Structured near-optimal channel-adapted quantum error correction

Andrew S. Fletcher,[1,2,*] Peter W. Shor,[3,†] and Moe Z. Win[1,‡]

[1]*Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massuchesetts 02139, USA*

[2]*MIT Lincoln Laboratory, 244 Wood Sreet, Lexington, Massachusetts 02420, USA*

[3]*Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA*

(Received 28 August 2007; published 17 January 2008)

We present a class of numerical algorithms which adapt a quantum error correction scheme to a channel model. Given an encoding and a channel model, it was previously shown that the quantum operation that maximizes the average entanglement fidelity may be calculated by a semidefinite program (SDP), which is a convex optimization. While optimal, this recovery operation is computationally difficult for long codes. Furthermore, the optimal recovery operation has no structure beyond the completely positive trace-preserving constraint. We derive methods to generate structured channel-adapted error recovery operations. Specifically, each recovery operation begins with a projective error syndrome measurement. The algorithms to compute the structured recovery operations are more scalable than the SDP and yield recovery operations with an intuitive physical form. Using Lagrange duality, we derive performance bounds to certify near-optimality.

DOI: 10.1103/PhysRevA.77.012320 PACS number(s): 03.67.Pp, 02.60.Pn

I. INTRODUCTION

All physical implementations of quantum-information-processing systems must incorporate a scheme to mitigate the effects of noise. The most common method for quantum error correction (QEC) is analogous to classical digital error correction schemes. The system of interest is encoded into a subspace of a larger quantum system by means of a quantum code. After passing through a noisy channel, a syndrome measurement projects errors onto orthogonal subspaces from which the original quantum state can be recovered. The first quantum error correcting codes demonstrated that such methods could correct arbitrary single-qubit errors [1–3]. These generic methods enabled a whole range of study in quantum error correction, particularly as it applies to fault-tolerant quantum computing.

The generic approach has its drawbacks, however. Most notably, quantum codes impose a severe amount of overhead to correct for arbitrary errors. As an example, the shortest block code that corrects an arbitrary qubit error embeds one qubit into five [4,5]. As scaling to many qubits is one of the principal barriers to building a working quantum computer, any efforts to improve the efficiency of error recovery are of great interest.

Several recent efforts have explored an optimization-based approach to quantum error recovery [6–9]. In each case, rather than correcting for arbitrary single-qubit errors, the error recovery scheme was adapted to a model for the noise, with the goal of maximizing the fidelity of the operation. In [6], a semidefinite program (SDP) was used to maximize the entanglement fidelity, given a fixed encoding and channel model. In [7] and [8], encodings and decodings were iteratively improved using the performance criteria of ensemble average fidelity and entanglement fidelity, respectively. A suboptimal method for minimum fidelity, using an SDP, was proposed in [9]. An analytical approach to channel-adapted recovery based on the pretty-good measurement and the average entanglement fidelity was derived in [10]. The main point of each scheme was to improve error corrective procedures by adapting to the physical noise process.

As in [6], we choose to focus our channel-adapted efforts on the recovery operation. While channel adaptation can be advantageous in both the encoding and the recovery operations, the optimization problem has a significantly nicer form when one of the two is held fixed. The numerical tools we develop can be used for either half of the problem; focusing on quantum error recovery (QER) operations illustrates nearly all of the important numerical procedures.

The optimization approach to quantum error recovery demonstrates the utility of channel adaptivity. Such efforts have shown that quantum error correction designed for generic errors can be inefficient in the face of a particular noise process. Since overhead in physical quantum computing devices is challenging, it is advantageous to maximize error recovery efficiency.

Recovery operations generated through convex optimization methods suffer two significant drawbacks. First, the dimensions of the optimization problem grow exponentially with the length of the code, limiting the technique to short codes. Second, the optimal operation, while physically legitimate, may be quite difficult to implement. The optimization routine is constrained to the set of completely positive, trace-preserving (CPTP) operations, but is not restricted to more easily implemented operations.

In this paper, we describe efforts to determine near-optimal channel-adapted quantum error recovery procedures that overcome the drawbacks of optimal recovery. We impose an intuitively satisfying structure on the recovery operation and seek to optimize performance. While still numerical procedures, the result is a class of algorithms that is less computationally intensive than the SDP and which

*fletcher@ll.mit.edu
†shor@math.mit.edu
‡moewin@mit.edu

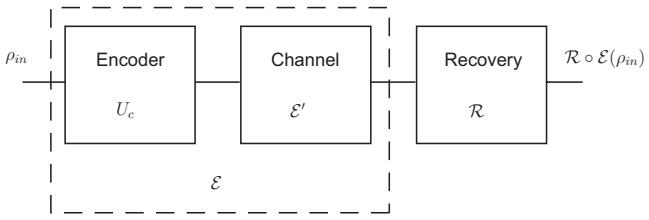

FIG. 1. Quantum error correction block diagram. For channel-adapted recovery, the encoding isometry U_C and the channel $\mathcal{E}'$ are considered as a fixed operation $\mathcal{E}$ and the recovery $\mathcal{R}$ is chosen according to the design criteria.

yields recovery operations of an intuitive and potentially realizable form.

II. CHANNEL-ADAPTED RECOVERY

To adapt quantum error recovery to a specific channel model, we must first determine a measure of performance. As detailed in [6,7], both entanglement fidelity (F_e) and ensemble average fidelity ($\bar{F}$) yield convex optimization problems.[1] As both measures may be of interest, we will use the *average entanglement fidelity* of [10], of which either of the above is a special case. Average entanglement fidelity is defined for a channel $\mathcal{A}$ and an ensemble E of states $\{\rho_i\}$ with prior probabilities p_i as

$$\bar{F}_e(E,\mathcal{A}) = \sum_i p_i F_e(\rho_i,\mathcal{A}) = \sum_{i,k} p_i |\mathrm{tr}(\rho_i A_k)|^2, \tag{1}$$

where $\{A_k\}$ are the Kraus elements for the CPTP map $\mathcal{A}$.

Consider the simple block diagram for a QEC system given in Fig. 1. We begin with a fixed model, labeled $\mathcal{E}'$, to describe the physical noise process. In designing a QEC procedure, we can choose the encoding U_C and the recovery operation $\mathcal{R}$. By holding either the encoding or the recovery operation as fixed, optimizing the other can be cast as a convex optimization problem [6–8]. As done in [7,8], one can iteratively optimize a encoding and recovery scheme. In this paper, we focus our attention on the efficacy of adapting the recovery operation and defer iterative optimization to subsequent work. We illustrate our emphasis on the recovery block by considering both the encoding U_C and the channel $\mathcal{E}'$ as a combined operation $\mathcal{E}$.

It is useful to note the dimensions of the various operations in Fig. 1. We define two Hilbert spaces $\mathcal{H}_S$ and $\mathcal{H}_C$, which refer to the source and the code spaces, respectively. These have dimensions d_S and d_C. The combined encoding and channel $\mathcal{E}$ therefore maps density matrices in $\mathcal{L}(\mathcal{H}_S)$ to $\mathcal{L}(\mathcal{H}_C)$, where $\mathcal{L}(\mathcal{H})$ refers to the space of bounded linear operators on $\mathcal{H}$. Our use of the fidelity implies that $\mathcal{R}$ maps from $\mathcal{L}(\mathcal{H}_C)$ to $\mathcal{L}(\mathcal{H}_S)$, i.e., $\mathcal{R}$ performs a decoding. This is mostly for computational convenience as $d_S < d_C$.

Channel-adapted recovery selects an operation $\mathcal{R}$ to maximize $\bar{F}_e(E,\mathcal{R}\circ\mathcal{E})$. As shown in [6], exact maximization can be accomplished via the convex optimization routine of semidefinite programming. For the remainder of the paper, we will discuss routines to approach the optimum channel-adapted recovery through a more computationally feasible method. In several cases, the routines also yield an intuitive form for the recovery operation.

We will make use of a convenient isomorphism in which bounded linear operators are represented by vectors and denoted with the symbol $|\cdot\rangle\rangle$. While there are several choices for this isomorphism [11,12], including most intuitively a "stacking" operation, we will follow the conventions of [13] (also [9]), which result in an isomorphism that is independent of the choice of basis. For convenience, we will restate the relevant results here.

Let $A = \Sigma_{ij} a_{ij} |i\rangle\langle j|$ be a bounded linear operator from $\mathcal{H}$ to $\mathcal{K}$ (i.e., $A \in \mathcal{L}(\mathcal{K},\mathcal{H})$), where $\{|i\rangle\}$ and $\{|j\rangle\}$ are bases for $\mathcal{K}$ and $\mathcal{H}$, respectively. Let $\mathcal{H}^*$ be the dual of $\mathcal{H}$. This is also a Hilbert space, generally understood as the space of *bras* $\langle j|$. If we relabel the elements as $\overline{|j\rangle} = \langle j|$, then we represent A as a vector in the space $\mathcal{K}\otimes\mathcal{H}^*$ as

$$|A\rangle\rangle = \sum_{ij} a_{ij} |i\rangle|\overline{j}\rangle. \tag{2}$$

It is useful to note the following facts. The inner product $\langle\langle A|B\rangle\rangle$ is the Hilbert-Schmidt inner product $\mathrm{tr}A^\dagger B$. Also, the partial trace $\mathrm{tr}_{\mathcal{K}} |A\rangle\rangle\langle\langle B| = \overline{AB^\dagger}$. Finally, index manipulation yields the relation $A\otimes\bar{B}\,|C\rangle\rangle = |ACB^\dagger\rangle\rangle$, where $\bar{B}$ is the conjugate of B such that $\overline{B|\psi\rangle} = \bar{B}|\psi\rangle$ for all $|\psi\rangle$.

These relations lead directly to a convenient representation of a CPTP operation $\mathcal{A}:\mathcal{L}(\mathcal{H})\mapsto\mathcal{L}(\mathcal{K})$ in terms of a positive semidefinite (PSD) operator $X_{\mathcal{A}} \in \mathcal{L}(\mathcal{K}\otimes\mathcal{H}^*)$ [11,12,14–16]. The PSD operator is calculated from the Kraus elements $\{A_k\}$ of $\mathcal{A}$ as

$$X_{\mathcal{A}} = \sum_k |A_k\rangle\rangle\langle\langle A_k| . \tag{3}$$

We will refer to $X_{\mathcal{A}}$ as the Choi matrix for $\mathcal{A}$, although most derivations do not use the basis-free double ket of (2). The operation output is given by $\mathcal{A}(\rho) = \mathrm{tr}_{\mathcal{H}^*} I\otimes\bar{\rho} X_{\mathcal{A}}$ and the CPTP constraint requires that $X_{\mathcal{A}} \geq 0$ and $\mathrm{tr}_{\mathcal{K}} X_A = I$.

In terms of the Choi matrix, the average entanglement fidelity can be written as $\bar{F}_e(E,\mathcal{A}) = \Sigma_i p_i \langle\langle\rho_i| X_{\mathcal{A}} |\rho_i\rangle\rangle$. From this expression, we can derive the dependence of $\bar{F}_e(E,\mathcal{R}\circ\mathcal{E})$ on $\mathcal{R}$ as

$$\begin{aligned}\bar{F}_e(E,\mathcal{R}\circ\mathcal{E}) &= \sum_i p_i \langle\langle\rho_i| X_{\mathcal{R}\circ\mathcal{E}} |\rho_i\rangle\rangle \\ &= \sum_i p_i \mathrm{tr} X_{\mathcal{R}} \left(\sum_k |\rho_i E_k^\dagger\rangle\rangle\langle\langle\rho_i E_k^\dagger| \right) \\ &= \mathrm{tr} X_{\mathcal{R}} C_{E,\mathcal{E}},\end{aligned} \tag{4}$$

where $C_{E,\mathcal{E}} = \Sigma_{ik} p_i |\rho_i E_k^\dagger\rangle\rangle\langle\langle\rho_i E_k^\dagger|$ encapsulates both the input ensemble E and the channel (with encoding) $\mathcal{E}$. It was shown in [6,7] the the optimum $X_{\mathcal{R}}$ satisfying the CPTP constraint can be calculated via semidefinite programming.

[1] Ensemble average fidelity yields a convex optimization problem if and only if the states in the ensemble are pure.

III. EIGQER ALGORITHM

To achieve a near-optimal QER operation, an algorithm must have a methodology to approach optimality while still satisfying the CPTP constraints. Furthermore, to ease implementation of such a recovery, we can impose structure to maintain relative simplicity.

Let us begin by considering the structure of a standard QEC recovery operation. QEC begins by defining a set of correctable errors, i.e., errors that satisfy the quantum error correction conditions. To correct this set, we construct the recovery operation by defining a projective syndrome measurement. Based on the detected syndrome, the appropriate unitary rotation restores the information to the code space, thereby correcting the error. This intuitive structure—projective measurement followed by unitary syndrome recovery—provides a simple geometric picture of error correction. Furthermore, it is a relatively straightforward task to translate such a recovery operation into a quantum circuit representation.

