|1)] + (1-f)h[p(0|0)] \\ &\geqslant 0, \end{aligned} \tag{32}$$

where the probabilities $p(b|a)$ refer to the probability that Bob measures a state he decides is from ρ_b given that Alice sent a state from ρ_a. We will derive expressions for these quantities below.

2. *Pairwise method*

We can also consider this approach from the perspective of using a finite set of states, what we called the "pairwise" protocol earlier. If we use a set of size M, with fM drawn from ρ_1 and $(1-f)M$ drawn from ρ_0, we should still maintain a fidelity scaling of $F \geqslant 1 - \frac{\bar{n}}{2M}$, since the overall statistics are still the same as before.

If we add factors of $1-f$ in front of $p(1|0)$ and f in front of $p(0|0)$ in the expression we previously derived in (30) for p_{err} under homodyne measurement, with these weightings we now have

$$m_c \approx \beta(r_0 + r_1) + \frac{\beta \ln\left(\frac{1}{f} - 1\right)}{2(r_0 - r_1)} \tag{33}$$

and

$$p_{\text{err}} = \frac{1}{4}\left\{ f\left[\text{erf}\left(\frac{m_c - 2\beta r_0}{\sqrt{2\beta^2 + r_0^2}} \right) + 1 \right] - (1-f)\left[\text{erf}\left(\frac{m_c - 2\beta r_1}{\sqrt{2\beta^2 + r_1^2}} \right) + 1 \right] \right\}. \tag{34}$$

We see that for $f = 1/2$ the value of m_c reverts to $\beta(r_0 + r_1)$.

For optimal generalized (Helstrom) measurements, we consider the quantity

$$M = (1-f)\,|\alpha_0\rangle\langle\alpha_0| - f\,|\alpha_1\rangle\langle\alpha_1|\,. \tag{35}$$

If we orthogonalize the basis using $\eta = \langle\alpha_0|\alpha_1\rangle \in \mathbb{R}$, we can express this as

$$\begin{aligned} M = (1 - f - f|\eta^2|)\,|\alpha_0\rangle\langle\alpha_0| - f(1-|\eta|^2)\,|\alpha_0^\perp\rangle\langle\alpha_0^\perp| \\ - f\eta\sqrt{1-|\eta|^2}\,|\alpha_0\rangle\langle\alpha_0^\perp| - f\eta\sqrt{1-|\eta|^2}\,|\alpha_0^\perp\rangle\langle\alpha_0|\,, \end{aligned} \tag{36}$$

with

$$|\alpha_1\rangle = \eta\,|\alpha_0\rangle + \sqrt{1-\eta^2}\,|\alpha_0^\perp\rangle\,. \tag{37}$$

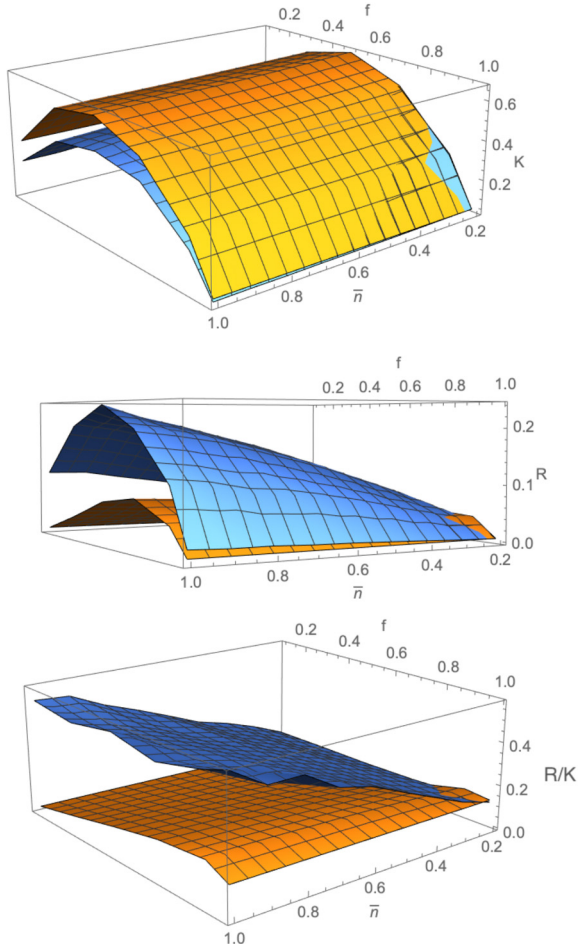

FIG. 5. The communication rate R, key rate K, and quotient $\frac{R}{K}$ for encoded transmission using homodyne-type (orange) and Helstrom-type (blue) measurements at a variety of f and $\bar{n}$ values.

