Degenerate Quantum Codes for Pauli Channels

Graeme Smith\textsuperscript{1,2} and John A. Smolin\textsuperscript{2}

\textsuperscript{1}Institute for Quantum Information, Caltech 107–81, Pasadena, CA 91125, USA
\textsuperscript{2}IBM T.J. Watson Research Center, Yorktown Heights, NY 10598

(Dated: October 3, 2018)

A striking feature of quantum error correcting codes is that they can sometimes be used to correct more errors than they can uniquely identify. Such degenerate codes have long been known, but have remained poorly understood. We provide a heuristic for designing degenerate quantum codes for high noise rates, which is applied to generate codes that can be used to communicate over almost any Pauli channel at rates that are impossible for a nondegenerate code. The gap between nondegenerate and degenerate code performance is quite large, in contrast to the tiny magnitude of the only previous demonstration of this effect. We also identify a channel for which none of our codes outperform the best nondegenerate code and show that it is nevertheless quite unlike any channel for which nondegenerate codes are known to be optimal.

It was Shannon\cite{Shannon} who discovered, by a random coding argument, the beautiful fact that the capacity of a noisy channel $\mathcal{N}$ is equal to the maximal mutual information between an input variable, $X$, and its image under the action of the channel:

$$ C = \max_X I(X; \mathcal{N}(X)). \quad (1) $$

It is remarkable that this maximization is over a single input to the channel; it does not require consideration of inputs correlated over many channel uses.

One would hope that, as in the classical case, there is some measure of quantum correlations that can be maximized over inputs to a quantum channel to give the capacity. Unfortunately, this appears not to be the case. The natural generalization of Eq. (1) is to replace the random variable $X$ with a quantum state $\rho$ and the mutual information with the coherent information $I^c$ giving

$$ Q_1 = \max_\rho I^c(\mathcal{N}, \rho), \quad (2) $$

where

$$ I^c(\mathcal{N}, \rho) = I^c \left( I \otimes \mathcal{N}(|\phi^{AB}\rangle \langle \phi^{AB}|) \right). \quad (3) $$

Here $|\phi^{AB}\rangle$ is a purification of $\rho$. Its use reflects the fact that unlike in the classical case, there can be no remaining copy of the channel input with which to compare correlations—instead we must consider the quantum state as a whole. The coherent information is defined by $I^c(\rho_{AB}) = S(\rho_B) - S(\rho_{AB})$ with $S(\rho) = - \text{Tr}(\rho \log \rho)$.

While we can achieve $Q_1$ using a random code on the typical subspace of the maximizing $\rho$, it has long been known that this rate is not always optimal\cite{emerson2005,bravyi2005}. They exhibit codes with rates larger than $Q_1$ for very noisy depolarizing channels which have $Q_1$ small or even zero.

The correct quantum capacity formula is not $Q_1$, but instead is given by\cite{emerson2005,bravyi2005,bravyi2007}

$$ Q = \lim_{n \to \infty} \frac{1}{n} \max_\rho I^c \left( \mathcal{N}^{\otimes n}, \rho_n \right), \quad (4) $$

where taking the limit $n \to \infty$ means that we must consider the behavior of the channel on inputs entangled across many uses. This multi-letter formula is an expression of our ignorance about the structure of capacity achieving codes for a quantum channel.

The difference between these single- and multi-letter formulas is intimately related to the existence of degenerate quantum codes. Strictly speaking, degeneracy is not a property of a quantum code alone, but a property of a code together with a family of errors it is designed to correct. More formally, one usually says that a code $\mathcal{C}$ degenerately corrects a set of errors $\mathcal{E}$ if in addition to correcting $\mathcal{E}$, there are multiple errors in $\mathcal{E}$ that are mapped to the same error syndrome. In the context of probabilistic noise, which we will be concerned with exclusively, we say that a code $\mathcal{C}$ degenerately corrects the noise due to a channel $\mathcal{N}$ if it can be decoded with a high fidelity and furthermore multiple errors in the set of typical errors of $\mathcal{N}$ are mapped to the same error syndrome. For the most part, we will be concerned with grossly degenerate codes, which have the further property that the number of typical errors mapped to each syndrome is exponential in the code’s block-length.

