Reversing quantum dynamics with near-optimal quantum and classical fidelity

H. Barnum1, E. Knill2
1Dept. of Computer Science, University of Bristol, Bristol BS8 1UB, UK. E-mail: hbarnum@cs.bris.ac.uk
2Los Alamos National Laboratories, Mail Stop B265, Los Alamos, NM 87545. E-mail: knill@lanl.gov.
(November 4, 2018)

We consider the problem of reversing quantum dynamics, with the goal of preserving an initial state’s quantum entanglement or classical correlation with a reference system. We exhibit an approximate reversal operation, adapted to the initial density operator and the “noise” dynamics to be reversed. We show that its error in preserving either quantum or classical information is no more than twice that of the optimal reversal operation. Applications to quantum algorithms and information transmission are discussed.

PACS numbers: 03.67-a, 03.67.Hk, 5.40.Ca, 89.70.+c

Introduction. Counteracting the effects on quantum systems of noise generated by interaction with an environment is a central problem for the emerging field of quantum information processing. It’s solution can be expected to have applications in quantum computation, precision measurement and information transmission. In this Letter we exhibit a reversal operation which takes account of both the noise and the initial density operator, to achieve near-optimal preservation of the initial density operator’s quantum entanglement or classical correlation with a reference system.

We model quantum noise in a system $Q$ by the most general dynamics that can arise via a unitary interaction $U_{QE}$ with an environment $E$ in initial state $|0^E\rangle$. This is a trace-preserving completely positive map $\mathcal{A}: \rho \rightarrow \sum_i A_i \rho A_i^\dagger$, where $A_i := \langle i^E| U_{QE}|i^E\rangle$. The $A_i$ form a decomposition of $\mathcal{A}$, $\mathcal{A} \sim \{A_i\}$. When such a map $\mathcal{A}$ acts on $Q$, an entangled state $|\psi_{0^Q}\rangle := \sum_i \sqrt{p_i}|i^R\rangle|i^Q\rangle$ of $Q$ with a reference system $R$ evolves as:

$$|\Psi_0\rangle := \sum_i \sqrt{p_i}|i^R\rangle|i^Q\rangle|0^E\rangle \rightarrow |\Psi_f\rangle := \sum_{ij} \sqrt{p_i p_j} |i^R\rangle A_i |j^Q\rangle|j^E\rangle.$$  \hfill (1)

The entanglement fidelity $F_c(\rho, \mathcal{A})$ is defined as $\|P_0 \otimes I_E|\Psi_f\rangle\|^2$, where $P_0 := |\psi_0^Q\rangle|\psi_0^Q\rangle \otimes I_E$. Thus $F_c$ is the squared norm of the projection of the final state in $E$ onto the subspace associated with the initial entangled state $|\psi_0^{RQ}\rangle$. It depends only on $\rho := \sum_i p_i |i^S\rangle|i^Q\rangle$, and is given by $\sum_i |\text{tr} A_i \rho|^2$. When $\rho = |\psi\rangle|\psi\rangle$, it is equal to the input-output fidelity $\langle \psi|\mathcal{A}|\psi\rangle$. For an ensemble $E = \{\rho_i\}$ where state $\rho_i$ occurs with probability $p_i$, we define the average entanglement fidelity by $F_c(E, \mathcal{A}) := \sum_i p_i F_c(\rho_i, \mathcal{A})$. A special case is

$$F_{cl}(\rho, \mathcal{A}) := \sum_i p_i \langle i|\mathcal{A}|i\rangle \langle i|i\rangle = F_c(\{p_i|i\rangle|\langle i|\}, \mathcal{A}),$$  \hfill (2)

where the $|i\rangle$ form an eigenbasis of $\rho$ and $\rho = \sum_i p_i|i\rangle|i\rangle$. This is the classical fidelity for the classical information (quantified by $S(\rho)$) of the ensemble of orthogonal eigenstates of the input density operator $\rho$. Another special case is an ensemble consisting of a single density operator $\rho$. In this case it is just $F_c(\rho, \mathcal{A})$ and may be viewed as the fidelity for transmission of the amount $S(\rho)$ of quantum information.

