Low-Entanglement Remote State Preparation

Igor Devetak* and Toby Berger
Department of Electrical and Computer Engineering
Cornell University, Ithaca, New York 14850

March 5, 2002

Abstract

An outer bound on the low-entanglement remote state preparation (RSP) ebits vs. bits tradeoff curve [1] is found using techniques of classical information theory. We show this bound to be optimal among an important class of protocols and conjecture optimality even without this restriction.

We all know what state preparation is: Alice, having complete classical knowledge of a quantum state, prepares it in her lab. Remote state preparation (RSP) refers to the case where Alice, again having a classical description of the state, wishes to prepare a physical instance of it in Bob's lab, Bob being far away. It seems natural to ask about how this situation differs from quantum teleporation [2] where Alice has no classical knowledge of state, but has a physical instance of it. This was first addressed by Pati [3] and Lo [4] who considered special ensembles of states. The general case was investigated by Bennett et al. [1] where they posed the question of quantifying the resources necessary and sufficient for asymptotically perfect RSP. Asymptotic perfection means that the average fidelity between the resulting states in Bob's lab and the corresponding states Alice intended him to prepare tends to 1 as the number of states to be remotely prepared is taken to infinity. The resources are the same as for teleportation: entanglement (ebits) between Alice and Bob and classical bits of forward communication from Alice to Bob. They also allow classical back-communication from Bob to Alice, this extra resource being unhelpful for teleportation. For the case of qubit states Bennett et al. found outer bounds on the achievable (b,e) pairs by explicit construction of RSP protocols (see Fig.1). The teleportation point $(2, 1)$ naturally divides the plane into a high and low-entanglement region where the number of ebits per remotely prepared state is greater than and less than 1, respectively; there is a large qualitative difference in the methods used for these two cases. The high-entanglement region is accessed by Alice performing certain generalized measurements on her ebit halves that possibly depend on her classical knowledge of the state, and sending classical information about the measurement results to Bob. The low-entanglement protocols described in [1] (which we refer to as *teleportation based*) involve sending classical information about the states themselves causing a reduction in the posterior von Neumann entropy from Bob's point of view, and teleportation of Schumacher compressed states. Here we concentrate on the latter case, pushing these ideas to their information theoretical limit. The main result is an analytic expresssion for the best teleportation based outer bound on the low-entanglement region. Our approach borrows heavily from Shannon's classical rate-distortion theory [5] [6], and we will emphasize the key concepts and ideas, relegating technical details to a future publication [8].

*Electronic address: igor@ece.cornell.edu

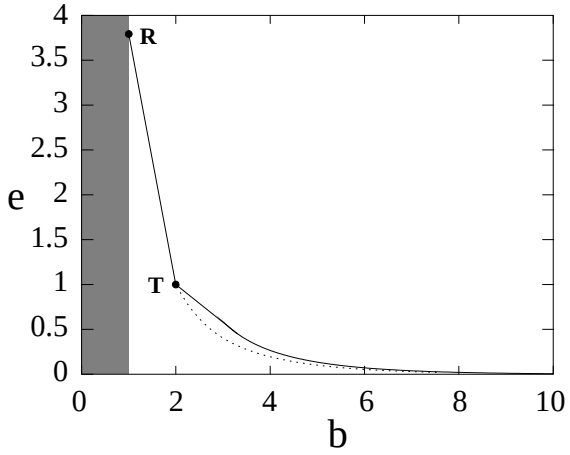

FIG. 1 Ebits vs. bits for remote state preparation (from [1]). The dotted curve represents our low-entanglement outer bound. The solid curve is the previous outer bound by Bennett et al. The shaded region is forbidden by causality.

