Quantum steganography over noisy channels: Achievability and bounds

Chris Sutherland and Todd A. Brun

Phys. Rev. A **100**, 052312 — Published 11 November 2019

DOI: 10.1103/PhysRevA.100.052312

Quantum Steganography over Noisy Channels: Achievability and Bounds

Chris Sutherland[1] and Todd A. Brun[1,2]
[1]*Department of Physics,* [2]*Ming Hsieh Department of Electrical Engineering,*
University of Southern California, Los Angeles, California

(Dated: October 15, 2019)

Characterizing secret communication over noisy quantum channels is an interesting problem from both a practical and theoretical perspective. Suppose Alice and Bob wish to communicate secret information so that an eavesdropper Eve will not suspect any type of encoded communication between the two. Classical or quantum cryptography will not suffice since it is always clear secret communication is taking place. Therefore Alice and Bob must execute what is known as a quantum steganographic protocol. Assuming Eve only has partial knowledge of the channel connecting Alice and Bob, we show that for the bit-flip and depolarizing channels Alice can use Eve's lack of knowledge of the channel parameter to encode quantum information steganographically. We give an explicit encoding procedure and calculate the rate at which Alice and Bob can communicate secretly. We also show that our encoding is optimal for nondegenerate quantum codes. We calculate the rate at which secret key must be consumed. Finally, we discuss the possibility of steganographic communication over more general quantum channels, and conjecture a general formula for the steganographic rate.

I. INTRODUCTION

Suppose Alice and Bob are the respective leaders of two countries and they wish to communicate highly classified information with each other over a public channel. They do not want other countries to know they are communicating secret information, perhaps because they have a history of shady international political relations. Simple cryptography would not be good enough here beacause it would alert a potential eavesdropper (Eve) that secret communication is taking place, even if she cannot read it. Therefore, if Alice and Bob wish to keep their conversations secret, they must employ a steganographic protocol.

Both cryptography and steganography are interesting and well-developed subjects, the studies of which date back millenia [1, 2]. In cryptography, a secret message is encrypted using a shared secret key between Alice and Bob, and Alice sends the resulting *ciphertext* across a channel to be decoded. Should Eve observe this ciphertext, she would not be able to decode it without the secret key. However, she would undoubtedly become suspicious if she weren't already, due to Alice sending encrypted messages to Bob.

Steganography solves this problem of secrecy. Although cryptography allows for *secure* communication, in this paper we are interested in *secret* communication. In steganography, a secret message is hidden into a larger *covertext*, which appears to Eve as an innocuous message. This seeming innocuousness of the message is what makes the protocol secret. The hidden message may also be encrypted itself to make the protocol not only secret but secure, so that even if Eve were tipped off to there being secret communication between Alice and Bob, she would not be able to decode the hidden message. For example, digital audio, video, and pictures are increasingly furnished with distinguishing but imperceptible marks, which may contain a hidden copyright notice or encrypted serial number [3].

Ever since Shor's remarkable discovery that a quantum computer could solve the prime factorization problem efficiently, hence cracking one of the internet's most common encryption schemes [4], interest in quantum cryptography has been intense. Quantum steganography is of more recent development [5–7]. The protocol we will be considering is to encode quantum information steganographically as error syndromes of a quantum error-correcting code. This was detailed extensively by Shaw and Brun [8, 9], where it was shown that such schemes can hide both quantum and classical information, with a quantitative measure of secrecy, even in the presence of a noisy physical channel.

A more precise analysis of this quantum steganographic protocol over noiseless channels was done by the present authors in terms of achievability and converse proofs [10]. In this work, we treated the case where Eve believes the channel connecting Alice and Bob to be some noisy quantum channel, but the actual physical channel is noiseless, and we gave optimal rates of steganographic communication.

A related field is known as covert quantum communication [11–15]. In covert communications, it is often assumed that the channel between Alice and Bob is a noisy optical channel, which is modelled by a beamsplitter with some transmissivity parameter that characterizes how many photons are lost to Eve. Covert quantum communication can be seen as a special case of quantum steganography over noisy quantum channels in the case where the eavesdropper has exact knowledge of the channel, and where Eve assumes the channel is idle (so only noise is being transmitted). Similarly, quantum steganography is a type of covert quantum communication where Eve knows about the covertext communication but not the hidden stegotext, and where Eve may not have perfect knowledge of the channel. In covert quantum communication, it has been shown that in general one can secretly communicate an amount of classical information which scales like the square root of the number of channel uses.

The goal of this paper is to extend the previous work on quantum steganography over noiseless channels in [10] to the scenario where the channel Alice and Bob share is noisy. We assume that Eve believes the channel to be noisier than it is, which allows Alice and Bob to communicate at a linear rate in the number of channel uses. This assumption is not unreasonable, especially when Alice and Bob have been systemetically deceiving Eve by adding extra noise. We also assume that Alice and Bob are using an error-correcting code powerful enough to correct errors induced by the channel Eve believes to be connecting them, or else she would become suspicious. Eve would also become suspicious if the pattern of errors Alice uses to encode her secret information does not match the typical errors induced by the channel that Eve expects.

In Section II we formalize our notion of quantum steganography where secret messages are hidden in the syndromes of an error-correcting code. We outline a specific steganographic encoding where Alice is able to emulate a bit flip or depolarizing channel $\mathcal{N}_{p+\delta p}$ (the channel Eve believes to be connecting them) on her encoded secret message and covertext, where the actual physical channel is $\mathcal{N}_p$. We also calculate the amount of key consumed in our protocol. In Section III we prove upper bounds on the amount of steganographic communication possible over these channels, and show that these bounds are asymptotically equal to the rates achieved in the previous section. Finally, in Section IV we summarize our results, and discuss quantum steganography for general quantum channels, conjecturing a capacity formula for general quantum steganographic communication.