The average entanglement fidelity can equivalently be defined as the norm squared of the projection of the overall final state onto the subspace in which entangled states $|\psi_{0^Q}^{RQ}\rangle$ representing the initial ensemble are correctly correlated with orthogonal states of an additional reference system $S$:

$$F_c = \|P_c \otimes I_E|\Psi_f\rangle\|^2 = \text{tr} P_c(I_{RS} \otimes \mathcal{A})|\psi_0^{RS}\rangle|\psi_0^{RS}\rangle,$$  \hfill (3)

where $P_c := \sum_i |i^S\rangle|i^S\rangle \otimes |\psi_i^{RQ}\rangle|\psi_i^{RQ}\rangle$. The initial state of $RQS$ may be $|\psi_0^{RQS}\rangle := \sum_i \sqrt{p_i} |i^S\rangle|\psi_i^{RQ}\rangle$, with the ensemble of entangled states in $RQ$ produced by entanglement with $S$. Alternatively, the initial state of $RQS$ may be mixed, with perfect classical correlation rather than entanglement, between a basis of $S$ and the different entangled states of $RQ$. In this case $|\psi_0^{RQS}\rangle$ is given by $\sum_i p_i |i^S\rangle|i^S\rangle \otimes |\psi_i^{RQ}\rangle|\psi_i^{RQ}\rangle$. The average entanglement measure is insensitive to whether entanglement, or merely classical correlation, exists between $S$ and $RQ$.

For example, consider the special case of $F_{cl}$. Here the reference system $R$ plays no role as the $p_i$ are pure. After suppressing $R$, $P_c = \sum_i |i^S\rangle|i^S\rangle \otimes |\psi_i^{Q}\rangle|\psi_i^{Q}\rangle$. $S$ contains a record of the classical information sent. $S$ and $Q$ may be supposed to be either entangled or classically correlated, with $\rho_0^Q = \sum_{i,j} \sqrt{p_i p_j} |i^S\rangle|i^S\rangle \otimes |\psi_j^{Q}\rangle|\psi_j^{Q}\rangle$, for orthonormal system and reference bases $|i\rangle$, where the system basis is the eigenbasis of $\rho$. In either case, computing the probability $\text{tr} (P_c \rho_0^{RQ})$ that the final system-reference state falls into the subspace in which system and reference exhibit perfect classical correlation in the desired bases, gives the classical fidelity $F_{cl}(\rho, \mathcal{A})$. 

---

1. Barnum, H., Knill, E. (2000). Reversing quantum dynamics with near-optimal quantum and classical fidelity. arXiv:quant-ph/0004088v1.
The reversal operation. We motivate the definition of the near-optimal reversal operation \( \mathcal{R}_{A,\rho} \) by considering operations \( \mathcal{A} \) that are perfectly reversible on a “code” subspace \( C \). Let \( P_C \) be the projector onto \( C \). Perfectly reversible operations have a decomposition \( A_i \) for which \( A_i P_C = \sqrt{p_i} W_i \), where the \( W_i \) are partial isometries from \( C \) into orthogonal subspaces, which means that \( W_i W_j = \delta_{ij} P_C \). Without loss of generality, assume that the ranges of the \( W_i \) span the state space. The reversal operation has a decomposition consisting of the operators \( W_i^\dagger = P_C A_i^\dagger / \sqrt{p_i} \). This resembles the adjoint (defined using the Hilbert-Schmidt inner product \( (A, B) := \text{tr} A^\dagger B \) on operators) \( A_C^\dagger \) of the restriction \( A_C \) of \( A \) to \( C \), which is given by \( A_C^\dagger \sim \{ P_C A_i^\dagger \} = \{ \sqrt{p_i} W_i^\dagger \} \). The \( \sqrt{p_i} \) need to be removed to ensure that the reversal operation is trace preserving. The general definition of \( \mathcal{R}_{A,\rho} \) is also based on the adjoint, suitably corrected to ensure that it is trace preserving (and continuous in the density operator):

\[
\mathcal{R}_{A,\rho} \sim \{ \rho^{1/2} A_i^\dagger A(\rho)^{-1/2} \} .
\]  

(4)

This agrees with the reversal operation for codes given earlier, using the uniform input state \( \rho = P_C / \text{tr} P_C \).

The notation \( \mathcal{R}_{A,\rho} \) is justified by:

Lemma 1. The definition of the reversal operation \( \mathcal{R}_{A,\rho} \) is independent of the decomposition \( \{ A_i \} \) of \( \mathcal{A} \).

