Entanglement purification of unknown quantum states

Todd A. Brun,[1,*] Carlton M. Caves,[2] and Rüdiger Schack[3]
[1]*Physics Department, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213*
[2]*Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131-1156*
[3]*Department of Mathematics, Royal Holloway, University of London, Egham, Surrey TW20 0EX, United Kingdom*
(Received 9 October 2000; published 21 March 2001)

A concern has been expressed that ''the Jaynes principle can produce fake entanglement'' [R. Horodecki *et al.*, Phys. Rev. A **59**, 1799 (1999)]. In this paper we discuss the general problem of distilling maximally entangled states from N copies of a bipartite quantum system about which only partial information is known, for instance, in the form of a given expectation value. We point out that there is indeed a problem with applying the Jaynes principle of maximum entropy to more than one copy of a system, but the nature of this problem is classical and was discussed extensively by Jaynes. Under the additional assumption that the state $\rho^{(N)}$ of the N copies of the quantum system is *exchangeable*, one can write down a simple general expression for $\rho^{(N)}$. By measuring one or more of the subsystems, one can gain information and update the state estimate for the remaining subsystems with the quantum version of the Bayes rule. Using this rule, we show how to modify two standard entanglement purification protocols, one-way hashing and recurrence, so that they can be applied to exchangeable states. We thus give an explicit algorithm for distilling entanglement from an unknown or partially known quantum state.

DOI: 10.1103/PhysRevA.63.042309 PACS number(s): 03.67.−a

I. INTRODUCTION

Entanglement is a quantum-mechanical resource that can be used for a number of tasks, including quantum teleportation, quantum cryptography, and quantum dense coding. Since real quantum channels are noisy, it is very difficult to create perfect entanglement directly between two distant parties. There is thus a need to purify (or distill) partial entanglement. Suppose two parties share N pairs of qubits such that each pair is in the same entangled but mixed state ρ, the total state $\rho^{(N)}$ of all N pairs thus being the N-fold tensor product $\rho^{(N)}=\rho^{\otimes N}\equiv\rho\otimes\ldots\otimes\rho$. There exist protocols [1–4], using only local operations and classical communication, which allow the two parties to transform $M<N$ of the pairs into maximally entangled states, for instance, singlet states. In the limit $N\to\infty$, the fidelity of the singlets approaches one and the fraction M/N a fixed limit, called the asymptotic yield.

In this paper, we consider the more general case in which the initial state $\rho^{(N)}$ is not a tensor-product state. This corresponds to the realistic situation that the state ρ of each individual pair is not perfectly known, for instance, because one of the particles has been sent through a channel with only partially known characteristics. In Secs. III and IV, we apply the entanglement purification methods known as one-way hashing [2] and recurrence [2,4] to partially known, including completely unknown, quantum states. It turns out that the generalization of the recurrence method is straightforward, whereas the hashing method as it is described in Ref. [2] depends on the initial state being of tensor-product form and therefore requires a more careful analysis. Unlike Giedke *et al.* [5], who have studied entanglement purification with imperfect quantum operations, we assume that all operations are error-free. A paper related to ours is Ref. [6] by Eisert *et al.*, who study how distillable entanglement decreases when information about a quantum state is lost.

Before we turn to the actual entanglement purification protocols, we discuss, in Sec. II, the problem of what density operator $\rho^{(N)}$ to assign to N pairs of qubits if only partial information is available. This is an unsolved problem, and we do not attempt to give a general solution. We show, however, that under the additional assumption of *exchangeability*, the state $\rho^{(N)}$ must have a certain simple form, which is amenable to entanglement purification. Our discussion also provides a resolution of the apparent paradox found by Horodecki *et al.* [7], who give an example where applying the Jaynes principle of maximum entropy [8,9] leads to a state with more distillable entanglement than seems to be warranted by the available information. We conclude in Sec. V.