For the depolarizing channel considered in\cite{emerson2005,bravyi2005}, as well as the Pauli channels considered below, $Q_1$ is exactly the maximum rate achievable with a nondegenerate code. That $Q > Q_1$ is then established by the construction of a massively degenerate code. While this was accomplished in the work of\cite{emerson2005,bravyi2005}, the difference found was over a minuscule range of noise parameters and extremely small in magnitude. As a result, one may have thought that Eq. (2) is “essentially correct”, with some minor modifications in the very noisy regime. In the decade since the appearance of these two works, there has been almost no progress on understanding the difference between the single- and multi-letter expressions above, a failure which has to some extent been tempered by the hope that the smallness of the effect would make it amenable to a perturbative analysis. We will show that this cannot be the case and in fact that the smallness of the effect found in
is more accidental than fundamental.

Until now, very little has been understood about why
the degenerate codes of do work, besides that they
seem to be highly degenerate. The main contribution
of this paper is to provide a conceptual explanation of why
degenerate codes of this type work, along with a related
heuristic for designing codes for more general channels.
Using this heuristic, we find better codes for almost all
Pauli channels, and exhibit cases where the effect of de-
generacy can be quite large. This large gap between the
performance of nondegenerate and degenerate codes im-
plies that a perturbative approach to is unlikely to be
useful.

A secondary contribution we believe to be no less im-
portant, but which lies on the periphery of the current
work, is the identification of the two Pauli channel as an
important piece of the degenerate coding puzzle. This
channel derives no benefit from the degenerate codes we
study, but is also quite different from any of the degrada-
dable channels, a set of channels including the dephasing
and erasure channels, and comprising the only channels
for which which nondegenerate codes are known to be op-
timal . Therefore either there is some other sort of de-
generate code that will beat Q1 or Q1 can be optimal for
nondegradable channels. Either outcome seems plausi-
ble, and progress on resolving this dichotomy would nec-
ecessarily deepen our understanding of the quantum coding
problem in general. Along a similar line, we have intro-
duced a general method for showing that a channel is not
degradable, taking one of the first steps in the program
to classify all degradable channels.

Cat Codes for Pauli Channels.—The codes we will con-
sider are m-qubit repetition code, sometimes called a
“cat codes” because the code space is spanned by |0⟩⊗m
and |1⟩⊗m. These have stabilizers Ζ1Ζ2, Ζ1Ζ3, . . . , Ζ1Ζm
and logical operators X = X⊗m and Z = Z1, so
that an error of the form XZν leads to syndrome
{ u1 ⊕ u2, . . . , u1 ⊕ um } and in the absence of a recovery
operation gives a logical error of XνZνξ. By encoding half of [φ+]AB = ⟨(00) + (11)⟩/ √2 in our repetition code,
we get the state for which the coherent information in Eq. [4]
will be more than m times Q1. Sending the B
system of the resulting |φ⁺⟩AB through N′⊗m and sub-
sequently measuring the stabilizers { Z1,Zi }i=2..m
leads to the state ρABm = ∑r∈{0,1}m−1 Pr(r)I⊗N′t[|φ⁺⟩⟩⊗|r⟩⟩],
where r is the syndrome measured, N′t is the induced
channel given r (which is also a Pauli channel), and Pr(r)
is the probability of measuring r. Concatenating our rep-
etion code with a random stabilizer code allows com-
munication with high fidelity at a rate of

\[ \frac{1}{m} I(\rho_{ABm}) = \frac{1}{m} \sum_r Pr(r) I(\rho_{AB,m}(\rho_{AB,m})^{φ⁺}) \tag{5} \]

Because the repetition code is highly symmetric we can find explicit formulas for both Pr(r) and N′t, and
thus a fairly compact expression for I(ρABm). The joint probabilities of logical errors and syndrome outcomes are

Pr( XνZν, r ) = \( \frac{1}{2} \left( (px+p_y)^{u(m−2r)+s} (1−px−p_y)^{(1−u)(m−2r)+s} + (−1)^s(px−p_y)^{u(m−2r)+s} (1−px−p_y−2p_z)^{(1−u)(m−2r)+s} \right) \),

where r = |r|, the Hamming weight of r. Eq. (6) allows us to find both Pr(r) and the error probabilities of the induced channels N′t. This formula depends on r but has
no other dependence on r, which dramatically reduces the
number of induced channels that need to be considered.