Let us first consider an example (attributed to H.-K. Lo [1] [4]) illustrating the way classical information about a qubit state reduces its von Neumann entropy. It is important to appreciate the fact that, in the scenario we are dealing with here, the density matrix is not a property of the qubit, but rather reflects *knowledge* about the actual pure state the qubit is in. Alice knows her states *exactly* prior to remotely preparing them, hence the individual density matrices have zero entropy, from her point of view. At the same time Bob is completely ignorant of the qubit states; for all he knows Alice could have chosen them from anywhere on the Bloch sphere. More formally, if we denote the Bloch sphere by $\mathcal{X}$, parametrized by spherical polar coordinates $x \equiv (\theta_x, \phi_x) \in [0, \pi] \times [0, 2\pi]$ (for convenience we will refer to the north pole $\theta_x = 0$ as $x = 0$), then the probability density corresponding to picking x is simply $p(x) = \frac{1}{4\pi}$. The corresponding quantum state is $|x\rangle = \sqrt{\frac{1+\cos\theta_x}{2}}|0\rangle + e^{i\phi_x}\sqrt{\frac{1-\cos\theta_x}{2}}|1\rangle$. The resulting density matrix from Bob's point of view is $\rho = \int dx p(x)|x\rangle\langle x| = \frac{1}{2}I$, and the von Neumann entropy is $S(\rho) = 1$, as one would expect from such a random distribution. Now, let us assume Alice gives Bob 1 bit of classical information about the state, e.g., tells him whether the state is in the upper or lower Bloch hemisphere. The *posterior* distribution is now uniform in the upper (lower) hemisphere, i.e. $p'(x) = \frac{1}{2\pi}$ for x in the upper(lower) hemisphere and zero otherwise. The density matrix ρ' is computed as above, and the posterior von Neumann entropy becomes $S(\rho') \approx 0.81$ in either case. Schumacher's theorem [7] now tells us that we have reduced the amount of quantum information to be conveyed to Bob, at the expense of an additional classical rate of 1 bit per letter. Based on this observation a protocol may be devised as follows [1].

• Alice sends classical information to Bob at a rate $R = 1$ bit per remotely prepared state about which hemisphere the state lies in.

• This reduces the von Neumann entropy of the source (as viewed by Bob) to $S \approx 0.81$. However, the density matrices now depend on the hemisphere. So Alice rotates, say, all the states in the lower hemisphere by a preagreed unitary transformation that maps the lower onto the upper hemisphere (any rotation sending the south pole to the north accomplishes this). Now the qubits are i.i.d. from Bob's point of view, and Schumacher's theorem applies. Alice prepares these rotated states, and Schumacher compresses them to S qubits per letter.

• Alice teleports the compressed qubit states at a rate of $2S$ bits and S ebits per remotely prepared state.

• Bob simply reverses Alice's steps in his laboratory, thus recovering asymptotically faithful instances of her states.

This teleportation based protocol yields the point $(2S + R, S)$ in the (b,e)-plane. The property of being asymptotically faithful is inherited from Schumacher compression, this being based on

classical Shannon compression. It is a low-entanglement protocol since $e = S \leq 1$. It is now evident that, if we restrict attention to teleportation based protocols, the problem reduces to finding the optimum *rate-entropy curve*, i.e. the frontier of (R, S) pairs attainable in this way. One may wonder, for example, how it is possible to further reduce S while keeping $R = 1$. The answer lies in exploiting the asymptotic formulation of the problem and processing blocks of states, now minimizing the entropy per remotely prepared state.

We proceed to formulate the source coding problem. The *source* is described by a random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, and we take the individual X_i to be independent and identically distributed (i.i.d.), each taking values x on the Bloch sphere $\mathcal{X}$ with probability density $p(x) = \frac{1}{4\pi}$. Thus the probability density distribution for $\mathbf{X}$ is $p(\mathbf{x}) = \prod_i p(x_i)$. This reflects Bob's view before he receives any classical information. Elements $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ of $\mathcal{X}^n$ are called *source words* of length n, and the x_i are called *letters*. We map the source $\mathbf{X}$ onto a set $B_n = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_K\}$, $\mathbf{y}_k \in \mathcal{X}^n$, called a *source code* of *size* K and *blocklength* n, of reproducing *codewords*. The *rate* of the code is formally defined as $R = n^{-1} \log_2 K$, and it signifies the number of bits per source letter needed to specify the index of the reproducing codeword. When Bob recieves these R bits, he knows the reproducing codeword, which is an approximation to the actual source word. In Lo's simple example $n = 1$, $K = 2$, $R = 1$ and B_1 consists of two codewords corresponding to the north pole y_1 and south pole y_2, respectively. There each source word gets mapped onto the closest pole, and knowledge of the codeword is equivalent to specifying the hemisphere. The goal is to minimize the von Neumann entropy of the source word as viewed by Bob upon receiving the reproducing codeword. Formally, each source word $\mathbf{x}$ gets mapped into a unique $\mathbf{y} \in B_n$ in such a way that the posterior von Neumann entropy of the source