II. ACHIEVABILITY

A. Bit Flip Channel

Suppose that Alice wishes to communicate steganographically to Bob by secretly sending him a message m drawn from a set of possible messages $\mathcal{M}$, assumed to all be equally likely. Alice and Bob can communicate via a quantum channel, but the eavesdropper Eve can monitor their communications over this channel if she chooses. They cannot communicate classically without Eve intercepting their communications, but before the protocol began they exchanged a secret key, in the form of an arbitrarily long string of random bits, unknown to Eve. We assume that Eve believes the quantum channel shared between Alice and Bob to be a bit flip channel with error rate $p+\delta p$, i.e.,

$$\mathcal{N}^{BF}_{p+\delta p}(\rho) = (1-(p+\delta p))\rho + (p+\delta p)X\rho X. \qquad (1)$$

The δp parameter represents Eve's ignorance of the actual physical channel connecting Alice and Bob. If Eve has based her belief about the channel on her observation of its error rates, which Alice and Bob have been systematically increasing, her posterior distribution would likely be close to a Gaussian. If the width of this Gaussian is fairly narrow, however, then in practice it seems unlikely to give results much different from those assuming a fixed value of δp, while adding an extra layer of complexity to the analysis, so we won't consider that here.

Recall, the actual physical channel between Alice and Bob is $\mathcal{N}^{BF}_p$. First, Alice encodes an innocent state, i.e. the covertext ρ_c, into a nondegenerate quantum error-correcting code (QECC) on N qubits. This code should be able to correct typical errors induced by the channel $(\mathcal{N}^{BF}_{p+\delta p})^{\otimes N}$. Next, depending on the secret key $k \in \mathcal{K}$ and the message $m \in \mathcal{M}$ that she would like to send, she applies the error

$$E^N(k,m) = E_1(k,m) \otimes ... \otimes E_N(k,m) \qquad (2)$$

to her state. This produces the codeword corresponding to the message m. If her message to Bob is a quantum state, she can prepare the system in a superposition of these codewords. This encoding can be done by a suitable quantum circuit. These codewords are generated by applying errors drawn randomly from the channel $(\mathcal{N}^{BF}_q)^{\otimes N}$, using the shared secret key k as the source of randomness. That is, the errors X or I on each qubit are drawn from the product distribution $p_{E^N}(e^N)$, where $p_E(e)$ is given by

$$\begin{aligned} p_E(X) &= q, \\ p_E(I) &= 1-q, \end{aligned} \qquad (3)$$

where $q = \delta p/(1-2p)$. Since the set of errors is selected using the shared secret key k, Bob knows which codeword corresponds to each message m.

The errors given by Eq. (2) are typical errors associated with the channel $(\mathcal{N}^{BF}_q)^{\otimes N}$ [16, 17]. By the asymptotic equipartition theorem [18], for large enough N, it is highly likely that each of these codewords that Alice generates is a typical sequence with a sample entropy close to $H(E) = -(1-q)\log(1-q) - q\log q = h(q)$. Furthermore, it follows from a simple calculation that $\mathcal{N}_p \circ \mathcal{N}_q = \mathcal{N}_{p+\delta p}$ if we set $q = \delta p/(1-2p)$. This will become important later when we discuss the secrecy of this protocol.

Alice then sends her state through the channel $(\mathcal{N}^{BF}_p)^{\otimes N}$. We are now essentially in the scenario of classical random coding over a classical bit-flip channel with parameter p. By the asymptotic equipartition theorem for conditionally typical sequences [18], for each input sequence (i.e., error $E^N(k,m)$ applied to the encoded covertext) there is a corresponding conditionally typical set of errors $\{F^N(k,m)\}$ which has the following properties: its total probability is close to 1, its size is $\approx 2^{nH(F|E)}$, and the probability of each conditionally typical error given knowledge of the input error $E^N(k,m)$ is $\approx 2^{-nH(F|E)}$.

With high probability, the error F^N Bob observes will be a typical error of the channel $(\mathcal{N}^{BF}_{p+\delta p})^{\otimes N}$. We know from Shannon's noisy channel coding theorem that if Alice and Bob set the number of messages $|\mathcal{M}| = 2^{NR}$ such that

$$2^{NR} \approx \frac{2^{NH(F)}}{2^{NH(F|E)}} = 2^{N(H(F)-H(F|E))}, \tag{4}$$

then Bob is able to decode correctly with high probability [18–20] which error $E^N(k,m)$ was applied by Alice, as long as the code is nondegenerate. For our protocol, it is straightforward to calculate that $H(F) = h(p+q-2pq) = h(p+\delta p)$ for $q = \delta p/(1-2p)$, and $H(F|E) = h(p)$. Hence Alice can communicate

$$M = \log|\mathcal{M}| \approx N(h(p+\delta p) - h(p)) \tag{5}$$

bits of information to Bob steganographically.

Moreover, this protocol does not arouse suspicion from Eve. We say that this protocol is *secret*, because the state passing through the channel is to good approximation the state Eve would expect to see. To see this, note that

$$\begin{aligned}
&\sum_{k\in\mathcal{K}}\sum_{m\in\mathcal{M}} p_k(\mathcal{N}_p^{BF})^{\otimes N}(E^N(k,m)V\rho_c V^\dagger E^N(k,m)) \\
&= (\mathcal{N}_p^{BF})^{\otimes N}(\sum_{k\in\mathcal{K}}\sum_{m\in\mathcal{M}} p_k E^N(k,m)V\rho_c V^\dagger E^N(k,m)) \\
&\approx (\mathcal{N}_p^{BF}\circ\mathcal{N}_{\delta p/(1-2p)}^{BF})^{\otimes N}(V\rho_c V^\dagger) \\
&= (\mathcal{N}_{p+\delta p}^{BF})^{\otimes N}(V\rho_c V^\dagger),
\end{aligned} \tag{6}$$

where V is the isometry corresponding to the QECC Alice and Bob are using. The first equality follows from linearity of quantum operations. The approximate equality follows from the fact that when we average the transmitted codeword over the key and all possible messages, we are applying all the typical errors of the channel $(\mathcal{N}_q^{BF})^{\otimes N}$ with their correct probabilities, and hence to good approximation [16, 17] we are simply applying the full channel. The final equality follows from calculating the composition of these quantum operations. This is exactly the state Eve expects to observe, hence our steganographic protocol is secret to an arbitrarily good approximation.