By evaluating (5) on the probabilities of (6), we find
that for almost all Pauli channels there is a repetition
code with nonzero rate at the hashing point. When px ≥ p_z
the best code is in the Z basis with length scaling
like 1/p_z, which we’ll study in detail in the next section.
For px ≥ p_z ≥ p_y it is a good rule of thumb to use a
Z repetition code of length m ≈ 1/p_z, with the largest
increase in rate for fairly asymmetrical channels (Fig. 1).

Repetition Lengths for Almost Bitflip Channels.—To il-
\[ \text{Pr}(X^uZ^v, r) = \frac{1}{2} \left( (p_x+p_y)^u(m-2r)+s (1-p_x-p_y)^{(1-u)(m-2r)+s} + (-1)^s(p_x-p_y)^u(m-2r)+s (1-p_x-p_y-2p_z)^{(1-u)(m-2r)+s} \right), \]
concentrated near \( r_o \equiv (m-1)q_2 \) and \( r_1 \equiv (m-1)(1-q_2) \).

This is because there are typically \((m-1)q_2\) X errors on qubits 2 through m and these qubits all get flipped if qubit 1 has an X error. So, the measured value of \( r \) tells us whether or not a logical X error has occurred, at least with high probability. One can see from this, together with the \( q_2 \) above, that as \( m \) increases we learn more about the logical X error at the expense of knowing less about the logical Z.

Indeed, the optimal repetition length will minimize the entropy in the logical qubits conditioned on \( r \), which near the hashing point occurs when the repetition length is around \( 1/q_2 \), at which point almost all of the X entropy has been removed. If we increase \( m \) beyond this the gain in information about the logical X is less than the resulting decrease in our knowledge of the logical Z’s. The overall rate thus achieved at the hashing point is \( 2q_2 \ln(1/q_2)/\ln(1/q_z) \).

Note that essentially all of the entropy in the X errors is removed by the best code, with the optimal length determined by a tradeoff between the reduction of entropy in the X errors and the increase of entropy in the Z errors. This sort of tradeoff also determines the optimal repetition code length for a general Pauli channel.

**Concatenated Repetition Codes.**—We can immediately apply this analysis to design even better codes by using concatenation. By adapting a second level of repetition code to the error probabilities of the channels induced by the first level we can exceed the performance of any single level cat code. We have used this approach for the depolarizing channel with the results shown in Fig. 2 where we plot the probabilities at which the rate of a concatenated 3 in \( m \) and 5 in \( m \) code goes to zero as a function of \( m \), the size of the outer cat code. If we first use a 3-cat code in the Z basis, followed by an \( m \)-cat code in the X basis, we find the highest threshold for a 3 in 19 code, with a nonzero rate up to \( p \approx 0.19086 \), surpassing the codes of \[3\]. Starting with a 5-cat code the threshold increases up to \( p \approx 0.19088 \) for \( m = 16 \), the best known code for this channel, but for higher values of \( m \) the computation of this probability is quite slow. Based on the character of the channels induced by the inner repetition code, together with the behavior for \( m \leq 16 \) we expect that the threshold increases until something like 5 in 25, at which point a larger \( m \) begins to reduce the effectiveness of the code.

**Two-Pauli Channels are Special.**—Besides the one-Pauli channels, the only channels for which we can find no code offering an advantage near the hashing point are tightly concentrated near \( N_p^o(p) \equiv (1-p)\rho + \frac{1}{2}X\rho X + \frac{1}{2}Z\rho Z \). While hashing is optimal for one-Pauli channels \[2\], \( N_p^o \) is not known to have additive coherent information, which is equivalent to the optimality of hashing. Furthermore, we will show that unlike all channels known to be additive this channel is not degradable \[3\].