$$S(B_n) = \frac{1}{n} E_{\mathbf{Y}} S(E_{\mathbf{X}|\mathbf{Y}} |\mathbf{X}\rangle\langle\mathbf{X}|) \tag{1}$$

is minimized. Here $\mathbf{Y}$ is the random vector associated with the probability distribution on the set of codewords B_n induced by our map. $E_{\mathbf{Y}}$ denotes the expectation value over the random vector $\mathbf{Y}$, and $E_{\mathbf{X}|\mathbf{Y}}$ is the conditional expectation over $\mathbf{X}$ given the value of $\mathbf{Y}$. Let us analyze the above expression. Let $\mathcal{M}_{\mathbf{y}}$ be the set of all values of $\mathbf{X}$ that get mapped into $\mathbf{Y} = \mathbf{y}$. When Bob learns that $\mathbf{Y} = \mathbf{y}$ he knows that $\mathbf{X}$ must have come from the set $\mathcal{M}_{\mathbf{y}}$. The density matrix he sees is now an average over all the $\mathbf{X}$'s from $\mathcal{M}_{\mathbf{y}}$ and is denoted by the expectation value $E_{\mathbf{X}|\mathbf{Y}=\mathbf{y}} |\mathbf{X}\rangle\langle\mathbf{X}|$. We average the corresponding von Neumann entropy over all the possible $\mathbf{Y}$'s Bob could have received, and divide by n to get a per letter result, thus giving rise to (1). In Lo's example the random variable Y takes on the values y_1 and y_2 with probabilities $\frac{1}{2}$ each, depending on the hemisphere of X. The distribution of X given Y is uniform over the hemisphere indicated by the value of Y. Thus (1) indeed yields the entropy obtained before.

Formally, a rate-entropy pair (R, S) is called (asymptotically) achievable iff there exists a sequence of source codes B_n of rate R and increasing blocklength n such that

$$\lim_{n\to\infty} S(B_n) \leq S \tag{2}$$

We now define the rate-entropy function $R(S)$ as the infimum of all R for which (R, S) is achievable. The way such a coding problem can be solved exactly is by first finding an information-theoretical lower bound on $R(S)$ and then producing a coding scheme that achieves said bound. Firstly, note that $\mathbf{Y}$ is completely determined by the corresponding value of $\mathbf{X}$, and hence the conditional probability density $Q(\mathbf{y}|\mathbf{x})$ is a δ-function. However, for the purpose of finding a lower bound we relax this constraint. Secondly, observe the following string of inequalities

$$R = \frac{1}{n} \log_2 K \geq \frac{1}{n} H(\mathbf{Y}) \geq \frac{1}{n} I(\mathbf{X}; \mathbf{Y}) \tag{3}$$

The first inequality is saying that the entropy of $\mathbf{Y}$ is maximum when the codewords occur with equal probability K^{-1} in which case the entropy is simply $\log_2 K$. Intuitively, this is the number of bits needed to specify one of K equiprobable codewords. The second one follows from the definition of mutual information $I(\mathbf{X}; \mathbf{Y}) \equiv H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{X})$. For the purpose of finding a lower bound, we consider minimizing the mutual information per letter instead of the rate, while keeping the von

Neumann entropy fixed. This leads to the following information-theoretical optimization problem. Given n and the random vector $\mathbf{X}$ as defined above, we wish to find

$$R_n(S) = \frac{1}{n} \inf_{Q(\mathbf{y}|\mathbf{x}):S(Q)=S} I(Q) \tag{4}$$

where $I(Q)$ is the mutual information

$$I(Q) = \iint d\mathbf{x}d\mathbf{y}p(\mathbf{x})Q(\mathbf{y}|\mathbf{x}) \log \frac{Q(\mathbf{y}|\mathbf{x})}{q(\mathbf{y})} = \iint d\mathbf{x}d\mathbf{y}q(\mathbf{y})P(\mathbf{x}|\mathbf{y}) \log \frac{P(\mathbf{x}|\mathbf{y})}{p(\mathbf{x})} \tag{5}$$