As described above, this protocol allows Alice to transmit a classical message m secretly to Bob. But in fact, by making this protocol coherent Alice can equally well transmit a quantum state—that is, a superposition of possible messages m. So we see that this protocol can transmit either classical or quantum information at the same rate $h(p+\delta p) - h(p)$. The one significant difference between these two cases is that if Eve actually carries out a measurement of the error on the transmitted state, this would destroy the superpositions of a quantum message, but not affect the ability to transmit classical messages. So this protocol works for secret quantum communication if it is assumed that Eve only sometimes checks the code blocks transmitted from Alice to Bob. As we did in the case of steganographic communication over a noiseless channel [10], we can show this by considering a protocol in which Alice sends a subsystem M to Bob which is maximally entangled with a reference subsystem R (see Figure 1).

B. Depolarizing Channel

Suppose that Eve believes the quantum channel shared between Alice and Bob is a depolarizing channel with error rate $p + \delta p$, i.e.,

$$\begin{aligned}
\mathcal{N}_{p+\delta p}^{DC}(\rho) &= (1-(p+\delta p))\rho \\
&\quad + \frac{p+\delta p}{3}(X\rho X + Y\rho Y + Z\rho Z),
\end{aligned} \tag{7}$$

where the actual physical channel between Alice and Bob is $\mathcal{N}_p^{DC}$. The protocol for Alice and Bob to communicate steganographically in this scenario is nearly identical to the protocol described in the previous subsection. First, Alice encodes an innocent state, i.e,. the covertext ρ_c, into a nondegenerate quantum error-correcting code (QECC) on N qubits. This code should be able to correct typical errors induced by the channel $(\mathcal{N}_{p+\delta p}^{DC})^{\otimes N}$. Next, depending on the secret key $k \in \mathcal{K}$ and the message $m \in \mathcal{M}$ that she would like to send, she applies the error

$$G^N(k,m) = G_1(k,m) \otimes ... \otimes G_N(k,m) \tag{8}$$

to her state. This produces the codeword corresponding to the message m. If her message to Bob is a quantum state, she can prepare the system in a superposition of these codewords. These codewords are generated by applying errors drawn randomly from the channel $(\mathcal{N}_q^{DC})^{\otimes N}$, using the shared secret key k as the source of randomness. That is, the errors X, Y, Z or I on each qubit are drawn from the product distribution $p_{G^N}(g^N)$, where $p_G(g)$ is given by

$$\begin{aligned}
&p_G(X) = p_G(Y) = p_G(Z) = q/3, \\
&p_G(I) = 1-q,
\end{aligned} \tag{9}$$

and $q = \delta p/(1-4p/3)$. Since the errors are selected using the shared secret key k, Bob knows what codeword corresponds to each message m.

The errors given by Eq. (8) are typical errors associated with the channel $(\mathcal{N}_q^{DC})^{\otimes N}$ [16, 17]. By the asymptotic equipartition theorem [18], for large enough N, it is highly likely that each of these codewords that Alice generates is a typical sequence with a sample entropy close to $H(G) = -(1-q)\log(1-q) - q\log(q/3) \equiv s(q)$, where $s(q)$ is the classical Shannon entropy of the depolarizing channel on one qubit with error parameter q in the Pauli representation. Furthermore, it follows from a simple calculation that $\mathcal{N}_p \circ \mathcal{N}_q = \mathcal{N}_{p+\delta p}$ if we set $q = \delta p/(1-4p/3)$, which is important for secrecy as discussed in the previous subsection.

Alice then sends her state through the channel $(\mathcal{N}_p^{DC})^{\otimes N}$. Following the same random coding argument described for the bit-flip channel, with high probability, the error J^N Bob observes will be a typical error of the channel $(\mathcal{N}^{DC}_{p+\delta p})^{\otimes N}$. We know from Shannon's noisy channel coding theorem that if Alice and Bob set the number of messages $|\mathcal{M}| = 2^{NR}$ such that

$$2^{NR} \approx \frac{2^{NH(J)}}{2^{NH(J|G)}} = 2^{N(H(J)-H(J|G))}, \tag{10}$$

then Bob is able to decode correctly with high probability [18–20] which error $G^N(k,m)$ was applied by Alice, as long as the code is nondegenerate. For our protocol, it is straightforward to calculate that $H(J) = s(p+q-4qp/3) = s(p+\delta p)$ for $q = \delta p/(1-4p/3)$, and $H(J|G) = s(p)$. Hence Alice can communicate

$$M = \log|\mathcal{M}| \approx N(s(p+\delta p) - s(p)) \tag{11}$$

classical or quantum bits of information to Bob steganographically. The proof of secrecy of this protocol is nearly identical to the one given in Eq. (6).

Note that the assumption of a nondegenerate code is quite natural in the case of the bit-flip channel, which is essentially classical; but not as much so for the depolarizing channel, where the errors do not commute. We believe that this general procedure for encoding will work for degenerate codes as well, but the achievable rate may be lower, and will require an analysis specific to the code in question. We will return to this point at the end of the paper, where we conjecture a general formula for the steganographic rate of a quantum channel using general quantum codes.