Proof. Let \( \mathcal{A} \sim \{ B_i \} \) be another decomposition. By adding null operators to one of the two decompositions, we can ensure that both have the same number of operators. Note that adding null operators to \( \{ A_i \} \) does not change the action of \( \mathcal{R}_{A,\rho} \). Then there exist \( u_{ij} \) such that \( B_i = \sum_j u_{ij} A_j \), where the matrix \( u \) defined by \( u_{ij} \) is unitary. The decomposition of \( \mathcal{R}_{A,\rho} \) given in (4) transforms via the coefficients of \( u \) into a decomposition given in terms of the \( B_i^\dagger \). As \( u^\dagger \) is also unitary, the result is another decomposition of the same operation.

The operation \( \mathcal{R}_{A,\rho} \) is near-optimal in the sense given by the following theorem.

Theorem 2. Let \( E = \{ p_i, \rho_i \} \) be an ensemble of commuting density matrices, and let \( \rho := \sum_i p_i \rho_i \). Then for any trace-preserving completely positive map \( \mathcal{R} \), \( \mathcal{F}_c(E, \mathcal{R},\rho) \geq \mathcal{F}_c(E, \mathcal{R},\rho)^2 \).

As a corollary, if \( \mathcal{F}_c(E, \mathcal{R},\rho) = 1 - \eta \), then \( \mathcal{F}_c(E, \mathcal{R},\rho)^2 \geq (1 - \eta)^2 \geq 1 - 2\eta \). That is, \( \mathcal{R}_{A,\rho} \)'s error is never greater than twice that of the best reversal operation.

Proof. Without loss of generality, assume that \( \mathcal{R} \sim \{ R_i \} \)'s domain is the algebra of operators on \( \text{supp}(\mathcal{A}(\rho)) \) and its range is \( \text{supp}(\rho) \). Allowing more general reversal operations cannot increase entanglement fidelity. Then there exist operators \( B_i \) such that

\[
R_i = \rho^{1/2} B_i^\dagger A(\rho)^{-1/2} ,
\]  

(5)

namely those defined by \( B_i^\dagger := \rho^{-1/2} R_i A(\rho)^{1/2} \). (Generalized inverses are to be understood here.) Let \( \mathcal{B} \sim \{ B_i \} \). We have

\[
\mathcal{F}_c(E, \mathcal{R},\rho) = \sum_l p_l \sum_{ij} |\text{tr} \rho^{1/2} B_i^\dagger A(\rho)^{-1/2} A_j^\dagger |^2 .
\]  

(6)

Define \( X_{ij}^l := \text{tr} \rho^{1/2} B_i^\dagger A(\rho)^{-1/2} A_j^\dagger \). By proper (l-dependent) choice of operator decompositions \( C \sim \{ C_i \} \) and \( \mathcal{A} \sim \{ A_i \} \) (corresponding to singular value decompositions—cf. [6], §3.3—of the matrices \( X_i^l \)), we may obtain the same expression, but with the inner sum having just one index. Then applying the cyclicity of the trace, the fact that \( [\rho, \rho] = 0 \), and defining \( Y_{ij} := \rho^{1/4} A(\rho)^{-1/4} A_i^\dagger B_i^\dagger \rho^{1/4} \rho_{ij}^l \), gives:

\[
\mathcal{F}_c(E, \mathcal{R},\rho) = \sum_l p_l \sum_{ij} |\text{tr} \rho^{1/2} B_i^\dagger A(\rho)^{-1/2} A_j^\dagger |^2 \\
= \sum_l \{ |\text{tr} X_{ij}^l |^2 \} \leq \sum_l \{ |\text{tr} X_{ij}^l |^2 \} \text{tr} Y_{ij}^l Y_{ij}^l \\
\leq \sum_l \{ |\text{tr} X_{ij}^l |^2 \} \leq \sum_l \{ |\text{tr} X_{ij}^l |^2 \} \text{tr} Y_{ij}^l Y_{ij}^l \}
= \sum_l \sum_{ij} \sum_{l} \{ |\text{tr} X_{ij}^l |^2 \} \text{tr} A_i^\dagger A(\rho)^{-1/2} A_j^\dagger \rho_{ij}^l |^2 \\
= \mathcal{F}_c(E, \mathcal{R},\rho)^2 .
\]  