II. STATE ASSIGNMENT BASED ON PARTIAL INFORMATION

Let us consider the example given by Horodecki *et al.* [7]. The authors consider a system composed of a single pair of qubits and define an operator

$$B=\frac{1}{2}(\sigma_x\otimes\sigma_x+\sigma_z\otimes\sigma_z)=(\mathbf{\Phi}_+-\mathbf{\Psi}_-), \tag{1}$$

where $\mathbf{\Psi}_\pm=|\Psi_\pm\rangle\langle\Psi_\pm|$, $\mathbf{\Phi}_\pm=|\Phi_\pm\rangle\langle\Phi_\pm|$ are projectors onto the Bell states,

$$|\Phi_\pm\rangle=\frac{1}{\sqrt{2}}(|00\rangle\pm|11\rangle),$$

*Present address: Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540.

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

Our definition of B differs from that of Ref. [7] by a constant factor to simplify the expressions. If *all that is known* about the system state is the expectation value $\langle B\rangle = 1/2$, then Jaynes's principle of maximum entropy stipulates that one should assign the state of maximum von Neumann entropy compatible with the constraint $\langle B\rangle = 1/2$, which in this case is

$$\rho_J = \frac{9}{16}\mathbf{\Phi}_+ + \frac{1}{16}\mathbf{\Psi}_- + \frac{3}{16}(\mathbf{\Psi}_+ + \mathbf{\Phi}_-). \tag{3}$$

This state has distillable entanglement. Horodecki *et al.* [7] point out that the state

$$\rho_H = \frac{1}{2}\mathbf{\Phi}_+ + \frac{1}{4}(\mathbf{\Psi}_+ + \mathbf{\Phi}_-) \tag{4}$$

also satisfies the constraint $\langle B\rangle = 1/2$, but is separable and, hence, unentangled. They conclude that the entanglement in the maximum entropy state ρ_J is ''fake,'' because it violates the condition that an inference scheme ''should not give us an inseparable estimated state if only theoretically there exists a separable state consistent with the measured data.'' As an alternative to the Jaynes principle, they propose first to minimize the entanglement and then to find the state of maximum entropy among those states that have minimal entanglement. For the constraint $\langle B\rangle = 1/2$, this alternative scheme results in the state ρ_H given above.

A simple defense of the Jaynes principle would be the following (see also Refs. [10,11]). The alternative procedure proposed by Horodecki *et al.* assumes additional information about the two qubits, namely that entanglement is *a priori* unlikely. This would be reasonable, e.g., in a situation where the parties know that the state has been prepared by an adversary whose objective is to let them have as little entanglement as possible. But then more is known about the state than just the given expectation value and, hence, the assumptions behind the Jaynes procedure are not fulfilled.

If there is no specific additional information, however, the maximum entropy state assignment ρ_J is preferable to the minimum entanglement assignment ρ_H. Indeed, if a projective measurement in the Bell basis is performed, assigning ρ_H corresponds to assigning zero probability to the measurement outcome $\mathbf{\Psi}_-$, an outcome that is not ruled out by the constraint $\langle B\rangle = 1/2$. In this sense, the minimum entanglement assignment is inconsistent with the prior information.

By contrast, no inconsistency of this kind can arise from the maximum entropy assignment in the absence of prior information beyond the given expectation value. Any measurement outcome that can be obtained from any state consistent with the constraint can also be obtained from ρ_J. Indeed, for the case at hand, the ρ_J of Eq. (3) has a nonzero probability for any measurement outcome whatever (this is generally true for maximum entropy state assignments, except in singular cases). No measurement on a single system in an unknown state can tell if that system is entangled or not; the only way to do so is to look at the outcomes produced by measurements on *multiple* copies. In particular, there is no way to turn a single ρ_J state into a maximally entangled state, even probabilistically [12], by local operations and classical communication. Thus, the prediction of ''fake entanglement'' for ρ_J causes no practical difficulty. It merely represents the fact that, in some sense, *most* states that are consistent with the constraint are also entangled.