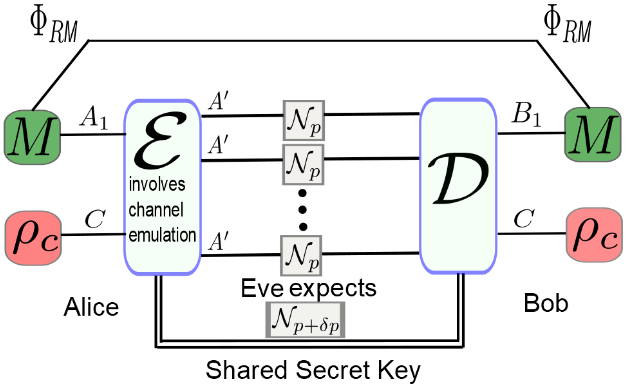

FIG. 1. The information processing task we consider, of Alice sending M stego qubits to Bob over a quantum channel $\mathcal{N}_p$ (which is either the quantum bit flip channel or depolarizing channel), which Eve believes to be noisier. Eve's ignorance of the channel is characterized by the parameter δp. Dependent on the secret key k, Alice encodes her message subsystem M and an innocent covertext ρ_c into a suitable quantum error-correcting code in such a way that once passed through the physical channel, it looks as though typical errors of the channel $\mathcal{N}_{p+\delta p}$ have been applied. Bob then decodes the message and covertext using his copy of the shared secret key k. Alice's message is entangled with a reference system R. The ability to transmit entanglement can be used to bound the ability to do general quantum communication.

C. Secret key consumption

Here we analyze how much secret key is used by the encodings outlined above. The details of the encoding—that is, how each message m is mapped to a codeword for a particular key element k—we assume have been decided between Alice and Bob ahead of time. Therefore secret key is required to pick the subsets of errors used in the encoding, but it is not needed otherwise.

Before the protocol begins, Alice and Bob divide the set of typical errors of the channel $(\mathcal{N}^{BF}_{\delta p/(1-2p)})^{\otimes N}$ into n nonoverlapping subsets of size $|\mathcal{M}| = 2^{N(h(p+\delta p)-h(p))}$ each, where

$$n = \frac{2^{Nh(\delta p/(1-2p))}}{2^{N(h(p+\delta p)-h(p))}}. \tag{12}$$

For each transmitted block, Alice and Bob must randomly choose one of these n subsets to encode her messages. This requires a number of bits K of secret key,

$$K = \log_2 n = N(h(\delta p/(1-2p)) - h(p+\delta p) + h(p)), \tag{13}$$

which is positive for $p + \delta p < 0.5$. Therefore the key consumption scales linearly with N. Notice in the limit where the physical channel is noiseless i.e., $p = 0$, we have that $K = 0$, which agrees with our result in [10] where it was shown that only a sublinear amount of key is needed for encoding across noiseless channels.

As discussed in [10], using this amount K of shared secret is key is sufficient to make the steganographic protocol secret, but not necessarily secure. That is, Eve should not become suspicious if she observes the state passing through the channel. However, if for some reason she knew a message was being sent, she would be able to deduce significant information about the message.

This can be prevented by first encrypting the message before doing the steganographic encoding. If we wish to make the protocol both secret and secure, encryption would require M bits of secret key in the case of an M-bit classical message (using a one-time pad), or $2M$ bits of secret key in the case of an M-qubit quantum message (by twirling). Thus the rate of key consumption would be increased by R (classical) or $2R$ (quantum, where R is the steganographic rate).

III. SECRECY, RELIABILITY, AND BOUNDS

A. The information processing task

Now we wish to put a bound on the amount of information that can be sent with the steganographic scenario outlined above. Recall that Alice is using Eve's ignorance of the actual noise rate of the physical channel to hide her message. We will consider the quantum information processing task known as *entanglement transmission* (visualized in Figure 1) in this section. The maximum rate of entanglement transmission is clearly an upper bound on the maximum rate of quantum communication and so we will use this to derive a bound on the steganographic transmission rate. Alice prepares a secret message of $M = \log_2 |A_1|$ qubits along with an innocent covertext ρ_c. Her secret message qubits are maximally entangled with a reference system R. Her covertext will be encoded into the N-qubit quantum error-correcting code. Therefore her encoded state, which is dependent on the secret key element k, can be written as:

$$\omega_{k,A'^nR} \equiv \mathcal{E}_{k,A_1C\to A'^n}(\rho_c \otimes \Phi_{A_1R}). \tag{14}$$

The dependence of the encoding on the secret key corresponds to choosing among the different sets of typical errors of the channel $\mathcal{N}_q$ in the protocols from the previous section. To someone (like Eve) who does not know the secret key k, the state is effectively

$$\sum_k p_k \mathcal{N}_p^{\otimes N}(\omega_{k,A'^nR}) = \mathcal{N}_p^{\otimes N}(\omega_{A'^nR}), \tag{15}$$

where we have used linearity of quantum operations and $\omega_{A'^nR} \equiv \sum_k p_k \omega_{k,A'^nR}$ is the state averaged over all possible values of the secret key k with probabilities p_k. (We can choose this probability to be uniform for simplicity, if we so desire.)

What is a good way to guarentee secrecy from Eve? We propose the following *secrecy* condition:

$$\frac{1}{2}||\mathcal{N}_p^{\otimes N}(\mathrm{Tr}_R[\omega_{A'^nR}]) - \mathcal{N}_{p+\delta p}^{\otimes N}(V\rho_c V^\dagger)||_1 \le \delta, \tag{16}$$

where $\mathcal{N}_{p+\delta p}$ is what Eve believes the physical channel to be, V is an isometry representing the encoding of the covertext into a suitably chosen codeword (one which can correct typical errors induced by the channel $\mathcal{N}_{p+\delta p}$) and $\delta > 0$ is some small parameter. This condition means that if Eve observes the quantum state, it will be effectively indistinguishable from an encoded covertext being sent through the noisy quantum channel $\mathcal{N}_{p+\delta p}$.