(7)

Here the first two inequalities are Schwarz inequalities, the third uses the fact that

\[
\sum_i |\text{tr} \rho^{1/2} B_i^\dagger A(\rho)^{-1/2} B_i^\dagger \rho_{ij}^l |^2 \leq \mathcal{F}_c(\rho, \mathcal{B}) \leq 1 ,
\]  

(8)

the fourth just adds positive terms inside the square root. The last identity depends on Lemma 3. Since \( \mathcal{B} \) is not necessarily trace-preserving, (8) is not immediate, but from (7) and the trace-preserving condition on \( \mathcal{R} \),

\[
\sum_l R_i^\dagger R_i = \sum_i \mathcal{A}(\rho)^{-1/2} B_i^\dagger B_i A(\rho)^{-1/2} = I .
\]  

(9)

It follows that \( \mathcal{B}(\rho) = \mathcal{A}(\rho) \), a normalized density operator. Let \( \psi^{\text{RO}} \) be a purification of \( \rho \). Then the states \( (\mathcal{I}^R \otimes \mathcal{B}^Q)|\psi^{\text{RO}}\rangle\langle \psi^{\text{RO}}| \) and \( (\mathcal{I}^R \otimes \mathcal{B}^Q)|\psi^{\text{RO}}\rangle\langle \psi^{\text{RO}}| \) are also normalized density matrices, whence \( \mathcal{F}_c(\rho, \mathcal{B}) = \text{tr} P_0 (\mathcal{I}^R \otimes \mathcal{B}^Q)|\psi^{\text{RO}}\rangle\langle \psi^{\text{RO}}| \leq 1 \).

Important special cases of Theorem 2 are noted in:

Corollary 3. For any trace-preserving completely positive \( \mathcal{R} \), \( \mathcal{F}_c(\rho, \mathcal{R},\rho) \leq \sqrt{\mathcal{F}_c(\rho, \mathcal{R},\rho)} \) and \( \mathcal{F}_c(\rho, \mathcal{R},\rho) \leq \sqrt{\mathcal{F}_c(\rho, \mathcal{R},\rho)} \).

When the members of the input ensemble \( \rho_i \) do not commute, we do not know whether \( \mathcal{R}_{A,\rho} \), for \( \rho := \sum_i p_i \rho_i \), is still near-optimal.

Relationship to the “pretty good measurement”. The above analysis of the fidelity of reversal makes it clear
that $\mathcal{R}_{A,\rho}$ provides a method for distinguishing, with close to optimal average error, density matrices from the ensemble $\{\rho_j, \hat{\rho}_j\}$, where $\hat{\rho}_j := A(\hat{\rho}_j)$, and $\rho = \sum_j p_j |j\rangle \langle j|$. This provides a near-optimal method for distinguishing density matrices in an arbitrary ensemble, for any ensemble $\{\rho_j, \hat{\rho}_j\}$ may be constructed by an operation

$$ \mathcal{A} \sim \{ \sqrt{\lambda_j} |v_{ij}\rangle \langle j| \} , \quad (10) $$

where $\hat{\rho}_j = \sum_j \lambda_{ij} |v_{ij}\rangle \langle v_{ij}|$ are the spectral decompositions of the density matrices to be distinguished. The operation $\mathcal{A}$ may be thought of as measuring in the orthogonal basis $|j\rangle$, and then producing the corresponding $\hat{\rho}_j$, for example by randomly applying, with probabilities $\lambda_{ij}$, unitary rotations taking $|j\rangle$ into $|v_{ij}\rangle$. With $\mathcal{A}$ as defined above,

$$ \mathcal{R}_{A,\rho} \sim \{ R_{ij} \} = \{ \sqrt{\lambda_{ij}} |v_{ij}\rangle \langle v_{ij}| \rho_{\text{out}}^{-1/2} \} . \quad (11) $$