We now turn to the case in which the parties share not just one, but N qubit pairs. We denote by $\rho^{(N)}$ the total state of the N pairs and assume that the N pairs are known to satisfy the constraints $\langle B\rangle_k \equiv \mathrm{Tr}(\rho_k B) = 1/2$ for $k = 1, \ldots, N$, where ρ_k is the reduced density operator of the kth pair. In this case, the state assignment $\rho^{(N)} = \rho_J^{\otimes N}$ is not supported by the prior information, even though this is the state of maximum entropy compatible with the given expectation values. For large N, this state assignment corresponds to the definite prediction that a nonzero number of perfect singlets can be distilled, which is certainly not implied by the given expectation values. The alternative state assignment $\rho^{(N)} = \rho_H^{\otimes N}$ would, however, be equally unsupported by the prior information. It corresponds to the definite prediction that *no* singlets can be distilled from the N pairs, which is the minimum number of distillable singlets compatible with the *a priori* knowledge. Although this is a very cautious prediction, it is also not implied by the given expectation values.

The fact that a naïve application of the principle of maximum entropy to many copies of a system fails is essentially of classical origin and is not unique to problems involving entanglement. Jaynes [13] has given a thorough discussion of this problem, which can be explained by a simple example. Consider a possibly loaded die. All that is known about the die is the mean value $\langle n\rangle \equiv \Sigma_n n p(n) = 3.5$, where $p(n)$ is the probability of the outcome n, $n = 1, \ldots, 6$. The probability distribution of maximum entropy compatible with the given mean-value constraint is $p(n) = 1/6$ for $n = 1, \ldots, 6$. Now, consider throwing the die N times. A naïve application of the maximum entropy principle would predict that the N dice throws were independent and identically distributed (iid) according to the single-trial distribution $p(n)$. This would lead to the prediction that the fraction of throws showing any particular outcome would approximate 1/6 with arbitrary precision as N tended to infinity. This prediction, however, is not implied by the prior knowledge, which is compatible with many possible outcome sequences, including sequences in which only the events $n = 1$ and $n = 6$ ever occur—quite possible, if the die is loaded. Moreover, with an i.i.d. distribution, the results of earlier throws imply nothing about the probability of later outcomes. Even the most gullible gambler might become suspicious if one and six were the only outcomes after thousands of throws.

In Ref. [13], Jaynes discusses how to choose the multitrial distribution in the classical case. The starting point of his discussion is the assumption that the probability distribution of the N dice throws is *exchangeable*. The same assumption is the starting point for our quantum analysis. If exchangeability is assumed, the task of assigning a state of N qubit pairs compatible with the constraints given above is much simplified. A state $\rho^{(N)}$ of N copies of a system is exchange-

able if it is a member of an exchangeable sequence $\rho^{(k)}$, $k=1,2,\ldots$. An exchangeable sequence is defined by

(i) $\rho^{(k)}=\mathrm{Tr}_{k+1}\rho^{(k+1)}$ for all k, where Tr_{k+1} denotes the partial trace over the $(k+1)$th system, and

(ii) each $\rho^{(k)}$ is invariant under permutations of the k systems on which it is defined.

This definition is the quantum generalization of de Finetti's [14] definition of exchangeable sequences of classical random variables.

A state $\rho^{(N)}$ is exchangeable if and only if it can be written in the form

$$\rho^{(N)}=\int d\rho\, p(\rho)\rho^{\otimes N}, \tag{5}$$

where $d\rho$ is a measure on density operator space, and $p(\rho)$ is a normalized generating function, $\int d\rho\, p(\rho)=1$. This is a consequence of the quantum de Finetti theorem, the quantum version of the fundamental representation theorem due to de Finetti [14]. The quantum theorem was first proved by Hudson and Moody [15] after pioneering work by Størmer [16]; for an elementary proof, see Ref. [17].

How, in general, do we pick $p(\rho)d\rho$? To our knowledge, there is no universal rule for this task, although there exist a number of proposals for unbiased measures $d\rho$ on density operator space [18–21]. These can be interpreted as proposals for state assignments for N systems under the sole assumption of exchangeability, i.e., using a generating function $p(\rho)=1$.

