Quantum Steganography

Bilal A. Shaw,¹* Todd A. Brun²

¹Computer Science Department, University of Southern California,
²Electrical Engineering Department, University of Southern California
Los Angeles, CA 90089, USA

*To whom correspondence should be addressed; E-mail: bilalshaw@gmail.com.

Steganography is the process of hiding secret information by embedding it in an “innocent” message. We present protocols for hiding quantum information in a codeword of a quantum error-correcting code passing through a channel. Using either a shared classical secret key or shared entanglement the sender (Alice) disguises her information as errors in the channel. The receiver (Bob) can retrieve the hidden information, but an eavesdropper (Eve) with the power to monitor the channel, but without the secret key, cannot distinguish the message from channel noise. We analyze how difficult it is for Eve to detect the presence of secret messages, and estimate rates of steganographic communication and secret key consumption for certain protocols.

Steganography is the science of hiding a message within a larger innocent-looking plain-text message, and communicating the resulting data over a communications channel or by a courier so that the steganographic message is readable only by the intended receiver. The word comes from the Greek words steganos which means “covered,” and graphia which means “writing.” The art of information hiding dates back to 440 B.C. to the Greeks (I). The term steganography
was first used in 1499 by Johannes Trithemius in his *Steganographia*, which was one of the first treatises on the use of cryptographic and steganographic techniques (2).

The modern study of steganography was initiated by Simmons and the paradigm can be stated as follows (3). Alice and Bob are imprisoned in two different cells that are far apart. They would like to devise an escape plan, but the only way they can communicate with each other is through a courier who is under the command of the warden (Eve, the adversary) of the penitentiary. The courier leaks all information to the warden. If the warden suspects that either Alice or Bob are conspiring to escape from the penitentiary, she will cut off all communication between them, and move both of them to a maximum security cell. Prior to their incarceration Alice and Bob had access to a shared secret key—assumed to be a sufficiently long string of random bits—which they later exploit to send secret messages hidden in a cover text. Can Alice and Bob devise an escape plan without arousing the suspicion of the warden?

Julio Gea-Banacloche (4) introduced the idea of hiding secret messages in the form of error syndromes by deliberately applying correctable errors to a quantum state encoded in the three-bit repetition quantum error-correcting code (QECC). In his paper, however, he did not address the issue of an innocent-looking message—in the protocol he proposed, the messages would not resemble a plausible quantum channel. The latter is one of the major contributions of our work. Curty et. al. propose three different quantum steganographic protocols (5). However, none of these protocols address the issue of communicating an innocent message over a noisy classical channel or a general quantum channel, or give key-consumption rates. Natori provides a rudimentary treatment of quantum steganography which is a modification of super-dense coding (6). Martin introduced a notion of quantum steganographic communication in (7). His protocol is a variation of Bennett and Brassard’s quantum-key distribution protocol (QKD), in which he hides a steganographic channel in the QKD protocol.

Our treatment of quantum steganography is more general than those above. We provide a
protocol (Fig. 1) for hiding quantum information using typical sequences of errors for general quantum channels. We begin by showing how quantum information can be hidden in the noise of a depolarizing channel, using a shared classical secret key between Alice and Bob. In our first quantum steganographic protocol the channel is intrinsically noiseless (i.e., all noise is controlled by Alice), and in the second case the channel has its own intrinsic noise (not controlled by Alice and Bob). We calculate the amount of secret key consumed. We later present a quantum steganographic protocol for general quantum channels. We also discuss whether Alice and Bob can send a finite amount of hidden information, or can actually communicate at a nonzero asymptotic rate (given an arbitrarily large secret key). This depends on Eve’s knowledge of the physical channel, and Alice and Bob’s knowledge of Eve’s expectations. Finally, we address the question of security. This is two-fold: first, can Eve detect that a secret message has been sent? And second, can she read the message?

The quantum analog of the classical binary symmetric channel (BSC) is the depolarizing channel (DC) which is one of the most widely used quantum channel models:

\[
\rho \rightarrow \mathcal{N}\rho = (1 - p)\rho + \frac{p}{3}X\rho X + \frac{p}{3}Y\rho Y + \frac{p}{3}Z\rho Z .
\]

That is, each qubit has an equal probability of undergoing an \(X\), \(Y\), or \(Z\) error. Applying this channel repeatedly to a qubit will map it eventually to the maximally mixed state \(I/2\). We can rewrite this channel in a different but equivalent form:

\[
\mathcal{N} = (1 - 4p/3)\mathcal{I} + (4p/3)\mathcal{T} .
\]

where \(\mathcal{I}\rho = \rho\) and \(\mathcal{T}\rho = (1/4)(\rho + X\rho X + Y\rho Y + Z\rho Z)\). The operation \(\mathcal{T}\) is twirling: it takes a qubit in any state \(\rho\) to the maximally mixed state \(I/2\). If we rewrite the channel in this way, instead of applying \(X\), \(Y\), or \(Z\) errors with probability \(p/3\), we can think of removing the qubit with probability \(4p/3\), and replacing it with a maximally mixed state. This picture makes the steganographic protocol more transparent. We will first assume that the actual physical channel
between Alice and Bob is noiseless. All the noise that Eve sees is due to deliberate errors that Alice applies to her codewords.