The “pretty good measurement” (PGM) was introduced by Holevo [5] (the term “pretty good measurement” is from [6]) for the case of linearly independent pure states, in which case the PGM is a measurement of orthogonal projectors, and as Holevo showed, the optimal such measurement. For an ensemble of unnormalized density matrices $\rho_j := p_j \rho_j$, where $\rho_{\text{out}} := \sum_j \rho_j$ a normalized density operator, the PGM is defined by the set of operators consisting of the

$$ X_j := \rho_{\text{out}}^{-1/2} \rho_j \rho_{\text{out}}^{-1/2} . \quad (12) $$

For pure $\rho_j \propto |j\rangle \langle j|$, these are just the operators corresponding to the “$\rho$-distorted” [7] states $\rho_{\text{out}}^{-1/2} |j\rangle$. Note that for a doubly indexed ensemble of unnormalized states $\rho_{ij}$, $\sum_i X_{ij} = X_j$, where the $X_j$ are the PGM for the $\rho_j := \sum_i \rho_{ij}$. The operation $\{ X_j \}$ may be viewed, via the given representation, as performing the PGM for the ensemble of unnormalized states $\{ \sqrt{\lambda_{ij}} |v_{ij}\rangle \}$, and returning $|j\rangle$ when the measurement result $ij$ is obtained. Indeed, for that ensemble $R_{ij}^\dagger R_{ij} = X_j$, and therefore $\sum_i R_{ij}^\dagger R_{ij} = X_j$. Thus the operation may also be viewed as doing the PGM for the $\rho_j$, and returning $|j\rangle$ when the measurement result is $j$. However, a given ensemble $\rho_j$ may in general arise from orthogonal states $|j\rangle$ by actions of channels different from the “classifying” one $\{ X_j \}$, which completely decoheres the orthogonal states $|j\rangle$ before producing $\hat{\rho}_j$. For example, if the $\rho_j$ are orthogonal and pure they may be produced either by measurement in the basis $|j\rangle$ followed by an appropriate unitary operator $U$, or by applying $U$ without prior measurement. In the first case quantum coherence is completely destroyed, while in the second case it is perfectly preserved. When the channel producing the $\hat{\rho}_j$ is not of the form $\{ X_j \}$, the reversal operation $\mathcal{R}_{A,\rho}$ will be different from $\{ X_j \}$. Although $\{ X_j \}$ still gives near-optimal classical fidelity, it will not necessarily give good entanglement fidelity, since in some sense it decoheres the states $\hat{\rho}_j$. $\mathcal{R}_{A,\rho}$, however, will have near-optimal entanglement fidelity while retaining near-optimal classical fidelity. $\mathcal{R}_{A,\rho}$ thus takes advantage of whatever coherence remains between the $\hat{\rho}_j$; it avoids decohering the $\hat{\rho}_j$ if the channel has not decohered them already.

**A bound on the classical fidelity of reversal.** To bound a fidelity of reversal it is sufficient to bound the fidelity for the near optimal reversal operation and apply Theorem 3. Here we have a look at such bounds for classical fidelities of reversal for $\mathcal{A}$ of the form $\{ X_j \}$. In this case, the classical fidelities are probabilities of success for measurements that attempt to infer which of the $|j\rangle |j\rangle$ of the input state $\rho = \sum_j p_j |j\rangle \langle j|$ was actually transmitted. The expression for the PGM gives the following bound on the optimal probability of success $F_{cl}$ (with the definitions of the previous section):

$$ F_{cl}^2 \leq F_{cl}(\rho, \mathcal{R}_{A,\rho}A) = \sum_j \text{tr} \left( \rho_{\text{out}}^{-1/2} \rho_j \rho_{\text{out}}^{-1/2} \rho_j \right) = 1 - \sum_{i,j;i \neq j} \text{tr} \left( \rho_{\text{out}}^{-1/2} \rho_j \rho_{\text{out}}^{-1/2} \rho_j \right), \quad (13) $$