If, in addition to exchangeability, there is a mean-value constraint $\langle O\rangle=o$, the naïve Jaynes maximum entropy state assignment leads to a generating function of the form $p(\rho)=\delta(\rho-\rho_J)$, where ρ_J is the single-system state of maximum entropy, subject to the constraint; this generating function is unacceptable for the reasons given above. A good choice of $p(\rho)d\rho$ should be nonzero for all ρ that are compatible with the prior information—we should never arbitrarily rule out any possibility. Similarly, $p(\rho)d\rho$ ought to vanish for any ρ that is actually ruled out by the prior information. We therefore would expect a multisystem generalization of Jaynes's maximum entropy procedure to have the form

$$p_{\mathrm{MAXENT}}(\rho)d\rho=\mathcal{N}\delta[o-\mathrm{Tr}(O\rho)]f(\rho)d\rho, \tag{6}$$

where $\mathcal{N}$ is a normalization constant and $f(\rho)d\rho$ is strictly positive. The exact form of the function $f(\rho)$ and of the measure $d\rho$ is the subject of ongoing research. In the spirit of the single-system Jaynes principle, $p(\rho)d\rho$ should favor states ρ with higher von Neumann entropy $S(\rho)$ and should give the usual ρ_J when $N=1$ [22].

Given an initial state assignment of the form (5), additional information can be obtained, e.g., by making measurements on individual subsystems. Suppose a measurement outcome k is represented by a positive single-system operator F_k, with $\Sigma_k F_k=1$; i.e., the F_k form a positive-operator valued measure (POVM) [23]. Given that the subsystem is in state ρ, the probability of getting outcome k is $p(k|\rho)=\mathrm{Tr}[F_k\rho]$. If the total state is given by Eq. (5), the probability of outcome k in a measurement on a single subsystem is then

$$p_k=\int d\rho\, p(\rho)p(k|\rho). \tag{7}$$

After the measurement, we must update the state of the remaining $N-1$ systems by the Bayes rule,

$$\rho^{(N-1)}=\int d\rho\, p(\rho|k)\rho^{\otimes(N-1)}, \tag{8}$$

where [24]

$$p(\rho|k)=\frac{p(\rho)p(k|\rho)}{p_k}. \tag{9}$$

By doing different measurements on several subsystems, we acquire more and more data; if these measurements are chosen well, the resulting posterior $p_{\mathrm{post}}(\rho)$ becomes more and more peaked and has less and less dependence on the choice of prior $p(\rho)$. This procedure is a straightforward Bayesian version of quantum-state tomography [25–27].

The condition of exchangeability in combination with the quantum de Finetti theorem provides only a partial solution to the problem of state assignment in the presence of partial information, but we show in the next two sections that exchangeability alone is sufficient to guarantee that the entanglement purification procedures known as one-way hashing and recurrence can be carried out. The probability of distilling a positive yield of maximally entangled states depends on the exact form of $p(\rho)d\rho$ in Eq. (5). It is exchangeable states of this type that are provided by the generalization (6) of the Jaynes construction [22], so we conclude that at least in principle, it is indeed possible to purify entanglement from partially known states.

III. ENTANGLEMENT PURIFICATION BY ONE-WAY HASHING

In this section, we first present a version of the one-way hashing algorithm that proceeds by Bayesian updating of the probabilities for products of Bell states and that can, in principle, be applied to general exchangeable states. We then briefly sketch the argument given in Ref. [2] that for a product state $\rho^{\otimes N}$, where ρ is Bell diagonal with von Neumann entropy S, the asymptotic yield of pure singlets is given by $N(1-S)$. We show how to modify this argument so that it can be applied to general exchangeable states. Finally, we give a simplified Bayesian hashing algorithm for exchangeable states and discuss its asymptotic yield. Our analysis is restricted to pairs of qubits, but the method generalizes straightforwardly to arbitrary Hilbert space dimensions.