1. Alice encodes a covertex of $k_c$ qubits into $N$ qubits with an $[[N, k_c]]$ quantum error-correcting code (QECC).

2. From (2), the DC would maximally mix $Q$ qubits with probability $p_Q$ where

$$p_Q = \binom{N}{Q} (4p/3)^Q (1 - 4p/3)^{N-Q}.$$  \hspace{1cm} (3)

For large $N$, Alice can send $M = (4/3)pN(1 - \delta)$ stego qubits, where $1 \gg \delta \gg \sqrt{(1 - 4p/3)/(4p/3)N}$. (The chance of fewer than $M$ errors is negligibly small.)

3. Using the shared random key (or shared ebits), Alice chooses a random subset of $M$ qubits out of the $N$, and swaps her $M$ stego qubits for those qubits of the codeword. She also replaces a random number $m$ of qubits outside this subset with maximally mixed qubits, so that the total $Q = M + m$ matches the binomial distribution (3) to high accuracy.

4. Alice “twirls” her $M$ stego qubits using $2M$ bits of secret key or $2M$ shared ebits. To each qubit she applies one of $I$, $X$, $Y$, or $Z$ chosen at random, so $\rho \rightarrow T \rho$. To Eve, who does not have the key, these qubits appear maximally mixed. (Twirling can be thought of as the quantum equivalent of a one-time pad.)

5. Alice transmits the codeword to Bob. From the secret key, he knows the correct subset of $M$ qubits, and the one-time pad to decode them.

This protocol transmits $(4/3)pN(1 - \delta)$ secret qubits from Alice to Bob (Fig. 2).

If the channel contains intrinsic noise, Alice will first have to encode her $k_s$ stego qubits in an $[[M, k_s]]$ QECC, swap those $M$ qubits for a random subset of $M$ qubits in the codeword, and apply the twirling procedure. This twirling does not interfere with the error-correcting
power of the QECC if Bob knows the key. Assuming the physical channel is also a DC with error rate $p$, and that Alice emulates a DC with error rate $q$, the effective channel will appear to Eve like a DC with error rate $p + q(1 - 4p/3) \equiv p + \delta p$. The rate of transmission $k_s/N$ will depend on the rate of the QECC used to protect the stego qubits. For a BSC this would be $(1 - \delta)(1 - h(p))\delta p/(1 - 2p)$. However, for most quantum channels (including the DC) the achievable rate is not known.

The secret key is used at two points in these protocols. First, in step 3 Alice chooses a random subset of $M$ qubits out of the $N$-qubit codeword. There are $C(N,M)$ subsets, so roughly $\log_2 C(N,M)$ bits are needed to choose one. Next, in step 4, $2M$ bits of key are used for twirling. This gives us

$$n_k \approx \log_2 \left( \frac{N}{M} \right) + 2M$$

bits of secret key used. Define the key consumption rate $K = n_k/N$ to be the number of bits of key consumed per qubit that Alice sends through the channel. We use $M \approx 4qN/3$ and $q \approx \delta p/(1 - 4p/3)$ to express $K$ in terms of $p, \delta p,$ and $N$ (Fig. 3):

$$K \approx \log_2 \left[ \frac{(4/\beta)^\beta (1 - \beta N)^{\beta - 1}}{\beta} \right], \quad \beta \equiv 4\delta p/(3 - 4p).$$

Alice can consume fewer bits of key if Bob and she have access to a source that averages to a maximally mixed state. This would allow them to bypass the twirling procedure. The protocols given above perform well in emulating a depolarizing channel. However, there are far more general channels than these, and the protocols may not work well, or at all, in these cases. If one has a channel that can be written

$$\rho \to \mathcal{N}\rho = (1 - p_T + p_E)I\rho + p_T T\rho + p_E E\rho$$

where $E$ is an arbitrary error operation, one can still use the above protocols to hide approximately $p_T N$ stego bits or qubits, while generating $p_E N$ random errors of type $E$. But for some
channels, $p_T$ may be very small or zero. How should we proceed? Moreover, hiding stego qubits locally as apparently maximally-mixed qubits sacrifices some potential information. The location of the error—that is, the choice of the subset holding the errors—could also be used to convey information, potentially increasing the rate and reducing the amount of secret key or shared entanglement required.

A different approach is instead to encode information in the error syndromes. For simplicity, we consider the case when $N$ is large. In this case, it suffices to consider only typical errors. We begin with the case where the physical channel is noise-free.