It is also important to discuss the requirement of *recoverability.* When Bob receives the state, he applies his decoder $\mathcal{D}_{k,A'^n\to B_1C}$ to obtain the original state $\rho_c \otimes \Phi_{B_1R}$. The recoverability condition can be written as follows:

$$\frac{1}{2}||\mathcal{D}_{k,A'^n\to B_1C}(\mathcal{N}_p^{\otimes N} \otimes I_R(\omega_{k,A'^nR})) - \rho_c \otimes \Phi_{B_1R}||_1 \le \epsilon \tag{17}$$

for all k, where $\epsilon > 0$ is a small parameter.

B. Upper bound on steganographic rate

We are now in a position to put a bound on the number of qubits M that can be sent reliably and steganographically from Alice to Bob. First we define $\sigma_E = \mathcal{N}_{p+\delta p}^{\otimes N}(V\rho_c V^\dagger)$ and apply the Fannes-Audeneart inequality [21] to the secrecy condition in Eq. (16):

$$H\big(\mathcal{N}_p^{\otimes N}(\mathrm{Tr}_R[\omega_{A'^nR}])\big) \le H(\sigma_E) + g(N,\delta) \tag{18}$$

where $g(N,\delta) \equiv \delta N + h_2(\delta)$, and $h_2(\cdot)$ is the binary entropy function. Also, from the recoverability condition we have

$$\begin{aligned} M &= \log|A_1| = I(R\rangle B_1)_{\Phi_{B_1R}} \\ &\le I(R\rangle B_1)_{D_k(\mathcal{N}_p^{\otimes N}(\omega_k))} + \epsilon N + (1+\epsilon)h_2(\epsilon/[1+\epsilon]) \\ &\le I(R\rangle A'^n)_{\mathcal{N}_p^{\otimes N}(\omega_k)} + f(N,\epsilon) \\ &= H\big(\mathcal{N}_p^{\otimes N}(\mathrm{Tr}_R[\omega_{k,A'^nR}])\big) \\ &\qquad - H\big(\mathcal{N}_p^{\otimes N} \otimes I_R(\omega_{k,A'^nR})\big) + f(N,\epsilon), \end{aligned} \tag{19}$$

where $f(N,\epsilon) \equiv \epsilon N + (1+\epsilon)h_2(\epsilon/[1+\epsilon])$. The first equality follows from the fact that the coherent information of a maximally entangled state is just the logarithm of the dimension of one of the subsystems. The first inequality follows from the Alicki-Fannes-Audeneart inequality [22] applied to the recoverability condition given in Eq. (17). The second inequality is a quantum data processing inequality [18]. The last equality follows from the definition of the coherent information.

Furthermore, using the concavity of the von Neumann entropy and linearity of quantum operations we have that

$$\begin{aligned} &\min_{k\in\mathcal{K}} H\big(\mathcal{N}_p^{\otimes N}(\mathrm{Tr}_R[\omega_{k,A'^nR}])\big) \\ &\le \sum_k p_k H\big(\mathcal{N}_p^{\otimes N}(\mathrm{Tr}_R[\omega_{k,A'^nR}])\big) \\ &\le H\big(\mathcal{N}_p^{\otimes N}(\mathrm{Tr}_R[\sum_k p_k \omega_{k,A'^nR}])\big) \\ &= H\big(\mathcal{N}_p^{\otimes N}(\mathrm{Tr}_R[\omega_{A'^nR}])\big), \end{aligned} \tag{20}$$

and for many cases we expect $H(\mathcal{N}_p^{\otimes N}(\mathrm{Tr}_R[\omega_{k,A'^nR}]))$ to be roughly the same for every k (see Sec. III C 1 and III C 2). Thus

$$H\big(\mathcal{N}_p^{\otimes N}(\mathrm{Tr}_R[\omega_{k,A'^nR}])\big) \le H\big(\mathcal{N}_p^{\otimes N}(\mathrm{Tr}_R[\omega_{A'^nR}])\big) \tag{21}$$

for all k. Now putting Eq. (18), (19), and (21) together we arrive at our main result for this section, which states that Alice can secretly and reliably send M stego qubits to Bob, where M is bounded above by

$$\begin{aligned} M &\le H(\sigma_E) \\ &\quad - H\big((\mathcal{N}_p^{\otimes N} \otimes I_R)(\omega_{k,A'^nR})\big) + g(N,\delta) + f(N,\epsilon). \end{aligned} \tag{22}$$

Thus, if we can compute a maximum for $H(\sigma_E) = H(\mathcal{N}_{p+\delta p}^{\otimes N}(\rho))$ where ρ is pure (because V is an isometric encoding and ρ_c is pure), and also compute a lower bound for $H\big((\mathcal{N}_p^{\otimes N} \otimes I_R)(\omega_{k,A'^nR})\big)$ (or compute it explictly, recalling that ω_{k,A'^nR} is a pure state), then we have a tight upper bound on the number of qubits M that can be sent steganographically over a noisy quantum channel $\mathcal{N}_p$.

C. Upper bounds for specific channels

For the channels discussed in the achievability section of this paper, we can now apply our result given in Eq. (22), where we make the implicit assumption that Alice is using a nondegenerate code. Though our result given by Eq. (22) is true in general, for a degenerate code the number of distinct error syndromes is smaller (depending on the code), and the bounds discussed here and achievable rates discussed in the previous section would be adjusted.