Let us impose the same constraint on the channel-adapted recovery operation. We construct an operation with operator elements that are a projective syndrome measurement followed by a classically controlled unitary operation. Thus the operator elements can be written $\{R_k=U_kP_k\}$, where P_k is a projection operator. While we could merely constrain U_k to be unitary, we will instead continue with the convention that the recovery operation performs a decoding: $\mathcal{R}:\mathcal{L}(\mathcal{H}_C)\mapsto\mathcal{L}(\mathcal{H}_S)$. Under this convention, $U_k\in\mathcal{L}(\mathcal{H}_C,\mathcal{H}_S)$ and $U_k^\dagger U_k=I$. In words, both $U_k^\dagger$ and $R_k^\dagger$ are isometries.

The CPTP constraint

$$I=\sum_k R_k^\dagger R_k=\sum_k P_kU_k^\dagger U_kP_k=\sum_k P_k \tag{5}$$

is satisfied if and only if the projectors span $\mathcal{H}_C$. To satisfy the CPTP constraint, therefore, $\{P_k\}$ must partition $\mathcal{H}_C$ into orthogonal subspaces, each identified with a correction isometry[2] U_k.

Since the $\{P_k\}$ project onto orthogonal subspaces, we see that $R_j^\dagger R_k=\delta_{jk}P_k$. From this we conclude that $\{|R_k\rangle\rangle\}$ are an orthogonal set and thus are eigenvectors of the Choi matrix $X_\mathcal{R}$. The eigenvalue λ_k associated with $|R_k\rangle\rangle$ is the rank of P_k and is thus constrained to be an integer. Furthermore, since U_k restores the kth syndrome to $\mathcal{H}_S$, $\lambda_k\leq d_S$.

We can conceive of a "greedy" algorithm to construct a recovery operation $\mathcal{R}$. The average entanglement fidelity can be decomposed into the contributions of each individual operator element as $\langle\langle R_k|C_{E,\mathcal{E}}|R_k\rangle\rangle$. We can construct $\mathcal{R}$ by successively choosing the syndrome subspace to maximize the fidelity contribution. As long as each syndrome is orthogonal to the previously selected subspaces, the resulting operation will be CPTP and will satisfy our additional constraints. In fact, this greediest algorithm has no immediate method for computation; the selection of the syndrome subspace to maximize the fidelity contribution has no simple form. We propose instead a greedy algorithm to approximate this procedure.

We motivate our proposed algorithm in terms of eigenanalysis. Let us assume for the moment that the rank of each syndrome subspace is exactly d_S which is the case for QEC recoveries for stabilizer codes. By such an assumption, we know that there will be d_C/d_S recovery operator elements. Consider now the average entanglement fidelity, in terms of the eigenvectors of $X_\mathcal{R}$:

$$\bar{F}(E,\mathcal{R}\circ\mathcal{E})=\sum_{k=1}^{d_C/d_S}\langle\langle R_k|C_{E,\mathcal{E}}|R_k\rangle\rangle. \tag{6}$$

If we were to maximize the above expression with the only constraint being a fixed number of orthonormal vectors $|R_k\rangle\rangle$, the solution would be the eigenvectors associated with the d_C/d_S largest eigenvalues of $C_{E,\mathcal{E}}$. In fact, the actual constraint differs slightly from this simplification, as we further must constrain $R_k^\dagger$ to be an isometry (i.e., $R_kR_k^\dagger=I$). The analogy to eigenanalysis, however, suggests a computational algorithm which we dub EIGQER (for eigen quantum error recovery). We use the eigenvectors of $C_{E,\mathcal{E}}$ to determine a syndrome subspace with a large fidelity contribution.

The algorithm proceeds as follows.

(1) Initialize $C_1=C_{E,\mathcal{E}}$.

For the kth iteration,

(2) determine $|X_k\rangle\rangle$, the eigenvector associated with the largest eigenvalue of C_k.

(3) Calculate $R_k^\dagger$, the isometry "closest" to $X_k^\dagger$ via the singular value decomposition. Call R_k an operator element of $\mathcal{R}$.

(4) Determine C_{k+1} by projecting out of C_k the support of R_k.

We return to step (2) and iterate until the recovery operation is complete.

The EIGQER algorithm is guaranteed to generate a CPTP recovery operation, and will satisfy the criterion that it can be implemented by a projective syndrome measurement followed by a syndrome dependent unitary operation.

Steps 2 and 3 in the above algorithm require further exposition. Given an operator $X\in\mathcal{L}(\mathcal{H}_C,\mathcal{H}_S)$, what is the closest isometry R_k? A straightforward answer uses the norm derived from the Hilbert-Schmidt inner product where $\|A\|^2=\mathrm{tr}A^\dagger A$. We will now allow the rank of the kth subspace to be $d_k\leq d_S$.[3] Thus $R_kR_k^\dagger=I_{d_k}$ where I_{d_k} is a diagonal operator with the 1 as the first d_k diagonal matrix elements and 0 for the rest. We have the minimization problem

$$\min_{R_k}\mathrm{tr}(X-R_k)^\dagger(X-R_k)\quad\text{such that}\quad R_kR_k^\dagger=I_{d_k}. \tag{7}$$

We will state the solution as the following lemma.

Lemma 1. Let X be an operator with singular value de-

[2] In fact, $U_k^\dagger$ is the isometry. For ease of explication, we will refer to U_k as an isometry as well.

[3] Inclusion of reduced rank subspaces may seem unnecessary or even undesirable—after all, such a projection would collapse superpositions within the encoded information. We allow the possibility since such operator elements are observed in the optimal recovery operations of [6].

composition $X=U\Sigma V^\dagger$. The rank-d isometry R that minimizes the Hilbert-Schmidt norm difference $\|X-R\|$ is given by $R=UI_{d_l}V^\dagger$.

Proof. Let $\mathcal{U}_d$ be the set of rank d isometries; that is, $\mathcal{U}_d=\{U|U^\dagger U=I_d\}$. We wish to find the $R^\dagger\in\mathcal{U}$ that minimizes $\mathrm{tr}(X-R)^\dagger(X-R)$. Since this can be written as

$$\mathrm{tr}(X-R)^\dagger(X-R)=\mathrm{tr}X^\dagger X+\mathrm{tr}R^\dagger R-\mathrm{tr}(X^\dagger R+R^\dagger X) \quad (8)$$

and $\mathrm{tr}R^\dagger R=d$, an equivalent problem is

$$\max_{R\in\mathcal{U}}\mathrm{tr}(X^\dagger R+R^\dagger X)=\max_{R\in\mathcal{U}}\mathrm{tr}(V\Sigma U^\dagger R+R^\dagger U\Sigma V^\dagger), \quad (9)$$

where we have replaced X with its singular value decomposition.

We can simplify the above expression by noting that $C^\dagger=U^\dagger R\in\mathcal{U}$. We can thus equivalently maximize the following expression over $C^\dagger\in\mathcal{U}$:

$$\mathrm{tr}(V\Sigma C^\dagger+C\Sigma V^\dagger)=\mathrm{tr}\Sigma(C^\dagger V+V^\dagger C)=\sum_{i=1}^{d}\sigma_i(c_i^\dagger v_i+v_i^\dagger c_i) \quad (10)$$

$$=2\sum_{i=1}^{d}\sigma_i\,\mathrm{Re}\{v_i^\dagger c_i\}\le 2\sum_{i=1}^{d}\sigma_i|v_i^\dagger c_i| \quad (11)$$

$$\le 2\sum_{i=1}^{d}\sigma_i\|v_i\|\|c_i\|=2\sum_{i=1}^{d}\sigma_i. \quad (12)$$

In (10), σ_i is the ith largest singular value of X and v_i and c_i are the ith columns of V and C, respectively. We have used the fact that Σ is a diagonal matrix of the singular values in descending order. The inequality is saturated when $c_i=v_i$, which also implies that $C=VI_d\Rightarrow R=UI_dV^\dagger$. ■

One item not mentioned above is the determination of the desired rank d_k. In our implementation of EIGQER, this is accomplished by setting a relatively high threshold on the singular values of X. We considered only singular values such that $\sigma^2\ge 0.05$. This *ad hoc* value was chosen as it led to acceptable numerical results in the examples.

We turn now to step 3 of the EIGQER algorithm. Recall that the CPTP constraint as written in (5) requires that the syndrome subspaces are mutually orthogonal. Thus, the syndrome measurement for the kth iteration must be orthogonal to the first $k-1$ iterations: $P_kP_i=0$ for $i<k$. We satisfy this constraint by updating the data matrix C_{k-1}.

To understand the update to C_{k-1}, recall that the first step of the kth iteration is the computation of the dominant eigenvector $|X_k\rangle\rangle$. To satisfy the constraint, we require that

$$X_kP_i=0\Leftrightarrow|X_kP_i\rangle\rangle=I\otimes\overline{P_i}\,|X_k\rangle\rangle=0 \quad (13)$$

for $i<k$. All $|X\rangle\rangle$ for which this is not satisfied should be in the null space of C_k. Thus, after each iteration we update the data matrix as

$$C_k=(I-I\otimes\overline{P_{k-1}})C_{k-1}(I-I\otimes\overline{P_{k-1}}). \quad (14)$$

The algorithm terminates when the recovery operation is complete, i.e., $\Sigma_kR_k^\dagger R_k=\Sigma_kP_k=I$. Given the structure of the recovery operations, this can be determined with a simple counter that is increased by d_k at each step k. When the counter reaches d_C, the recovery is complete.

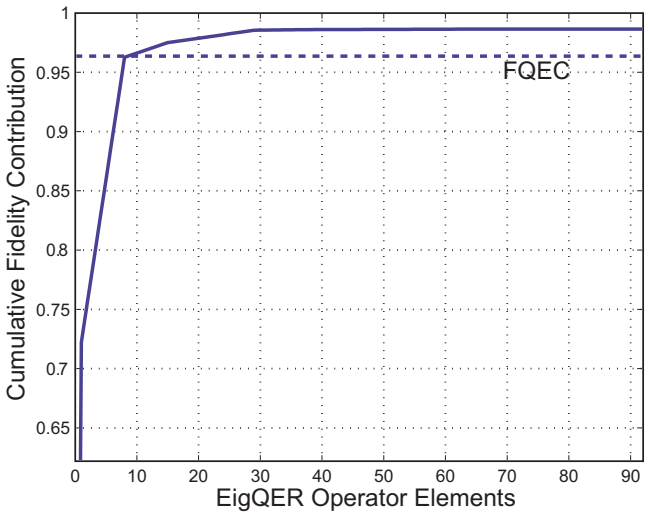

FIG. 2. (Color online) Fidelity contribution of EIGQER recovery operators for the amplitude-damping channel ($\gamma=.09$) and the Steane code. Notice that the QEC performance is equaled with only eight operator elements, and the relative benefit of additional operators goes nearly to zero after 30.

In fact, the greedy nature of EIGQER allows early termination of the above algorithm. Each R_k contributes $\langle\langle R_k|\,C_{E,\mathcal{E}}\,|R_k\rangle\rangle$ to the average entanglement fidelity. Since the algorithm seeks to maximize its gain at each step, the performance return of each R_k diminishes as k grows. This is illustrated in Fig. 2, where we show the cumulative contribution for each recovery operator element with the Steane code and the amplitude-damping channel. The greedy construction results in simplifications in both computation and implementation. When the contribution $\langle\langle R_k|\,C_{E,\mathcal{E}}\,|R_k\rangle\rangle$ passes below some selected threshold, the algorithm may terminate and thus reduce the computational burden. This results in an undercomplete recovery operation where $\Sigma_kR_k^\dagger R_k\le I$. An undercomplete specification for the recovery operation may significantly reduce the difficulty in physically implementing the recovery operation. In essence, an undercomplete recovery operation will have syndrome subspaces whose occurrence is sufficiently rare that the recovery operation may be left as a "don't care."

Before we consider examples of EIGQER recovery performance, we should say a few words about the algorithm complexity when channel-adapting an $[n,k]$ code. The SDP of [6,7] to calculate the optimal recovery operation has 4^{n+k} complex optimization variables constrained to a semidefinite cone with a further 4^k equality constraints. From [17], a SDP with n variables and a $p\times p$ semidefinite matrix constraint requires $O(\max\{np^3,n^2p^2,n^3\})$ flops per iteration (with typically 10–100 iterations necessary). For our case, this yields $O(2^{5(n+k)})$ flops per iteration.

For the EIGQER operation, the dominant computation is the calculation of $|X_k\rangle\rangle$, the eigenvector associated with the largest eigenvalue of C_k. C_k is a $(2^{n+k}\times 2^{n+k})$-dimensional matrix, but the eigenvector has only 2^{n+k} dimensions. Using

the *power method* for calculating the dominant eigenvector requires $O(2^{2(n+k)})$ flops for each iteration of the power method. While both problems grow exponentially with n, the reduced size of the eigenvector problem has a significant impact on the computational burden.

We should note that the eigenvector computation must be repeated for each operator element of $\mathcal{R}$. If we were to compute all of them, not truncating early due to the diminishing returns of the greedy algorithm, this would require iterating the algorithm approximately $d_C/d_S=2^{n-k}$ times. In fact, we have a further reduction as the algorithm iterates. At the jth iteration we are calculating the dominant eigenvector of C_j which lives on a $[(d_C-jd_S)d_S=2^k(2^n-j2^k)]$-dimensional subspace. We can therefore reduce the size of the eigenvector problem at each iteration of EIGQER.

