Quantum-channel capacity of very noisy channels

David P. DiVincenzo,[1] Peter W. Shor,[2] and John A. Smolin[1]
[1]IBM T. J. Watson Research Center, Yorktown Heights, New York 10598
[2]AT&T Research Center, Murray Hill, New Jersey 07974

(Received 30 June 1997)

We present a family of additive quantum error-correcting codes whose capacities exceed those of quantum random coding (hashing) for very noisy channels. These codes provide nonzero capacity in a depolarizing channel for fidelity parameters f when $f>0.809\,44$. Random coding has nonzero capacity only for $f>0.810\,71$; by analogy to the classical Shannon coding limit, this value had previously been conjectured to be a lower bound. We use the method introduced by Shor and Smolin of concatenating a nonrandom repetition (cat) code within a random code to obtain good codes. The cat code with block size five is shown to be optimal for single concatenation. The best known multiple-concatenated code we found has a block size of 25. We derive a general relation between the capacity attainable by these concatenation schemes and the coherent information of the inner code states. [S1050-2947(98)06901-7]

PACS number(s): 03.67.Hk, 3.65.Bz, 89.80.+h, 89.70.+c

I. INTRODUCTION

It still comes as a surprise to many physicists that the error correction techniques that we now know to exist for quantum states are basically digital and not analog. Even after the discovery of quantum error correcting codes [1], it was felt that the analog metaphor must be more appropriate; after all, a quantum state is specified by a continuous set of complex numbers, and the fundamental physical process which we consider as ''noise'' on the quantum state, unitary transformations involving the state and its environment, also are drawn from a continuous and not a discrete set. Nevertheless, the entangled structure of quantum states, which has no analogy in classical mechanics, permits an essentially digital treatment of errors.

In fact, quantum error correction as we presently understand it [1–18] is required to be oblivious to the continuous nature of the quantum state: error correction is accomplished by using a coded subspace such that the effect of the errors and error-correction scheme are both independent of the direction of the state vector in that subspace. Furthermore, these error-correcting actions are not continuous, but are drawn from a discrete set. This is related to the fact that the continuous action of the environment can also be ''digitized,'' in the following sense: quite generally, noise on a quantum state can be described as a transformation which takes a pure quantum state Ψ to a mixed state ρ given as an ensemble of pure states $\{A_i\Psi\}$. Each of the set of operators A_i can be written as a linear combination of some fixed operator basis, $A_i=\Sigma_j a_{ij}E_j$. The fixed set of operators E_j are the ''error operators'' of the quantum channel. There form a finite set with d^2 elements, where d is the dimension of the Hilbert space of Ψ. For qubits ($d=2$) it has become conventional to use the error basis $E_j=I$ (identity), σ_x, σ_y, σ_z (the Pauli matrices).

These four operators have a very ''digital'' interpretation. In the simplest memoryless noisy bitwise channel, the ''binary symmetric channel,'' each bit is either left alone or is flipped. This corresponds to the actions of the operators I and σ_x on the z-basis quantum bit (qubit) states. In the quantum case there are just two other actions; we refer to the action of σ_z as a ''phase flip,'' and that of σ_y as ''both a bit flip and a phase flip.'' An alternative point of view is to represent the quantum state as a two-bit object (see [3,13]), or an object in a four-element Galois field GF(4) [15], so that the four operators are just the four possible digital noise actions on them. The only ''analog'' features left in the description of the quantum channel are the continuous amplitudes a_{ij}; but these play a very similar role in the quantum noisy channel to the bit-flip probabilities in the classical digital channel.

In this paper we will concern ourselves with the quantum capacity of a simple qubit channel, the *depolarizing* channel. This channel is completely characterized by one fidelity parameter f; with probability f the qubit passes through the channel undisturbed ($A_1=\sqrt{f}I$), while with equal probability $g=(1-f)/3$ the qubit is subjected to a rotation by one of the three Pauli matrix operators $\sigma_{x,y,z}$ ($A_{2,3,4}=\sqrt{g}\sigma_{x,y,z}$).

Defining the quantum capacity requires a discussion of the quantum error correction codes mentioned above. This discussion will be given in detail in Sec. II. Suffice it to say now that many quantum codes $[n,k,d]$ are now known [1,3,4,7–9,11–15], in which an arbitrary state of k qubits are coded into a state of $n>k$ qubits in such a way that if no more than $t\equiv\lfloor d/2\rfloor$ of the n qubits are subjected to an error, the original k-qubit state can nevertheless be perfectly recovered. The *rate* of this code is $r=k/n$.

With this, the *quantum capacity* $Q(\chi)$ of a quantum channel χ can be defined: $Q(\chi)$ is the maximum number Q such that for any rate $R<Q$ and any $\delta>0$ there exists a quantum code $\mathcal{C}$ with rate $k/n\geqslant R$ such that after the action of χ any state ψ encoded by $\mathcal{C}$ can be recovered with fidelity at least $1-\delta$ at the receiving end of the channel [3,17,18].

Naively one might expect there to be a relationship between the achievable capacity Q for a depolarizing channel of fidelity f and the rate r of a code in which $1-t/n\approx f$, since $1-f$ is the expected fraction of qubits on which errors will occur. In fact, there is no direct relationship because the definition of capacity does not require that all errors of weight less than t are correctable, but only that the fraction of uncorrectable errors vanishes for large block sizes. For example, Rains has shown [19] that all families of codes for which $1-t/n<\frac{5}{6}$ have vanishing rate r; nevertheless, Q is known to be greater than zero for a range of $f<\frac{5}{6}$. Indeed, in our previous work on ''one-way hashing,'' [2,3] we identified a method for which a nonzero capacity could be attained down to about $f=0.81$ [20]. The capacity attained has the form of one minus a von Neumann entropy [see Eq. (11)]. This quantum expression bears a close resemblance to the result of classical information theory, where the maximum information reliably transferable though a noisy channel is limited by the Shannon bound [21]

$$C\leqslant 1-H(\chi), \tag{1}$$

where $H(\chi)$ is the average entropy introduced in a bit by classical channel χ.

In the classical problem the Shannon bound is achieved by a random coding procedure; the one-way hashing protocol which we invented is the natural quantum analog of random coding. Thus, it was natural to expect that Eq. (11) would also be the upper bound on the quantum capacity for the depolarizing channel.

However, the quantum coding problem has not proved to be exactly parallel to the classical one; recent work has identified several important properties (the pipelining inequality, and subadditivity of mutual information) [22] which are true for classical capacity measures, but are not for the quantum version. More concretely, it was recognized early that quantum error correcting codes can have a property referred to as ''degeneracy'' which is not permitted in the quantum case: two different errors may be indistinguishable by their error syndromes, but may nevertheless be both correctable (see Sec. VI). Spurred by intuitive ideas of how this degeneracy might improve the capacity of the quantum channel, Shor and Smolin [14] explored some *nonrandom* coding strategies, and found a range of depolarizing channels (very noisy ones) for which the obvious analog of the Shannon bound is violated; a higher capacity is attained than for random codes.