Every channel \( \mathcal{N} \) can be expressed as an isometry followed by a partial trace, which is to say there is an isometry \( U_N : A \to BE \) such that \( \mathcal{N}(\rho) = \text{Tr}_E U_N \rho U_N^\dagger \). The complementary channel of \( \mathcal{N} \), called \( \mathcal{N}^C \), results by tracing out system \( B \) rather than \( E \): \( \mathcal{N}^C(\rho) = \text{Tr}_B U_N \rho U_N^\dagger \). A channel is called degradable if there is a completely positive map, \( \mathcal{D} : B \to E \), which “degrades” \( \mathcal{N} \) to \( \mathcal{N}^C \), so that \( \mathcal{D} \circ \mathcal{N} = \mathcal{N}^C \). The existence of such a map immediately implies the additivity of \( I^E[\mathcal{N}] \), which can be seen by noting that \( I^E(\mathcal{N}^o(n_1+n_2), \rho_{n_1 n_2}) \leq I^E(\mathcal{N}^o_{n_1}, \rho_{n_1}) + I^E(\mathcal{N}^o_{n_2}, \rho_{n_2}) \) ex-
must be a CPTP map and recalling that $I(E_{n_1};E_{n_2}) \leq I(B_{n_1};B_{n_2})$ and that $I(B_{n_1};B_{n_2})$ cannot increase under local operations. We now show there is no such $D$ for $N^{\text{pp}}_p$ when $0 < p < 1$.

Letting $N^{\text{pp}}_p(i|j) = \sum_{kl} N_{ik;jl}|k\rangle\langle l|$ define $N$ and $N^{\text{pc}}_p(i|j) = \sum_{kl} N^{\text{pc}}_{ik;jl}|k\rangle\langle l|$ define $N^{\text{pc}}$, we find

$$N = \begin{pmatrix}
1-p/2 & 0 & 0 & p/2 \\
0 & 1-3p/2 & p/2 & 0 \\
p/2 & 0 & 1-3p/2 & 0 \\
p/2 & 0 & 0 & 1-p/2
\end{pmatrix}$$

and

$$N^{\text{pc}} = \begin{pmatrix}
p/2 & 0 & 0 & p/2 & \alpha & 0 & \alpha & 1-p \\
0 & -p/2 & \alpha & p/2 & 0 & 0 & 0 & 0 \\
p/2 & 0 & -p/2 & 0 & 0 & 0 & 0 & 0 \\
p/2 & 0 & 0 & p/2 & -\alpha & 0 & -\alpha & 1-p
\end{pmatrix},$$

where $\alpha = \sqrt{p(1-p)/2}$. If $N^{\text{pp}}_p$ is degradable, there must be a CPTP map $D$ such that $D \circ N = N^{\text{pc}}$, which is equivalent to $ND = N^{\text{pc}}$, with $D$ defined by $D(|s\rangle\langle t|) = \sum_{uv} D_{st;uv}|u\rangle\langle v|$. For $N$ and $N^{\text{pc}}$ as above, this gives

$$D = \begin{pmatrix}
p/2 & 0 & 0 & 0 & \beta & 0 & \beta & 1-p \\
0 & -\gamma & \beta & \gamma & 0 & 0 & 0 & 0 \\
0 & \gamma & -\beta & \gamma & 0 & 0 & 0 & 0 \\
p/2 & 0 & 0 & 0 & \beta & -\beta & 0 & 1-p
\end{pmatrix},$$

with $\beta = \sqrt{p/(2-2p)}$ and $\gamma = p/(2-4p)$. The Choi matrix $[3]$ of $D$, $C^{\text{pp}}_{ik;jl} = D_{ij;kl}$, is thus