A. EIGQER examples

To demonstrate the value of the EIGQER algorithm, we consider several channels and codes; we would like to consider common codes and channels with nontrivial channel-adapted recoveries. It will be shown in the Appendix that channels represented by scaled Pauli group operators yield straightforward channel-adapted recovery operations; it is therefore useful to consider non-Pauli channels. The most common and useful such channel is the amplitude-damping channel, which we will denote $\mathcal{E}_a$. Amplitude damping was the example used in [6] to illustrate optimal QER, as well as the example for channel-adapted code design of [18]. The channel is a commonly encountered model, where the parameter γ indicates the probability of decaying from state $|1\rangle$ to $|0\rangle$ (i.e., the probability of losing a photon). For a single qubit, $\mathcal{E}_a$ has operator elements

$$E_0=\begin{bmatrix}1 & 0\\ 0 & \sqrt{1-\gamma}\end{bmatrix} \quad \text{and} \quad E_1=\begin{bmatrix}0 & \sqrt{\gamma}\\ 0 & 0\end{bmatrix}. \tag{15}$$

The EIGQER algorithm does not require a channel as simple to model as the amplitude-damping channel; the optimization routine is general to any channel. To illustrate, we consider a qubit channel that is less familiar, though with a straightforward geometric description. We will call this the "pure state rotation" channel and label it as $\mathcal{E}_{ps}$. To describe the channel, we define a pure state by its angle in the xz plane: $|\theta\rangle=\cos\theta|0\rangle+\sin\theta|1\rangle$. The channel mapping is defined by its action on two pure states an angle θ apart, symmetric about the z axis. When $|\pm\theta/2\rangle$ is input to the channel, the result is $|\pm(\theta-\phi)/2\rangle$, also as a pure state. Thus, these two states are rotated toward each other by ϕ. Any other state input to the channel will emerge mixed. The operator elements for this channel can be written as

$$\mathcal{E}_{ps}\sim\left\{\alpha\begin{bmatrix}\cos\frac{\theta-\phi}{2}\sin\frac{\theta}{2} & \pm\cos\frac{\theta-\phi}{2}\cos\frac{\theta}{2}\\ \pm\sin\frac{\theta-\phi}{2}\sin\frac{\theta}{2} & \sin\frac{\theta-\phi}{2}\cos\frac{\theta}{2}\end{bmatrix},\ \beta\begin{bmatrix}\frac{\cos\frac{\theta-\phi}{2}}{\cos\frac{\theta}{2}} & 0\\ 0 & \frac{\sin\frac{\theta-\phi}{2}}{\sin\frac{\theta}{2}}\end{bmatrix}\right\}, \tag{16}$$

where α and β are constants chosen to satisfy the CPTP constraint.

The pure state rotation channel has multiple parameters which characterize its behavior. θ indicates the initial separation of the targeted states. ϕ, the amount of rotation, clearly parametrizes the "noise strength," as $\phi=0$ indicates no decoherence while $\phi=\theta$ is strong decoherence. Furthermore, we have chosen the target states to be symmetric about the z axis, but this is only for clarity in stating the channel; any alternate symmetry axis may be defined. Furthermore, a similar channel with asymmetric rotations ϕ_1 and ϕ_2 may be defined. This, however, corresponds to a symmetric channel followed by a unitary rotation. While less physically motivated than amplitude damping, the pure state rotation channel model provides an extended set of qubit channels which are not represented with Pauli group operator elements. We will look at examples of this channel where $\theta=5\pi/12$. There is no particular significance to this choice; it merely illustrates well the principles of channel-adapted QEC.

Since the EIGQER algorithm is more computationally scalable than the SDP, we can consider channel-adapted QER for several codes. We compare the EIGQER recovery performance to the optimal channel-adapted recovery performance for the five-qubit stabilizer code [4,5]. We also compare the EIGQER performance for the five-qubit code, the seven-qubit Steane code [2,3], and the nine-qubit Shor code [1]. All comparisons consider an ensemble E of qubit states that are in the completely mixed state $\rho=I/2$.

Figure 3 compares the performance of the EIGQER algorithm to the optimal QER recovery for the case of the five-qubit stabilizer code and the amplitude-damping channel. Also included are the generic QEC recovery and the entanglement fidelity of a single qubit acted upon by $\mathcal{E}_a$ (i.e., no error correction performed). From this example we observe that the EIGQER performance nearly achieves the optimum, especially for the values of γ below 0.4. For higher γ, the

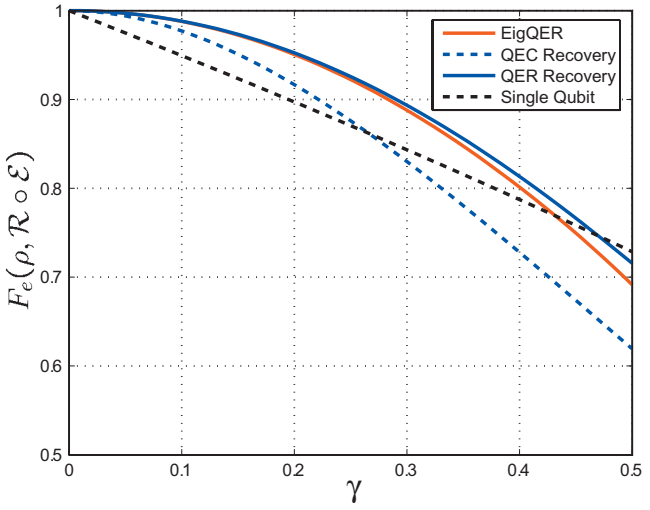

FIG. 3. (Color) EIGQER and optimal QER for the amplitude-damping channel and the five-qubit stabilizer code. EIGQER nearly duplicates the optimal channel-adapted performance, especially for lower-noise channels (small γ).

EIGQER performance begins to diverge, but this is less important as that region is one in which even the optimal QER lies below the fidelity of a single qubit obtainable with no error correction.

Figure 4 compares EIGQER and optimal QER for the five-qubit stabilizer code and the pure state rotation channel with $\theta=5\pi/12$. We see again that the EIGQER algorithm achieves a recovery performance nearly equivalent to the optimum, especially as the noise level approaches 0.

Figure 5 demonstrates the performance of several codes and the amplitude-damping channel. We compare the EIGQER performance for the five-, seven-, and nine-qubit codes, contrasting each with the generic QEC performance. Notice first the pattern with the standard QEC recovery: the entanglement fidelity decreases with increasing length of the code.

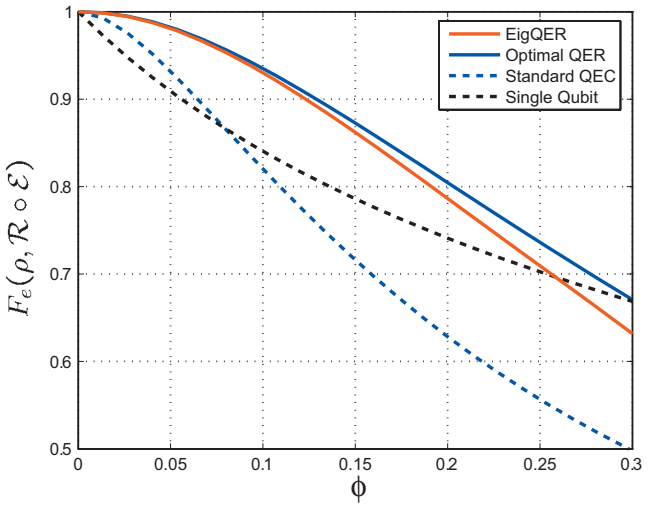

FIG. 4. (Color) EIGQER and optimal QER for the pure state rotation channel with $\theta=5\pi/12$ and the five-qubit stabilizer code. EIGQER nearly duplicates the optimal channel-adapted performance, especially for lower-noise channels (small ϕ).

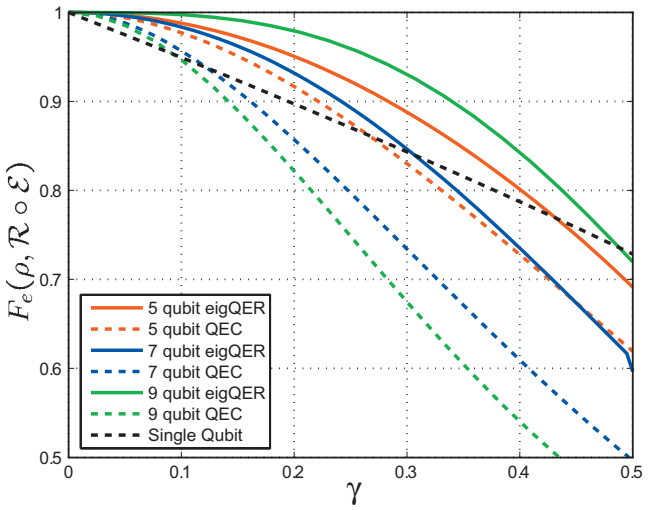

FIG. 5. (Color) EIGQER and standard QEC recovery performance for the five-, seven-, and nine-qubit codes and the amplitude-damping channel. Note that generic QEC performance decreases for longer codes, as multiple-qubit errors become more likely. While the EIGQER performance for the nine-qubit Shor code is excellent, the seven-qubit Steane code shows only modest improvement, with performance similar to the generic five-qubit QEC recovery.

The five-qubit stabilizer code, the Steane code, and the Shor code are all designed to correct a single error on an arbitrary qubit, and fail only if multiple qubits are corrupted. For a fixed γ, the probability of a multiple-qubit error rises as the number of physical qubits n increases.

The QEC performance degradation with code length is a further illustration of the value of channel adaptivity. All three codes in Fig. 5 contain one qubit of information, so longer codes include more redundant qubits. Intuitively, this should better protect the source from error. When we channel adapt, this intuition is confirmed for the Shor code, but not for the Steane code. In fact, the EIGQER entanglement fidelity for the Steane code is only slightly higher than the generic QEC recovery for the five-qubit code. From this example, it appears that the Steane code is not particularly well suited for adapting to amplitude-damping errors. We see that the choice of encoding significantly impacts channel-adapted recovery.

The effect is even more dramatically (and puzzlingly) illustrated in the pure state rotation channel. Figure 6 compares the EIGQER recoveries for the five-qubit, Steane, and Shor codes with $\theta=5\pi/12$. It is interesting to see that the five-qubit code outperforms each of the others despite having less redundancy to protect the information. Furthermore, both the standard QEC and channel-adapted recoveries for the Steane code perform worse than the generic recovery of the Shor code! This suggests that the five-qubit code is particularly well suited to adapt to errors of this type, while the Steane code is particularly ill suited. (We suspect that the Shor code with QEC recovery outperforms the Steane due to its degenerate structure.)

IV. BLOCK SDP QER

The recovery operation generated by the EIGQER algorithm of the preceding section is one of a broader class of

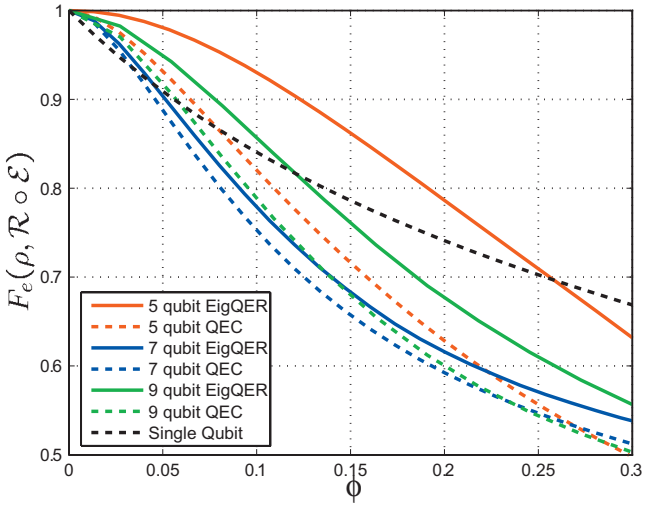

FIG. 6. (Color) EIGQER and standard QEC recovery performance for the five-, seven-, and nine-qubit codes and the pure state rotation channel with $\theta=5\pi/12$. Despite the least redundancy, the five-qubit code has the best channel-adapted performance. The Steane code appears particularly poor for this channel: both the generic QEC and the adapted recovery have lower fidelity than the other codes.

quantum error recoveries. The class is characterized by an initial projective syndrome measurement, followed by a syndrome-specific recovery operation. The projective measurement partitions $\mathcal{H}_C$ and provides some knowledge about the observed noise process.

Projective syndrome measurements for quantum error correction are tricky to design. We wish to learn as much as possible about the error while learning as little as possible about the input state, so as not to destroy quantum superposition. The EIGQER algorithm aggressively designs the syndrome measurement, as the $R_k=U_kP_k$ structure of the operator elements implies a finality about the syndrome selection. The outcome of the syndrome measurement completely determines the correction term U_k.