The main point of this paper is to present the Shor-Smolin discovery using the more modern and streamlined tools for describing quantum coding which have been developed recently. We will formulate the capacity calculation in terms of the orthogonal-group formalism [12,13] which has proved very successful in systematizing almost all known quantum codes (the additive codes) [23–25]. We identify new quantum weight enumerators (see [16]), for the stabilizer *cosets*, with which a compact expression for the capacity can be given. We show that the Shor-Smolin codes can be understood in the language of code *concatenation* which has been very popular for the discussion of fault-tolerant quantum computation [26–30]. We establish a rather general relation between concatenated-code capacities and the quantum *coherent information* [5,6], a quantity believed to be (but not proved to be in general) the quantum capacity [31,22]. Finally, we show that, using the coset-enumerator formulation, closed-form expressions can be derived for the capacities of the original Shor-Smolin protocols, which permit us to perform a more extensive quantitative exploration of the performance of these codes. We do not yet know what the actual attainable capacity of the depolarizing channel is, but hopefully the techniques explored here may provide a clue of how to obtain this result.

This paper is organized as follows. Section II briefly reviews the orthogonal-geometry group theory which has been introduced for the classification of quantum codes, and introduces the necessary coset weight polynomials. Section III derives the average-entropy expression for concatenated codes as originally obtained by Shor and Smolin. Section IV shows that the capacity attained by the concatenation procedure is equal to the quantum coherent information. Section V presents the compact expressions which we have obtained for the concatenation using the ''cat'' code, which give capacities exceeding the random-coding bound. Section VI presents our conclusions, and some thoughts about the use of degeneracies to attain improved capacities using concatenated codes.

II. GROUP-THEORETIC CHARACTERIZATION OF CODES

We consider the group $\overline{E}$ introduced in [13] which describes all possible standard errors on n uses of the channel (described by products of Pauli-matrix operators on a set of n qubits). The bar indicates that the group is understood to be taken modulo phases ± 1, $\pm i$. The dimension of $\overline{E}$ is 2^{2n}. As [13] showed, the Abelian subgroups of E play a central role in the theory of quantum error correcting codes. Consider such an Abelian subgroup S; again, we will work only with $\overline{S}$, from which phases have been removed. We define $k<n$ by specifying that $\overline{S}$ has $n-k$ generators (and thus is of dimension 2^{n-k}); then any of the 2^k-dimensional eigenspaces (denoted $\mathcal{C}_i$, $0\leqslant i<2^k$) of this set of Pauli-matrix operators forms a quantum code. We now introduce $S^{\perp}$ and $\overline{S^{\perp}}$ where $S^{\perp}$ consists of all elements in E which commute with all the elements of S and $\overline{S^{\perp}}$ is $S^{\perp}$ modulo phases. The dimension of the set $\overline{S^{\perp}}$ is 2^{n+k}. To analyze the error-correction capability of the code, we define the weight of an operator e in E, wt(e), as the number of Pauli-matrix operators (either σ_x, σ_y, or σ_z) appearing in e. If the minimal-weight element of $\overline{S^{\perp}}\setminus\overline{S}$ (i.e., the set $\overline{S^{\perp}}$ excluding the elements in $\overline{S}$) has weight d, then the correct state can be restored after $d-1$ erasures, which means also that it can correct arbitrary errors on any $t=\lfloor d/2\rfloor$ qubits. d is referred to as the ''distance'' of the quantum code, and the notation for the code is $[n,k,d]$.

All these facts have been discussed previously for quantum codes; but for the present purposes we need to introduce some additional mathematical objects, the *cosets* of $\overline{S}$ in $\overline{E}$. To understand why these cosets might be natural objects to consider for quantum error correcting codes, we recall that the elements of $\overline{S}$ are the *stabilizers* of the code; this means that if the error suffered by the set of n qubits is any member of $\overline{S}$, then the code state is unaffected. Note that in the same way any element of the coset $\overline{s}_{\alpha}\overline{S}$ acts identically on the code; all such elements act as if just the error $\overline{s}_{\alpha}$ had occurred. For this reason, these cosets will play a central role in the analysis below. Note that because $\overline{E}$ is Abelian (N.B. E

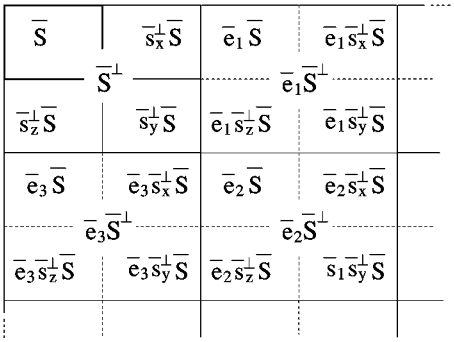
$\overline{S}$
$\overline{s_x^\perp}\,\overline{S}$
$\overline{e_1}\,\overline{S}$
$\overline{e_1}\,\overline{s_x^\perp}\,\overline{S}$
$\overline{S}^\perp$
$\overline{e_1}\overline{S}^\perp$
$\overline{s_z^\perp}\,\overline{S}$
$\overline{s_y^\perp}\,\overline{S}$
$\overline{e_1}\,\overline{s_z^\perp}\,\overline{S}$
$\overline{e_1}\,\overline{s_y^\perp}\,\overline{S}$
$\overline{e_3}\,\overline{S}$
$\overline{e_3}\,\overline{s_x^\perp}\,\overline{S}$
$\overline{e_2}\,\overline{S}$
$\overline{e_2}\,\overline{s_x^\perp}\,\overline{S}$
$\overline{e_3}\overline{S}^\perp$
$\overline{e_2}\overline{S}^\perp$
$\overline{e_3}\,\overline{s_z^\perp}\,\overline{S}$
$\overline{e_3}\,\overline{s_y^\perp}\,\overline{S}$
$\overline{e_2}\,\overline{s_z^\perp}\,\overline{S}$
$\overline{s_1}\,\overline{s_y^\perp}\,\overline{S}$

FIG. 1. Hierarchical partitioning of the set $\overline{E}$ into cosets of $\overline{S^\perp}$, and those in turn into cosets of $\overline{S}$. The case of $\overline{S^\perp}$ dividing into four cosets is special to the case of coding a single qubit.

is *not* Abelian) no distinction need be made between left and right cosets.

We will need to consider three different coset partitionings: (i) The cosets of $\overline{S^\perp}$ in $\overline{E}$. Consider the "transversal" of $\overline{E}$, the set $G=\{\overline{\alpha}\}\subset\overline{E}$ which generates the coset decomposition of $\overline{S^\perp}$ in $\overline{E}$. That is, $\overline{e_\alpha}\,\overline{S^\perp}\cup\overline{e_\beta}\,\overline{S^\perp}=\emptyset$ if $\alpha\neq\beta$, and $\cup_\alpha\overline{e_\alpha}\,\overline{S^\perp}=\overline{E}$. The dimension of G is 2^{n-k}. We indicate the αth coset of the coset decomposition as $\overline{e_\alpha}\,\overline{S^\perp}$. (ii) The cosets of $\overline{S}$ in $\overline{S^\perp}$. In this case we denote the set generating all the distinct cosets as $G^\perp=\{\overline{s_\alpha^\perp}\}\subset\overline{E}$. The dimension of this set is 2^{2k}. A typical coset is indicated as $\overline{s_\alpha^\perp}\,\overline{S}$. (iii) The cosets of $\overline{S}$ in $\overline{E}$. This is just the direct product of (i) and (ii), the generating set is $G\otimes G^\perp$, and the (α,β) coset is denoted $\overline{e_\alpha}\,\overline{s_\beta^\perp}\,\overline{S}$. This hierarchy of coset decompositions is indicated in Fig. 1 for $k=1$.

We introduce a weight-enumerator polynomial P for set C as

$$P(C)\equiv\sum_{e\in C}f^{n-\mathrm{wt}(e)}g^{\mathrm{wt}(e)},\tag{2}$$

where $C\subset\overline{E}$.