For large $N$, almost all (probability $1 - \epsilon$) combinations of errors on the individual qubits will correspond to one of the set of typical errors. There are roughly $2^s N$ of these, and their probabilities $p_e$ are all bounded within a range $2^{-N(s+\delta)} \leq p_e \leq 2^{-N(s-\delta)}$. The number $s$ is the entropy of the channel on one qubit; for the BSC $s = h(p) = -p \log_2 p - (1 - p) \log_2 (1 - p)$, and for the DC $s = -(1 - p) \log_2 (1 - p) - p \log_2 p / 3$. We label the typical error operators $E_0, E_1, \ldots, E_{2^s N - 1}$, and their corresponding probabilities are $p_j$. A good choice of QECC for the cover text will be able to correct all these errors. We make the simplifying assumption that the QECC is nondegenerate, so each typical error $E_j$ has a distinct error syndrome labelled $s_j$.

Ahead of time, Alice and Bob partition the typical errors into $C$ roughly equiprobable sets $S_k$, so that

$$\sum_{E_j \in S_k} p_j \approx \frac{1}{C}, \forall k.$$  \hspace{1cm} (7)

As far as possible, the errors in a given set should be chosen to have roughly equal probabilities. The maximum of $C$ is roughly $C \approx 2^{N(s-\delta)}$, and $k = 0, \ldots, C - 1$. We can now present a new quantum steganographic protocol, using error syndromes to store information.

1. Alice prepares $k_c$ qubits of cover text in a state $|\psi_c\rangle$. 

6
2. Alice’s secret message is a string of \( \log_2 C \approx N(s - \delta) \) qubits, in a state

\[
|\psi_s\rangle = \sum_{k=0}^{C-1} \alpha_k |k\rangle.
\] (8)

She “twirls” each qubit of this string, using \( 2N(s - \delta) \) bits of the secret key or shared ebits, to get a maximally mixed state. To this, she appends \( N - k_c - (s - \delta)N \) extra ancilla qubits in the state \( |0\rangle \) to make up a total register of \( N - k_c \) qubits.

3. Using the shared secret key, Alice chooses from each set \( S_k \) a typical error \( E_{jk} \) with syndrome \( s_{jk} \). She applies a unitary \( U_S \) to the register of \( N - k_c \) qubits, that maps

\[
U_S \left( |k\rangle \otimes |0\rangle^{\otimes N-k_c-sN} \right) = |s_{jk}\rangle.
\]

She appends this register to the cover qubits in state \( |\psi_c\rangle \), then applies the encoding unitary \( U_E \). Averaging over the secret key, the resulting state will appear to Eve like \( \rho \approx \sum_{j=0}^{2^{nS}-1} p_j |\psi_c\rangle\langle\psi_c| E_j^\dagger \), which is effectively indistinguishable from the channel being emulated acting on the encoded cover text.

4. Alice sends this codeword to Bob. If Eve examines its syndrome, she will find a typical error for the channel being emulated.

5. Bob applies the decoding unitary \( U_D = U_E^\dagger \), and then applies \( U_S^\dagger \) (which he knows using the shared secret key). He discards the cover text and the last \( N - k_c - sN \) ancilla qubits, and undoes the twirling operation on the remaining qubits, again using the secret key. If Eve has not measured the qubits, he will have recovered the state encoded by Alice (8).

This protocol may easily be used to send classical information by using a single basis state rather than a superposition like (8). The steganographic transmission rate \( R \) is roughly \( R \approx s - \delta \rightarrow s \).

The rate of transmission \( s \) is higher than the rate \( 4p/3 \) of our first protocol. This protocol used \( 2N(s - \delta) \) bits of secret key (or ebits) for twirling in step 2, and roughly \( N\delta \) bits of secret key in choosing representative errors \( E_{jk} \) from each set \( S_k \) in step 3. So the key rate is roughly \( K \approx 2s - \delta \rightarrow 2s \), better than the first protocol in key usage per stego qubit transmitted. Since
almost all the key usage goes to the twirling operation, for sources that are maximally mixed on average the rate of key usage can actually go to zero as \( N \to \infty \). However, this encoding is much trickier in the case where the channel contains intrinsic noise.

In principle this quantum steganographic protocol can be used when the channel contains noise. The steganographic qubits are first encoded in a QECC to protect them against the noise in the channel. In practice, for many channels this can be difficult: the effects of errors on the space of syndromes look quite different from a usual additive error channel. Also, unlike the depolarizing channel, general channels when composed together may change their type. However, by drawing on codes with suitable properties, the problem of designing steganographic protocols for general channels may be simplified. We discuss a simple example in the supporting online material (SOM), but the solution for a general channel is a problem for future work.

What is the standard of security for a stego protocol? There are two obvious considerations. First, if Eve becomes suspicious, can she read the message? At the cost of using one-time pads or twirling, Alice and Bob can prevent this from happening.