and

$$S(Q) = \frac{1}{n} \int d\mathbf{y}q(\mathbf{y})S\left(\int d\mathbf{x}P(\mathbf{x}|\mathbf{y})|\mathbf{x}\rangle\langle\mathbf{x}|\right) \tag{6}$$

is the posterior von Neumann entropy, as in (1). The probability density for the marginal $\mathbf{Y}$ distribution is given by $q(\mathbf{y}) = \int d\mathbf{x}p(\mathbf{x})Q(\mathbf{y}|\mathbf{x})$ and the conditional distribution for $\mathbf{X}$ given $\mathbf{Y}$ is $P(\mathbf{x}|\mathbf{y}) = p(\mathbf{x})Q(\mathbf{y}|\mathbf{x})/q(\mathbf{y})$. The minimization should be done for a general length n of $\mathbf{x}$. We have found a local extremum of this problem [8], which we conjecture to be global, for which the conditional distribution factorizes, i.e. $Q(\mathbf{y}|\mathbf{x}) = \prod_i Q^\lambda(y_i|x_i)$ where

$$Q^\lambda(y|x) = P^\lambda(x|y) = \frac{1}{4\pi} \frac{\lambda}{e^\lambda - 1} e^{\lambda|\langle x|y\rangle|^2} \tag{7}$$

so that $n = 1$ suffices. Here λ plays the role of a Lagrange multiplier. Some light may be shed on this result by noticing that there are two competing efffects. One comes from subadditivity of von Neumann entropy, which says that the von Neumann entropy of the whole is no greater than the sum of the von Neumann entropies of the parts. This favors large n in order to decrease the von Neumann entropy per letter. The other comes from superadditivity of mutual information, valid only when $\mathbf{X}$ is i.i.d. (as in our case) which states that the mutual information between $\mathbf{X}$ and $\mathbf{Y}$ is no less than the sum of the mutual informations between the corresponding components X_i and Y_i. This favours $n = 1$. The latter effect apparently wins. The corresponding $R_1(S)$ is parametrized as follows:

$$R_1(\lambda) = \frac{\lambda}{e^\lambda - 1} - 1 + \log\left(\frac{\lambda e^\lambda}{e^\lambda - 1}\right) \tag{8}$$

$$S(\lambda) = h_2\left(\frac{1}{\lambda} - \frac{1}{e^\lambda - 1}\right) \tag{9}$$

where the $\lambda \in (0, \infty)$ and $h_2(p) = -p\log_2 p - (1-p)\log_2(1-p)$ is the binary Shannon entropy function. $R_1(\lambda)$ is given in nats, and should be converted into bits by dividing by $\log 2$. The curve is readily found to be convex, and is shown in Fig 2.

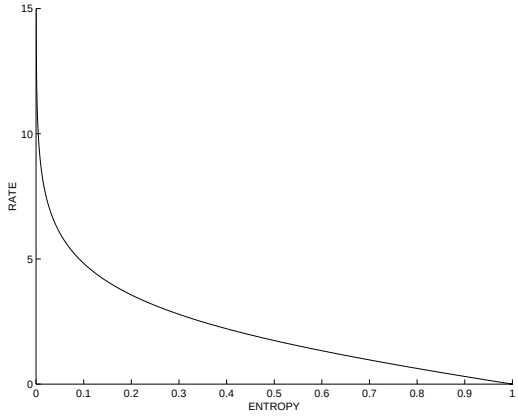

FIG. 2 The rate-entropy function $R(S)$.