We note that this weight polynomial is directly related to the Shor-Laflamme weight enumerator [16]. Their weight function A_d is (we use the normalization choice of Rains [32])

$$A_d(\mathcal{O}_1,\mathcal{O}_2)=\sum_{\substack{e'\in\overline{E}\,|\\ \mathrm{wt}(e')=d}}\mathrm{tr}(e'\mathcal{O}_1)\mathrm{tr}(e'\mathcal{O}_2)\tag{3}$$

and their weight-enumerator polynomial is

$$A(z,\mathcal{O}_1,\mathcal{O}_2)=\sum_{d=0}^{n}A_d(\mathcal{O}_1,\mathcal{O}_2)z^d.\tag{4}$$

To relate these to Eq. (2), note that

$$A_d(e,e)=2^{2n}\delta_{\mathrm{wt}(e),d},\tag{5}$$

from which we see that

$$P(C)\equiv\sum_{e\in C}\sum_{d=0}^{n}A_d(e,e)f^{n-d}g^d=\frac{f^n}{2^{2n}}\sum_{e\in C}A\left(\frac{g}{f},e,e\right)$$
$$=\left(\frac{f}{4}\right)^n\sum_{e\in C}A\left(\frac{1-f}{3f},e,e\right).\tag{6}$$

So our new polynomials are simple functions of those which have been introduced in previous work in quantum error correcting codes.

Our weight polynomial $P(C)$ has particular significance for the depolarizing channel with fidelity f when C is the one of the various cosets which we introduced above.

$P(\overline{S})$ is the probability that the coded quantum state will leave the depolarizing channel without error. Each distinct operator $e\in\overline{S}$ is an action of the channel which has this property, and $f^{n-\mathrm{wt}(e)}g^{\mathrm{wt}(e)}$ is the probability of that action.

$P(\overline{s_\alpha^\perp}\,\overline{S})$ gives the probability that the coded state leaving the quantum channel is *detected* to have no error, but has actually been rotated inside the code eigenspace by $\overline{s_\alpha^\perp}$. This is so from the definition of $\overline{s_\alpha^\perp}$: since it commutes with all elements of S, it does not change the eigenvalues of $\overline{S}$ which are detected in the channel-decoding operation [27]; and, since by definition $\overline{s_\alpha^\perp}\notin\overline{S}$, it performs a nonidentity rotation of the coded state inside the code eigenspace. In fact, every member of the coset $\overline{s_\alpha^\perp}\,\overline{S}$ performs the same rotation.

$P(\overline{S^\perp})$ gives the probability that no error will be detected upon decoding, regardless of whether the final quantum state is correct or not.

$P(\overline{e_\alpha}\,\overline{S^\perp})$ is the probability that decoding detects error $\overline{e_\alpha}$, regardless of the rotation of the coded state.

$P(\overline{e_\beta}\,\overline{s_\alpha^\perp}\,\overline{S})$ is the probability that decoding detects error $\overline{e_\beta}$, and the coded state is rotated by $\overline{e_\beta}\,\overline{s_\alpha^\perp}$.

These will be the essential tools for developing a compact formula for the attainable capacity for code states, and establishing the identity of the coherent information with the Shor-Smolin quantum channel capacity.

III. SHOR-SMOLIN CONCATENATION PROCEDURE

In order to formulate the main result, we first review the Shor-Smolin procedure [14] for sending reliable qubit states, with a finite capacity, over a depolarizing channel. Just as in conventional channel coding, it involves an additive code specified above by $\overline{S}$. In conventional channel coding shown in Fig. 2, the additive code is used as follows: the state $|\xi\rangle$ to be transmitted (we specialize in the figure to a single-qubit state) is rotated by the encoding unitary transformation $\mathcal{E}$ into the eigenspace $\mathcal{C}_0$ of the operators in $\overline{S}$. When this state passes through the depolarizing channel, it is rotated into one of the other eigenspaces $\mathcal{C}_m$ with some probability: we will analyze this process in detail later. Then after passage through the noisy channel, the decoding transformation $\overline{D}$ places $n-k$ of the qubits (the lower $n-1$ in the figure) in a state such that, when they are measured in the standard basis, they give the eigenvalue of each of the $n-k$ generators of $\overline{S}$, that is, it determines which of the spaces $\mathcal{C}_m$ the state had been placed into by the noise [27]. So long as the errors

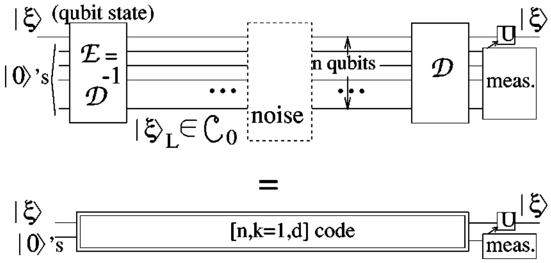

FIG. 2. Quantum-channel coding, in which (top) the state to be transmitted $|\xi\rangle$ is encoded by $\mathcal{E}$, transmitted through the noisy channel, decoded by $\mathcal{D}$ and restored by U after syndrome measurement. The entire encode-transmit-decode process can be thought of as a module (double box, below) to be used in concatenation (see Fig. 4).

produced by the channel are restricted to have weight no greater than $\lfloor d/2 \rfloor$, then a rotation U can always be determined which restores the state to its noiseless form $|\xi\rangle$.

The discussion below uses another protocol for coding shown in Fig. 3, the purification protocol of [2,3]. For the depolarizing channel the two procedures of Figs. 2 and 3 are completely equivalent. Reference [3] gives a detailed derivation of the mapping of the first protocol to the second. In the protocol of Fig. 3, the sender begins with n completely entangled states, in this example the Bell state

$$\Phi^{+}=\frac{1}{\sqrt{2}}(|00\rangle+|11\rangle). \tag{7}$$

The sender keeps one half of each of the n Bell states, and the other n particles are sent through the depolarizing channel to the receiver. When we are sending halves of EPR-Bell particles through the channel, we no longer discuss the action of the channel in terms of rotations among different code spaces (no coding transformation has yet been applied to these states); rather, the state of the system at slice $X1$ in Fig. 3 (which in general is at two different times for the two different sets of particles) is one in which the set of Bell states has been rotated to a set of some of the other Bell states with various probabilities which we will discuss shortly. The full set of Bell states is

$$\Phi^{\pm}=\frac{1}{\sqrt{2}}(|00\rangle\pm|11\rangle), \tag{8}$$

$$\Psi^{\pm}=\frac{1}{\sqrt{2}}(|01\rangle\pm|10\rangle). \tag{9}$$

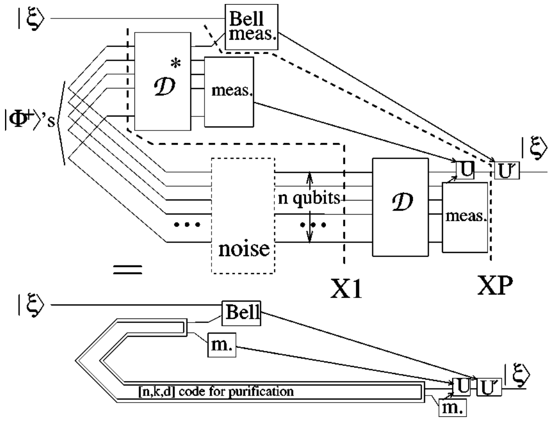

FIG. 3. A protocol for transmitting through the noisy channel equivalent to Fig. 2 which uses entanglement purification and teleportation. The sender passes halves of Bell states (Φ^{+}) through the channel to the receiver; the degraded pairs are purified (after slice $X1$) by $\mathcal{D}$ and $\mathcal{D}$* (same $\mathcal{D}$ as in Fig. 2). The purified pairs at slice XP can then be used to transmit the state $|\xi\rangle$ from sender to receiver by teleportation. Below: the Bell-state distribution and processing with $\mathcal{D}$ and $\mathcal{D}$* may be used as a module (double box) for the concatenation of Fig. 5.