The more important question is, can Alice and Bob avoid arousing Eve’s suspicions in the first place? To do this, the messages that Alice sends must emulate as closely as possible the channel that Eve expects. We can make this condition quantitative. Let \( \mathcal{E}_C \) be the channel on \( N \) qubits that Eve expects, and let \( \mathcal{E}_S \) be the effective channel that Alice and Bob produce with their steganographic protocol. Then the protocol is secure if \( \mathcal{E}_S \) is \( \epsilon \)-close to \( \mathcal{E}_C \) in the diamond norm

\[
\| \mathcal{E}_S - \mathcal{E}_C \|_\diamond \leq \epsilon
\]

for some small \( \epsilon > 0 \). The diamond norm is directly related to the probability for Eve to distinguish \( \mathcal{E}_C \) from \( \mathcal{E}_S \) under ideal circumstances (i.e., when she controls both inputs and outputs), and so puts an upper bound on her ability to distinguish them in practice.

For a simple example, the difference between two DCs applied to \( N \) qubits has norm

\[
\| \mathcal{N}_r \otimes N - \mathcal{N}_p \otimes N \|_\diamond = \sum_{j=0}^N \binom{N}{j} \left| r^j (1 - r)^{N-j} - p^j (1 - p)^{N-j} \right|,
\]  \(8\)
where $p$ is the error-rate of the channel Eve expects and $r = p + \delta p$ is the error-rate of the steganographic channel that emulates Eve’s expected channel. If we make $\delta p < \epsilon \sqrt{p(1-p)/N}$ then we can make this norm as small as we like, while communicating $O(\delta p N) = O(\epsilon \sqrt{N})$ secret qubits. This indicates that even if Eve has exact knowledge of the channel, Alice and Bob can in principle send an arbitrarily large (but finite) amount of information without arousing Eve’s suspicion, by choosing a sufficiently small $\delta p$ and large $N$ (8). If Eve’s knowledge of the channel is imperfect, Alice and Bob can do even better, communicating steganographic information at a nonzero rate. If Eve is constantly monitoring the channel over a long period of time, and if she has exact knowledge of the channel then she will eventually learn that Alice and Bob are communicating with each other steganographically. Moreover, with constant measurement Eve can disrupt the superpositions of the steganographic qubits and prevent any information from ever reaching Bob, effectively flooding the quantum channel with noise.

If Alice and Bob have shared ebits, they can perform measurements on each of their halves and distill correlated random bits. Moreover, with shared ebits Alice can send her quantum information to Bob via quantum teleportation by sending only classical bits through the channel. These classical bits are the result of her measurement on her half of the ebits and her stego qubits. To Eve who may be monitoring the channel, these bits will look maximally mixed (random). For her to change the outcome of what Bob receives on his end, Eve would have to disrupt the bits. So if Eve is measuring the channel continuously, Alice and Bob can still send quantum information to each other using their shared ebits.

References and Notes

1. Herodotus, The Histories (Penguin Books, London, 1996).

2. J. Trithemius, Steganographia (First printed edition, Frankfurt, 1606).
3. G. J. Simmons, *Advances in Cryptology - CRYPTO 83*, 51-67 (1983).

4. J. Gea-Banacloche, *J. Math. Phys.* 43, 4531 (2002).

5. M. Curty, D. J. Santos, paper presented at the 2nd Bielefeld Workshop on Quantum Information and Complexity, Bielefeld, Germany, 10 October (2004).

6. S. Natori, *Quantum Computation and Information*, 102 (2006).

7. K. Martin, *Lecture Notes in Computer Science*, 4569 (2008).

8. Further details can be found in the supporting online material on *Science Online*.

9. T. Mittelholzer, *Lecture Notes in Computer Science*, 1768 (2000).

10. Jeremiah J. Harmsen, William A. Pearlman, Capacity of Steganographic Channels, IEEE Transactions on Information Theory, 55, No. 4, (2009).

11. Amir Bennatan, David Burshtein, Giuseppe Caire, Shlomo Shamai, Superposition Coding for Side-Information Channels, IEEE Transactions on Information Theory, 52, No. 5, (2006).

12. Acknowledgement. BAS and TAB would like to thank Daniel Gottesman, Patrick Hayden, Debbie Leung, John Preskill, Stephanie Wehner and Mark Wilde for their useful comments and suggestions at various stages of this work. This work was funded in part by NSF Grant No. CCF-0448658. TAB also acknowledges the hospitality and support of the Kavli Institute for Theoretical Physics.