Then we have that in this basis

$$p(0|0) = \mathrm{Tr}(|\alpha_0\rangle\langle\alpha_0|v_0\rangle\langle v_0|) = |\langle\alpha_0|v_0\rangle|^2,$$
$$p(0|1) = \mathrm{Tr}(|\alpha_1\rangle\langle\alpha_1|v_0\rangle\langle v_0|) = |\langle\alpha_1|v_0\rangle|^2, \quad (38)$$

where v_i is the eigenvector of M corresponding to the α_i eigenspace.

From these quantities we can roughly determine the rate of this communication method, which is limited by the entropy difference between the encoded and decoded information

$$R \approx h(q) - h(q|x_i), \quad (39)$$

and the key rate is set by the entropy difference between a simulated binary symmetric channel and the mutual information of the actual channel

$$K = h(f) - R, \quad (40)$$

where $q = (1-f)p(0|0) + fp(0|1) = \sum_i p_i p(0|x_i)$. Then, we can plot the associated quantities, such as K, R, and $\frac{R}{K}$, and we see in Fig. 5 that the ratio is fairly small for the homodyne measurement case but not for the Helstrom case.

We will also need additional shared secret key to specify which pair of states is being used for each transmitted bit. To specify a pair of states, one for each bit value, requires an additional $\ln[(1-f)(f)M^2]$ bits of key. The communication rate, however, remains the same as described above. Note that the asymptotic communication rate per bit of key can be demonstrated analytically to exceed 1 by examining the limiting behavior of the expression for $\frac{R}{K}$.

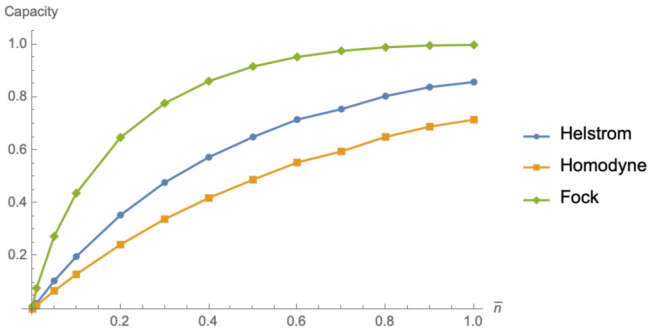

FIG. 6. A comparison of the communication rates using the Fock and coherent state encodings. In the Fock case this results from the theoretical optimum measurement based on the binary symmetric channel capacity and in the coherent state from optimal homodyne measurements and generalized measurements that approach the Helstrom bound. In the coherent state cases, the results are derived from sampling from the constituent distributions since an analytical result is not as straightforward to obtain.

Note that the value of $\bar{n}$ manifests itself in the specific coherent states that will be drawn from the distributions, rather than being directly visible in the binary transmitted bits as in the Fock case. If the message does not contain a roughly equal number of 1s and 0s, we can use an encoding, such as the one described in the Appendix, to compensate for that. However, in that case the message should first be compressed, which increases the entropy of the transmitted string; and if need be, a sublinear amount of secret key can be used to make the message string indistinguishable from a purely random bit string [27,28].

VII. DISCUSSION AND FUTURE WORK

As we can see in Fig. 6, the Fock state encoding is superior in the noiseless case and also does not require a key. Moreover, it is straightforward and the encoding has a clean visual representation based on Pascal's triangle (as shown in the Appendix). However, coherent states are more resistant to noise and are easier to generate in the laboratory, although they perform noticeably worse under either ideal (Helstrom limit) or homodyne measurements. In both cases, however, it is viable to communicate information covertly using the above-described methods for any channel parameter $\bar{n}$, and even in the coherent case the amount of secret key required is low compared to, for example, a one-time pad.

There are many possible future directions to explore to build on this work. An interesting problem is that of modeling noisy channels (for example, one with an existing thermal noise background) and the transmission of coherent states through such a channel (which is a well-understood problem) [21]. Such a study would provide a more thorough grounding for this work's study of coherent state methods, which are suboptimal in the noiseless case we examine.

Another important problem is that of transmitting *quantum* information, including entanglement. In this work we have only considered classical information transmission. It is unclear what kinds of encodings can be used for quantum information and how well they preserve the entanglement of the system, both in the noisy and the noiseless case. Such a work might also delve into the potential applications of this communication method to teleportation and superdense coding protocols, for example, and methods of making those more secure under the type of schemes covered in this paper.

A final promising area of study could be different key utilization protocols, with the goal of utilizing the inherent noise protection assumptions of steganography to efficiently scale the encryption process, and using other quantum communications as a vehicle for encoding steganographic information. This aims to get around this work's requirement that secret key is needed for nonorthogonal state discrimination by exploiting the information difference between Alice, Bob, and Eve to communicate a secret key seed, without compromising the communication rate derived above.

ACKNOWLEDGMENT

The authors would like to thank Jonathan Habif and Haley Weinstein for helpful discussions and insights. This research was supported in part by NSF Grants No. 1719778 and No. 1911089.