We restrict attention to Bell-diagonal states, i.e., mixtures of the Bell states,

$$\rho_{\vec{w}}=w_1\mathbf{\Psi}_-+w_2\mathbf{\Psi}_+ +w_3\mathbf{\Phi}_-+w_4\mathbf{\Phi}_+, \tag{10}$$

where we denote the weights by $\vec{w}=\{w_1,w_2,w_3,w_4\}$, $w_1+w_2+w_3+w_4=1$, $w_j\geq 0$ for $j=1,\ldots,4$. Most existing entanglement purification procedures begin by making this assumption. If it does not hold, it is possible to put any state in this form by ''twirling,'' that is, by randomly rotating both spins of an entangled pair. The final yield of maximally entangled states cannot be diminished by omitting this step, however, so it is better to think of twirling as a conceptual, rather than a physical procedure. After twirling, the initial, exchangeable state (5) of our N pairs of qubits becomes

$$\rho^{(N)}=\int d\vec{w}\, p(\vec{w})\rho_{\vec{w}}^{\otimes N}, \tag{11}$$

where

$$\int d\vec{w}\, p(\vec{w})=1. \tag{12}$$

We now define the set of labeled states

$$\rho_{00}=\mathbf{\Psi}_-, \quad \rho_{01}=\mathbf{\Psi}_+,$$
$$\rho_{10}=\mathbf{\Phi}_-, \quad \rho_{11}=\mathbf{\Phi}_+. \tag{13}$$

The first bit in the label tells us whether the pair is in a Ψ or a Φ state; the second bit tells us whether it is in a $+$ or $-$ state. If we are restricted to local measurements and classical communication on a single pair, the best we can do is to determine one of these two bits, but not both, and the pair will be left in an unentangled state. Bennett *et al.* have shown, however, that if we can manipulate the qubits collectively, much more interesting measurements are possible [2].

The first step is to rewrite the state (11) as a probability distribution over strings of bits, with each qubit pair associated with two bits in the string. For this, we define the product distribution

$$p(i_1i_2\cdots i_{2N}|\vec{w})=w_{i_1i_2}w_{i_3i_4}\cdots w_{i_{2N-1}i_{2N}}, \tag{14}$$

where $w_{00}\equiv w_1$, $w_{01}\equiv w_2$, $w_{10}\equiv w_3$, and $w_{11}\equiv w_4$. Using this notation,

$$\rho^{(N)}=\sum_{i_1,i_2,\ldots,i_{2N}} p(i_1i_2\cdots i_{2N})\rho_{i_1i_2}\otimes\cdots\otimes\rho_{i_{2N-1}i_{2N}}, \tag{15}$$

where

$$p(i_1i_2\cdots i_{2N})=\int d\vec{w}\, p(\vec{w})p(i_1i_2\cdots i_{2N}|\vec{w}). \tag{16}$$

We now select a random subset of the bits $i_1\cdots i_{2N}$ and list all the qubit pairs that have at least one associated bit in the subset. From this list we choose one qubit pair to be the *target*. For each of the other qubit pairs in the list, Alice and Bob both perform one of a set of three unitary transformations on their half of the pair, followed by a bilateral controlled–NOT operation onto the target pair. This sequence of operations is equivalent to replacing one of the bits of the target pair with the parity of a subset of all the bits. The choice of unitary transformation corresponds to including the first, second, or both of the bits from a particular pair in the parity calculation. Then a measurement is performed on the target pair. (The details of this procedure are given in [2].) By carrying out such a procedure, one bit of joint information is acquired about all the pairs, at the expense of sacrificing one entangled pair (that is, two bits). The unmeasured pairs in general undergo an invertible transformation among the Bell states, but they do not become entangled with each other, and this transformation can, if one chooses, be undone, leaving the sequence of bits for the unmeasured pairs unaltered. Bennett *et al.* have shown that such a procedure can be equivalent to finding the parity of any subset of the $2N$ bits. This parity bit then allows one to update the probability distribution for the remaining $2(N-1)$-bit string.