The probability of a particular set of n Bell states at slice $X1$ is determined by the rule that the Bell state remains a Φ^{+} with probability f, and becomes one of the three other states Φ^{-}, $\Psi^{\pm}$ with probability $g=(1-f)/3$.

Using the decoding transformations $\mathcal{D}$ and $\mathcal{D}$* in Fig. 3, followed by measurements on both ends, classical communication from the sender to the receiver, and the final unitary transformation U, the sender and receiver can come into possession of a ''purified'' Φ^{+} pair, which is then used to send the qubit state $|\xi\rangle$ by teleportation [33] (for details see [3]).

The two methods of employing the channel shown in Figs. 2 and 3 are completely equivalent. But it will be useful to use both points of view for explaining the generalized channel transmission protocol of Shor and Smolin [14], and we will continue our review using both languages.

We will need to apply our capacity definition of Sec. I to the purification picture. The fidelity $F^{\mathcal{D}}$ of the depolarizing channel output can be most simply defined in this picture in the following way: at the end of purification (slice XP in Fig. 3) the output is desired to be a collection of k Φ^{+} states; if the code scheme is a successful one, then the overlap between the actual state at this slice ρ_{XP} and the desired Bell state will be high; thus the fidelity for an encoding $\mathcal{D}$ is

$$F^{\mathcal{D}}=\langle(\Phi^{+})^{k}|\rho_{XP}|(\Phi^{+})^{k}\rangle. \tag{10}$$

The capacity Q is simply the best rate k/n for a $\mathcal{D}$ for which this fidelity approaches unity, since each high-fidelity EPR pair can be used to teleport one qubit.

The maximization of Q has proved to be difficult. But a variety of code families have been introduced for which finite Q's are known, establishing useful lower bounds on the attainable capacity. One of the most useful is the sequence of *random additive codes*, referred to in the original papers [2,3] as ''one-way hashing.'' As the name suggests, these sequences are built by selecting, at random, an Abelian subgroup S from the group of all Pauli matrices E for successively larger block sizes n. Bennett *et al.* [3] show that almost all such sequences attain the ''hashing capacity''

$$Q_{H}\equiv\lim_{n\to\infty}\frac{k}{n}=1-S_{W}(f). \tag{11}$$

Thus, $Q_{H}(f)$ is a lower bound on the attainable capacity. $S_{W}(f)$ is the von Neumann entropy of one Bell state after one of its particles has been passed through the depolarizing channel, and it is given by

$$S_{W}(f)=-f\log f-3g\log g. \tag{12}$$

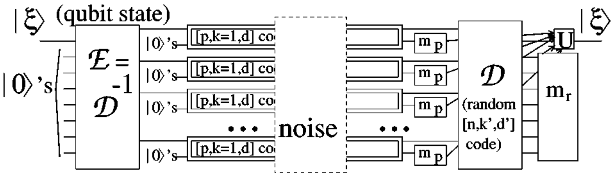

FIG. 4. Concatented coding for channel transmission. The inner code (double box) is the encode-transmit-decode module of Fig. 2. In the Shor-Smolin procedure the outer part is a random code.

Here is a brief explanation of why one-way hashing achieves the capacity of Eq. (11). The entropy of the mixture of Bell states at slice $X1$ is just $nS_W(f)$. The decoding can be simply thought of as a sequence of measurements of the $n-k$ operators which are the generators of $\overline{S}$. Each of these measurements has two outcomes, splitting the set of possible remaining states in two; thus, it has the potential for reducing the entropy of the state by one bit. Reference [3] provides arguments for why, for almost all choices of $\overline{S}$ and for large n, each measurement in fact succeeds in extracting one bit of entropy. The total state remains a mixture of Bell states, so that if k is chosen so that the entropy is reduced to zero, i.e., if $nS(W)-(n-k)=0$, then the Bell mixture becomes a pure state, which is to say that the final state is one particular set of known Bell states, which can always be rotated with U to become a set of Φ^+ states. Thus, purification has succeeded, and the ratio k/n attains the value given in Eq. (11).

This result naturally raises the question of whether there exist any *nonrandomly* chosen sequences of codes which could attain a capacity exceeding Eq. (11). While appeal to analogous classical results and other thinking suggested that random coding would be optimal, the Shor-Smolin construction which we now review shows that higher capacities are attainable. Their construction involves what is known as concatenation; it is illustrated, for both versions of the quantum coding protocols, in Figs. 4 and 5. In the language of Fig. 4, the idea is that instead of sending the qubits as encoded by the random encoder $\mathcal{E}$ directly into the channel, they are encoded once again in another additive code $[p,k,d]$, and it is these $n\times p$ qubits that are finally sent through the channel. The codes whose capacity we will consider involve $n\rightarrow\infty$, but fixed p. While we tend to associate ''good'' (i.e., high-capacity) codes with large distance d, we will find that the desirable inner $[p,k,d]$ codes actually have a *small* distance d. As we discuss at the end, it may be the ''degeneracy'' of this code which is relevant.

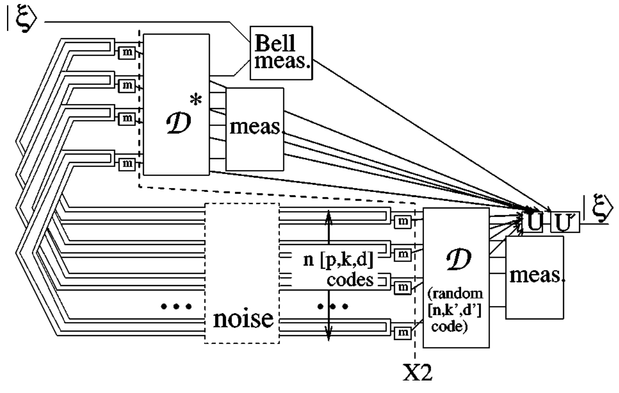

FIG. 5. Same as Fig. 4, but using the purification-teleportation protocol.