1. The bit flip channel

For the bit flip channel with parameter $p+\delta p$ given by Eq. (1), the maximum of $H((\mathcal{N}^{BF}_{p+\delta p})^{\otimes N}(\rho))$ over all N-qubit pure states ρ is $Nh(p+\delta p)$ where $h(p+\delta p) = -(p+\delta p)\log(p+\delta p) - (1-(p+\delta p))\log(1-(p+\delta p))$ is the entropy of a single qubit sent through this bit flip channel. To prove this, consider some pure state $\rho = |\psi\rangle\langle\psi|$. Then

$$(\mathcal{N}^{BF}_{p+\delta p})^{\otimes N}(|\psi\rangle\langle\psi|) = \sum_s p(s) X^s |\psi\rangle\langle\psi| X^s \tag{23}$$

where we are summing over all binary strings s of length N; X^s is the operator acting on N qubits with an X acting at every location where s has a 1 and an I where s has a 0. The probability $p(s)$ is given by

$$p(s) = p^{w(s)}(1-p)^{N-w(s)}, \tag{24}$$

where $w(s)$ is the weight of string s. The Shannon entropy of this distribution is $Nh(p+\delta p)$ since it is a binomial distribution. The von Neumann entropy is the minimum Shannon entropy over all possible ensemble decompositions of the given state, and it is not hard to check that it is achieved when $|\psi\rangle$ is a Z eigenstate. Thus we have computed a maximum for the first term on the right hand side of Eq. (22).

Now we compute a lower bound for the second term on the right hand side of Eq. (22) in the case of the bit flip channel, i.e., $H((\mathcal{N}^{BF}_p)^{\otimes N} \otimes I_R)(\omega_{k,A'^nR}))$. First note that we can write

$$\begin{aligned}
&((\mathcal{N}^{BF}_p)^{\otimes N} \otimes I_R)(\omega_{k,A'^nR}) \\
&= \sum_{\underline{i}\in\mathcal{T}} (E_{\underline{i}} \otimes I_R)\omega_{k,A'^nR}(E^\dagger_{\underline{i}} \otimes I_R) \\
&+ \sum_{\underline{i}\notin\mathcal{T}} (E_{\underline{i}} \otimes I_R)\omega_{k,A'^nR}(E^\dagger_{\underline{i}} \otimes I_R),
\end{aligned} \tag{25}$$

where the index is $\underline{i} = i_1 i_2 ... i_N$, the errors $E_{\underline{i}}$ are given by

$$E_{\underline{i}} = A_{i_1} \otimes ... \otimes A_{i_N}, \tag{26}$$

and $\mathcal{T}$ is the set of typical sequences $\underline{i}$ corresponding to the typical errors of the channel $(\mathcal{N}^{BF}_p)^{\otimes N}$ [16, 17].

Recall we are making the assumption that the QECC Alice is using to correct the typical errors of the channel $(\mathcal{N}^{BF}_{p+\delta p})^{\otimes N}$ is nondegenerate. Because of this, we can infer that for each k, her encoded state ω_{k,A'^nR} forms a nondegenerate code for the channel $(\mathcal{N}^{BF}_p)^{\otimes N}$. This follows from the discussion in Section II A. This means that on a valid codeword in the QECC, for $\underline{i} \in \mathcal{T}$ the typical errors $E_{\underline{i}}$ all have distinct error syndromes, and act as unitaries that move the code space to a distinct, orthogonal subspace labeled by $\underline{i}$. So an error $E_{\underline{i}}$ occurs with a fixed probability $p_{\underline{i}}$ for all valid codewords of the QECC. Recall also that since these errors are typical, they have almost all the probability, i.e.

$$\sum_{i\in\mathcal{T}} p_i = 1-\epsilon \tag{27}$$

for arbitrarily small $\epsilon > 0$ (in the limit of large N). From Eq. (25) we have that

$$\begin{aligned}
&H\big(((\mathcal{N}^{BF}_p)^{\otimes N} \otimes I_R)(\omega_{k,A'^nR})\big) \\
&\geq (1-\epsilon) H\big(\frac{1}{1-\epsilon}\sum_{\underline{i}\in\mathcal{T}} (E_{\underline{i}} \otimes I_R)\omega_{k,A'^nR}(E^\dagger_{\underline{i}} \otimes I_R)\big) \\
&+ \epsilon H\big(\frac{1}{\epsilon}\sum_{\underline{i}\notin\mathcal{T}} (E_{\underline{i}} \otimes I_R)\omega_{k,A'^nR}(E^\dagger_{\underline{i}} \otimes I_R)\big) \\
&\geq (1-\epsilon) H\big(\frac{1}{1-\epsilon}\sum_{\underline{i}\in\mathcal{T}} (E_{\underline{i}} \otimes I_R)\omega_{k,A'^nR}(E^\dagger_{\underline{i}} \otimes I_R)\big)
\end{aligned} \tag{28}$$

where the first inequality follows from the concavity of the von Neumann entropy and Eq. (27) , and the second inequality follows because the term proportional to ϵ is positive. Continuing, we have

$$\begin{aligned}
&(1-\epsilon) H\big(\frac{1}{1-\epsilon}\sum_{\underline{i}\in\mathcal{T}} (E_{\underline{i}} \otimes I_R)\omega_{k,A'^nR}(E^\dagger_{\underline{i}} \otimes I_R)\big) \\
&= -\sum_{\underline{i}\in\mathcal{T}} p_{\underline{i}} \log\frac{p_{\underline{i}}}{1-\epsilon} = \sum_{\underline{i}\in\mathcal{T}} \big(p_{\underline{i}}\log(1-\epsilon) - p_{\underline{i}}\log p_{\underline{i}}\big) \\
&= (1-\epsilon)\log(1-\epsilon) - \sum_{\underline{i}\in\mathcal{T}} p_{\underline{i}}\log p_{\underline{i}} \\
&\geq Nh(p) - \mathcal{O}(\epsilon).
\end{aligned} \tag{29}$$