where we used the identity $\sum_{i,j} \text{tr} \left( \rho_{\text{out}}^{-1/2} \rho_j \rho_{\text{out}}^{-1/2} \rho_j \right) = 1$. Thus the probability of error $E_{cl}$ is bounded by $2 \sum_{i,j;i \neq j} \text{tr} \left( \rho_{\text{out}}^{-1/2} \rho_j \rho_{\text{out}}^{-1/2} \rho_j \right)$, which is a multiple of the sum of the Hilbert-Schmidt inner products of the different $\rho_{\text{out}}^{-1/2} \rho_j \rho_{\text{out}}^{-1/2}$. When $\rho_{\text{out}}$ is uniform (proportional to a projection), this sum can be easy to estimate. An often used measure of overlap between density matrices is the Bures-Uhlmann fidelity $\mathcal{F}_{BU}$. This measure depends only on the pair of density matrices, not on an overall average, and is defined by $\mathcal{F}_{BU}(\sigma_1, \sigma_2) := \text{tr} \sqrt{\sigma_1^{1/2} \sigma_2 \sigma_1^{1/2}}$. The expression for the optimal reversal given in (11) can be used to derive a bound on the probability of error in terms of the Bures-Uhlmann fidelities.

**Theorem 4.**

$$ F_{cl}^2 \geq 1 - \sum_{i,j;i \neq j} \sqrt{p_i p_j} \mathcal{F}_{BU}(\hat{\rho}_i, \hat{\rho}_j) . \quad (14) $$

**Proof.** Let $A_j$ be the matrix whose $i\text{th}$ column is $\sqrt{p_j} \sqrt{\lambda_{ij}} |v_{ij}\rangle$, and $A$ the matrix obtained by placing the $A_j$ in a row. Then $AA^\dagger = \rho_{\text{out}}$. Let $R_j$ be the matrix whose $i\text{th}$ row is $\sqrt{p_j} \sqrt{\lambda_{ij}} |v_{ij}\rangle \rho_{\text{out}}^{-1/2}$ and $R$ the matrix obtained by placing the $R_j$ in a column. $R$ is simply an explicit matrix form of $\mathcal{R}_{A,\rho}$. Since $R = A^\dagger (AA^\dagger)^{-1/2}$, $R$ is a matrix with the property that $RA$ is positive semi-definite. (In fact, this gives an alternative approach to defining $\mathcal{R}_{A,\rho}$.) The matrix $RA$ has a natural block structure that mirrors that used to define $R$ and $A$. It is readily verified that $F_{cl}$ for $\mathcal{R}_{A,\rho}$ is given by the sum of the
squared Frobenius norms \(|(RA)_{ij}|^2 := \text{tr} (RA)_{ij}^\dagger (RA)_{ij}\) of the diagonal blocks. Since \(|RA|_{ij}^2\) is one, it suffices to estimate the sum of the squared Frobenius norms of the off-diagonal blocks to bound the optimal \(F_{cl}\). To do so, observe that \((RA)^2 = A^1 A\). The block at block position \(i,k\) in \(B := A^1 A\) is given by \(B_{ik} = A_{i}^k A_k\). Since \(A_i A_k = \rho_i\),

\[
B_{ik}^\dagger B_{ik} = A_k^\dagger \rho_i A_k = U \rho_k^{1/2} \rho_i \rho_k^{1/2} U^\dagger
\]  

(by polar decomposition of \(A_k\)) for some unitary operator \(U\). Consequently, the \(L_1\)-norm of \(B_{ik}\), defined by \(|B_{ik}|_1 := \text{tr} \sqrt{B_{ik}^\dagger B_{ik}}\) is given by \(\sqrt{\text{tr}\rho_i \rho_k F_{BU}(\rho_i, \rho_k)}\). It therefore suffices to relate the Frobenius norms of the off-diagonal blocks of a positive semi-definite matrix to the \(L_1\) norms of the off-diagonal blocks of its square.

**Lemma 5.** Let \(M = \begin{pmatrix} a & b^\dagger \\ b & c \end{pmatrix}\) be positive semi-definite, with \(a, b, c\) matrices. Write \(M^{1/2} = \begin{pmatrix} x & y^\dagger \\ y & z \end{pmatrix}\) with the same block structure. Then \(|y|_2^2 \leq |b|_2^2\).  