Shor and Smolin showed [14] that the following capacity is attainable by this concatenated scheme:

$$Q_{SS}=\frac{1}{p}(1-S_{X2}). \tag{13}$$

The $1/p$ just comes from the fact that the whole scheme requires $p\times n$ bits rather than just n bits to be sent through the channel. S_{X2} is the average entropy of each bipartite state at slice $X2$ in Fig. 5 (the total entropy at slice $X2$ is nS_{X2}). Shor and Smolin noted that this entropy is *not* given by the von Neumann entropy of the quantum state at this slice, because of the presence of the results of the classical measurements. Rather it is the *average* of the von Neumann entropies of the quantum states conditional on the measurement outcomes:

$$S_{X2}=\sum_{\substack{i\in\text{meas}\\ \text{outcomes}}}\Pr(i)S(\rho|i)=\sum_{\substack{i\in\text{meas}\\ \text{outcomes}}}\Pr(i)h_4(\{\Pr(B_j|i)\}). \tag{14}$$

It is this entropy that is to be reduced to zero by the random-hashing stage of the decoding. In the second part of Eq. (14) we have specialized to the case where the inner code has $k=1$ (and thus produces just one-qubit-pair state in Fig. 5). In this case the output is a mixture of the four Bell states $\{B_j\}=\Phi^\pm,\ \Psi^\pm$, so that the entropy just involves the probability of Bell state B_j conditional on the particular measurement outcome i:

$$\Pr(B_j|i)=\frac{\Pr(B_j,i)}{\Pr(i)},\qquad \Pr(i)=\sum_{j=1}^{4}\Pr(B_j,i). \tag{15}$$

The h_4 function in Eq. (14) on the set $\{x_i\}$ is defined by

$$h_n(\{x_j\})\equiv-\sum_{j=1}^{n}x_j\log_2 x_j,\qquad \sum_{j=1}^{n}x_j=1. \tag{16}$$

By using the elementary algebraic properties of the h_n function S_{X2} may be simplified so that Q_{SS} is expressed as

$$Q_{SS}=\frac{1}{p}[1+h_N(\{\Pr(i)\})-h_{4N}(\{\Pr(B_j,i)\})]. \tag{17}$$

Here N is the number of distinct measurement outcomes; for an additive $[p,k=1,d]$ code, $N=2^{p-k}=2^{p-1}$.

The probabilities appearing in Eq. (17) have appeared above; in fact they are equal to

$$\Pr(i)=P(\overline{e_iS^\perp}), \tag{18}$$

$$\Pr(B_j,i)=P(\overline{s_j^\perp}\,\overline{e_iS}). \tag{19}$$

Equation (18) follows from the fact that the members of the set $\overline{S^\perp}$ are, by definition, those errors which all lead to the measurement which indicates the ''no-error'' condition; thus, its cosets in $\overline{E}$, $\overline{e_iS^\perp}$, each contain the errors which all lead to the same measurement i. Finally, the weight polynomials are, as discussed above, constructed so as to enumerate

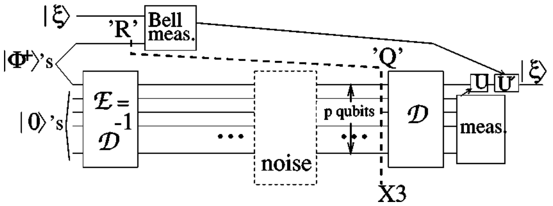

FIG. 6. The channel protocol as considered by [6] in its treatment of coherent information. The Q subsystem in the one transmitted through the channel, while the single qubit R remains behind.

properly the probabilities of these sets. Equation (19) follows similarly: The set $\overline{S}$ indicates those errors which lead to the no-error measurement *and* leave the Bell state B_i in the correct Φ^+ state. Furthermore, the coset $\overline{s_j^\perp}\,\overline{e_i S}$ contains those errors which lead to measurement i and Bell state B_j. It should be noted that the error operations $\overline{s_j^\perp}$ have the effect of performing a unitary operation of the coded qubit; the four operations are (i) $\overline{s_0^\perp}=I$ (the identity), which leaves the Bell state Φ^+ unaffected, (ii) $\overline{s_x^\perp}$, which performs a coded σ_x, leading to a final Bell state $B_x=\Psi^+$, (iii) $\overline{s_y^\perp}$ which performs σ_y and leads to $B_y=\Psi^-$, and (iv) $\overline{s_z^\perp}$ which performs σ_z and leads to $B_z=\Phi^-$. So, the weight polynomial in Eq. (19) is constructed to evaluate the probability that a member of the coset occurs.

Finally we may rewrite the capacity equation as

$$Q_{SS}=\frac{1}{p}[1+h_N(\{P(\overline{e_i S^\perp})\})-h_{4N}(\{P(\overline{s_j^\perp}\,\overline{e_i S})\})]. \tag{20}$$

IV. RELATION OF Q_{SS} TO QUANTUM COHERENT INFORMATION

The two noisy-channel transmission constructions which we have discussed above are equivalent to yet a third one shown in Fig. 6, which has been extensively discussed in the literature [5,6,22]. The rationale of introducing the one-qubit ancillary system R is that it is the minimum-size ancilla required to ‘‘purify’’ the input of the channel, that is, to make it part of a larger pure state [34] (this is a different sense of the word ‘‘purification’’ than used in [2]). In this scenario there is an important information-theoretic measure, the *coherent information*; at slice $X3$ this is given by the difference of two von Neumann entropies:

$$I_e\equiv\frac{1}{p}[S(\rho_Q)-S(\rho_{RQ})]. \tag{21}$$

Refs. [22,35] show that I_e provides an upper bound for the quantum channel capacity when maximized over all possible input-state ensembles and quantum codes. What we will show is that the achievable Shor-Smolin capacity Q_{SS} in fact *attains* the coherent information for the same additive quantum code, and for the input as in Fig. 6. To establish this we need to show the following two equalities:

$$S(\rho_Q)=1+h_N(\{P(\overline{e_i S^\perp})\}), \tag{22}$$

and

$$S(\rho_{RQ})=h_{4N}(\{P(\overline{s_j^\perp}\,\overline{e_i S})\}). \tag{23}$$

Establishing these just requires a consideration of how the noise acts on the input state in Fig. 6. For Eq. (22), we note that the density matrix ρ_Q *before* the action of the noise is just an equal mixture of the $|0\rangle_{L0}$ and $|1\rangle_{L0}$ states, where the subscript 0 indicates that these vectors lie in the eigenspace $\mathcal{C}_0$. Each eigenspace $\mathcal{C}_i$, $0\leqslant i<2^{n-k}$, is spanned by a pair of vectors $|0\rangle_{Li}$, $|1\rangle_{Li}$, where we can define the 0 and 1 vectors by

$$|0\rangle_{Li}=\overline{e_i}|0\rangle_{L0},\quad |1\rangle_{Li}=\overline{e_i}|1\rangle_{L0}, \tag{24}$$

where $\overline{e_i}$ is the coset-generating operator (see Fig. 1). The importance of the basis $|0,1\rangle_{Li}$ is that the density operator ρ_Q *after* the action of the depolarizing noise is diagonal in it. The diagonal matrix elements (i.e., the probabilities) for each vector is evaluated by noting that the state $|0\rangle_{Li}$ is reached in four possible ways: (i) the initial state is $|0\rangle_{L0}$ (with probability $\frac{1}{2}$) and an operator of the coset $\overline{e_i S}$ is applied by the channel, (ii) the initial state is $|0\rangle_{L0}$ and an operator of the coset $\overline{e_i s_z^\perp}\,\overline{S}$ is applied by the channel, (iii) the initial state is $|1\rangle_{L0}$ (also with probability $\frac{1}{2}$) and an operator of the coset $\overline{e_i s_x^\perp}\,\overline{S}$ is applied by the channel, or (iv) the initial state is $|1\rangle_{L0}$ and an operator of the coset $\overline{e_i s_y^\perp}\,\overline{S}$ is applied by the channel. Each of these is given by the appropriate weight polynomial, so

$$\langle 0|\rho_Q|0\rangle_{Li}=\frac{1}{2}P(\overline{e_i S})+\frac{1}{2}P(\overline{e_i s_z^\perp}\,\overline{S})+\frac{1}{2}P(\overline{e_i s_x^\perp}\,\overline{S})$$

$$+\frac{1}{2}P(\overline{e_i s_y^\perp}\,\overline{S}) \tag{25}$$

$$=\frac{1}{2}P(\overline{e_i S^\perp}). \tag{26}$$

The enumeration of the ways that the state $|1\rangle_{Li}$ can be arrived at is identical, with 0's and 1's interchanged; so we find that this matrix element is identical:

$$\langle 1|\rho_Q|1\rangle_{Li}=\langle 0|\rho_Q|0\rangle_{Li}. \tag{27}$$

Because it is diagonal, the von Neumann entropy of ρ_Q is just the ordinary entropy of the probability distribution

$$S(\rho_Q)=h_{2N}\left(\left\{\frac{1}{2}P(\overline{e_i S^\perp}),\frac{1}{2}P(\overline{e_i S^\perp})\right\}\right)$$

$$=1+h_N(\{P(\overline{e_i S^\perp})\}). \tag{28}$$

And thus Eq. (22) is established. The reasoning needed to establish Eq. (23) is very similar: the joint state of systems R and Q after encoding but before the noise is

$$\frac{1}{2}|0_R\rangle|0_Q\rangle_{L0}+\frac{1}{2}|1_R\rangle|1_Q\rangle_{L0}\equiv\Phi_0^+. \tag{29}$$