APPENDIX

1. Proof of Eq. (14)

We start from the fidelity between

$$\rho_c = \rho = \frac{1}{M}\sum_{j=1}^{M} e^{-r_j^2}\sum_n \frac{r_j^{2n}}{n!}\,|n\rangle\langle n| \tag{A1}$$

and the thermal state ρ_{th}:

$$\sqrt{F(\rho,\rho_t h)} = \text{Tr}\sqrt{\rho_{\text{th}}^{1/2}\rho\rho_{\text{th}}^{1/2}}. \tag{A2}$$

If we define the quantity $\Delta\rho = \rho - \rho_{\text{th}}$, then we know immediately that $E(\Delta\rho) = 0$ since $E(\rho) = \rho_{\text{th}}$. Then since $[\rho_{th}, \rho] = 0$, we can write

$$\sqrt{F} = \text{Tr}\big(\rho_{\text{th}}\sqrt{I + \rho_{\text{th}}^{-1}\Delta\rho}\big), \tag{A3}$$

which can be lower-bounded by taking the binomial expansion of the square root. We can then take the ensemble average

$$\begin{aligned} &E\big\{\text{Tr}\big[\rho_{\text{th}}\big(I + \tfrac{1}{2}\rho_{th}^{-1}\Delta\rho - \tfrac{1}{2}(\rho_{\text{th}}^{-1}\Delta\rho)^2\big)\big]\big\} \\ &\quad = 1 - \tfrac{1}{2}\text{Tr}\big[\rho_{\text{th}}^{-1}E(\Delta\rho^2)\big] \\ &\quad = \tfrac{3}{2} - \tfrac{1}{2}\text{Tr}\big[\rho_{\text{th}}^{-1}E(\rho^2)\big], \end{aligned} \tag{A4}$$

where here

$$\rho^2 = \left(\frac{1}{M}\sum_{j=1}^{M} e^{-r_j^2}\sum_n \frac{r_j^{2n}}{n!}\,|n\rangle\langle n|\right)^2 = \frac{1}{M^2}\sum_{jk}\rho_j\rho_k. \tag{A5}$$

We have that the ρ_j are independent and that for each one $E(\rho_j) = \rho_{\text{th}}$, so if we consider the cases only where all r_j are equal, which we will call ρ_1, we have

$$E(\rho^2) = \frac{M^2 - M}{M}\rho_{th}^2 + \frac{1}{M}E\big(\rho_1^2\big), \tag{A6}$$

which makes the result of Eq. (A4)

$$\sqrt{F} \geqslant 1 + \frac{1}{2M} - \frac{1}{2M}\text{Tr}\big[\rho_{\text{th}}^{-1}E\big(\rho_1^2\big)\big]. \tag{A7}$$

We can now finally evaluate this trace term

$$\begin{aligned} &\text{Tr}\big[\rho_{\text{th}}^{-1}E\big(\rho_1^2\big)\big] \\ &\quad = (\bar{n}+1)E\left[e^{-2r^2}\sum_{n=0}^{\infty}\left(\frac{\bar{n}+1}{\bar{n}}\right)^n\frac{r^{4n}}{(2n)!^2}\right] \\ &\quad = \sum_{n=0}^{\infty}\frac{2}{(n!)^2}\left(\frac{\bar{n}+1}{\bar{n}}\right)^{n+1}\int_0^{\infty} e^{-(2+1/\bar{n})r^2}r^{4n}dr \\ &\quad = \sum_{n=0}^{\infty}\frac{(2n)!}{(n!)^2}\frac{\bar{n}+1}{2\bar{n}+1}\left(\frac{\bar{n}+1}{\bar{n}(2+1/\bar{n})^2}\right)^n \\ &\quad = \frac{1}{2\pi}\frac{\bar{n}+1}{2\bar{n}+1}\sum_{n=0}^{\infty}\left(\frac{\bar{n}+1}{\bar{n}(2+1/\bar{n})^2}\right)^n\int_0^{2\pi}(4\cos^2\phi)^n d\phi \\ &\quad = \frac{1}{2\pi}\frac{\bar{n}+1}{2\bar{n}+1}\int_0^{2\pi}\sum_{n=0}^{\infty}\left(\frac{(\bar{n}+1)4\cos^2\phi}{\bar{n}(2+1/\bar{n})^2}\right)^n d\phi \\ &\quad = \bar{n}+1, \end{aligned} \tag{A8}$$

where we use that

$$\frac{(2n)!}{(n!)^2} = \frac{1}{2\pi}\int_0^{2\pi}(4\cos^2\phi)^n d\phi. \tag{A9}$$

This makes the final result of Eq. (A4)

$$\sqrt{F} \geqslant 1 - \frac{\bar{n}}{2M} \tag{A10}$$

in the average case.