We can conceive of a less aggressive projective measurement. If we projected onto larger subspaces of $\mathcal{H}_C$, we would learn less about the noise but perhaps have less chance of destroying the superposition of the input state. We could consider this an intermediate syndrome measurement, a preliminary step to further error correction. To design a recovery operation of this type, we must have a strategy to select a projective measurement. Given the outcome P_k, we must further design the syndrome recovery operation $\mathcal{R}_k$.

Consider the projective syndrome measurement operator P_k. For the EIGQER algorithm, $P_k=R_k^\dagger R_k$ always projects onto a subspace of dimension less than or equal to the source space: $\text{rank}(P_k)\leq d_S$. This is an aggressive condition that arises from constraining the subsequent syndrome recovery to be a unitary operator. We will relax this constraint and allow an arbitrary syndrome recovery $\mathcal{R}_k$ for the kth syndrome measurement. It turns out that we can determine the optimum such recovery $\mathcal{R}_k^{opt}$ via semidefinite programming, just as in [6]. The intermediate syndrome measurement P_k reduces the dimension of the SDP, and thus the technique is still applicable to long codes where computing the global optimum recovery is impractical.

We will demonstrate how the optimum syndrome recovery $\mathcal{R}_k$ can be calculated via a semidefinite program. Let $\{P_k\}_{k=1}^K$ be a set of projectors such that $\Sigma_k P_k=I\in\mathcal{H}_C$ that constitute an error syndrome measurement. Let $\mathcal{S}_k$ be the support of P_k with dimension d_k; it is clear that $\mathcal{S}_1\oplus\mathcal{S}_2\oplus\cdots\oplus\mathcal{S}_K=\mathcal{H}_C$. Given the occurrence of syndrome k, we must now design a recovery operation $\mathcal{R}_k:\mathcal{S}_k\mapsto\mathcal{H}_S$. $\mathcal{R}_k$ is subject to the standard CPTP constraint on quantum operations, but only has support on $\mathcal{S}_k$. We may calculate the recovery $\mathcal{R}_k$ that maximizes the average entanglement fidelity using the SDP in a structure identical to that of [6] while accounting for the reduced input space:

$$X_{\mathcal{R}_k}=\arg\max_X \text{tr}X(C_{E,\mathcal{E}})_k$$
$$\text{such that } X\geq 0,\ \text{tr}_{\mathcal{H}_S}X=I\in\mathcal{S}_k. \tag{17}$$

Here, $(C_{E,\mathcal{E}})_k=I\otimes\overline{P_k}C_{E,\mathcal{E}}I\otimes\overline{P_k}$ is the data matrix projected into the kth subspace. Notice that $X_{\mathcal{R}_k}$ and $(C_{E,\mathcal{E}})_k$ are operators on $\mathcal{H}_S\otimes\mathcal{S}_k^*$. In contrast to $C_{E,\mathcal{E}}$, which requires $d_S^2d_C^2$ matrix elements, $(C_{E,\mathcal{E}})_k$ is fully specified by $d_S^2d_k^2$ matrix elements. By partitioning $\mathcal{H}_C$ into subspaces $\{\mathcal{S}_k\}$ through a careful choice of a syndrome measurement $\{P_k\}$, we may apply semidefinite programming to high-dimensional channels without incurring the full computational burden of computing the optimal recovery. In the following sections we discuss two strategies for determining the syndrome measurement.

A. Block EIGQER

The first step of an iteration of EIGQER computes the dominant eigenvalue and corresponding eigenvector of $C_{E,\mathcal{E}}$. This eigenvector corresponds to the operator that maximizes the average entanglement fidelity gain at a single step. While such an operator may violate the CPTP constraint for the recovery operation, it serves to identify an important subspace onto which we may project. Indeed, the good performance of the EIGQER algorithm rests on the successful identification of suitable syndrome subspaces via eigenanalysis.

An intuitive extension of this concept is to use multiple eigenvectors to specify a higher-dimension subspace. If $\{|X_m\rangle\rangle\}_{m=1}^M$ are the eigenvectors corresponding to the M largest eigenvalues of $C_{E,\mathcal{E}}$, then it is reasonable to define the subspace $\mathcal{S}_1$ as the union of the support of the operators $\{X_m\}$. We define the corresponding projector P_1 and calculate the syndrome recovery $\mathcal{R}_1$ via the SDP of (17). As in the EIGQER algorithm, we update the data matrix C by projecting out the subspace $\mathcal{S}_1$, at which point we select another set of eigenvectors. We will refer to this algorithm as BLOCK-EIGQER.

How many eigenvectors should be selected to define a block? A simple solution is for a fixed block size, say M, to be processed until the recovery is complete. For $M=1$, BLOCKEIGQER is identical to EIGQER. For $M=d_Sd_C$, BLOCK-EIGQER computes the optimal recovery operation, as the syndrome measurement is simply the identity operator. For values in between, one would expect to trade off performance for computational burden. While there is no guarantee that

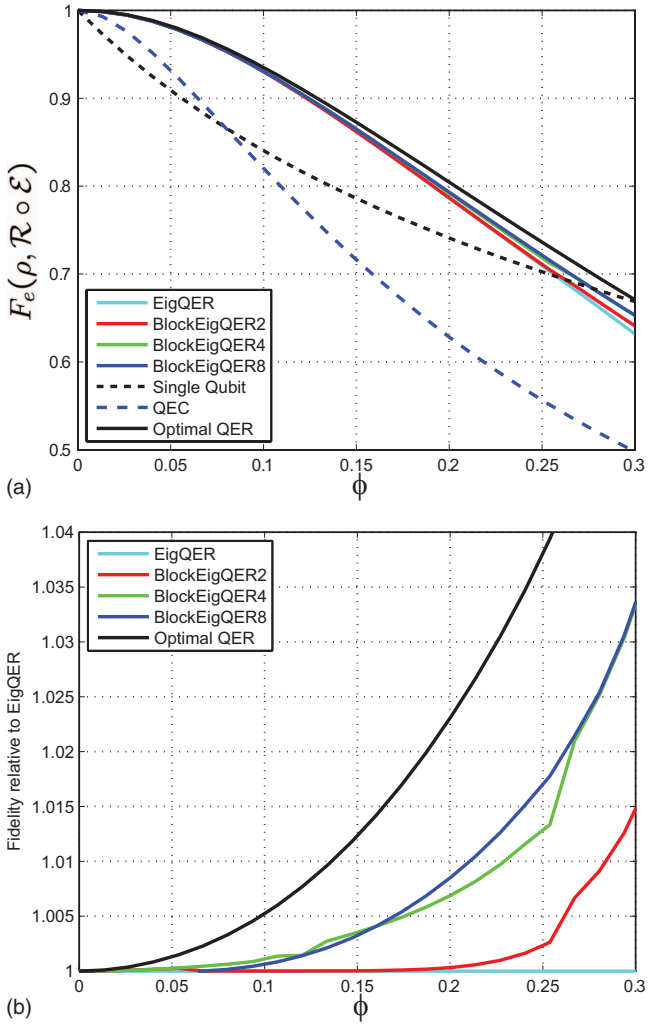

FIG. 7. (Color) BLOCKEIGQER performance for the five-qubit code and the pure state rotation channel with $\theta = 5\pi/12$. BLOCKEIGQER is computed with fixed block lengths of 2, 4, and 8. In (a) we compare the entanglement fidelity to the EIGQER recovery, standard QEC recovery, and single-qubit baseline. The different block lengths have nearly indistinguishable performance from EIGQER. In (b), we compute the fidelity relative to the EIGQER recovery and show that the fidelity improves by less than 4% for the displayed region. We can note, however, that longer block lengths tend to better performance.

performance will improve monotonically, we would anticipate improved performance as M increases.

We illustrate the performance for several choices of M in Fig. 7. We use the pure state rotation channel ($\theta = 5\pi/12$) and the five-qubit code with block sizes of 2, 4, and 8. The expected improvement as M increases is evident, though the gain is quite modest for noise levels of interest (below the crossover with the single-qubit recovery) and is not strictly monotonic. The variations in performance, including the nonmonotonicity, are likely the result of syndrome measurements that collapse the input superpositions. While the eigenvectors of $C_{E,\mathcal{E}}$ that identify the syndrome subspace generally avoid collapsing the input state, the mechanism is imperfect.

While BLOCKEIGQER outperforms EIGQER in the[5,1] code, we see in Fig. 7(b) that the improvement is less than 5% within the ϕ of interest. We see more significant gains when we encode multiple qubits. Consider a random [6,2] encoding for the amplitude-damping channel, shown in Fig. 8. In this case we see a distinct performance gain as M increases and the difference is nontrivial.

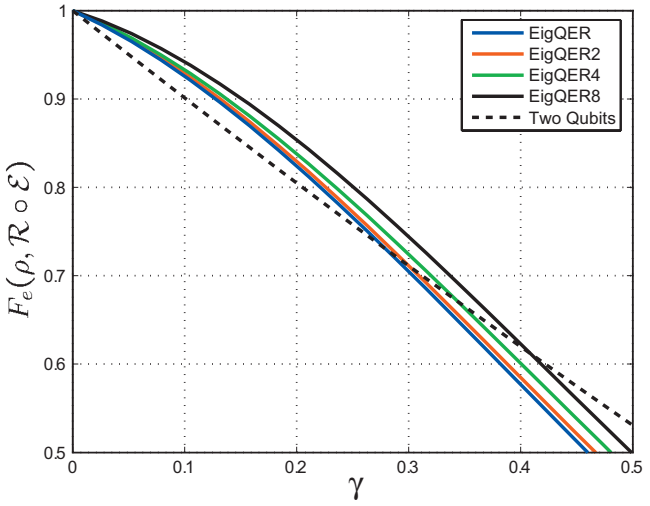

FIG. 8. (Color) BLOCKEIGQER for the amplitude-damping channel and a random [6,2] code. We compare the BLOCKEIGQER algorithm for block sizes of 2, 4, and 8 with the EIGQER algorithm. We see significant performance improvement for larger block sizes, at the cost of computational and recovery complexity. Baseline in this case is the entanglement fidelity for two qubits input to the channel without error correction.

Fixing the block size M ignores some of the inherent symmetries in the channel and encoding. In particular, it is quite common for $C_{E,\mathcal{E}}$ to have degenerate eigenvalues. By fixing the number of eigenvectors to simultaneously consider, one may inadvertently partition such a degenerate subspace according to the numerical precision of the eigenanalysis software. To avoid this unwanted circumstance, we may select a variable block size based on the magnitude of the eigenvalues. This approach necessitates a strategy for parsing the eigenvalues into variable size blocks, which can be a tricky procedure. Due to the modest returns of such an attempt, we have not pursued such a strategy.

While BLOCKEIGQER shows modest performance improvements when compared to EIGQER, it has one significant drawback. Unlike EIGQER, the recovery operation from BLOCKEIGQER is not constrained to a collection of isometries. Once the initial projective syndrome measurement is performed, the subsequent correction terms are arbitrary CPTP maps. This may complicate attempts to physically implement such an operation. Furthermore, BLOCKEIGQER does not provide much more intuition for recovery design than EIGQER. For this reason, we consider BLOCKEIGQER a numerical tool whose principal value is its incremental improvement approaching optimality. It also proves useful for the performance bounds derived in Sec. V.

B. ORDERQER

We now consider a block QER algorithm that provides intuition for error recovery design. We are often interested in

channels where each qubit is independently corrupted; thus the overall channel is the tensor product of single-qubit channels. We can use this structure to design an intuitive projective measurement. We illustrate using the classical bit-flip channel with probability of error p. If a single bit of the codeword is flipped, we label this a "first-order error" as the probability of such an error is $O(p)$. If two codeword bits are flipped, this is a "second-order error," which occurs with probability $O(p^2)$.

This intuition can easily yield a choice of syndrome subspaces $\{\mathcal{S}_k\}$. Consider, for example, the amplitude-damping channel given in (15). Recognizing E_1 as the error event, we declare first-order errors to be of the form $E_k^1 = E_0 \otimes \cdots \otimes E_1 \otimes E_0 \otimes \cdots$ where the error is on the kth qubit. In this case we can declare the first-order syndrome subspace to be

$$\mathcal{S}_1 = \text{span}(\{|E_0^{\otimes n} 0_L\rangle, |E_0^{\otimes n} 1_L\rangle, |E_1^1 0_L\rangle, |E_1^1 1_L\rangle, \ldots, |E_n^1 1_L\rangle\}), \tag{18}$$

where $|0_L\rangle$ and $|1_L\rangle$ are the logical codewords for an n-length code. We include the "no-error" term as numerical experience suggests that the code projector P_C is not always an optimal syndrome measurement. By parallel construction, we can define the second-order syndrome subspace $\mathcal{S}_2$. While these two will probably not complete the space $\mathcal{H}_C$, quite possibly we may neglect any higher orders. Alternatively, we can analyze the remaining subspace with either the SDP or the numerically simpler EIGQER algorithm. We will refer to this block SDP algorithm as ORDERQER.