$$C^{\text{pp}} = \begin{pmatrix}
p/2 & 0 & 0 & -\gamma \\
p/2 & 0 & \beta & 0 \\
0 & -p/2 & \gamma & 0 \\
0 & 0 & \beta & 1-p \\
0 & \gamma & 0 & \beta \\
0 & -\gamma & 0 & 0 \\
0 & 0 & 0 & p/2 \\
-\gamma & 0 & 0 & 0 \\
\beta & 0 & 0 & -\beta \\
\beta & 0 & 0 & -\beta & 1-p
\end{pmatrix},$$

which contains the subblock $\left(\begin{smallmatrix} p/2 \\ -\gamma \\ p/2 \end{smallmatrix}\right)$. This has a negative eigenvalue for all $0 < p < 1$, so that $C^{\text{pp}}$ cannot be nonnegative and thus $D$ is not CP.

Besides repetition codes, we have explored concatenated repetition codes for $N^{\text{pp}}_p$, all of which performed worse than the hashing rate of $1-H(p)-p$. This suggests the capacity of $N^{\text{pp}}_p$ is exactly $1-H(p)-p$, and in light of its nondegradability we hope a proof of this conjecture will point towards a new sufficient criterion for the additivity of coherent information.

Discussion.—It is tempting to ask if there is a simpler characterization of the quantum channel capacity than is provided by Eq. [4]. In particular, contrary to what is sometimes claimed, the results of [2, 3] and this work do not rule out a single letter formula for the capacity—what is ruled out is the possibility that the single letter optimized coherent information is the correct formula. It could be that there is a single letter formula for the capacity, or less ambitiously simply an efficiently calculable expression, which takes degeneracy into account. The characterization of capacity in terms of coherent information is fundamentally nondegenerate, and it may be this which leads to the necessity of regularization, rather than an inherent superadditivity of quantum information.

More concretely, the two-Pauli channel with equal probabilities seems to be somewhat different from other Pauli channels. Given their success with almost all other Pauli channels, the failure of cat codes to beat $Q_1$ in this case suggests that hashing is optimal. Resolving this conjecture seems to be a manageable problem whose solution may lead to a better understanding of additivity questions for quantum channels in general.

The ideas explored here are also useful for quantum key distribution. In particular, using highly structured codes for information reconciliation improves the noise threshold of BB84 with one-way classical post-processing from 12.4% to 12.9% [10].

Finally, we hope the coding approach suggested by the almost bitflip channel will lead to codes with rates beyond what we have presented here. Focusing on reducing the amplitude error rate with an inner code while trying to avoid scrambling the phase errors more than necessary and following this up with a random stabilizer code (or perhaps a second similarly chosen code reversing the roles of amplitude and phase) offers an appealing heuristic for code design. Viewed in this way, the inner codes we have considered are quite primitive—a repetition code is the simplest code there is—and it seems likely more sophisticated codes will perform better.

In summary, we have provided a toolset for studying degenerate codes on Pauli channels. We have demonstrated channels and codes for which the gap between the degenerate and nondegenerate performance is quite large compared to previous results, and improved the threshold for the more generally applicable depolarizing channel. Whether the capacity of the two-Pauli channel can be improved by degenerate codes remains an open question, the solution to which will likely prove illuminating.

We are grateful to DP DiVincenzo, D Leung, and MB Ruskai for valuable discussions. GS thanks NSF PHY-0456720 and NSERC of Canada. JAS thanks ARO contract DAAD19-01-C-0056.

[1] C. Shannon, Bell Syst. Tech. J. 27, 379 (1948).
[2] P. Shor and J. Smolin, lecture notes, MSRI Workshop on Quantum Computation, 2002. [http://www.msri.org/publications/ln/msri/2002/quantumcrypto/shor/1/].
[3] S. Lloyd, Phys. Rev. A 55, 1613 (1997).
[4] E. Rains, Phys. Rev. A 60, 179 (1999).
[5] I. Devetak and P. W. Shor, arXiv quant-ph/0311131.
[6] M. Choi, Linear Algebra and Appl. 10, 285 (1975).
[7] G. Smith, J. M. Renes, and J. A. Smolin, arXiv quant-ph/0607018.