2. Discretizing the Circle

The second bound is when the circle is discretized over θ. This encoding gives us that for $j = 1\ldots M$ and $k = 0\ldots L-1$ and $N = ML$

$$|\alpha_{jk}\rangle = |r_j e^{\frac{2\pi ik}{L}}\rangle = \sum_{jk}\sum_{n=0}^{\infty} e^{-r_j^2/2}\frac{(r_j e^{2\pi ik/L})^n}{\sqrt{n!}}\,|n\rangle\,. \tag{A11}$$

Then we have

$$\begin{aligned} \rho_c &= \frac{1}{N}\sum_{n,n'}\sum_{jk} e^{-r_j^2}\frac{r_j^{n+n'}}{\sqrt{n!n'!}}e^{2\pi ik(n-n')/L}\,|n\rangle\langle n'| \\ &= \frac{1}{N}\sum_{n,n'}\sum_{j} e^{-r_j^2}\frac{r_j^{n+n'}}{\sqrt{n!n'!}}\frac{1-e^{2\pi i(n-n')}}{1-e^{2\pi i(n-n')/L}}\,|n\rangle\langle n'| \\ &= \frac{1}{N}\sum_{n=0}^{\infty}\sum_{j} e^{-r_j^2}\frac{r_j^{2n}}{n!}\,|n\rangle\langle n| \end{aligned} \tag{A12}$$

since the geometric series sums to $\delta_{nn'}$. We can now evaluate the fidelity

$$\sqrt{F} = \mathrm{Tr}\left(\sqrt{\sum_{n=0}^{\infty} \frac{1}{N(\bar{n}+1)}\left(\frac{\bar{n}}{\bar{n}+1}\right)^n \frac{1}{n!}\sum_j e^{-r_j^2} r_j^{2n} |n\rangle\langle n|}\right). \tag{A13}$$

This gives the same bound as before:

$$\sqrt{F} \geqslant 1 - \frac{\bar{n}}{2N}. \tag{A14}$$

If we instead consider the distributions ρ_1 and ρ_0, we can easily show, at least, that $\sqrt{F} \to 1$ as $N \to \infty$ (since the integrals over the distributions are hard to compute) starting from Eq. (A4):

$$\begin{aligned}
\sqrt{F} &= 1 + \frac{1}{2N} - \frac{1}{2N}\mathrm{Tr}\left(\rho_{th}^{-1} E\left[\rho_i^2\right]\right) \\
E[\rho_i] &= \frac{2}{\bar{n}}\left(\int_0^{r_{1/2}} r^{4n+1} e^{-(2+1/\bar{n})r^2} dr \right. \\
&\quad \left. + \int_{-r_{1/2}}^{-\infty} r^{4n+1} e^{-(2+1/\bar{n})r^2} dr\right) \\
&\leqslant \frac{2}{\bar{n}} \int_0^{\infty} r^{4n+1} e^{-(2+1/\bar{n})r^2} dr \\
&\Rightarrow -\bar{n} - 1 \leqslant \mathrm{Tr}\left(\rho_{\mathrm{th}}^{-1} E[\rho_i]\right) \leqslant \bar{n} + 1 \\
&\Rightarrow \sqrt{F} \geqslant 1 - \frac{\bar{n}}{2N}.
\end{aligned} \tag{A15}$$

3. Derivation of Vertical Angle Bound with no Key

We start with the two states, given $c = 1 + 1/\bar{n}$:

$$\begin{aligned}
\rho_0 &= \frac{4}{\bar{n}} \int_0^{r_{1/2}} \sum_{n=0}^{\infty} e^{-(1+1/\bar{n})r^2} \frac{r^{2n+1}}{n!} |n\rangle\langle n| \\
&= \frac{2^{-\bar{n}}}{\bar{n}+1} \sum_{n=0}^{\infty} \frac{1}{c^n}\left(\sum_{k=0}^{n} \frac{(cr_{1/2}^2)^k}{k!}\right) |n\rangle\langle n|,
\end{aligned} \tag{A16}$$

and

$$\begin{aligned}
\rho_1 &= \frac{4}{\bar{n}} \int_{r_{1/2}}^{\infty} \sum_{n=0}^{\infty} e^{-(1+1/\bar{n})r^2} \frac{r^{2n+1}}{n!} |n\rangle\langle n| \\
&= \frac{2^{-\bar{n}}}{\bar{n}+1} \sum_{n=0}^{\infty} \frac{1}{c^n}\left(\sum_{k=n+1}^{\infty} \frac{(cr_{1/2}^2)^k}{k!}\right) |n\rangle\langle n|.
\end{aligned} \tag{A17}$$

The fidelity between these two states is given by

$$\sqrt{F} = \mathrm{Tr}\sqrt{\rho_0\rho_1} = \frac{2}{\bar{n}+1} \sum_{n=0}^{\infty} \left(\frac{\bar{n}}{\bar{n}+1}\right)^n \sqrt{Q_n(1-Q_n)}, \tag{A18}$$

where $Q_n = 2^{-(\bar{n}+1)} \sum_{k=0}^{n} \frac{(cr_{1/2}^2)^k}{k!}$ is the nth cumulant of the Poisson process with parameter $\lambda = cr_{1/2}^2$.