The first equality follows from the definition of the von Neumann entropy and Eq. (25). The inequality follows from performing a Taylor expansion on $\log(1-\epsilon)$ and from the fact that for the bit flip channel $(\mathcal{N}^{BF}_p)^{\otimes N}$:

$$\begin{aligned}
-\sum_{\underline{i}\in\mathcal{T}} p_{\underline{i}}\log p_{\underline{i}} &\approx -\sum_{\underline{i}\in\mathcal{T}} 2^{-Nh(p)}\log 2^{-Nh(p)} \\
&= Nh(p)\sum_{\underline{i}\in\mathcal{T}} 2^{-Nh(p)} \\
&\approx Nh(p) 2^{Nh(p)} 2^{-Nh(p)} = Nh(p),
\end{aligned} \tag{30}$$

where the approximate equalities follow directly from the theory of typical sequences. Thus

$$H\big(((\mathcal{N}^{BF}_p)^{\otimes N} \otimes I_R)(\omega_{k,A'^nR})\big) \geq Nh(p) - \mathcal{O}(\epsilon), \tag{31}$$

and combining this with the upper bound we computed for the first term in Eq. (22) gives

$$M \leq N(h(p+\delta p) - h(p)) \tag{32}$$

for a sufficiently reliable and secret protocol and large enough block size N. Comparing this to Eq. (5), we have that the encoding described in the previous section for steganography over the channel $\mathcal{N}_p^{BF}$ where Eve expects the channel to be $\mathcal{N}_{p+\delta p}^{BF}$ is essentially optimal.

2. *The depolarizing channel*

Unfortunately, for the depolarizing channel $\mathcal{N}^{DC}$ we do not know what N-qubit pure state ρ maximizes $H((\mathcal{N}_{p+\delta p}^{DC})^{\otimes N}(\rho))$. However, we can still bound this quantity, i.e., give an upper bound on the first term in Eq. (22). Consider the action of this channel on an N qubit pure state as follows:

$$(\mathcal{N}_{p+\delta p}^{DC})^{\otimes N}(\rho) \approx \sum_j E_j \rho E_j^\dagger \tag{33}$$

where $\{E_j\}$ is the set of typical errors associated with N applications of the channel $(\mathcal{N}_{p+\delta p}^{DC})^{\otimes N}$. Recall that we are choosing our isometric encoding to correct for typical errors of the channel Eve believes to be connecting Alice and Bob, i.e. $\mathcal{N}_{p+\delta p}^{DC}$. Furthermore, we're assuming that our code is nondegenerate. Therefore the states $E_j \rho E_j^\dagger$ are all orthogonal to each other for ρ in the codespace, and $\mathrm{Tr}[E_j \rho E_j^\dagger] = p_j$ where p_j are the typical probabilities associated with the errors E_j. The von Neumann entropy is the Shannon entropy minimized over all possible decompositions, so the entropy of this state is clearly

$$\begin{aligned} H(\sigma_E) &= H((\mathcal{N}_{p+\delta p}^{DC})^{\otimes N}(V\rho_c V^\dagger)) \\ &\leq -\sum_j p_j \log(p_j) \\ &= -\sum_j 2^{-Ns(p)} \log 2^{-Ns(p)} \\ &\approx Ns(p) 2^{Ns(p)} 2^{-Ns(p)} = Ns(p) \end{aligned} \tag{34}$$

In a similar argument to the one given in Sec. III C 1 , we can give a lower bound for the second term on the right hand side of Eq. (22) in the case of the depolarizing channel. We get that

$$H(((\mathcal{N}_p^{DC})^{\otimes N} \otimes I_R)(\omega_{k,A'^n R})) \geq Ns(p) - \mathcal{O}(\epsilon), \tag{35}$$

where ϵ becomes arbitrarily small for large N. Combining this with the upper bound given in Eq. (34) we have that

$$M \leq N(s(p+\delta p) - s(p)) \tag{36}$$

for a sufficiently reliable and secret protocol and large enough block size N. Comparing this to Eq. (11) we have that the encoding described in the previous section for steganography over the channel $\mathcal{N}_p^{DC}$ where Eve expects the channel to be $\mathcal{N}_{p+\delta p}^{DC}$ is essentially optimal, at least when we restrict ourselves to nondegenerate codes.

IV. CONCLUSION

Characterizing secret communication over noisy quantum channels is an interesting problem from both a practical and theoretical perspective. Here we have shown that two parties are able to communicate secretly with each other at a nonzero rate over a bit-flip or a depolarizing channel $\mathcal{N}_p$ using a shared secret key, without arousing suspicion from a potential eavesdropper Eve, so long as Eve believes the channel to be noisier than it really is. Eve can be made to believe this through Alice and Bob systematically adding extra noise to the channel prior to secret communication. In this paper we gave explicit bounds on the number of stego qubits that Alice can send to Bob by hiding her secret message in the syndromes of a nondegenerate error-correcting code designed to correct the typical errors of the channel Eve believes: $\mathcal{N}_{p+\delta p}$. We also gave explicit encodings that achieve these bounds.

Interesting future work should include a generalization of these results to steganography over general quantum channels $\mathcal{N}$. It is possible that in order to achieve the maximum possible rates in this scenario that degenerate codes are needed. For example, it is likely that the steganographic capacity we calculated for the depolarizing channel in this paper could be improved in this way. It is also possible that coding across multiple codeblocks using degenerate quantum codes could increase the steganographic capacity.

If the actual physical channel shared between Alice and Bob is $\mathcal{N}$, and the channel Eve believes is $\mathcal{N}'$, then what is the quantum steganographic capacity? In this paper we proved that for the bit-flip channel, the rate is the difference of quantum capacities, i.e., $Q(\mathcal{N}_p^{BF}) - Q(\mathcal{N}_{p+\delta p}^{BF}) = N(1-h(p)-1+h(p+\delta p)) = N(h(p+\delta p) - h(p))$. Also, allowing for our restriction to nondegenerate codes, this is true for the depolarizing channel as well. We conjecture that one might be able to prove that the steganographic rate in general will be $Q(\mathcal{N}) - Q(\mathcal{N}')$. This will require proof methods that go beyond those of the current paper, but we believe that this wll be an area of fruitful future study.

ACKNOWLEDGMENTS

Thanks to Yi-Hsiang Chen and Namit Anand for helpful discussions. This research was supported in part by NSF Grants CCF-1421078 and QIS-1719778, and by an IBM Einstein Fellowship at the Institute for Advanced Study.