The SDPs for first- and second-order subspaces significantly reduce the dimension from the full optimal SDP, though the effect is not as dramatic as with BLOCKEIGQER. Consider the case of the amplitude-damping channel which has only two operator elements for the single-qubit channel. For an $[n,k]$ code, there is one no-error operator and n first-order error operators. This suggests that $\mathcal{S}_1$ has dimension $(n+1)d_S = (n+1)2^k$. The SDP then has $(n+1)^2 2^{4k}$ optimization variables. Contrast this n^2 growth with the 4^n growth of the optimal SDP. For second-order errors, there are $\binom{n}{2} \approx \frac{n^2}{2}$ error operators. The subspace $\mathcal{S}_2$ has approximate dimensions of $n^2 2^{k-1}$ and thus the SDP has $n^4 2^{4k-2}$ optimization variables. For the [7,1] Steane code, computing the full optimal SDP requires an impractical $4^7 \times 4 = 65\,536$ variables. However, the first-order SDP requires $8^2 \times 2^4 = 1024$ variables and the actual second-order SDP has $42^2 \times 4 = 7056$ optimization variables. For contrast, the full SDP and the five-qubit code requires 1024 optimization variables. For the [9,1] Shor code, the second-order SDP has an impractical $72^2 \times 4 = 20\,736$ optimization variables. We therefore do not use ORDERQER for the Shor code.

While the scaling of ORDERQER grows quickly with n, making its use challenging for codes as long as nine qubits, ORDERQER results provide significant insight into the mechanism of channel adaptation. Consider the first- and second-order recovery performance for the Steane code and the amplitude-damping channel from Fig. 9. We note that the fidelity performance for the recovery from $\mathcal{S}_1$ is comparable to the performance of standard QEC, especially as γ approaches 0. This matches the intuition that standard QEC is correcting single qubit errors which are almost completely restricted to $\mathcal{S}_1$. For small γ, the most likely syndrome measurement will be a Pauli X or Y, as these characterize single-qubit dampings. These same errors are corrected by 1^{st} order ORDERQER. As γ grows, the distortion from the no-error term $E_0 \otimes \cdots \otimes E_0$ becomes more pronounced and the QEC outperforms first-order ORDERQER.

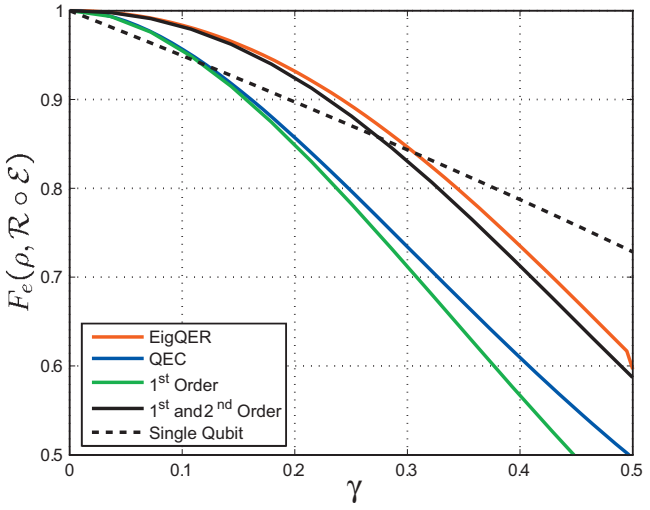

FIG. 9. (Color) ORDERQER recovery for the seven-qubit Steane code and the amplitude-damping channel. We compare the recovery fidelity of the first-order error to the standard QEC performance. The performance of the first- and second-order recoveries together are comparable to the EIGQER recovery, especially as γ approaches 0.

We see that first- and second-order recovery performance is quite comparable to the EIGQER performance. Thus, the performance gains observed for channel-adapted QER can be understood as corrections of higher-order errors. Since $\mathcal{S}_1$ has dimension significantly less than d_C and yet approximates the QEC recovery performance, it is only reasonable that the remaining redundancy of the code can be exploited to protect from further error.

V. QER PERFORMANCE UPPER BOUND

In the preceding sections, we imposed constraints on the recovery operations to provide structure and aid computation. While the resulting channel-adapted recoveries out perform the generic QEC recovery operation in all of the examples, the constraints essentially guarantee suboptimality. For the five-qubit code (where computation of the optimal QER operation is practical), we observe that the proposed algorithms (EIGQER, BLOCKEIGQER, and ORDERQER) closely approximate the optimal performance. This anecdotal evidence, however, is hardly sufficient to justify the bold description in the title of "near-optimal" channel-adapted QER. In this section, we more fully justify the near-optimal label by deriving channel-adapted performance bounds. We accomplish this by using the Lagrange dual function.

Every optimization problem has an associated dual problem [17]. Derived from the objective function and constraints of the original optimization problem (known as the *primal*

problem), the dual problem optimizes over a set of dual variables often subject to a set of dual constraints. The dual problem has several useful properties. First of all, the dual problem is always convex. In many cases, calculation of the dual function is a useful method for constructing optimization algorithms. Most important for our purposes, the dual function provides a bound for the value of the primal function. We define a *dual feasible point* as any set of dual variables satisfying the dual constraint. The dual function value for any dual feasible point is less than or equal to the primal function at any primal feasible point. (We have implicitly assumed the primal function to be a minimization problem, which is the canonical form.)

The dual function for channel-adapted recovery was derived in [7]; we will re-derive it here in a notation more convenient for our purposes.

The primal problem as given in [6] can be stated succinctly as

$$\min_X \; -\mathrm{tr}XC_{E,\mathcal{E}} \quad \text{such that } X \geq 0 \text{ and } \mathrm{tr}_{\mathcal{H}_S} X = I. \tag{19}$$

The negative sign on the $\mathrm{tr}XC_{E,\mathcal{E}}$ terms casts the primal problem as a minimization, which is the canonical form. The Lagrangian is given by

$$L(X,Y,Z) = -\mathrm{tr}XC_{E,\mathcal{E}} + \mathrm{tr}Y(\mathrm{tr}_{\mathcal{H}_S} X - I) - \mathrm{tr}ZX, \tag{20}$$

where Y and $Z \geq 0$ are operators that serve as the Lagrange multipliers for the equality and generalized inequality constraints, respectively. The dual function is the (unconstrained) infimum over X of the Lagrangian:

$$g(Y,Z) = \inf_X L(X,Y,Z) = \inf_X -\mathrm{tr}X(C_{E,\mathcal{E}} + Z - I \otimes Y) - \mathrm{tr}Y, \tag{21}$$

where we have used the fact that $\mathrm{tr}(Y\mathrm{tr}_{\mathcal{H}_S} X) = \mathrm{tr}(I \otimes Y)X$. Since X is unconstrained, note that $g(Y,Z) = -\infty$ unless $Z = I \otimes Y - C_{E,\mathcal{E}}$, in which case the dual function becomes $g(Y,Z) = -\mathrm{tr}Y$. Y and $Z \geq 0$ are the dual variables, but we see that the dual function depends only on Y. We can therefore remove Z from the function as long as we remember the constraint implied by $Z = I \otimes Y - C_{E,\mathcal{E}}$. Since Z is constrained to be positive semidefinite, this can be satisfied as long as $I \otimes Y - C_{E,\mathcal{E}} \geq 0$.

We now have the bounding relation $-\mathrm{tr}XC_{E,\mathcal{E}} \geq \mathrm{tr} - Y$ for all X and Y that are primal and dual feasible points, respectively. If we now reverse the signs so that we have a more natural fidelity maximization, we write

$$\bar{F}_e(E, \mathcal{R} \circ \mathcal{E}) = \mathrm{tr}X_{\mathcal{R}} C_{E,\mathcal{E}} \leq \mathrm{tr}Y, \tag{22}$$

where $\mathcal{R}$ is CPTP and $I \otimes Y - C_{E,\mathcal{E}} \geq 0$. To find the best bounding point Y, we solve the dual optimization problem

$$\min_Y \mathrm{tr}Y \quad \text{such that } I \otimes Y - C_{E,\mathcal{E}} \geq 0. \tag{23}$$

Notice that the constraint implies that $Y = Y^\dagger$. Note also that $Y \in \mathcal{L}(\mathcal{H}_C^*)$.

We will use the bounding property (22) of the dual function. Given any dual feasible point $Y \in \mathcal{L}(\mathcal{H}_C^*)$, we know that $\mathrm{tr}Y$ upper bounds $\bar{F}_e(E, \mathcal{R} \circ \mathcal{E})$ for all $\mathcal{R}$; Y is thus a certificate of convergence for a recovery operation.

To provide a good performance bound, it is desirable to find a dual feasible point with a small dual function value. Indeed, the best such bound is the solution to (23), that is, to find the dual feasible point with the smallest trace. However, finding the optimal Y is the equivalent of solving for the optimal recovery due to the strong duality of the SDP. As this suffers the same computational burden as computing the optimal recovery, we require an alternate method for generating useful dual feasible points. We will establish methods to convert the suboptimal recovery operations of the preceding sections into dual feasible points.

We need to determine a good dual feasible point beginning with one of the suboptimal recoveries computed by the EIGQER, BLOCKEIGQER, or ORDERQER algorithms. We utilize the structure of the suboptimal recovery operations to generate a dual feasible point. We present two methods that exploit the projective syndrome measurement to achieve performance bounds. The first bound is motivated by the proof of Theorem 3 in the Appendix, where the optimal dual feasible point is constructed for Pauli group errors. Beginning with this construction and the recovery generated by EIGQER, we use the Geršgorin disk theorem to generate a dual feasible point. The resulting dual function we denote the Geršgorin dual bound. The second construction iteratively generates dual feasible points given an initial infeasible point. While it is more computationally burdensome, it generates tighter bounds for the considered examples. We begin with a trial dual variable that may or may not be feasible and iteratively extend this point until it is feasible. We call this construction the iterative dual bound. We present several methods for providing an initial trial point.

Discussion of both bounding methods is facilitated by choosing an appropriate basis for $\mathcal{H}_S \otimes \mathcal{H}_C^*$. Both methods begin with a recovery operation generated by one of the structured suboptimal methods. As they all begin with a projective measurement, the recovery provides a partition of $\mathcal{H}_C$ into subspaces $\mathcal{S}_q$ of dimension d_q described by projection operators $\{P_q\} \in \mathcal{L}(\mathcal{H}_C)$. We are interested in a basis $\{|v_i\rangle\}_{i=1}^{2^{n+k}}$ where the first block of $d_S d_0$ basis vectors span $I \otimes \mathcal{S}_0^*$ and the qth block spans $I \otimes \mathcal{S}_q^*$. Let us define

$$(C_{E,\mathcal{E}})_{qq'} \equiv I \otimes \overline{P_q} C_{E,\mathcal{E}} I \otimes \overline{P_{q'}} \tag{24}$$

as we did in (17) and then write

$$C_{E,\mathcal{E}} = \begin{bmatrix} (C_{E,\mathcal{E}})_{00} & \cdots & (C_{E,\mathcal{E}})_{0q} & \cdots \\ \vdots & \ddots & \vdots & \\ (C_{E,\mathcal{E}})_{q0} & \cdots & (C_{E,\mathcal{E}})_{qq} & \\ \vdots & & & \ddots \end{bmatrix} \tag{25}$$

in our defined basis. This block structure delineates the relationship of the data operator $C_{E,\mathcal{E}}$ on each of the subspaces $\mathcal{S}_q$, which will be useful when discussing dual feasible points.

A. Geršgorin dual bound

The first method for constructing dual feasible points imposes a convenient structure on Y. In the case of Pauli group errors considered in [19], the optimal dual feasible point has the form

$$Y = \sum_q w_q \overline{P_q}, \tag{26}$$

where w_q are a set of weights corresponding to the probability of the most likely error resulting in the qth syndrome measurement. The form of (26) is appealing due its simplicity, especially for the EIGQER recovery operation, where the rank d_q of the P_q is constrained to be $\leq d_S$. While we cannot necessarily generate the optimal dual feasible point in this form for non-Pauli errors, we can use similar methods to generate a reasonable performance bound.

Before we state the Geršgorin dual bound, we take a second look at the optimal dual point for Pauli errors. For an $[n,k]$ stabilizer code, recall that $\mathcal{H}_C$ is partitioned into 2^{n-k} syndrome subspaces $\mathcal{S}_q$ and we establish a basis $\{|m\rangle_q\}$ for each subspace. We also determined that $|U_{Cq}^{\dagger} A_p\rangle\rangle$ is an eigenvector of $C_{E,\mathcal{E}}$. Note that $\{|U_{Cq}^{\dagger} A_p\rangle\rangle\}_{p=0}^{2^{2k}-1}$ span the space $I \otimes \overline{\mathcal{S}_q}$.