We can rewrite the trace distance as

$$\begin{aligned}
\frac{1}{2}||\rho_0 - \rho_1|| = \frac{1}{\bar{n}+1}\left(\sum_{n=0}^{N_{1/2}} \left(\frac{\bar{n}}{\bar{n}+1}\right)^n (1-2Q_n)\right. \\
\left. + \sum_{n=N_{1/2}}^{\infty} \left(\frac{\bar{n}}{\bar{n}+1}\right)^n (2Q_n - 1)\right),
\end{aligned} \tag{A19}$$

where $N_{1/2}$ is such that $Q_n < 1/2$ iff $n < N_{1/2}$.

We take each term in the sum in turn

$$\begin{aligned}
&\frac{1}{\bar{n}+1}\left(\sum_{n=0}^{N_{1/2}} \left(\frac{\bar{n}}{\bar{n}+1}\right)^n (1-2Q_n)\right) \\
&= \left(1 - \left(\frac{\bar{n}}{\bar{n}+1}\right)^{N_{1/2}+1}\right) \\
&\quad - \frac{2^{-\bar{n}}}{\bar{n}+1} \sum_{k=0}^{N_{1/2}} \frac{[(\bar{n}+1)\ln 2)]^k}{k!} \left(\frac{\left(\frac{\bar{n}}{\bar{n}+1}\right)^k - \left(\frac{\bar{n}}{\bar{n}+1}\right)^{N_{1/2}+1}}{1 - \frac{\bar{n}}{\bar{n}+1}}\right) \\
&= 1 + (2Q_{N_{1/2}} - 1)\left(\frac{\bar{n}}{\bar{n}+1}\right)^{N_{1/2}+1} - \tilde{Q}_{N_{1/2}},
\end{aligned} \tag{A20}$$

where $\tilde{Q}_n$ is the CDF for the Poisson process with $\lambda = \bar{n}\ln 2$.

The second sum is

$$\begin{aligned}
&\frac{1}{\bar{n}+1} \sum_{k=N_{1/2}+1}^{\infty} \left(\frac{\bar{n}}{\bar{n}+1}\right)^n (2Q_n - 1) \\
&= \sum_{n=N_{1/2}+1}^{\infty} \left(\frac{\bar{n}}{\bar{n}+1}\right)^n \sum_{k=0}^{n} \frac{[(\bar{n}+1)\ln 2]^k}{k!} - \left(\frac{\bar{n}}{\bar{n}+1}\right)^{N_{1/2}+1} \\
&= -\left(\frac{\bar{n}}{\bar{n}+1}\right)^{N_{1/2}+1} \\
&\quad + \frac{2^{-\bar{n}}}{\bar{n}+1}\left(\sum_{k=0}^{N_{1/2}} \frac{[(\bar{n}+1)\ln 2]^k}{k!} \sum_{n=N_{1/2}+1}^{\infty} \left(\frac{\bar{n}}{\bar{n}+1}\right)^n \right. \\
&\quad \left. + \sum_{k=N_{1/2}+1}^{\infty} \frac{[(\bar{n}+1)\ln 2]^k}{k!} \sum_{n=k}^{\infty} \left(\frac{\bar{n}}{\bar{n}+1}\right)^n\right) \\
&= 1 + (2Q_{N_{1/2}} - 1)\left(\frac{\bar{n}}{\bar{n}+1}\right)^{N_{1/2}+1} - \tilde{Q}_{N_{1/2}}.
\end{aligned} \tag{A21}$$

Combining these sums gives us the final result:

$$\frac{1}{2}||\rho_0 - \rho_1|| = 2\left(1 + (2Q_{N_{\frac{1}{2}}} - 1)\left(\frac{\bar{n}}{\bar{n}+1}\right)^{N_{\frac{1}{2}}+1} - \tilde{Q}_{N_{\frac{1}{2}}}\right). \tag{A22}$$

4. Encoding Method for Fock States

Elaborating on the results of Sec. V, we can imagine a particular encoding method for Fock state communication. Let's consider a particular example. Suppose $\bar{n} = 0.56$ and Alice wants to communicate the six-bit string 101001, which we can think of as the binary expression for the number 41. (Note that this protocol requires Bob to know the size of the message being communicated). If Alice wants to know