**Proof.** Without loss of generality, assume that \(y\) is non-negative diagonal. Otherwise, with a block-diagonal unitary \(U = \begin{pmatrix} u & 0 \\ 0 & v \end{pmatrix}\) with \(u\) and \(v\) chosen to implement the singular value decomposition of \(y\), we may transform \(M\) and \(M^{1/2}\) so that \(y\) is a diagonal matrix with non-negative diagonal entries (for rectangular \(y\), the upper or left-hand square portion is diagonalized). This does not affect the norms, since \(U\) transforms blocks independently and the \(L_1\) and Frobenius norms are both unitarily invariant. Let \(y_i\) be the diagonal entries of \(y\). Note that \(b = xy + zy\) and \(|b|_1 \geq \text{tr} b\) (cf. [3], p. 432). Now \(\text{tr}(xy + zy) = \sum_i y_i (x_{ii} + z_{ii})\). By the positivity of \(M^{1/2}\), \(y_i^2 \leq x_{ii} z_{ii}\), so \(y_i\) is less than at least one of \(x_{ii}, z_{ii}\). Thus \(|y_i|^2 = \sum_i y_i^2 \leq \text{tr} b \leq |b|_1\), as desired.  

To apply Lemma 5, consider first the block decomposition of \(RA\) and \(B\) determined by \(B_{11}\). The squared Frobenius norm of the first block row and column excluding \((RA)_{11}\) is bounded by the sum of the \(L_1\) norms of the corresponding block row and column in \(B_{11}\). By subadditivity of the norm, this is at most \(\sum_{i \neq 1} (|B_{11}|_1 + |B_{11}|)\). After a suitable permutation, the same argument applies to the \(i\) the row and column determined by \(B_{ii}\), for each \(i\). The proof of the theorem then follows by summing over the resulting inequalities and noting that each off-diagonal block occurs twice on both sides.

**Applications.** Due to its near optimality, the reversal operation \(\mathcal{R}_{A \rho}\) can be used in any situation where classical or quantum information has been corrupted by noise with known behavior. \(\mathcal{R}_{A \rho}\) has a simple definition, but whether it or a good approximation can be implemented efficiently depends on the details of the situation. Whether or not it can be efficiently implemented, because its error is at most twice the optimum, it can be used as a theoretical tool to obtain upper bounds on the achievable fidelities in a given situation. The upper bounds can then be compared to the fidelity achieved by simpler algorithms. An example of this occurs in the use of stabilizer codes for quantum error-correction. When the noise model is independent and depolarizing, classical coding theory immediately suggests a combinatorially straightforward error-correction algorithm based on maximum likelihood error syndrome decoding. Comparing this method to \(\mathcal{R}_{A \rho}\) suggests itself as a fruitful path of investigation with applications to asymptotic bounds in quantum coding theory.

Another application is to query complexity for quantum oracles. Here we are given a quantum black box implementing an unknown quantum operation from some set. A simple method for attempting to determine which operation we are given is to apply it to copies of some input state and attempt to distinguish the output state. A bound on the probability of success can then be obtained by using bounds such as the one of Theorem 4. This was how the fact that the hidden subgroup problem has low query complexity was first realized.

**Acknowledgments.** We thank the following for support: The ONR (N00014-93-1-0116, H.B.), the NSF (PHY-9722614, H.B.), the ISI Foundation (Turin, Italy, H.B.), Elsag-Bailey (H.B.), the ITP at UC Santa Barbara (NSF PHY94-07194, H.B. and E.K.), the NSA (E.K.) and the DOE (W-7405-ENG-36, E.K.).

---

[1] B. Schumacher, Phys. Rev. A 54, 2614 (1996).
[2] E. Knill and R. Laflamme, Phys. Rev. A 55, 900 (1997).
[3] M. A. Nielsen, C. M. Caves, B. W. Schumacher, and H. Barnum, Proc. Royal Soc. London A 454, 277 (1998).
[4] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 1985).
[5] A. S. Holevo, Theor. Probab. and Appl. 23, 411 (1978).
[6] P. H. Maass, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters, Phys. Rev. A 54, 1896 (1996).
[7] L. P. Hughston, R. Jozsa, and W. K. Wootters, Phys. Lett. A 183, 14 (1993).
[8] D. Bures, Trans. Am. Math. Soc. 135, 199 (1969).
[9] A. Uhlmann, Rep. Math. Phys. 9, 273 (1976).
[10] A. Ashikhmin, A. Barg, E. Knill, L. S. Litsyn, (1999), to appear in *IEEE Tr. Inf. Th.*, Los Alamos ArXiV Preprint Archive quant-ph/9906123.
[11] J. M. Ettinger, P. Hoyer, and E. Knill (1999), Los Alamos ArXiV Preprint Archive quant-ph/9901034.