If we write out the operator $(C_{E,\mathcal{E}})_{qq}$ in this basis, we have

$$(C_{E,\mathcal{E}})_{qq} = \begin{bmatrix} a_{0q} & & \\ & \ddots & \\ & & a_{(2^{2k}-1)q} \end{bmatrix} \tag{27}$$

which is diagonal because $\{|m\rangle_q\}$ are eigenvectors of $C_{E,\mathcal{E}}$. This also implies that all of the off-diagonal blocks $(C_{E,\mathcal{E}})_{qq'}$ where $q \neq q'$ are also 0. We can now see that $Y = \Sigma_q \tilde{a}_q P_q$ where $\tilde{a}_q = \max_p |a_{pq}|$ is a dual feasible point, since

$$I \otimes Y^{\star} = \begin{bmatrix} \tilde{a}_0 I & 0 & \cdots & 0 \\ 0 & \tilde{a}_1 I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{a}_{2^{n-k}-1} I \end{bmatrix} \tag{28}$$

is diagonal in the chosen basis.

We return now to the general case. Unlike in the case of a Pauli error channel and a stabilizer code, we cannot guarantee that $C_{E,\mathcal{E}}$ will be either diagonal or block diagonal in this basis. However, if our suboptimal recovery $\mathcal{R}$ is generated from the EIGQER algorithm, then the subspaces $\mathcal{S}_q$ are selected based on the eigenvectors of $C_{E,\mathcal{E}}$, and we can expect $C_{E,\mathcal{E}}$ to be approximately block diagonal when we partition according to the subspaces $I \otimes \mathcal{S}_q^*$. We say that $C_{E,\mathcal{E}}$ is approximately block diagonal in this basis if $\|(C_{E,\mathcal{E}})_{qq}\| \gg \|(C_{E,\mathcal{E}})_{qq'}\|$ for $q \neq q'$.

To generate a dual feasible point of the form $Y = \Sigma_q w_q \overline{P_q}$, we need to choose w_q so that $I \otimes Y - C_{E,\mathcal{E}} \geq 0$. If $C_{E,\mathcal{E}}$ were exactly block diagonal in this basis, we could accomplish this by setting $w_q = \lambda_{\max}((C_{E,\mathcal{E}})_{qq})$. Since the block terms off the diagonal are not strictly 0, we must account for their contributions in the location of the eigenvalues of $C_{E,\mathcal{E}}$.

We will make use of a linear algebra theorem known as the Geršgorin disk theorem. This theorem provides bounds on the location in the complex plane of the eigenvalues of an arbitrary matrix. As will be evident, the theorem is most valuable when the matrix is dominated by its diagonal entries. We state the theorem as it is given in [20], Sec. 6.1

Theorem 2. Let $A = [a_{ij}] \in \mathbb{C}^{n \times n}$, and let

$$R_i'(A) \equiv \sum_{j=1, j \neq i}^{n} |a_{ij}|, \quad 1 \leq i \leq n, \tag{29}$$

denote the deleted absolute row sums of A. Then all the eigenvalues of A are located in the union of n disks

$$\bigcup_{i=1}^{n} \{z \in \mathbb{C} : |z - a_{ii}| \leq R_i'(A)\} \equiv G(A). \tag{30}$$

Furthermore, if a union of k of these n disks forms a connected region that is disjoint from all the remaining $n-k$ disks, then there are precisely k eigenvalues of A in this region.

Theorem 2 is particularly useful for proving the positivity of a matrix. The $R_i'(A)$ are the radii of disks centered at the diagonal entries a_{ii} and the eigenvalues are constrained to lie within the union of these disks. If A is a Hermitian matrix, then we can be certain it is positive semidefinite if $a_{ii} \geq R_i'(A)$ for all i as all of the eigenvalues would be constrained to lie to the right of the origin (or on the origin) on the real line.

We can apply Theorem 2 to generating a dual feasible point structured like (26). In this case we use the weights w_q to ensure that the diagonal entries of $I \otimes Y - C_{E,\mathcal{E}}$ are greater than the deleted absolute row sums. Let c_{ij} denote the matrix elements of $C_{E,\mathcal{E}}$ in our defined basis and let the basis vector $|v_i\rangle$ lie in the subspace $\mathcal{S}_q$. We then the have the ith diagonal element $[I \otimes Y - C_{E,\mathcal{E}}]_{ii} = w_q - c_{ii}$ and the ith deleted absolute row sum is $\Sigma_{i \neq j} |c_{ij}|$. We can assure non-negativity if

$$w_q \geq \sum_j |c_{ij}| \quad \text{for all } i \text{ such that } |v_i\rangle \in \mathcal{S}_q. \tag{31}$$

Thus, we can guarantee a dual feasible point if w_q is set to be the maximum absolute row sum for all rows i such that $|v_i\rangle \in \mathcal{S}_q$. We may express w_q concisely in terms of the induced ∞-norm ([20], Sec. 5.6.5), denoted $\|\cdot\|_\infty$:

$$w_q = \|[(C_{E,\mathcal{E}})_{q0} \cdots (C_{E,\mathcal{E}})_{qq} \cdots]\|_\infty \tag{32}$$

$$= \|I \otimes \overline{P_q} C_{E,\mathcal{E}}\|_\infty. \tag{33}$$

The Geršgorin disk theorem is a computationally simple way to guarantee construction of a dual feasible point given a partition of $\mathcal{H}_C$ into subspaces $\{\mathcal{S}_q\}$. Unfortunately, the induced infinity norm does not provide a particularly useful performance bound as can be seen in Fig. 10. When we compare to the optimal recovery performance for the five-qubit code and the amplitude-damping channel, we see that the dual bound is far from tight. In fact, for many values of γ, the bound is greater than 1, which is truly useless for upper bounding fidelities. While we have generated a dual point Y that is guaranteed to be feasible, such a guarantee imposes too strict a cost to have a useful bounding property.

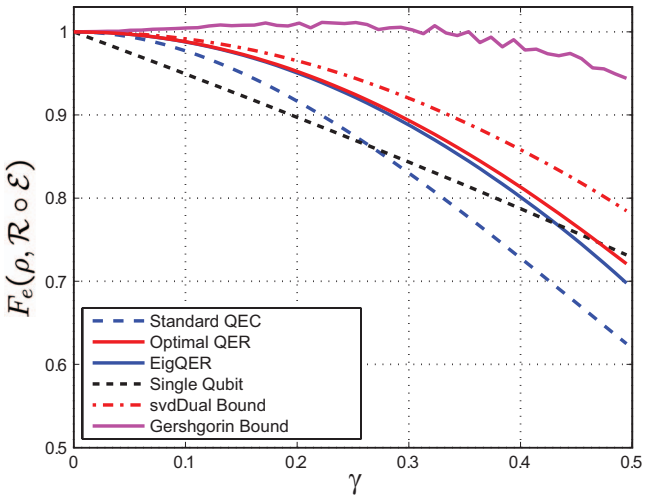

FIG. 10. (Color) Geršgorin and SVD dual bound for the amplitude-damping channel and the five-qubit stabilizer code. The Geršogrin bound is clearly not very useful as in some cases it is greater than 1. The SVD dual bound clearly tracks the optimal performance, although the departure from optimal of the bound exceeds the EIGQER recovery.

The Geršgorin dual bound provides useful insight for a tighter dual construction. If we replace the induced infinity norm with the induced 2-norm, we generate a dual point that is often dual feasible. That is, choose

$$w_q = \|I \otimes \overline{P}_q C_{E,\mathcal{E}}\|_2 \tag{34}$$

$$= \max_{|x\rangle\rangle} \langle\langle x| I \otimes \overline{P}_q C_{E,\mathcal{E}} |x\rangle\rangle \tag{35}$$

$$= \sigma_{\max}(I \otimes \overline{P}_q C_{E,\mathcal{E}}), \tag{36}$$

where $\sigma_{\max}(\cdot)$ in (36) indicates the maximum singular value and is the computational method for the induced 2-norm. We will refer to this construction as the singular value decomposition (SVD) dual point. The Y generated in this way is not guaranteed to be dual feasible as was the case with the ∞-norm, but has proven to be dual feasible in all of the examples that we have tried. If for some circumstance the SVD dual point is not feasible, it can be iteratively adjusted to become dual feasible in a manner we present in the following section.

B. Iterative dual bound

We now present an iterative procedure to generate a dual feasible point given an initial dual point $Y^{(0)}$ that is presumably not dual feasible. After presenting the algorithm, we will discuss choices for the initial dual point.

At the kth iteration, we update the dual point to produce $Y^{(k)}$ until we achieve feasibility. For convenience we will define

$$Z^{(k)} \equiv I \otimes Y^{(k)} - C_{E,\mathcal{E}}. \tag{37}$$

Let x and $|x\rangle\rangle$ be the smallest eigenvalue and associated eigenvector of $Z^{(k)}$. If $x \geq 0$, we may stop, as $Y^{(k)}$ is already dual feasible. If $x \leq 0$, we wish to update $Y^{(k)}$ a small amount to ensure that $\langle\langle x| Z^{(k+1)} |x\rangle\rangle \geq 0$. Essentially, we are replacing a negative eigenvalue with a 0 eigenvalue. Given no constraints on the update, we could accomplish this as $Z^{(k+1)} = Z^{(k)} + x |x\rangle\rangle\langle\langle x|$ but we must instead update $Y^{(k)}$ with the tensor product structure implicit.

We determine the properly constrained update by means of the Schmidt decomposition of the eigenvector:

$$|x\rangle\rangle = \sum_i \lambda_i |\hat{x}_i\rangle_{\mathcal{H}_S} |\tilde{x}_i\rangle_{\mathcal{H}_C^*}. \tag{38}$$

As we can only perturb $Z^{(k)}$ in the $\mathcal{H}_C^*$ slot, we choose the smallest perturbation guaranteed to achieve $\langle\langle x| Z^{(k+1)} |x\rangle\rangle \geq 0$. Let

$$Y^{(k+1)} = Y^{(k)} + \frac{|x|}{|\lambda_1|^2} |\tilde{x}_1\rangle\langle\tilde{x}_1|. \tag{39}$$

Then

$$\langle\langle x| Z^{(k+1)} |x\rangle\rangle = x + \frac{|x|}{|\lambda_1|^2} \langle\langle x| (I \otimes |\tilde{x}_1\rangle\langle\tilde{x}_1|) |x\rangle\rangle \tag{40}$$

$$= x + \frac{|x|}{|\lambda_1|^2} |\lambda_1|^2 \tag{41}$$

$$= 0, \tag{42}$$

since $x < 0$. While we have not yet guaranteed that $Z^{(k+1)} \geq 0$, $|x\rangle\rangle$ is no longer associated with a negative eigenvalue. By repeatedly perturbing $Y^{(k)}$ in this manner, we iteratively approach a dual feasible point while adding as little as possible to the dual function value $\mathrm{tr} Y^{(k)}$.

As a final point, we demonstrate that the iterative procedure will converge to a dual feasible point. Let us consider the effect of the kth iteration on the space orthogonal to $|x\rangle\rangle$. Let $|y\rangle\rangle \in \mathcal{H}_S \otimes \mathcal{H}_C^*$ be orthogonal to $|x\rangle\rangle$. Then, for $Z^{(k+1)}$ we see that

$$\langle\langle y| Z^{(k+1)} |y\rangle\rangle = \langle\langle y| Z^{(k)} |y\rangle\rangle + \frac{|x|}{|\lambda_1|^2} \langle\langle y| (I \otimes |\tilde{x}_1\rangle\langle\tilde{x}_1|) |y\rangle\rangle. \tag{43}$$

But since $I \otimes |\tilde{x}_1\rangle\langle\tilde{x}_1| \geq 0$ we see that

$$\langle\langle y| Z^{(k+1)} |y\rangle\rangle \geq \langle\langle y| Z^{(k)} |y\rangle\rangle \tag{44}$$

for all $|y\rangle\rangle \in \mathcal{H}_S \otimes \mathcal{H}_C^*$. We see that the update to $Y^{(k)}$ moved one negative eigenvalue to 0 while no new negative eigenvalues can be created. Thus the procedure will eventually converge to a dual feasible point.

C. Initial dual points

Having established a procedure to generate a dual feasible point given an arbitrary intial point $Y^{(0)}$, we now present initialization options. While we can start with any Hermitian operator in $\mathcal{L}(\mathcal{H}_C^*)$ including 0, we do not recommend such an unstructured choice as each iteration is imperfect. Each iteration adds $|x|/|\lambda_1|^2$ to the dual function value. If $|\lambda_1|$ is

not close to 1, the iteration is not efficient. We will use more educated initializations to begin closer to feasibility, thus minimizing the number of iterations and improving the bounding properties of the resulting dual feasible point.

We have already presented one method for initialization with the SVD dual point. In most cases we have seen, this point is already feasible and in fact is a relatively loose bound. Its advantage lies in its easy computation, but other choices provide better bounding properties. We would prefer an initial $Y^{(0)}$ such that $Z^{(0)}$ is nonpositive with eigenvalues very close to 0. If this is the case, we will require only small perturbations (and thus a small dual function value) to achieve a positive semidefinite $Z^{(k)}$.