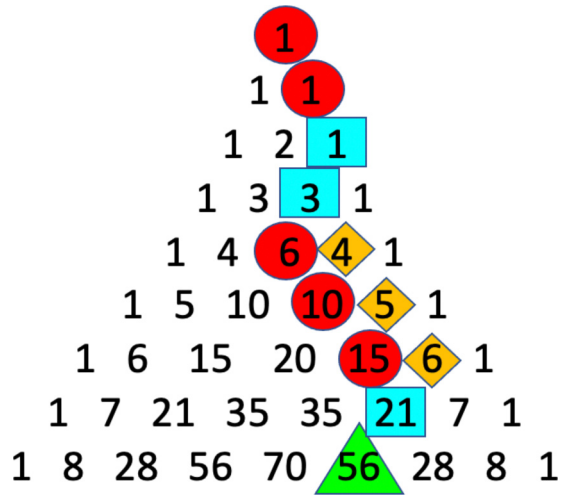

FIG. 7. This figure illustrates the process Alice uses to construct the encoded message as described above on Pascal's triangle. First, she calculates that the message is transmissible using eight bits since $41 < 56$ (green), as above. Then, she determines the message (41) is in the largest of 21 numbers, and so she boxes 21 and the first digit is 1 so she moves right. Then she determines 41 is neither in the largest 6, nor $6+5$, nor $6+5+4$ messages, and so circles the numbers 6, 10, and 15, and moves left each time, with each circle signifying a zero (the intermediate numbers 4, 5, and 6 are also highlighted since it is useful to keep track of their values). These steps are repeated until Alice reaches the right edge of the triangle at which point she knows all remaining digits are 0. As this figure shows, this is an efficient encoding as only a linear amount of combinations in the starting position need to be computed.

how many bits she needs to encode an x-bit number with $\frac{M}{\bar{n}+1}$ zeroes, she solves the equation

$$\binom{M}{\frac{M}{\bar{n}+1}} = 2^{x+1} - 1. \tag{A23}$$

Alice can transmit her string with $M = 8$ bits: the condition is that

$$\binom{M}{\frac{M}{\bar{n}+1}} = 2^{Mh(\frac{1}{\bar{n}+1})} \geqslant |m|, \tag{A24}$$

where we use $|m|$ to denote the value of the message, which is 41. Therefore, Alice's message should consist of 5 bits of 0, 3 bits of 1, and we can verify that $\binom{8}{3} = 56 > 41$.

As we will show, 41 is encoded by the binary string 10 001 100, which is the binary expression for 140, the 41st-smallest number with an 8 bit string representation (possibly with leading zeroes) with the appropriate ratio of 1s to 0s. To find this string, we follow a process derived from the "Christmas Stocking Theorem" mentioned above, which we detail in Fig. 7. First, consider that if we are using N bits, n_z of which are 0, there are $\binom{N-1}{n_z}$ strings with 1 in the first spot. There are $\binom{N-2}{n_z-1}$ strings with 0 in the first spot and 1 in the second spot, and so on. Alice can subtract these from 56 total possible strings to find the string in question. For example, the first $\binom{7}{5} = 21$ bits have a 1 at the start, and $41 > 56 - 21$ so Alice's message must start with 1, denoted by the blue square over the 21 in the figure. Then there are $\binom{6}{5} = 6$ strings with 11, so the second bit she sends must be 0, since $41 \leqslant 56 - 6$ (denoted by the red square over the 6 in the figure), $56 - 6 - 5 - 4 = 41$ so the string so far is 10 001. Actually, at this point we have reached the 41st largest string so Alice's message is the largest such string with that prefix, 10 001 100 = 140, and the remaining digits are in red in the figure since they correspond to the trailing 0s now that we have allocated all the 1s.

Once Alice knows that this is the encoded string she wants to send, she generateseight Fock states in a ratio of 5 : 3 zeroes to ones (where "one" here refers to a mode greater than 0, with the appropriate statistics to emulate the thermal state). Then she sends one of the $|1+\rangle$ states for every position corresponding to 1 in the string 10 001 100 and a $|0\rangle$ state for each 0. Bob measures the Fock states and should receive 1 001 100 as the most likely string, after inverting the above algorithm, which is straightforward. Eve should see something that looks like a thermal state: After all, it is bitwise random and has the required overall statistics.

5. Practical Fidelity Bounds for the Fock Encoding

In practice, the Fock state encoding is not exactly equivalent to a thermal state. While we would like to send the state

$$\rho = \frac{1}{\bar{n}+1} |\bar{0}\rangle \langle\bar{0}| + \frac{\bar{n}}{\bar{n}+1} |\bar{1}\rangle \langle\bar{1}|, \tag{A25}$$

in practice we have a finite number of bits, so we are sending either the state

$$\rho' = \frac{1}{N}\left(\left\lfloor \frac{N}{\bar{n}+1} \right\rfloor |\bar{0}\rangle \langle\bar{0}| + \left\lceil \frac{N\bar{n}}{\bar{n}+1} \right\rceil |\bar{1}\rangle \langle\bar{1}|\right), \tag{A26}$$

or the state

$$\rho'' = \frac{1}{N}\left(\left\lceil \frac{N}{\bar{n}+1} \right\rceil |\bar{0}\rangle \langle\bar{0}| + \left\lfloor \frac{N\bar{n}}{\bar{n}+1} \right\rfloor |\bar{1}\rangle \langle\bar{1}|\right) \tag{A27}$$