In this notation the i in Φ_i^+ means that the state in the Q subsystem lies in the $\mathcal{C}_i$ eigenspace. After the noise the density matrix ρ_{RQ} is diagonal in this generalized Bell basis, with the probability of the state being B_{ji} given by

$$\langle B_{ji}|\rho_{RQ}|B_{ji}\rangle = P(\overline{e_i s_j^{\perp} \overline{S}}), \tag{30}$$

since it is again only members of a particular coset that will produce a final B_{ji} state. (This discussion can equivalently be given in terms of the behavior of the $[p+1,k=0,d]$ code to which the composite system belongs.) From Eq. (30), the desired result Eq. (23) follows immediately, so the identity between the Shor-Smolin capacity and the coherent information is established for any code.

V. Q_{SS} FOR THE CAT CODE

A. Closed-form evaluation

It has not proved easy to evaluate the Shor-Smolin capacity Eq. (20) (or the equivalent coherent information) for a general concatenation. But a closed-form evaluation has proved possible for one important family of inner $[p,1,d]$ codes which we refer to as ''cat'' codes. In the cat code for $p \geqslant 2$ the stabilizer group $\overline{S}$ is generated by the operators

$$\sigma_{z1}\sigma_{z2},\ \ \sigma_{z1}\sigma_{z3},\ \ldots,\ \ \sigma_{z1}\sigma_{zp}. \tag{31}$$

For this code the code space $\mathcal{C}_0$ is spanned by

$$|0\rangle_{L0} = |\overbrace{000\ldots}^{p\text{ qubits}}\rangle \tag{32}$$

and

$$|1\rangle_{L0} = |111\cdots\rangle. \tag{33}$$

Thus, the source density matrix before passage through the channel is, using the Schumacher-Nielsen notation (Fig. 6) [6],

$$\rho_{Q(in)} = \frac{1}{2}|000\cdots\rangle\langle 000\cdots| + \frac{1}{2}|111\cdots\rangle\langle 111\cdots|. \tag{34}$$

A purification of this density matrix involving just one qubit in the subsystem R is

$$\Psi_{RQ} = \frac{1}{\sqrt{2}}(|0\,\overbrace{000\ldots}^{p\text{ qubits}}\rangle + |1\,\overbrace{111\ldots}^{p\text{ qubits}}\rangle). \tag{35}$$

Here the first qubit is the one belonging to system R. This wave function has been referred to as the cat state in the literature.

The decoding network $\mathcal{D}$ for this code is extremely simple, just consisting of the sequence of XOR gates shown in Fig. 7. Shor and Smolin provide a detailed argument [14] for counting all the probabilities in Eq. (14) by determining how each different type of error process is modified by the XOR circuit. We summarize their results here: consider counting the probabilities of the cases (including all members of one of the cosets of $\overline{S^{\perp}}$) in which the measurements give exactly r 1s, in particular when the measurements of qubits 2 through $p-r$ give zero, and qubits $p-r+1$ through p give one. It is obvious that the counting is the same for any permutation of the qubits; this means that there are $\binom{n}{p}$ equivalent cosets being counted. It is this high multiplicity that permits the calculation to be tractable, despite the fact that there are exponentially many (in p) coset weight polynomials to be evaluated.

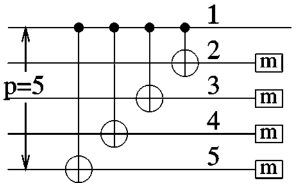

FIG. 7. Quantum network for decoding the cat code, shown for $p=5$. The same network is used for encoding.

The further four subcases (i.e., the cosets of $\overline{S}$; see Fig. 1) to be evaluated are as follows.

(i) The remaining qubits (qubit 1 of Q and the qubit of R) are in the state Ψ^+. The error processes for which this occurs are those where there are amplitude (σ_x) errors on qubits 1 through $p-r$, and an *even* number of phase (σ_z) errors on any of the qubits. We may forthwith calculate the probability of this occurrence:

$$\begin{aligned}\Pr(\Psi^+,r) &= \sum_{t\ (\text{even})}\sum_i \binom{p-r}{i}\binom{r}{t-i} g^{p-r+t-i} f^{r-t+i}\\ &= 2^{r-1} g^{p-r}(f+g)^r.\end{aligned} \tag{36}$$

Here t is the total number of phase errors and i is the number of these phase errors occurring on the qubits which already have amplitude errors (leading to a σ_y error process). The t and i sums go over the full range for which the binomial coefficients are nonzero.

(ii) The remaining state is Ψ^-. For this the error processes are those where there are amplitude (σ_x) errors on qubits 1 through $p-r$, and an *odd* number of phase (σ_z) errors on any of the qubits. In fact, it turns out that this count is exactly the same as for Ψ^+:

$$\begin{aligned}\Pr(\Psi^-,r) &= \sum_{t\ (\text{odd})}\sum_i \binom{p-r}{i}\binom{r}{t-i} g^{p-r+t-i} f^{r-t+i}\\ &= 2^{r-1} g^{p-r}(f+g)^r.\end{aligned} \tag{37}$$

(iii) The remaining state is Φ^+. In this case there must be amplitude errors on qubits $p-r+1$ to p (or no amplitude errors if $r=0$), and there must be an even number of phase errors. This gives

$$\Pr(\Phi^+,r)=\sum_{t\ (\mathrm{even})}\sum_i \binom{r}{i}\binom{p-r}{t-i} g^{r+t-i} f^{p-r-t+i}$$

$$=\begin{cases}\frac{1}{2}[(f+g)^p+(f-g)^p], & r=0,\\ 2^{r-1}g^r(f+g)^{p-r}, & r>0.\end{cases} \tag{38}$$

(iv) The remaining state is Φ^-. In this case there must be amplitude errors on qubits $p-r+1$ to p, and there must be an odd number of phase errors. The result is the same as for Φ^+ except for the $r=0$ case:

$$\Pr(\Phi^-,r)=\sum_{t\ (\mathrm{odd})}\sum_i \binom{r}{i}\binom{p-r}{t-i} g^{r+t-i} f^{p-r-t+i}$$

$$=\begin{cases}\frac{1}{2}[(f+g)^p-(f-g)^p], & r=0,\\ 2^{r-1}g^r(f+g)^{p-r}, & r>0.\end{cases} \tag{39}$$

Plugging these expressions into Eqs. (15)–(17) permits an efficient calculation of the Shor-Smolin capacity for the family of cat codes.