Let us examine this in a little more detail. Let $\vec{\imath}\equiv i_1i_2\cdots i_{2N}$ denote a sequence of bits. We can select a subset of these bits by giving another sequence $\vec{x}$, which includes a 1 for each bit to be included in the subset and a 0 for the rest. The parity of the subset is then

$$\pi_{\vec{x}}(\vec{\imath})\equiv\vec{x}\cdot\vec{\imath}\equiv\left(\sum_{m=1}^{2N}x_mi_m\right)\bmod 2. \tag{17}$$

For a given $\vec{\imath}$, the probability of getting a value $\pi_{\vec{x}}$ for the parity is either 0 or 1, so the probability of getting $\pi_{\vec{x}}$ as a measurement result is

$$p(\pi_{\vec{x}})=\sum_{\vec{\imath}} p(\pi_{\vec{x}}|\vec{\imath})p(\vec{\imath})=\sum_{\vec{\imath}} \delta_{\pi_{\vec{x}},\vec{x}\cdot\vec{\imath}}p(\vec{\imath}). \tag{18}$$

For simplicity, let us assume that the target pair is the last, so the last two bits are sacrificed; the new state for the $N-1$ remaining pairs is

$$\rho^{(N-1)}=\sum_{\vec{\imath}'} p(\vec{\imath}'|\pi_{\vec{x}})\rho_{i_1i_2}\otimes\rho_{i_3i_4}\otimes\cdots\otimes\rho_{i_{2N-3}i_{2N-2}}, \tag{19}$$

where $\vec{\imath}'\equiv i_1i_2\cdots i_{2N-2}$ and

$$p(\vec{\imath}'|\pi_{\vec{x}})=\sum_{i_{2N-1},i_{2N}} p(\vec{\imath}|\pi_{\vec{x}})=\sum_{i_{2N-1},i_{2N}}\frac{p(\vec{\imath})p(\pi_{\vec{x}}|\vec{\imath})}{p(\pi_{\vec{x}})}. \tag{20}$$

Note that while the initial probability distribution $p(\vec{\imath})$ is symmetric under interchanges of the pairs, this symmetry is lost after measurement.

The purification scheme follows simply from this. One chooses subsets of the bit string at random and measures their parity, sacrificing one pair with each measurement, but updating the probability distribution for the remaining strings. This procedure is repeated until one is left with only a single string, say $\vec{\imath}_0$, with probability $1-\delta$ for some small δ. Written more formally, the posterior probability p_{post} at the end of the procedure, conditioned on all measurement results, has the property $p_{\text{post}}(\vec{\imath}_0)=1-\delta$ for some sequence

$\vec{i}_0$. One then knows with high probability the maximally entangled state of each remaining pair, which can then be transformed into a standard state (such as Ψ_-) by local operations. The yield of this procedure is the number of entangled pairs left at the end.

It is clear that there are states for which the yield is zero. The obvious example is a state $\rho^{\otimes N}$ where ρ is unentangled. For states of the form $\rho_{\vec{w}}^{\otimes N}$, Bennett *et al.* have shown that asymptotically, the method gives a yield of $N(1-S_{\vec{w}})$ maximally entangled pairs with fidelity approaching 1, where

$$S_{\vec{w}} = -\sum_{j=1}^{4} w_j \log w_j = -\mathrm{Tr}(\rho_{\vec{w}} \log \rho_{\vec{w}}) \tag{21}$$

is the entropy of $\rho_{\vec{w}}$. The argument makes use of the theorem of typical sequences [28] (which is closely related to Shannon's noiseless coding theorem [29]), according to which, for any $\epsilon > 0$ and $\delta > 0$ and sufficiently large N, there exists a subset $\mathcal{S}_{\mathrm{TYP}}(N)$ of the set of all sequences $\vec{i}$ with the following properties:

$$p[\mathcal{S}_{\mathrm{TYP}}(N)] \equiv \sum_{\vec{i} \in \mathcal{S}_{\mathrm{TYP}}(N)} p(\vec{i}|\vec{w}) \geqslant 1-\epsilon, \tag{22}$$

i.e., the total probability of the set $\mathcal{S}_{\mathrm{TYP}}(N)$ is arbitrarily close to 1; and

$$|\mathcal{S}_{\mathrm{TYP}}(N)| \leqslant 2^{N(S_{\vec{w}}+\delta)}, \tag{23}$$