**Supporting Online Material**

SOM Text

SOM Figures

References
Figure 1: There are three different inputs to the steganographic encoder $\mathcal{E}$: a cover-message $|C\rangle$; the secret message that we would like to hide, which can be quantum $|S\rangle$ or classical $S$; a shared secret key which may be quantum (ebit) $|K\rangle$ or classical $K$. Eve can monitor some part of the noisy quantum channel $\mathcal{N}$ shown in the red box. Bob can decode the steganographic message using the decoder $\mathcal{D}$ and the shared secret key $|K\rangle$ or $K$ and recover $|C\rangle$, and $|S\rangle$ or $S$ with very high probability.
Figure 2: Alice hides her information qubit (solid brown circle) by swapping it in with a qubit of her quantum codeword. She uses her shared secret key with Bob to determine which qubit to swap. She uses the shared key again to twirl the information qubit. She further applies random depolarizing errors to the rest of the qubits of the codeword (shown in green). She sends the codeword through a depolarizing channel to Bob who uses the shared secret to correctly apply the untwirling operation, followed by locating and swapping out Alice’s original information qubit.
Figure 3: We plot the key consumption rate (KCR) as a function of the error-rate $p$ of the channel.
Supporting Online Material

We assume a basic knowledge of quantum information science at the level of (1). In this section we gather the definition of the diamond norm and some of its relevant properties to derive the norm of the difference between \( N \) uses of two binary-symmetric channels (BSC) and two depolarizing channels (DC). We refer the reader to John Watrous’s lecture notes for the definition and properties of the diamond norm (2). As mentioned in the main text the diamond norm give us a measure of how “close” or similar two channels can be when they transform an arbitrary density matrix from one Hilbert space to another. More formally let \( \mathcal{N} \) be some arbitrary super-operator, and let \( \mathcal{N} : L (\mathcal{V}) \to L (\mathcal{W}) \), where \( L (.) \) is a space of linear operators on the Hilbert spaces \( \mathcal{V} \) and \( \mathcal{W} \). Then one can define the diamond-norm of \( \mathcal{N} \) as:

\[
\| \mathcal{N} \|_\diamond \equiv \| I_{L (\mathcal{V})} \otimes \mathcal{N} \|_{tr},
\]

where \( \| \mathcal{N} \|_{tr} \) is defined as:

\[
\| \mathcal{N} \|_{tr} \equiv \max \{ \| \mathcal{N} (O) \|_{tr} : O \in L (\mathcal{V}) , \| O \|_{tr} = 1 \}.
\]

The maximization in (S2) is over all density matrices. When the Hilbert space is infinite dimensional we take the supremum of the set defined in (S2).

Binary Symmetric Channel

Let \( 0 < p < 1/2 \) be the rate at which Alice flips the qubits of her codeword. Let \( r \equiv p + \delta p \) be the rate at which the BSC flips qubits, where \( \delta p \) is some additional noise which is not under the control of either Alice or Bob. We assume that \( 0 < p < r < 1/2 \) because at \( p = 1/2 \) the channel has zero capacity to send information and \( p > 1/2 \) means that more qubits are being flipped which is unnatural for this channel. For a single qubit (\( N = 1 \)) let \( \mathcal{N}_p \) be the BSC that Alice applies to an arbitrary single-qubit density operator \( \rho \):

\[
\mathcal{N}_p \rho = (1 - p)\rho + pX \rho X,
\]
and let $\mathcal{N}_r$ be the actual BSC

$$\mathcal{N}_r\rho = (1 - r)\rho + rX\rho X.$$  \hfill (S4)

We can now express the difference of the two channels as:

$$(\mathcal{N}_r - \mathcal{N}_p)\rho = (p - r)\rho + (r - p)X\rho X$$  \hfill (S5)

We can express the diamond norm of the difference of the channels $\mathcal{N}_p$ and $\mathcal{N}_r$ as:

$$\|\mathcal{N}_r - \mathcal{N}_p\|_\diamond = \max_\rho \|(I \otimes (\mathcal{N}_r - \mathcal{N}_p))\rho\|_{tr}$$  \hfill (S6)

$$= (r - p)\max_\rho \|(I \otimes I)\rho(I \otimes I) - (I \otimes X)\rho(I \otimes X)\|_{tr}.$$  \hfill (S7)

When we substitute $\rho = \psi \otimes |0\rangle\langle 0| \ (\psi$ is some arbitrary density operator) in the above equation we achieve the maximum.

$$\|\mathcal{N}_r - \mathcal{N}_p\|_\diamond = (r - p)\|\psi \otimes |0\rangle\langle 0| - \psi \otimes |1\rangle\langle 1|\|_{tr}$$  \hfill (S8)

$$\leq \|\psi \otimes |0\rangle\langle 0|\|_{tr} + |1\|\|\psi \otimes |1\rangle\langle 1|\|_{tr}$$  \hfill (S9)

$$= (r - p)\|\psi\|_{tr} \|\langle 0|\langle 0|_{tr} + \|\psi\|_{tr} \|\langle 1|\langle 1|\|_{tr}$$  \hfill (S10)

$$= (r - p)(1 + 1)$$  \hfill (S11)

$$= 2(r - p)$$  \hfill (S12)

$$= 2(p + \delta p - p)$$  \hfill (S13)

$$= 2\delta p.$$  \hfill (S14)