(whichever maximizes the fidelity). Another way of writing this is, e.g.,

$$\rho' = \frac{\lfloor \frac{N}{\bar{n}+1} t \rfloor}{N} |0\rangle \langle 0| + \frac{1}{N(\bar{n}+1)} \sum_{n=1}^{\infty} \left\lceil \frac{N\bar{n}}{\bar{n}+1} \right\rceil \times \left(\frac{\bar{n}}{\bar{n}+1}\right)^{n-1} |n\rangle \langle n|. \tag{A28}$$

We would like to compute, as a function of $\bar{n}$,

$$\begin{aligned}
&\max_{\sigma\in\{\rho',\rho''\}} F(\rho,\sigma) \\
&= \max[\mathrm{Tr}(\sqrt{\rho\sigma})] \\
&= \max\left(\mathrm{Tr}\sqrt{\sum_n \frac{1}{\bar{n}+1}\left(\frac{\bar{n}}{\bar{n}+1}\right)^n c_n |n\rangle\langle n|}\right) \\
&= \text{e.g.,} \sqrt{\frac{\lfloor \frac{N}{\bar{n}+1} \rfloor}{N(\bar{n}+1)}} + \sum_{n=1}^{\infty} \frac{1}{\bar{n}+1} \sqrt{\frac{\lceil \frac{N\bar{n}}{\bar{n}+1} \rceil \bar{n}}{N(\bar{n}+1)}} \left(\frac{\bar{n}}{\bar{n}+1}\right)^n \\
&\geqslant 1 - \frac{1}{8}\left(\frac{\left(1 - \frac{\lfloor \frac{N}{\bar{n}+1} \rfloor (\bar{n}+1)}{N}\right)^2}{\bar{n}+1} + \frac{\bar{n}\left(1 - \frac{\lceil \frac{N\bar{n}}{\bar{n}+1} \rceil (\frac{\bar{n}+1}{\bar{n}})}{N}\right)^2}{\bar{n}+1}\right) \\
&\quad - \frac{(\frac{\bar{n}+1}{\bar{n}})^2}{16N^3}
\end{aligned}$$

$$
\begin{aligned}
&= 1 - \frac{1}{8}\left(\frac{\left\lceil \frac{N\bar{n}}{\bar{n}+1}\right\rceil^2 \frac{\bar{n}+1}{\bar{n}} + \left\lfloor \frac{N}{\bar{n}+1}\right\rfloor^2 (\bar{n}+1)}{N^2} - 1\right) - \frac{\left(\frac{\bar{n}+1}{\bar{n}}\right)^2}{16N^3} \\
&\geqslant 1 - \frac{1}{8}\left(\frac{\left(\frac{N\bar{n}}{\bar{n}+1}+1\right)^2\left(\frac{\bar{n}+1}{\bar{n}}\right) + \left(\frac{N}{\bar{n}+1}\right)^2(\bar{n}+1)}{N^2} - 1\right) \\
&\quad - \frac{\left(\frac{\bar{n}+1}{\bar{n}}\right)^2}{16N^3} \\
&= 1 - \frac{1}{8}\left(\frac{2N+1}{N^2} + \frac{1}{N^2\bar{n}}\right) - \frac{1}{16}\frac{\left(\frac{\bar{n}+1}{\bar{n}}\right)^2}{N^3}, \quad \bar{n} \geqslant 1, \\
&\geqslant 1 - \frac{1}{8}\left(\frac{2N+\bar{n}+1}{N^2}\right) - \frac{1}{16}\frac{(\bar{n}+1)^2}{N^3}, \quad \bar{n} < 1, \quad \text{(A29)}
\end{aligned}
$$

where the third-order correction is due to Taylor's theorem. The above chain of inequalities shows that when N is fairly large, the fidelity between the encoded state and the thermal state is very close to 1. At the cost of sharing some additional secret key, Alice and Bob can actually make the fidelity (almost) perfect. The gap in fidelity between the encoded state and the thermal state results from using integer approximations to $N/(\bar{n}+1)$ and $N\bar{n}/(\bar{n}+1)$. However, Alice and Bob can, in principle, randomly choose the numbers of 0s and 1s to be used in the encoding, from the binomial distribution with probabilities corresponding to the thermal state: $p(0) = 1/(\bar{n}+1)$ and $p(1) = \bar{n}/(\bar{n}+1)$. By averaging over this choice, one can match the thermal state perfectly.

[1] Wobst and Reinhard, *Cryptology Unlocked* (Wiley, New York, 2001), p. 19.

[2] G. J. Simmons, Vigenère cipher, encyclopedia britannica, 25 Feb. 2023, https://www.britannica.com/topic/Vigenere-cipher. Accessed 28 February 2023.