i.e., the number of sequences in $\mathcal{S}_{\mathrm{TYP}}(N)$ is not much larger than $2^{NS_{\vec{w}}}$. The set $\mathcal{S}_{\mathrm{TYP}}(N)$ is called the set of typical sequences. Since the parity measurement in each hashing round rules out half the typical sequences on average, and since essentially all the probability is concentrated on the typical sequences, it can be expected that after sacrificing approximately $NS_{\vec{w}}$ pairs, essentially all the probability is concentrated on a single typical sequence. Clearly this leads to a positive yield only if $S_{\vec{w}} < 1$.

The theorem of typical sequences does not hold in general for sequences corresponding to exchangeable states of the form (11). To apply the hashing method in this case, we rely on a generalization of the theorem of typical sequences due to Cziszár and Körner [30] (this theorem has recently been used by Jozsa *et al.* [31] to derive a universal quantum information compressing scheme). Applied to our setting, the theorem is that, given a fixed entropy S_0, then for any $\epsilon > 0$ and $\delta > 0$ and sufficiently large N, there exists a subset $\mathcal{S}_{\mathrm{CK}}(N)$ of the set of all sequences $\vec{i}$ with the following properties:

$$\sum_{\vec{i} \in \mathcal{S}_{\mathrm{CK}}(N)} p(\vec{i}|\vec{w}) \geqslant 1-\epsilon, \tag{24}$$

for all $\vec{w}$ such that $S_{\vec{w}} < S_0$, which means that the set $\mathcal{S}_{\mathrm{CK}}(N)$ is typical for all probability distributions with entropy less than S_0; and

$$|\mathcal{S}_{\mathrm{CK}}(N)| \leqslant 2^{N(S_0+\delta)}, \tag{25}$$

i.e., the number of sequences in $\mathcal{S}_{\mathrm{CK}}(N)$ is not much larger than 2^{NS_0}. In the following, when we write ''typical sequences,'' we mean sequences in $\mathcal{S}_{\mathrm{CK}}(N)$, whereas by ''atypical sequences'' we mean sequences in $\overline{\mathcal{S}}_{\mathrm{CK}}(N)$, the complement of $\mathcal{S}_{\mathrm{CK}}(N)$.

Now assume that we want to perform the hashing protocol on a state of N pairs of the form (11) with the property

$$\int_{S_{\vec{w}} > S_0} d\vec{w}\, p(\vec{w}) = \eta \ll 1 \tag{26}$$

for some entropy $S_0 < 1$; i.e., there is only a small *a priori* probability that the entropy of the unknown state exceeds the given value S_0. (The case of states that do not have this property will be discussed at the end of this section.) Furthermore, assume that N is large enough that there exists a Cziszár-Körner set $\mathcal{S}_{\mathrm{CK}}(N)$ with constants $\epsilon, \delta \ll 1$ in Eqs. (24) and (25). It then follows that

$$\begin{aligned}
p[\mathcal{S}_{\mathrm{CK}}(N)] &\equiv \sum_{\vec{i} \in \mathcal{S}_{\mathrm{CK}}(N)} p(\vec{i}) \\
&= \sum_{\vec{i} \in \mathcal{S}_{\mathrm{CK}}(N)} \int d\vec{w}\, p(\vec{w}) p(\vec{i}|\vec{w}) \\
&\geqslant \sum_{\vec{i} \in \mathcal{S}_{\mathrm{CK}}(N)} \int_{S_{\vec{w}} < S_0} d\vec{w}\, p(\vec{w}) p(\vec{i}|\vec{w}) \\
&= \int_{S_{\vec{w}} < S_0} d\vec{w}\, p(\vec{w}) \sum_{\vec{i} \in \mathcal{S}_{\mathrm{CK}}(N)} p(\vec{i}|\vec{w}) \\
&\geqslant \int_{S_{\vec{w}} < S_0} d\vec{w}\, p(\vec{w})(1-\epsilon) \\
&= (1-\eta)(1-\epsilon) \\
&\geqslant 1-\eta-\epsilon,
\end{aligned} \tag{27}$$

where Eqs. (12), (24), and (26) have been used.