In (S9) we use the triangle inequality and in (S10) we use the fact that for any two linear operators $A$ and $B$, the trace norm of their tensor product is equal to the product of their trace norms, i.e., $\|A \otimes B\|_{tr} = \|A\|_{tr} \|B\|_{tr}$. We would like an expression for the optimal probability to correctly distinguish two channels.

$$P_{opt} = \frac{1}{2} + \frac{1}{4}\|\mathcal{N}_r - \mathcal{N}_p\|_\diamond.$$  \hfill (S15)
So for a single-qubit use
\[ P_{\text{opt}} = \frac{1}{2}(1 + \delta p) . \] (S16)

For the case where we have two qubits, we can write Alice’s BSC as:
\[ (N_p \otimes N_p)\rho = (1 - p)^2 \rho + p(1 - p)X_1 \rho X_1 + p(1 - p)X_2 \rho X_2 + p^2 X_1 X_2 \rho X_1 X_2 , \] (S17)
where \( X_1 \equiv X \otimes I \) and \( X_2 \equiv I \otimes X \), and \( X_1 X_2 \equiv X \otimes X \). We can similarly calculate \( N_1 \otimes N_1 \).

We can now write the difference between the two channels as:
\[ (N_r \otimes N_r - N_p \otimes N_p)\rho = (r^2 - 2r + 2p - p^2) \]
\[ + (r - r^2 - p + p^2)(X_1 \rho X_1 + X_2 \rho X_2) \] (S18)
\[ + (r^2 - p^2)X_1 X_2 \rho X_1 X_2 . \]

The diamond norm of the difference between two BSC on two qubits can be expressed as:
\[ \|N_r \otimes N_r - N_p \otimes N_p\|_\diamond = \max_\rho \| (I \otimes (N_r \otimes N_r - N_p \otimes N_p))\rho \|_{tr} . \] (S19)

We use a similar construction from the single-qubit case to maximize the right side of (S17). Letting \( \rho = \psi \otimes |00\rangle \langle 00| \) in (S19), we get:
\[ \|N_r \otimes N_r - N_p \otimes N_p\|_\diamond = \left| (1 - r)^2 - (1 - p)^2 \right| + 2 \left| r(1 - r) - p(1 - p) \right| + \left| r^2 - p^2 \right| . \] (S20)

Given our constraints that \( 0 < p < r < 1/2 \), the first term on the right side of (S18) is negative while the second and third terms are positive. This give us:
\[ \|N_r \otimes N_r - N_p \otimes N_p\|_\diamond = 2(r - p)(2 - r - p) \] (S21)
\[ = 2\delta p(2 - 2p - 2\delta p) . \] (S22)

So in the double-qubit case \( P_{\text{opt}} \) is:
\[ P_{\text{opt}} = \frac{1}{2}(1 + \delta p(2 - 2p - 2r)) . \] (S23)
If we observe $S(20)$ carefully we find that the terms are distributed binomially. For the case where we have $N$ qubits, we can use $\rho = \psi \otimes |00\cdots0\rangle\langle00\cdots0|$ to maximize the diamond norm for $N$ uses of BSC to get:

$$\|N_r^{\otimes N} - N_p^{\otimes N}\|_\diamond = \sum_{j=0}^N \binom{N}{j} \left| r^j (1-r)^{N-j} - p^j (1-p)^{N-j} \right| .$$  \hspace{1cm} (S24)

**Depolarizing Channel**

The calculation of the diamond norm of the difference between $N$ uses of two depolarizing channels (DC) is similar to the calculation of BSC that we performed in the previous section. The expression for the channel is

$$\mathcal{N}_p \rho = (1-p)\rho + (p/3)(X\rho X + Y\rho Y + Z\rho Z) .$$  \hspace{1cm} (S25)

Eve sees a channel with a somewhat higher rate $r = p + \delta p$. As in the BSC case we assume that $0 < p < r < 1/2$. For $N = 2$ case the difference between the two depolarizing channels is:

$$(\mathcal{N}_r \otimes \mathcal{N}_r - \mathcal{N}_p \otimes \mathcal{N}_p)\rho = ((1-r)^2 - (1-p)^2)\rho$$

$$+ ((1-r)(r/3) - (1-p)(p/3))(X_1\rho X_1 + \cdots + Z_2\rho Z_2)$$

$$+ ((r/3)^2 - (p/3)^2)(X_1 X_2\rho X_1 X_2 + \cdots + Z_1 Z_2\rho Z_1 Z_2) .$$  \hspace{1cm} (S27)