So far we have only found a lower bound on $R(S)$. Now we will demonstrate achievability, and thus establish that $R(S) = R_1(S)$. It may appear that blocking was not needed after all, but this is due to the fact that we have not quite solved the coding problem. In particular, our solution $Q^\lambda(y|x)$ is not deterministic, as a code should be, but probabilistic. Given x, y is most likely to be x itself, and then as the arc distance from x increases the probability decreases, reaching a minimum at the antipode of x. It is only in the $\lambda \to \infty$ limit that $Q^\lambda(y|x)$ becomes a δ-function centered at x, which corresponds to the identity map. This also implies that the second inequality in (3) is not tight. However, one could expect it to become tight in the large blocklength limit, since $H(\mathbf{Y})$ is subadditive, and $I(\mathbf{X};\mathbf{Y})$ is superadditive. The idea is to simulate the *noisy* single letter channel defined by $P^\lambda(x|y)$ (acting in the reverse direction, i.e. from Y to X) by the average effect that a *deterministic* coding map (from $\mathbf{X}$ to $\mathbf{Y}$) involving large strings of letters has on the ith letter. To elaborate, let us assume that the ith letter in a given codeword $\mathbf{y}$ is some y_i. Then our code is such that the ith components x_i of all the $\mathbf{x}$'s that get mapped onto $\mathbf{y}$ are distributed as if randomly chosen according to the conditional distribution $P^\lambda(x_i|y_i)$. Since $P^\lambda(x|y)$ depends only on the overlap $\langle x|y\rangle$, when Alice rotates $\mathbf{x}$ by the map that sends $\mathbf{y}$ to $\mathbf{0}$, the block density matrix Bob sees after being told the codeword is the Schumacher compression friendly tensor product of single qubit density matrices $\rho' = \int dx P^\lambda(x|0)|x\rangle\langle x|$ with entropy per qubit given by $S(\lambda)$ (9). The way to construct such a coding map is by using joint typicality decoding, a technique well known in classical rate-distortion theory [6]. It is necessary first to coarse grain $\mathcal{X}$ into a disjoint union of small near-circular caps of diameter $\approx \epsilon$ and replace the probability densities $P^\lambda(x|y)$ etc. by discrete probabilities $\widehat{P}^\lambda(\widehat{x}|\widehat{y})$ etc. where $\widehat{x}$ and $\widehat{y}$ belong to $\widehat{\mathcal{X}}$, the set of cap centroids. A *δ-typical sequence* $\widehat{\mathbf{x}} \in \widehat{\mathcal{X}}^n$ with respect to the distribution $\widehat{p}(\widehat{x})$ is defined as one that satisfies

$$\left| \frac{N(\widehat{a}|\widehat{\mathbf{x}})}{n} - \widehat{p}(\widehat{a}) \right| < \frac{\delta}{|\widehat{\mathcal{X}}|} \tag{10}$$

where $N(\widehat{a}|\widehat{\mathbf{x}})$ is the number of occurences of $\widehat{a} \in \widehat{\mathcal{X}}$ in the sequence $\widehat{\mathbf{x}}$. We call the set of all such typical sequences the *typical set* $T_\delta(\widehat{p})$. In words, a sequence is typical if the fraction of appearances of any given letter in the sequence approximates the probability for that letter. Another way of putting it is that picking an element of the sequence at random approximatley simulates the probability distribution. Note that, by the law of large numbers, a sufficiently long sequence chosen according to the probability distribution will "almost always" be typical. One similarly defines the *jointly typical set* $T_\delta(\widehat{P}\widehat{q})$ of pairs of typical sequences $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) \in (\widehat{\mathcal{X}} \times \widehat{\mathcal{X}})^n$ with respect to the distribution $\widehat{P}^\lambda(\widehat{x}|\widehat{y})\widehat{q}(\widehat{y})$ [6]. The coding map is as follows:

• The codewords $\widehat{\mathbf{y}}$ are chosen at random. More precisely, each letter of each codeword is chosen according to $\widehat{q}(\widehat{y})$ (which mimics the uniform distribution). This ensures with high probability that the codewords will be typical of the distribution $\widehat{q}(\widehat{y})$.

• Mapping a given $\mathbf{x}$ onto a $\widehat{\mathbf{y}}$ with the property that the pair $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ is typical of the joint distribution $\widehat{P}^\lambda(\widehat{x}|\widehat{y})\widehat{q}(\widehat{y})$. Here $\widehat{\mathbf{x}}$ is the componentwise centroid of the cap that contains $\mathbf{x}$. This implies that if we randomly pick a $\widehat{\mathbf{x}}$ and its corresponding $\widehat{\mathbf{y}}$, the ith component pair will equal $(\widehat{x}_i, \widehat{y}_i)$ with probability $\widehat{P}^\lambda(\widehat{x}_i|\widehat{y}_i)\widehat{q}(\widehat{y}_i)$. Hence, given $\widehat{y}_i$, $\widehat{x}_i$ was the source letter with probability $P^\lambda(\widehat{x}_i|\widehat{y}_i)$. This is how the noisy channel $P^\lambda(x|y)$ is simulated.