Consider an initial $Y^{(0)}$ of the form given in (26). We choose an initial $Y^{(0)}$ in the same way that was used in the proof of Theorem 3:

$$w_q = \lambda_{\max}((C_{E,\mathcal{E}})_{qq}). \tag{45}$$

This is very simple to calculate, though it will not generally be dual feasible. This is the logical choice when we begin with the EIGQER recovery, as the only useful information we have is the projective syndrome measurement. This initialization often iterates to a better bound than the SVD dual point and requires no further information than the partition $\{\mathcal{S}_q\}$ provided by any of the structured QER methods. It has one drawback, however, in that $Z^{(0)}$ almost certainly has eigenvalues much greater than 0. For the $|v_i\rangle$ associated with the largest eigenvalue of $(C_{E,\mathcal{E}})_{qq}$, $\langle v_i|Z^{(0)}|v_i\rangle=0$. However, unless $(C_{E,\mathcal{E}})_{qq}$ has only one distinct eigenvalue there will be vectors $|x\rangle\rangle \in \mathcal{S}_q$ such that $\langle\langle x|\, Z^{(0)}\, |x\rangle\rangle \geq 0$, and perhaps quite large, relatively. Such vectors indicate portions of the Hilbert space where $Y^{(0)}$ is already greater than the optimal dual feasible point. While this likely cannot be avoided in the iterations, it seems wasteful to begin at such a point if not necessary.

We have an alternative choice for $Y^{(0)}$ arising from the block SDP QER algorithms of Sec. IV. These algorithms already provide information useful for generating a dual feasible point. When solving the SDP on a subspace $\mathcal{S}_q$ one can simultaneously generate the optimal dual function value $Y_q^\star \in \mathcal{L}(\mathcal{S}_q^*)$. Given such optimal subspace dual points, define the block diagonal operator

$$Y^{(0)} = \begin{bmatrix} Y_0^\star & & & \\ & \ddots & & \\ & & Y_q^\star & \\ & & & \ddots \end{bmatrix} \tag{46}$$

as the initial point. We know that $I \otimes Y_q^\star - (C_{E,\mathcal{E}})_{qq} \geq 0$, so there will be $|x\rangle\rangle$ for which $\langle\langle x|\, Z^{(0)}\, |x\rangle\rangle \geq 0$. However, since $Y_q^\star$ is optimal within $\mathcal{L}(\mathcal{S}_q^*)$, we know that we are not being overly wasteful with the initialization.

D. Iterated block dual

Let us consider the computational burden of the iterated dual bound. At each iteration we must compute the smallest eigenvalue and associated eigenvector of $Z^{(k)}$, a $2^{n+k} \times 2^{n+k}$ Hermitian matrix. (We can accomplish this by looking for the largest eigenvalue of $\eta I - Z^{(k)}$ where $\eta \geq 1$ is an arbitrary offset to ensure positivity.) This must be repeated at most 2^{n+k} times to ensure dual feasibility, though there may be significantly fewer iterations if the $Z^{(0)}$ is nearly positive semidefinite already. As mentioned in Sec. III, this can be accomplished in $O(2^{2(n+k)})$ flops by the power method. This is very costly if we must repeat the iteration many times.

The block diagonal structure of the initial points suggests a slightly modified alternative procedure with some computational advantage. Consider the optimal dual points Y_i and Y_j in $\mathcal{L}(\mathcal{S}_i^*)$ and $\mathcal{L}(\mathcal{S}_j^*)$. We can use the same iterative procedure as before to compute a dual feasible $Y_{ij} \in \mathcal{L}(\mathcal{S}_i^* \oplus \mathcal{S}_j^*)$ requiring only $O(2^{2k}(d_i+d_j)^2)$ flops per iteration with a maximum of $2^k(d_i+d_j)$ iterations. We can generate a dual feasible point on the whole space $\mathcal{L}(\mathcal{H}_C^*)$ by successively combining subspace blocks. Eventually we will have to iterate over the full space, but we will have done most of the work in the smaller blocks, and the full $2^{n+k} \times 2^{n+k}$ eigendecomposition will require few iterations.

In the examples we have processed, the iterated block dual procedure created nearly identical bounds (often within 10^{-5} of each other and never more than 10^{-4}) as the original algorithm. The computational burden is reduced by approximately 20%.

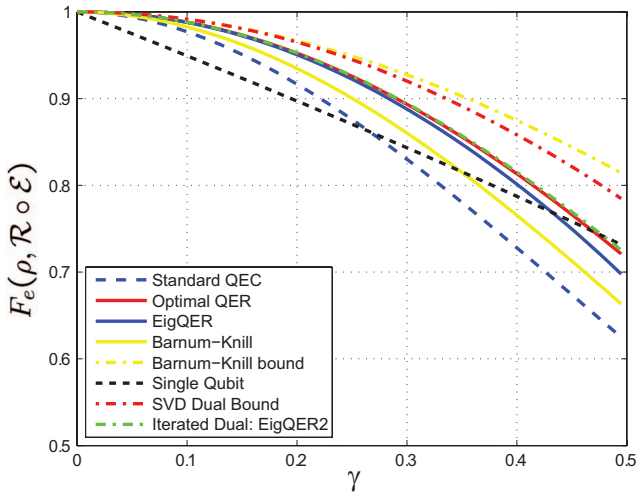

FIG. 11. (Color) Dual bound comparison for the amplitude-damping channel and the five-qubit code. The iterated dual initialized with the BLOCKEIGQER algorithm with $M=2$ is essentially indistinguishable from the optimal recovery performance, thus producing a very tight bound. Included for comparison are the EIGQER performance, the SVD dual bound, and both a channel-adapted recovery and associated bound derived by Barnum and Knill in [10].

E. Examples

We provide several examples to demonstrate the utility of the iterated dual bound. At the same time, we we illustrate the near-optimality of the structured QER algorithms. In Fig. 11, we show several bounds for channel-adapted QER for the amplitude-damping channel and the five-qubit code. In this

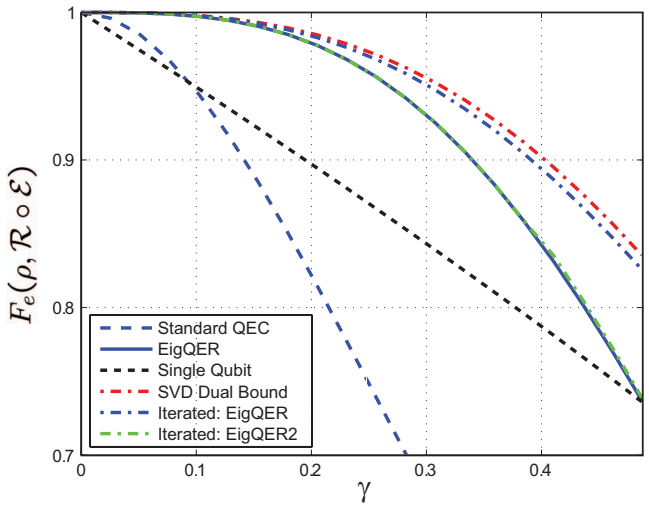

FIG. 12. (Color) Dual bound comparison for the amplitude-damping channel and the nine-qubit Steane code. The iterated dual bound initialized with the BLOCKEIGQER recovery with $M=2$ produces a bound that is tight to the EIGQER recovery operation. This demonstrates that the EIGQER recovery operation is essentially optimal in this case. Notice that the iterated bound initialized with the EIGQER recovery operation does not generate a tight bound.

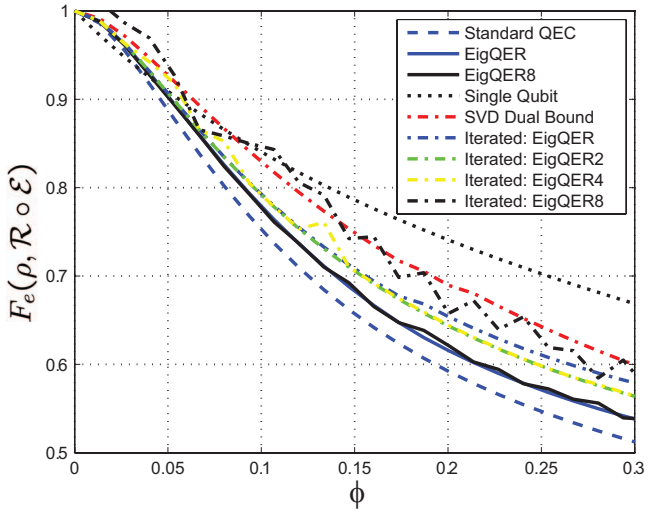

FIG. 13. (Color) Dual bound comparison for the pure state rotation channel with $\theta=5\pi/12$ and the seven-qubit Steane code. Note that the iterated bounds are generally, though not universally, better than the SVD dual bound. We also see that the shorter block lengths for the BLOCKEIGQER algorithm generally produce a tighter bound, despite slightly poorer recovery performance.

case, we know the optimal performance and can see that the iterated dual bound, beginning with the BLOCKEIGQER with $M=2$, is quite tight. This is in contrast to the SVD dual bound, which was also shown in Fig. 10. We have included in Fig. 11 the numerical channel-adapted recovery and performance bound from [10]. We see that this bound is looser than even the SVD dual bound for this example.

Figure 12 shows several dual bounds for the amplitude-damping channel and the nine-qubit Shor code. While we cannot compute the optimum directly, we see that the EIGQER performance curve and the iterated bound derived from BLOCKEIGQER with $M=2$ are essentially equivalent. We can conclude that the EIGQER operation is essentially optimal in this case. While not shown, iterations for BLOCKEIGQER with $M=4$ and $M=8$ achieved essentially the same bound. Note that neither the SVD dual bound nor the iterated bound beginning with the EIGQER recovery operation is tight, illustrating the importance of a good initialization for the dual iterations.

Our final example is the pure state rotation channel with $\theta=5\pi/12$ and the seven-qubit Steane code. In Fig. 13, we can distinguish between several initialization methods for the dual iterative bound. We see that none of the recovery operations approach the bound performance for large ϕ, though the performance is relatively tight as the noise level drops ($\phi \to 0$). Notice that in general the iterative bounds are better than the SVD dual bound; however, there are points, especially for the BLOCKEIGQER algorithm with $M=8$, where the iterated bound is poor. It is interesting to note that the longer block lengths (larger M) usually generate better recovery performance (which can be seen with slight improvement even in this case) yet often produce poorer bounds. Anecdotal experience suggests that the best iterative starting point is the BLOCKEIGQER recovery operation with $M=2$.

Finally, we should point out the gap for large ϕ between the recovery performance and the dual bounds. Absent a better recovery operation or a smaller performance bound, we have no way to know whether the bound or the recovery is further removed from the optimal. However, this region is below the baseline performance for a single unencoded qubit, and thus is not of serious concern.

VI. CONCLUSION AND FUTURE WORK

Adapting a quantum recovery operation to a physical channel can significantly improve the effectiveness of a quantum channel. In this way, quantum error correction can be made more efficient, which should aid in scaling physical implementations to a larger number of qubits. While the optimal recovery (in terms of average entanglement fidelity) may be calculated via convex optimization of a semidefinite program, we have derived a class of near-optimal algorithms that are less computationally intensive. Furthermore, these algorithms yield recovery operations of a particular form: they implement a projective error syndrome measurement followed by a syndrome recovery operation. This structure may prove easier to implement physically and provides intuition into the mechanism for channel adaptation.

Despite the reduction in computation from the SDP, even these algorithms grow exponentially in the length of (i.e., the number of qubits in) the code. For this reason, the next step toward practical application of channel-adapted quantum error correction must include analytical tools to supplement these numerical techniques. Furthermore, to apply channel-adapted methods to fault-tolerant quantum computing, we must show how errors propagate from block to block. These two open questions are likely closely linked. Despite these obstacles, the added efficiency of channel-adapted recovery

suggests significant value for practical efforts in quantum error correction.

ACKNOWLEDGMENTS

A.S.F. would like to thank the Department of the Air Force, who sponsored this work under AF Contract No. FA8721-05-C-0002. All authors thank the National Science Foundation for support through Grant No. CCF-0431787.

APPENDIX

The discussion of the Geršgorin and SVD dual bounds make use of a structured dual feasible point. This is motivated by the optimal dual feasible point for a stabilizer code and a Pauli error channel. Construction of this optimal dual feasible point proves the intuitive structure of the optimal recovery operation for Pauli error channels. This theorem was proven in [19] and will be restated here for reference.

We can construct the optimal recovery operation for a stabilizer code when the channel $\mathcal{E}'$ is characterized by Pauli group errors and the input ensemble is the completely mixed state. That is, E is given by $\rho=I/d_S$ with $p=1$ and the channel can be represented by Kraus operators $\{E_i\}$ where each E_i is a scaled element of the Pauli group. (Notice that this does not require every set of Kraus operators that characterize $\mathcal{E}'$ to be scaled elements of the Pauli group, since unitary combinations of Pauli group elements do not necessarily belong to the Pauli group.)

To state the optimal recovery, we carefully define the syndrome measurement subspaces and the Pauli group operators that connect the subspaces. We must do this in a way to consistently describe the normalizer operations of the code. Consider an $[n,k]$ stabilizer code with generators $\langle g_1,\ldots,g_{n-k}\rangle$ and logical $\bar{Z}$ operators $\bar{Z}_1,\ldots,\bar{Z}_k$ such that $\{g_1,\ldots,g_{n-k},\bar{Z}_1,\ldots,\bar{Z}_k\}$ form an independent and commuting set. Define logical $\bar{X}$ operators such that $[\bar{X}_i,g_j]=[\bar{X}_i,\bar{X}_j]=0 \ \forall \ i,j$, $[\bar{X}_i,\bar{Z}_j]=0$ for $i\neq j$, and $\{\bar{X}_i,\bar{Z}_i\}=0$.