Appendix: Relevant definitions and inequalities

First we give the definition of typical errors associated with a channel $\mathcal{N}^{\otimes N}$ as originally outlined in [16, 17].

Consider a quatum channel $\mathcal{N}$ with Kraus operators $E_1, ..., E_N$. Without loss of generality we can assume they are diagonal i.e. $\operatorname{Tr} E_i E_j^\dagger = 0$ for $i \neq j$. The probability of each Kraus operator is given by $p_i = \frac{1}{M} \operatorname{Tr} E_i E_i^\dagger$ where M is the dimension of the Hilbert space on which these operators act.

Now the operator $\mathcal{N}^{\otimes N}$ can be represented by N^m Kraus operators

$$E_{j_1} \otimes E_{j_2} \otimes ... \otimes E_{j_n} \equiv E_J \qquad (A.1)$$

where $j_i = 1, ..., N$. It is straightforward to verify that the probability associated with each of these operators E_J is given by

$$p_J = \frac{1}{M^n} \operatorname{Tr} E_J^\dagger E_J = p_{j_1} ... p_{j_n}. \qquad (A.2)$$

Hence the Kraus operators E_J of $\mathcal{N}^{\otimes N}$ are sequences of length n in which the symbols E_i of an alphabet $E_1, ..., E_N$ appear according to the probability distirbution $\{p_i\}$. Hence we are in the domain of classical random sequences. Thus we can take only the operators E_J that are ϵ-typical in the usual sense with respect to this probability distribution and write

$$\mathcal{N}^{\otimes N}(\rho) \approx \sum_{J \text{ typical}} E_J \rho E_J^\dagger. \qquad (A.3)$$

This strongly reduces the number of Kraus operators of $\mathcal{N}^{\otimes N}$ from N^m to roughly $2^{NH(\{p_i\})}$.

We now define the various inequalities which are used in Section III fo prove our converse theorem.

Definition 1 (Fannes-Audeneart inequality). *Let $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ be density operators and suppose that $\frac{1}{2}||\rho - \sigma||_1 \leq \epsilon \in [0, 1]$. Then the following inequality holds*

$$|H(\rho) - H(\sigma)| \leq \epsilon \log \dim(\mathcal{H}) + h_2(\epsilon)$$

Definition 2 (Alicki-Fannes-Audeneart inequality). *Let $\rho_{AB}, \sigma_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$. Suppose that*

$$\frac{1}{2}||\rho_{AB} - \sigma_{AB}||_1 \leq \epsilon,$$

for $\epsilon \in [0, 1]$. Then

$$|H(A|B)_\rho - H(A|B)_\sigma| \leq 2\epsilon \log \dim(\mathcal{H}_A) + g_2(\epsilon)$$

where $g_2(\epsilon) \equiv (\epsilon + 1) \log_2(\epsilon + 1) - \epsilon \log_2(\epsilon)$.

Definition 3 (Data Processsing for Coherent Information). *Let $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ and let $\mathcal{N} : \mathcal{L}(\mathcal{H}_B) \to \mathcal{L}(\mathcal{H}_{B'})$ be a quantum channel. Set $\sigma_{AB'} \equiv \mathcal{N}_{B \to B'}(\rho_{AB})$. Then the following quantum data-processing inequality holds*

$$I(A\rangle B)_\rho \geq I(A\rangle B')_\sigma.$$

[1] Herodotus, *The Histories* (Penguin Books, 1996).
[2] S. Singh, *The code book: the secret history of codes and code-breaking* (Fourth Estate, 2000).
[3] F. A. Petitcolas, R. J. Anderson, and M. G. Kuhn, Proceedings of the IEEE **87**, 1062 (1999).
[4] P. W. Shor, SIAM review **41**, 303 (1999).
[5] S. Natori, in *Quantum Computation and Information* (Springer, 2006) pp. 235–240.
[6] I. Banerjee, S. Bhattacharyya, and G. Sanyal, International Journal of Computer Network and Information Security **4**, 65 (2012).
[7] J. Gea-Banacloche, Journal of Mathematical Physics **43**, 4531 (2002).
[8] B. A. Shaw and T. A. Brun, arXiv preprint arXiv:1007.0793 (2010).
[9] B. A. Shaw and T. A. Brun, Physical Review A **83**, 022310 (2011).
[10] C. Sutherland and T. A. Brun, arXiv preprint arXiv:1805.01599 (2018).
[11] B. A. Bash, A. H. Gheorghe, M. Patel, J. L. Habif, D. Goeckel, D. Towsley, and S. Guha, Nature Communications **6** (2015).
[12] A. Sheikholeslami, B. A. Bash, D. Towsley, D. Goeckel, and S. Guha, in *Information Theory (ISIT), 2016 IEEE International Symposium on* (IEEE, 2016) pp. 2064–2068.
[13] L. Wang, in *Information Theory Workshop (ITW), 2016 IEEE* (IEEE, 2016) pp. 364–368.
[14] K. Bradler, T. Kalajdzievski, G. Siopsis, and C. Weedbrook, arXiv preprint arXiv:1607.05916 (2016).
[15] J. M. Arrazola and V. Scarani, Physical Review Letters **117**, 250503 (2016).
[16] R. Klesse, Physical Review A **75**, 062315 (2007).
[17] R. Klesse, Open Systems & Information Dynamics **15**, 21 (2008).
[18] M. M. Wilde, *Quantum information theory* (Cambridge University Press, 2013).
[19] T. M. Cover and J. A. Thomas, *Elements of information theory* (John Wiley & Sons, 2012).
[20] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge university press, 2010).
[21] K. M. Audenaert, Journal of Physics A: Mathematical and Theoretical **40**, 8127 (2007).
[22] R. Alicki and M. Fannes, Journal of Physics A: Mathematical and General **37**, L55 (2004).