The above map fails when there are not enough reproducing codewords to ensure that one can find a member of the code B_n jointly typical with a given $\widehat{\mathbf{x}}$. It turns out [6] that the minimal rate for which such an error "almost never" occurs is precisely the mutual information corresponding to $\widehat{P}(\widehat{x}|\widehat{y})\widehat{q}(\widehat{y})$, which is approximated by $R_1(\lambda)$ (8). Finally, it is necessary to take the $\epsilon, \delta \to 0$ and $n \to \infty$ limits carefully to ensure that the pair (R, S) indeed approaches the $R_1(S)$ curve arbitrarily closely [8]. Note that joint typicality decoding is suboptimal, strictly speaking. The actual optimal map makes no reference to coarse graining. The code B_n is chosen at random, and the coding map is the one that minimizes $S(B_n)$. However, stating it that way gives us little hope of computing $R(S)$.

Our RSP protocol is now analogous to the simple one described earlier. Alice wishes to remotely

prepare a string of n qubits using an (R, S) source code. She identifies the corresponding codeword and rotates the original string by the map that sends the codeword to $\mathbf{0}$ (this is analogous to mapping the south pole onto the north pole in Lo's example), and prepares these qubits in her laboratory. She may Schumacher compress them without additional blocking to Sn qubits. She teleports these to Bob using $2Sn$ classical bits and Sn ebits. A further Rn bits are sent in order to convey the codeword. Bob reverses Alice's steps in his laboratory, thus recovering an asymptotically faithful copy of the qubits to be prepared. The corresponding point in the (b,e)-plane is $(R+2S, S)$ per remotely prepared state. The ebits vs. bits tradeoff curve is shown by the dotted curve in Fig 1. and is parametrized by $(R_1(\lambda) + 2S(\lambda), S(\lambda))$.

It should be noted that our protocol does not require back-communication, since it is based on teleportation, which enjoys the same property. We conjecture that teleportation based protocols are optimal among all low-entanglement protocols, and hence that our result is exact. To show this formally it is crucial to understand the high-entanglement region, since we expect other candidates to be "generated" by special points in the high-entanglement region in the same way our upper bound was generated by the teleportation point via $R(S)$.

We are grateful to N.D.Mermin for bringing reference [1] to our attention. We also thank C.H.Bennett, D.P.DiVincenzo, P.W.Shor, B.M.Terhal and H.-K. Lo for useful discussions that revealed misstatements in an earlier version of the paper, and A.K.Pati for pointing us to reference [3]. This research was supported in part by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under Grant DAAD19-99-1-0215 and NSF Grant CCR-9980616.

References

[1] C.H.Bennett, D.P.DiVincenzo, J.A.Smolin, B.M.Terhal and W.K.Wooters, quant-ph/0006044 (2000)

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

[3] A. K. Pati *Phy. Rev. A* **63**, 14302 (2001)

[4] H.-K. Lo, quant-ph/9912009 (1999)

[5] C.E.Shannon, *IRE Nat'l Conv.Rec.*, part 4, 142 (1959); T.Berger, *Rate distortion theory, Prentice Hall* (1971)

[6] T.Cover and J.Thomas, *Elements of information theory, Wiley and Sons* (1991); T.Berger, "Multiterminal Source Coding," in *The Information Theory Approach to Communications*, G. Longo, Ed. CISM Courses and Lectures, **229**, *Springer-Verlag, Vienna and New York* (1978)

[7] B.Schumacher, *Phys.Rev.A* **51**, 2738 (1995); R.Jozsa and B.Schumacher, *J.Mod.Opt.* **41**, 2343 (1994)

[8] I.Devetak and T.Berger, "Bounds on Remote State Preparation", to be submitted to *IEEE Trans. Inf. Theory*