The threshold of the cat-code family may be computed exactly for $p\to\infty$ using an asymptotic analysis. Briefly, we find that the capacity Eq. (20) is dominated for large p by two contributions: (i) Those for the cosets of $\overline{S}$ with $r=0$ (recall that r is the number of ones in the measured syndrome). We find that this contribution goes as

$$Q_{SS}(r=0)=c\left(\frac{(f-g)^2}{f+g}\right)^p. \tag{40}$$

(ii) Those for cosets with $r\approx p/2$. This contribution has the form

$$Q_{SS}\left(r\approx\frac{p}{2}\right)=-\gamma(f)[\sqrt{8g(f+g)}]^p. \tag{41}$$

Here $\gamma(f)>0$ is a fairly complicated function of f. Nevertheless, the threshold for f specified by $Q_{SS}(f)=0$ is simply obtained by equation the bases of these two contributions:

$$\frac{(f-g)^2}{f+g}=\sqrt{8g(f+g)}. \tag{42}$$

The relevant root of this equation, $f\approx0.81808$, is the asymptotic threshold. We have not developed any simple intuitive understanding for why this threshold should remain finite as $p\to\infty$, but nevertheless remain worse than the threshold for finite p as we will now see.

B. Investigations of cat-code capacities

The simplest codes to calculate are the cat code family Eqs. (32), (33). Table I shows the results for values of p from one to fourteen. The capacities Q_{SS} of these codes near $f=0.81$ are shown in Fig. 8. We note that odd-p codes work better than nearby even-p codes; the lowest threshold fidelity in this family is achieved for $p=5$.

TABLE I. The value of the threshold fidelity f for cat codes of size p. Values of p not shown all work less well than the random coding method ($p=1$). The value for $p=\infty$ is analytic from Eq. (42).

p	f	p	f
1	0.81071	9	0.81002
2	0.81148	10	0.81028
3	0.80987	11	0.81032
4	0.81010	12	0.81056
5	0.80964 best	13	0.81062
6	0.80991	14	0.81085
7	0.80977		
8	0.81004	∞	0.81808

Generically, many other multiple-concatenation codes are possible and may lead to better thresholds. We explored the family of codes where the innermost code has a rotated cat code for which the stabilizers are

$$\sigma_{x1}\sigma_{x2},\ \ \sigma_{x1}\sigma_{x3},\ \ldots, \tag{43}$$

and the next-level code remains the ordinary cat code of Eq. (31). The best code we found was for both inner and outer cat codes having $p=5$. The capacity of this code was found to be nonzero down to a fidelity of $f\approx0.80944$, the best code known (see Fig. 9) [36]. This threshold is still far above the best known lower bound for the threshold of $f=\frac{3}{4}$ [37,3]. Unfortunately, larger codes become computationally intractable using our methods, because the number of distinct cosets scales exponentially with p. It is hoped that another approach, perhaps an approximation method for coset weights, will permit a more thorough exploration of concatenated codes.

VI. CONCLUSIONS

The obvious unanswered question which this work raises is, can any finite capacity be achieved for even noisier depolarizing channels, ones with f below the lowest value,

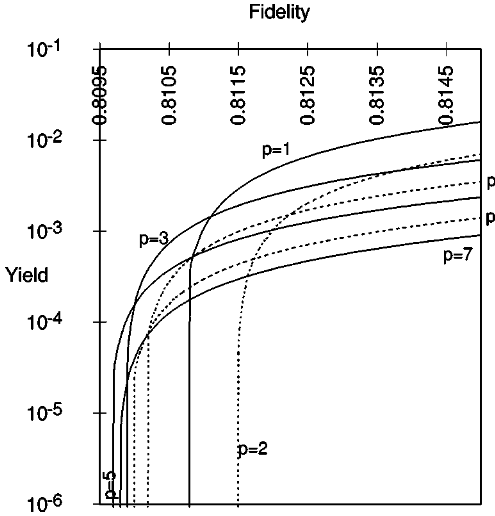

FIG. 8. The yield, i.e., capacity Q_{SS}, as a function of fidelity f for inner cat codes of size p for various values of p. Note that the curves are all in p order from $p=1$ to $p=7$ along the right side of the graph.

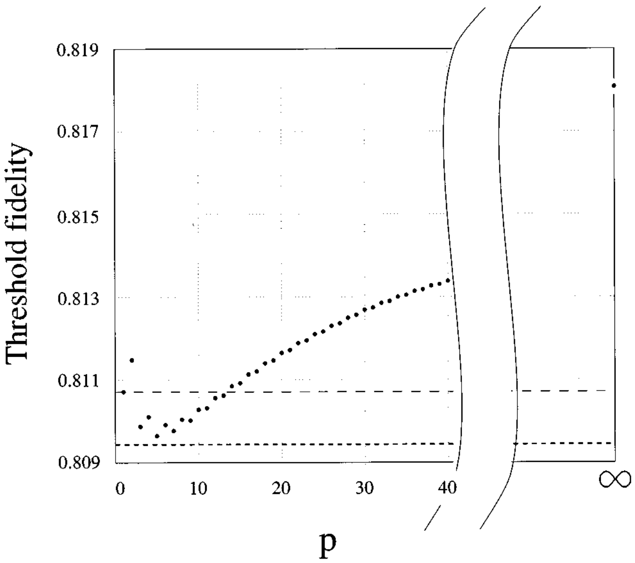

FIG. 9. The threshold value f for which $Q_{SS}=0$ vs p. Note that the points fall on two smooth curves, one for even p and one for odd p. The value at $p\rightarrow\infty$, $f\approx0.81808$, is obtained by asymptotic analysis Eq. (42). The heavy dashed line at $f\approx0.80944$ is the best known threshold achieved by the twice concatenated 25-bit scheme (Sec. V A). The light dashed line at $f\approx0.81071$ is the threshold for ordinary quantum random coding, equivalent to the $p=1$ cat code.

0.80944, achievable with the 25-bit inner code, but above the absolute minimum threshold $f=0.75$ set by the no-cloning argument [37]? In other words, do there exist even more clever nonrandom codes (recall [23]) for protecting qubits from high levels of noise?

It may be worthwhile to note here why we initially believed that the use of inner codes of the cat type was a promising direction for finding good codes for very noisy channels; this belief was based on the property of degeneracy mentioned earlier. While these motivations may end up having no more than historical interest, since they have not at present led us to any conclusive answer to the questions just posed, we hope that it might assist some reader who is interested in exploring these problems further.

Degeneracy is a property of quantum codes which has no analog for classical error correcting codes. Degeneracy arises from the fact that two different error patterns can have indistinguishable effects on a coded quantum state. This is obviously impossible for a coded binary (classical) string, but it is *obligatory* for additive quantum codes; indeed, the cosets of $\overline{S}$ introduced in Sec. II are precisely these groups of indistinguishable errors. A code is considered degenerate if some of the low-weight ($\leq\lfloor d/2\rfloor$ for a $[n,k,d]$ code) error patterns fall in the same coset of $\overline{S}$ and are therefore indistinguishable. The original 9-bit code of Shor [1] was degenerate; the 7-bit code [7–9] and the 5-bit code [11,3] are nondegenerate.