As in the BSC case we can express the diamond norm as in (S19). The density matrix that maximizes the trace norm is $\rho = \psi \otimes |\Phi^+\rangle\langle\Phi^+|$, where $|\Phi^+\rangle = 1/\sqrt{2}(|00\rangle + |11\rangle)$, and $\psi$ is some arbitrary single-qubit density operator.

$$\|\mathcal{N}_r \otimes \mathcal{N}_r - \mathcal{N}_p \otimes \mathcal{N}_p\|_\diamond = \left| (1-r)^2 - (1-p)^2 \right|$$

$$+ 6 \left| (1-r)(r/3) - (1-p)(p/3) \right|$$

$$+ 9 \left| (r/3)^2 - (p/3)^2 \right|$$

$$= \left| (1-r)^2 - (1-p)^2 \right| + 2 \left| (1-r)r - (1-p)p \right| + \left| r^2 - p^2 \right| .$$  \hspace{1cm} (S28)
After evaluating the absolute value terms, we get:

\[
\| N_r \otimes N_r - N_p \otimes N_p \|_\diamond = 2(r - p)(2 - r - p) \\
= 2\delta p \left(2 - 2p - \delta p\right). \tag{S29}
\]

So,

\[
P_{opt} = \frac{1}{2} + \frac{1}{2}\delta p \left(2 - 2p - \delta p\right). \tag{S30}
\]

For the general case for \( N \) uses of the depolarizing channel we may write the diamond norm as:

\[
\| N_r^{\otimes N} - N_p^{\otimes N} \|_\diamond = \sum_{j=0}^{N} \binom{N}{j} \left| r^j (1 - r)^{N-j} - p^j (1 - p)^{N-j} \right|, \tag{S31}
\]

which is exactly the same expression as for the BSC.

**Achievable Rate for Protocol 2**

We will work out the simplest example—the BSC in the case where the physical channel is noise-free. The errors in the codewords that Alice sends to Bob are binomially distributed. Let \( pN \) be the mean of this distribution and let the variance be \( pN\delta \), where \( 0 < \delta \ll 1 \). Here \( N \) is the length of each of codeword. Let

\[
p_k = \binom{N}{k} p^k (1 - p)^{N-k} \tag{S32}
\]

be the errors that Alice applies to her codewords. For each \( k \) from \( Np(1 - \delta) \) to \( Np(1 + \delta) \) choose \( C_k \) strings of weight \( k \). Let

\[
C = \sum_{k=Np(1-\delta)}^{Np(1+\delta)} C_k \tag{S33}
\]

Let these sets of strings be called \( S_k \), and

\[
S = \cup_k S_k \tag{S34}
\]
So the total number of strings in the set $S$ is $C$. Define the probability $q \equiv 1/C$. Then we want to satisfy $qC_k = C_k/C = p_k$. Clearly we must have $\binom{C_k}{k}$, for all $k$. This implies that:

$$C_k p^k (1-p)^{N-k} \leq \binom{N}{k} p^k (1-p)^{N-k}$$
$$\Rightarrow C_k p^k (1-p)^{N-k} \leq C_k q$$
$$\Rightarrow p^k (1-p)^{N-k} \leq q$$

We want $C$ to be as large as possible, which means we want $q$ to be as small as possible. This constraint then gives us

$$q = p^{Np(1-\delta)}(1-p)^{N(1-p+p\delta)}$$
$$\Rightarrow C = 1/q$$
$$\Rightarrow C = p^{-Np(1-\delta)}(1-p)^{-N(1-p+p\delta)}$$

The number of bits that Alice can send is, therefore

$$M = \log_2 C$$
$$= N(-p \log_2 p - (1-p) \log_2 (1-p) + \delta(p \log_2 p - p \log_2 (1-p)))$$
$$= N(h(p) - p\delta \log_2 ((1-p)/p))$$

(S35)

So with this encoding Alice can send almost $Nh(p)$ bits.

**Diamond norm for protocol 2**

Again we consider the simplest case of the BSC. Let $N$ be sufficiently large so that the total probability of the typical errors is $> 1 - \epsilon$, and these typical errors have weight $k$ in the range $Np(1-\delta) \leq k \leq Np(1+\delta)$. We divide up all errors of weight $k$ into $C_k$ partitions containing

$$n_k \approx \frac{\binom{N}{k}}{C_k} \approx \left(\frac{1-p}{p}\right)^{k-Np(1-\delta)}$$
errors each. Within each set the errors are all equally likely to be chosen. However, because the number of errors is unlikely to divide exactly evenly into $C_k$ sets, the probabilities $q_k$ of an error of weight $k$ will be slightly different from the probability $p_k = p^k(1 - p)^{N-k}$ of the binomial distribution. We can put a (not-very-tight) bound on this difference:

$$|q_k - p_k| < \frac{p_k}{(1-p)^{k-Np(1-\delta)}} - 1 < \frac{1-p}{1-2p} p^{2k}(1-p)^{N-2k}.$$ (S36)

Plugging this into the expression for the diamond norm, we get

$$||\mathcal{N}_p \otimes \mathcal{N} - \mathcal{N}_{\text{enc}}||_\diamond < \epsilon + \sum_{k=Np(1-\delta)+1}^{Np(1+\delta)} \binom{N}{k} |p_k - q_k| < \epsilon + \left( \frac{1-p}{1-2p} \right) \left( \frac{p}{1-p} \right)^{Np(1-\delta)} \left( \frac{1-2p+2p^2}{1-p} \right)^N,$$ (S37)

which is exponentially small in $N$.