[3] A. A. Bruen and M. A. Forcinito, *Cryptography, Information Theory, and Error-Correction: A Handbook for the 21st Century* (John Wiley & Sons, New York, 2011), p. 21.

[4] "Origin of Cryptography." Origin of Cryptography, https://www.tutorialspoint.com/cryptography/origin_of_cryptography.htm.

[5] S. T. Abdulrazzaq, M. M. Siddeq, and M. A. Rodrigues, A novel steganography approach for audio files, SN Comput. Sci. **1**, 97 (2020).

[6] M. Dušek, N. Lütkenhaus, and M. Hendrych, Quantum cryptography, Prog. Opt. **49**, 381 (2006).

[7] C. Elliott, D. Pearson, and G. Troxel, Quantum cryptography in practice, arXiv:quant-ph/0307049.

[8] S. Pirandola, U. L. Andersen, L. Banchi, M. Berta, D. Bunandar, R. Colbeck, D. Englund, T. Gehring, C. Lupo, C. Ottaviani, J. L. Pereira, M. Razavi, J. Shamsul Shaari, M. Tomamichel, V. C. Usenko, G. Vallone, P. Villoresi, and P. Wallden, Advances in quantum cryptography, Adv. Opt. Photon. **12**, 1012 (2020).

[9] B. Bash, D. Goeckel, and D. Towsley, Square root law for communication with low probability of detection on AWGN channels, in *Proceedings of the 2012 IEEE International Symposium On Information Theory* (IEEE, New York, 2012), pp. 448–452.

[10] B. A. Bash, D. Goeckel, and D. Towsley, Covert communication gains from adversary's ignorance of transmission time, arXiv:1403.1013.

[11] J. Gea-Banacloche, J. Math. Phys. **43**, 4531 (2002).

[12] M. Curty and D. J. Santos, *Paper Presented at the 2nd Bielefeld Workshop on Quantum Information and Complexity*, Bielefeld, Germany, 10 October (2004).

[13] M. Tahmasbi and M. R. Bloch, Steganography protocols for quantum channels, J. Math. Phys. **61**, 082201 (2020).

[14] B. A. Shaw and T. A. Brun, Hiding quantum information in the perfect code, arXiv:1007.0793.

[15] B. A. Shaw and T. A. Brun, Quantum steganography with noisy quantum channels, Phys. Rev. A **83**, 022310 (2011).

[16] B. Wu, Z. Wang, Y. Tian, M. P. Fok, B. J. Shastri, D. R. Kanoff, and P. R. Prucnal, Optical steganography based on amplified spontaneous emission noise, Opt. Express **21**, 2065 (2013).

[17] A. Riaz and M. Ali, Chaotic communications, their applications and advantages over traditional methods of communication, in *Proceedings of the 2008 6th International Symposium on Communication Systems, Networks and Digital Signal Processing* (IEEE, New York, 2008).

[18] An introduction to spread spectrum techniques. https://www.ausairpower.net/OSR-0597.html#::text=Security%20from%20Eavesdropping.almost%20perfect%20immunity%20to%20interception. Accessed 19 Aug. 2023.

[19] H. Shi, Z. Zhang, and Q. Zhuang, Practical route to entanglement-assisted communication over noisy bosonic channels, Phys. Rev. Appl. **13**, 034029 (2020).

[20] C. A. Fuchs and J. van de Graaf, Cryptographic distinguishability measures for quantum mechanical states, IEEE Trans. Inf. Theory **45**, 1216 (1998).

[21] A. Vourdas, Optical signals with thermal noise, Phys. Rev. A **39**, 206 (1989).

[22] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).

[23] J. H. Shapiro, *6.453 Quantum Optical Communication*. Fall 2016. Massachusetts Institute of Technology: MIT OpenCourseWare, https://ocw.mit.edu/. License: Creative Commons BY-NC-SA.

[24] E. W. Weisstein, Christmas stocking theorem, from *MathWorld*–a wolfram web resource. https://mathworld.wolfram.com/ChristmasStockingTheorem.html.

[25] M. Boesgaard, M. Vesterager, T. Pedersen, J. Christiansen, and O. Scavenius, Rabbit: A new high-performance stream cipher, in *Fast Software Encryption*, edited by T. Johansson, FSE 2003, Lecture Notes in Computer Science (Springer, Berlin, 2003), Vol. 2887.

[26] A. Arsal, A study on wireless channel models: Simulation of fading, shadowing and further applications, 10.13140/RG.2.2.13520.66562.

[27] C. Sutherland and T. Brun, Quantum steganography over noisy channels: Achievability and bounds, Phys. Rev. A **100**, 052312 (2019)

[28] C. Sutherland and T. Brun, Quantum steganography over noiseless channels: Achievability and bounds, Phys. Rev. A **101**, 052319 (2020)