It is known [10] that a Hamming-like bound could be easily derived on the maximum attainable distance for a quantum code, *provided* that it was nondegenerate. However, in this work the possibility remained open that degenerate codes could attain a greater distance. We were thus motivated to consider highly degenerate codes for the attainment of high capacity, given the qualitative relationship between code distance and capacity. This possibility of attaining large distance using degeneracy has subsequently been rendered unlikely by a recent result of Rains [19] who has obtained a bound on d which applies for both degenerate and nondegenerate codes and which is tighter than the Hamming bound for a substantial part of the (p,k,d) parameter space. Nevertheless the fact is that the cat codes, which we have used successfully to attain high capacity, are highly degenerate: single phase errors are all indistinguishable, and all pairs of amplitude errors are indistinguishable from the no-error process. All this is true despite the fact that the cat codes have very poor distance ($d=1$ for all p).

The best we can say about why this scheme succeeded is that the high degeneracy, by making many outcomes indistinguishable, ‘‘hides’’ the large amount of entropy which the very noisy channel adds to the quantum states, thus permitting the average entropy S_{X2} to be below one over a greater range of f. This reasoning is certainly not rigorous; nevertheless, in an extensive Monte-Carlo search of other additive codes, we found no other inner code with $p\leq5$ which does a better job than the cat code for reducing the average entropy and hence attaining any higher capacity. It was further the observation that the cat code ‘‘hides’’ phase error more effectively than amplitude error that motivated us to consider a second level of concatenation, in which the innermost code was a cat code with the role of amplitude and phase reversed. Of course, this is what led us to the 25-bit code described above which give the best capacity to date.

It is clear that further generalizations of this problem await exploration. The issue of attainable capacities for channels other than the depolarizing channel is largely untouched. It is fairly clear that for the *generalized* depolarizing channel, in which the error operators are still proportional to the Pauli matrices, but with unequal probability amplitudes, the formalism developed here (i.e., the weight polynomials, and the relation to coherent information) will go through with little modification, so that would be an easy direction for further study. For the much larger space of general channels, nothing better than our random-unitary-operation (‘‘twirling’’) arguments of [3] (which bounds the capacity of any arbitrary channel by that of a corresponding generalized depolarizing channel) is presently known. Further extensions of the formalism would obviously also be desirable; a generalization of the present approach for inner codes with $k>1$ would be desirable; also, asymptotic expressions for the capacity which would not require an exact evaluation of all the coset weight polynomials could lead to significant progress. Certainly there remains much to be done to fully characterize the usefulness of the very noisy quantum channel.

ACKNOWLEDGMENTS

We thank Charles H. Bennett, Artur Ekert, Daniel Gottesman, Emmanuel Knill, John Preskill, and Eric Rains for many helpful discussions. We are grateful to the Army Research Office for support under Contract No. DAAH04-96-C-0060.

[1] P. W. Shor, Phys. Rev. A **52**, 2493 (1995).

[2] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, Phys. Rev. Lett. **76**, 722 (1996).

[3] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).

[4] E. Knill and R. Laflamme, Phys. Rev. A **55**, 900 (1997).

[5] B. Schumacher, Phys. Rev. A **54**, 2614 (1996).

[6] B. Schumacher and M. A. Nielsen, Phys. Rev. A **54**, 2629 (1996).

[7] A. R. Calderbank and P. W. Shor, Phys. Rev. A **54**, 1098 (1996); quant-ph/9512032.

[8] A. M. Steane, Proc. R. Soc. London, Ser. A **452**, 2551 (1996); quant-ph/9601029.

[9] A. M. Steane, Phys. Rev. Lett. **77**, 793 (1996).

[10] A. Ekert and C. Macchiavello, Phys. Rev. Lett. **77**, 2585 (1996).

[11] R. Laflamme, C. Miquel, J. P. Paz, and W. H. Zurek, Phys. Rev. Lett. **77**, 198 (1996); quant-ph/9602019.

[12] D. Gottesman, Phys. Rev. A **54**, 1862 (1996).

[13] A. R. Calderbank, E. M. Rains, P. W. Shor, and N. J. A. Sloane, Phys. Rev. Lett. **78**, 405 (1996).

[14] P. W. Shor and J. A. Smolin, quant-ph/9604006.

[15] A. R. Calderbank, E. M. Rains, P. W. Shor, and N. J. A. Sloane, IEEE Trans. Inf. Theory (to be published); quant-ph/9608006.

[16] P. W. Shor and R. Laflamme, Phys. Rev. Lett. **78**, 1600 (1997).

[17] C. H. Bennett, D. P. DiVincenzo, and J. A. Smolin, Phys. Rev. Lett. **78**, 3217 (1997).

[18] R. Cleve, Phys. Rev. A **55**, 4054 (1997).

[19] E. M. Rains, IEEE Trans. Inf. Theory (to be published); quant-ph/9611001.

[20] Quantum hash coding is a protocol for which a vanishingly small fraction of errors are not corrected. The distance d for a typical hash code can be obtained by the following counting procedure: There are $2^{nS(f)}$ likely error patterns to be assigned to 2^{n-k} possible hash strings. It is highly likley that there will be cases where two error patterns are at random assigned to the same hash. A smaller set of error patterns can be assigned uniquely to hashes with a low probability of collision, provided that this number is smaller than $\sqrt{2^{n-k}}$ (coming from the usual birthday-paradox reasoning). This fixes the distance between such errors according to $2^{(n-k)/2}=2^{(n/2)S(f)}=2^{nS(1-d/2n)}$. Thus the distance scales as the block size, being approximately $d\approx 0.8(1-f)n$.

[21] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948); **27**, 623 (1948).

[22] H. Barnum, M. A. Nielsen, and B. Schumacher, quant-ph/9702049.

[23] Of course, even more exotic codes exist, ones that do not conform to the additive code classification at all. The only nonadditive codes of which we are aware are in [4,24,25]. We cannot rule out the possibility that a nonadditive code achieves the best possible capacity for depolarizing channels. Only the code in [25] is directly relevant the depolarizing channel discussed in most of this paper.

[24] D. W. Leung, M. A. Nielsen, I. L. Chuang, and Y. Yamamoto, Phys. Rev. A **56**, 2567 (1997).

[25] E. M. Rains, R. H. Hardin, P. W. Shor, and N. J. A. Sloane, Phys. Rev. Lett. **79**, 953 (1997).

[26] P. W. Shor, in *Proceedings of the Symposium on the Foundations of Computer Science* (IEEE Press, Los Alamitos, CA, 1996); quant-ph/9605011.

[27] D. P. DiVincenzo and P. W. Shor, Phys. Rev. Lett. **77**, 3260 (1996).

[28] D. Aharonov and M. Ben-Or, quant-ph/9611025.

[29] E. Knill and R. Laflamme, quant-ph/9608012; E. Knill, R. Laflamme, and W. H. Zurek, quant-ph/9610011.

[30] J. Preskill, quant-ph/9705031.

[31] S. Lloyd, Phys. Rev. A **55**, 1613 (1997).

[32] E. M. Rains, quant-ph/9612015.

[33] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Phys. Rev. Lett. **70**, 1895 (1993).

[34] R. Jozsa, J. Mod. Opt. **41**, 2315 (1994).

[35] A. E. Allahverdyan and D. B. Saakian, quant-ph/9702034.

[36] We did not search all possible cat-in-cat codes up to 25 qubits, but the next best code we identified was a $p=5$ inside a $p=3$ cat code which had a threshold at $f\approx 0.809\,52$. The $p=3$ inside a $p=5$ code was much worse, having a threshold at $f\approx 0.811\,12$, worse even than the regular hashing result for which $p=1$ and the threshold is at $f\approx 0.810\,71$.

[37] D. Bruss, D. P. DiVincenzo, A. Ekert, C. Fuchs, C. Macchiavello, and A. Smolin, quant-ph/9705038.