The syndrome subspaces correspond to the intersection of the ± 1 eigenspaces of each generator. Accordingly, we label each space $\mathcal{S}_q$ where $q=0,1,\ldots,2^{n-k}-1$, where $\mathcal{S}_0$ corresponds to the code subspace. Let P_q be the projection operator onto $\mathcal{S}_q$. Let $\{|i_1 i_2\cdots i_k\rangle_q\}$ form a basis for $\mathcal{S}_q$ such that

$$\bar{Z}_1\bar{Z}_2\cdots\bar{Z}_k|i_1 i_2\cdots i_k\rangle_q=(-1)^{i_1}(-1)^{i_2}\cdots(-1)^{i_k}|i_1 i_2\cdots i_k\rangle_q, \tag{A1}$$

where $i_j\in\{0,1\}$. In this way, we have a standardized basis for each syndrome subspace which can also be written as $\{|m\rangle_q\}$, $m=0,\ldots,2^k-1$.

Let us recall the effect of a unitary operator on a stabilizer state. If $|\psi\rangle$ is stabilized by $\langle g_1,\ldots,g_{n-k}\rangle$, then $U|\psi\rangle$ is stabilized by $\langle Ug_1U^\dagger,\ldots,Ug_{n-k}U^\dagger\rangle$. What happens if $U\in G_n$, the Pauli group on n qubits? In that case, since U either commutes or anticommutes with each stabilizer, $U|\psi\rangle$ is stabilized by $\langle\pm g_1,\ldots,\pm g_{n-k}\rangle$ where the sign of each generator g_i is determined by whether it commutes or anticommutes with U. Thus, a Pauli group operator acting on a state in the code subspace $\mathcal{S}_0$ will transform the state into one of the subspaces $\mathcal{S}_q$.

We have established that the Pauli group errors always rotate the code space onto one of the stabilizer subspaces, but this is not yet sufficient to determine the proper recovery. Given that the system has be transformed to subspace $\mathcal{S}_q$, we must still characterize the error by what happened within the subspace. That is to say, the error consists of a rotation to a syndrome subspace and a normalizer operation within that subspace.

Let us characterize these operations using the bases $\{|m\rangle_q\}$. Define $W_{qq'}\equiv\Sigma_m|m\rangle_{q'q}\langle m|$ as the operator that transforms $\mathcal{S}_q\mapsto\mathcal{S}_{q'}$ while maintaining the ordering of the basis. Define the encoding isometry $U_C\equiv\Sigma_m|n\rangle_{0S}\langle n|$ where $|n\rangle_S\in\mathcal{H}_S$, the source space. Further define $U_{cq}\equiv W_qU_C$, the isometry that encodes the q^{th} syndrome subspace. We will define the 4^k code normalizer operators as

$$A_p\equiv\bar{X}_1^{i_1}\bar{X}_2^{i_2}\cdots\bar{X}_k^{i_k}\bar{Z}_1^{j_1}\bar{Z}_2^{j_2}\cdots\bar{Z}_k^{j_k} \tag{A2}$$

where p is given in binary notation as $i_1i_2\cdots i_kj_1j_2\cdots j_k$. Notice that, if a similarly defined A_p^S is an element of the Pauli group $\mathcal{G}_k\in\mathcal{L}(\mathcal{H}_S)$ with generators $\langle X_1^S,\ldots,X_k^S,Z_1^S,\ldots,Z_k^S\rangle$, we can conclude that $A_pU_C=U_CA_p^S$.

The preceding definitions were chosen to illustrate the following facts. First, we can see by the definitions that $[W_{qq'},A_p]=0$. That is, $W_{qq'}$ characterizes a standard rotation from one syndrome subspace to another, and A_p characterizes a normalizer operation within the subspace. These have been defined so that they can occur in either order. Second, let $\mathcal{E}'$ be a quantum channel represented by operator elements that are scaled members of the Pauli group $\mathcal{G}_n$. Then the composite channel $\mathcal{E}$ which includes the encoding isometry U_C can be represented by operator elements of the form

$$\{E_{pq}=a_{pq}A_pW_qU_C=a_{pq}A_pU_{Cq}\}, \tag{A3}$$

where the CPTP constraint requires $\Sigma_{pq}|a_{pq}|^2=1$.

We can understand the amplitudes a_{pq} by noting that, with probability $|a_{pq}|^2$, the channel $\mathcal{E}$ transforms the original state to $\mathcal{S}_q$ and applies the normalizer operation A_p. To channel-adaptively recover, we project onto the stabilizer subspaces $\{\mathcal{S}_q\}$ and determine the most likely normalizer operation for each syndrome subspace $\mathcal{S}_q$. Let $p_q=\arg\max_p|a_{pq}|^2$, and let $\tilde{a}_q\equiv a_{p_qq}$. With these definitions in place, we can state the following theorem.

Theorem 3. Let $\mathcal{E}$ be a channel in the form of (A3), i.e., a stabilizer encoding and a channel with Pauli group error operators. For a source in the completely mixed state $\rho=I/d_S$, the optimal channel-adapted recovery operation is given by $\mathcal{R}\sim\{U_{Cq}^\dagger A_{p_q}\}$, which is the stabilizer syndrome measurement followed by maximum likelihood normalizer syndrome correction.

Proof. We prove Theorem 3 by constructing a dual feasible point Y such that the dual function value tr$Y$ is equal to the entanglement fidelity $F_e(\rho,\mathcal{R}\circ\mathcal{E})$.

We begin by calculating $F_e(\rho,\mathcal{R}\circ\mathcal{E})$. For later convenience, we will do this in terms of the Choi matrix $C_{E,\mathcal{E}}$ from (4). We write the entanglement fidelity in terms of the recovery operator elements $|U_{Cq}^\dagger A_{p_q}\rangle\rangle$:

$$F_e(\rho,\mathcal{R}\circ\mathcal{E}) = \mathrm{tr} X_{\mathcal{R}} C_{E,\mathcal{E}} \tag{A4}$$

$$= \sum_{q'} \langle\langle U_{Cq'}^{\dagger} A_{p_{q'}} | \, C_{E,\mathcal{E}} \, | U_{Cq'}^{\dagger} A_{p_{q'}} \rangle\rangle. \tag{A5}$$

To evaluate (A5), we note that

$$\langle\langle \rho U_{Cq}^{\dagger} A_p | U_{Cq'}^{\dagger} A_{p_{q'}} \rangle\rangle = \mathrm{tr} A_p U_{Cq} \rho U_{Cq'}^{\dagger} A_{p_{q'}} \tag{A6}$$

$$= \mathrm{tr} A_p W_q U_C \rho U_C^{\dagger} W_{q'}^{\dagger} A_{p_{q'}} \tag{A7}$$

$$= \mathrm{tr} A_p W_{q'}^{\dagger} W_q U_C \rho U_C^{\dagger} A_{p_{q'}} \tag{A8}$$

$$= \delta_{qq'} \mathrm{tr} A_p U_C \rho U_C^{\dagger} A_{p_{q'}} \tag{A9}$$

$$= \delta_{qq'} \mathrm{tr} A_p^C \rho A_{p_{q'}}^C. \tag{A10}$$

We have used the commutation relation $[W_{qq'}, A_p]=0$ to arrive at (A8) and the facts that $W_{q'}^{\dagger} W_q = \delta_{qq'} P_0$ and $P_0 U_C = U_C$ to conclude (A9). Since $\rho = I/d_S$ and $\mathrm{tr} A_p^C A_{p_{q'}}^C = \delta_{pp_{q'}} d_S$, we see that $\mathrm{tr} A_p^C \rho A_{p_{q'}}^C = \delta_{pp_{q'}}$. Thus,

$$\langle\langle \rho U_{Cq}^{\dagger} A_p | U_{Cq'}^{\dagger} A_{p_{q'}} P_{q'} \rangle\rangle = \delta_{pp_{q'}} \delta_{qq'}. \tag{A11}$$

Using (A11), it is straightforward to evaluate (A5):

$$F_e(\rho,\mathcal{R}\circ\mathcal{E}) = \sum_{pqq'} |a_{pq}|^2 \left| \langle\langle \rho U_{Cq}^{\dagger} A_p | U_{Cq'}^{\dagger} A_{p_{q'}} \rangle\rangle \right|^2 \tag{A12}$$

$$= \sum_{pqq'} |a_{pq}|^2 \delta_{qq'} \delta_{pp_{q'}} \tag{A13}$$

$$= \sum_{q} |\tilde{a}_q|^2. \tag{A14}$$

We now propose the dual point $Y = \Sigma_q |\tilde{a}_q|^2 \overline{P_q}/d_S$. Since

$$\mathrm{tr} Y = \sum_q |\tilde{a}_q|^2 \mathrm{tr} \overline{P_q}/d_S \tag{A15}$$

$$= \sum_q |\tilde{a}_q|^2 \tag{A16}$$

$$= F_e(\rho,\mathcal{R}\circ\mathcal{E}), \tag{A17}$$

we complete the proof by demonstrating that

$$I \otimes Y - C_{E,\mathcal{E}} \geq 0, \tag{A18}$$

i.e., Y is a dual feasible point. We show this by demonstrating that $I \otimes Y$ and $C_{E,\mathcal{E}}$ have the same eigenvectors, and that the associated eigenvalue is always greater for $I \otimes Y$.

By the same argument used for (A11), we note that

$$\langle\langle \rho U_{Cq}^{\dagger} A_p | \rho U_{Cq'}^{\dagger} A_{p'} \rangle\rangle = \delta_{pp'} \delta_{qq'} / d_S^2. \tag{A19}$$

This means that $|\rho U_{Cq}^{\dagger} A_p \rangle\rangle$ is an eigenvector of $C_{E,\mathcal{E}}$ with eigenvalue $|a_{pq}|^2/d_S$. We normalize the eigenvector to unit length and apply it to $I \otimes Y$:

$$I \otimes Y \, |\rho U_{Cq}^{\dagger} A_p/d_S \rangle\rangle = \sum_{q'} |\tilde{a}_{q'}|^2 \overline{P_{q'}}/d_S \, |\rho U_{Cq}^{\dagger} A_p/d_S \rangle\rangle \tag{A20}$$

$$= \frac{1}{d_S} \sum_{q'} |\tilde{a}_{q'}|^2 \, |\rho U_{Cq}^{\dagger} A_p P_{q'}/d_S \rangle\rangle \tag{A21}$$

$$= \frac{1}{d_S} |\tilde{a}_q|^2 \, |\rho U_{Cq}^{\dagger} A_p/d_S \rangle\rangle. \tag{A22}$$

Thus we see that $|\rho U_{Cq}^{\dagger} A_p \rangle\rangle$ is an eigenvector of $I \otimes Y$ with eigenvalue $|\tilde{a}_q|^2/d_S \geq |a_{pq}|^2/d_S \;\; \forall \;\; p$. Thus $I \otimes Y - C_{E,\mathcal{E}} \geq 0$ and Y is a dual feasible point. ■

[1] P. W. Shor, Phys. Rev. A **52**, R2493 (1995).
[2] A. M. Steane, Phys. Rev. Lett. **77**, 793 (1996).
[3] A. R. Calderbank and P. W. Shor, Phys. Rev. A **54**, 1098 (1996).
[4] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).
[5] R. Laflamme, C. Miquel, J.-P. Paz, and W. H. Zurek, Phys. Rev. Lett. **77**, 198 (1996).
[6] A. S. Fletcher, P. W. Shor, and M. Z. Win, Phys. Rev. A **75**, 012338 (2007).
[7] R. L. Kosut and D. A. Lidar, e-print arXiv:quant-ph/0606078.
[8] M. Reimpell and R. F. Werner, Phys. Rev. Lett. **94**, 080501 (2005).
[9] N. Yamamoto, S. Hara, and K. Tsumura, Phys. Rev. A **71**, 022322 (2005).
[10] H. Barnum and E. Knill, J. Math. Phys. **43**, 2097 (2002).
[11] G. M. D'Ariano and P. Lo Presti, Phys. Rev. A **64**, 042308 (2001).
[12] T. F. Havel, J. Math. Phys. **44**, 534 (2003).
[13] J. Tyson, J. Phys. A **36**, 10101 (2003).
[14] M.-D. Choi, Linear Algebr. Appl. **10**, 285 (1975).
[15] C. M. Caves, J. Supercond. **12**, 707 (1999).
[16] J. de Pillis, Pac. J. Math. **23**, 129 (1967).
[17] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, U.K., 2004).
[18] D. W. Leung, M. A. Nielsen, I. L. Chuang, and Y. Yamamoto, Phys. Rev. A **56**, 2567 (1997).
[19] A. S. Fletcher, Ph.D. thesis, Massachusetts Institute of Technology, 2007.
[20] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, U.K., 1985).