**Error-correction for protocol 2 with a noisy channel**

Since errors can act in a complicated manner on the space of syndromes, it is not entirely clear what the optimal encoding is even for a simple channel. Here we present one encoding for the BSC that gives an achievable rate in the limit of large $N$, but it is quite likely that higher rates are possible.

In the noiseless case, it is possible to use the $C(N, M)$ strings of weight $M$ as a code—each string represents one possible weight-$M$ error. If we then apply a BSC with probability $p$, on average $Np$ bits would be flipped. If $Np \ll M$ then one can keep only a subset of the weight-$M$ strings, separated by a distance $> 2Np$.

This encoding quickly becomes inefficient as $p$ gets larger. Using the shared secret key, Alice can instead chose only a subset of the $N$ bits to hold the codewords. If this subset includes $N'$ bits, then the errors on the remaining $N - N'$ bits are irrelevant and do not need to be corrected. The limit of this would be similar to encoding 1 in the paper, where $N' \approx 2M$. 

20
Let $N' = qN$ for some $0 < q \leq 1$. The number of strings of weight $M$ is $C(qN, M)$, and there will be an average number of bit flips $pqN$ on the relevant portion of the codeword. Keep a subset of these codewords separated by distance $2pqN$. Decoding is done by finding the closest codeword to the output string.

As $N, M \to \infty$ then the number of codewords will go like

$$C(N, M, p, q) \sim \binom{qN}{M}. \binom{qN}{pqN}.$$ 

The number of bits will be $\log_2 C(N, p, q)$. 

Since $q$ is a parameter we can choose freely, we choose it to maximize the rate $R(N, M, p, q) \equiv \frac{1}{N} \log_2 C(N, M, p, q)$. Using the Stirling approximation, differentiating with respect to $q$, and setting the result equal to 0, we can solve for $q$:

$$q = \frac{M}{N} \left( \frac{2^{h(p)}}{2^{h(p)} - 1} \right).$$

We can then plug this back into the formula for $R$. If the physical channel has error rate $p$ and Alice is attempting to emulate a channel with error rate $p + \delta p$, then $M = N\delta p/(1 - 2p)$. This gives us the following expression for the rate:

$$R(p, \delta p) = -\frac{\delta p}{1 - 2p} \log_2 \left( \frac{2^{h(p)}}{2^{h(p)} - 1} \right).$$

We can compare this to the rate from encoding 1, which for the BSC is $2\delta p(1 - h(p))/(1 - 2p)$. It is not hard to see that $R(p, \delta p)$ above approaches this rate as $p \to 1/2$ (and both rates go to zero), but as $p \to 0$ this encoding does considerably better than encoding 1. It is quite likely, however, that there may be even more efficient encodings.

**Shared classical secret key vs. shared ebits**

If Alice and Bob share a secret, random key, they can use the steganographic encodings described in the paper. Shared entanglement (ebits) can act as a resource in the same way—by
measuring the two halves of a maximally entangled pair of qubits $\frac{|00\rangle + |11\rangle}{\sqrt{2}}$ Alice and Bob can generate a shared secret bit.

However, the use of ebits does open up an additional possibility beyond what can be done with a classical key. Instead of sending quantum information through the channel, Alice can instead teleport qubits to Bob. Teleportation consumes one ebit and requires the transmission of two classical bits for each qubit teleported. These classical bits can be sent through the channel steganographically. Because these bits are perfectly random, no one-time pad or twirling is needed. And because they are purely classical information, they are not disrupted if Eve chooses to measure the error syndromes, as a general quantum state would be. In this sense, quantum steganography with shared ebits is more powerful than quantum steganography with a shared classical key.

References and Notes

1. I. Chuang, M. Nielsen, *Quantum Computation and Quantum Information* (Cambridge University Press, ed. 1, 2000).

2. J. Watrous, lecture notes, Theory of Quantum Information, (2008).

3. Dorit Aharonov, Alexei Kitaev, Noam Nisan, Quantum Circuits with Mixed States, *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, 20-30, (1998).

4. Ben-Aroya *et. al.* (available at [http://arxiv.org/abs/0902.3397](http://arxiv.org/abs/0902.3397)).