We use this inequality, in combination with Eq. (25), to derive the asymptotic yield of the hashing algorithm applied to exchangeable states. We restrict our analysis to a simplified protocol, in which we choose a number r, somewhat larger than $N(S_0+\delta)$, such that

$$\zeta \equiv 2^{N(S_0+\delta)-r} \ll 1. \tag{28}$$

We begin with input strings $\vec{i}$ that have probability $p(\vec{i})$. Let h denote a sequence of r parity checks on random subsets, and let $\vec{o} = o_1, \ldots, o_r$ denote the r-bit string of parity checks, or outcomes. (Note that we denote all strings of bits as vectors, even though they are not all of the same length.) The probability distribution $p(h)$ on parity-check sequences is weighted uniformly on all sequences. For a given input string $\vec{i}$ and a given parity-check sequence h, the outcome $\vec{o}$ is determined; we denote this deterministic outcome by $\vec{o}_{h;\vec{i}}$. We can express this deterministic outcome in terms of a

probability for outcome string $\vec{o}$, given parity-check sequence h and input string $\vec{\imath}$:

$$p(\vec{o}|h,\vec{\imath})=\delta_{\vec{o},\vec{o}_{h;\vec{\imath}}}. \tag{29}$$

Since for each parity-check bit obtained, two bits of the input string are discarded, two strings with the same parity check, which differ only on those two bits, become the same after that step. After r steps of a parity-check sequence h, there will be only $N-r$ entangled pairs, corresponding to a string of $2(N-r)$ bits. If one starts with a string $\vec{\imath}$, one will be left with a shorter substring $\vec{\imath}_h(\vec{\imath})$. Different initial strings $\vec{\imath}$ that generate the same outcome $\vec{o}$ and lead to the same final substring $\vec{\imath}_h(\vec{\imath})$ are equivalent for practical purposes. Let us denote the set of all *input* strings $\vec{\imath}$ that lead to outcome $\vec{o}$ and to *output* substring $\vec{\imath}_h$ by $I_h(\vec{o},\vec{\imath}_h)\equiv\{\vec{\imath}|\vec{o}_{h;\vec{\imath}}=\vec{o},\vec{\imath}_h(\vec{\imath})=\vec{\imath}_h\}$.

For parity-check sequence h, we are interested in outcomes $\vec{o}$ such that all *typical* input strings $\vec{\imath}$ that lead to $\vec{o}$ produce the same output string $\vec{\imath}_h(\vec{\imath})$. For outcomes where this is the case, the procedure picks out a unique output string from among all those that could be produced by a typical input string. In this case, we say that we *accept* the outcome $\vec{o}$ and the corresponding unique output string, which we denote by $\vec{\imath}_{h;\vec{o}}$. In this way, we divide the outcomes for a parity-check sequence h into two sets, the set of accepted outcomes, A_h, and its complement. For an outcome that we accept and for a typical input string, we can write the conditional probability (29) as

$$\begin{aligned}p(\vec{o}|h,\vec{\imath},\vec{\imath}\in\mathcal{S}_{\mathrm{CK}}(N))&=\begin{cases}1, & \text{if } \vec{\imath}\in I_h(\vec{o},\vec{\imath}_{h;\vec{o}}),\\ 0, & \text{if } \vec{\imath}\notin I_h(\vec{o},\vec{\imath}_{h;\vec{o}}),\end{cases}\\ &=\delta_{\vec{o},\vec{o}_{h;\vec{\imath}}}\delta_{\vec{\imath}_h(\vec{\imath}),\vec{\imath}_{h;\vec{o}}}\ \text{ for } \vec{o}\in A_h.\end{aligned} \tag{30}$$

Though the additional Kronecker delta in this expression is redundant, it reminds one that any *typical* input string $\vec{\imath}$ that leads to an *accepted* outcome $\vec{o}$ produces output string $\vec{\imath}_{h;\vec{o}}$. Notice that this is not true for atypical input strings: an atypical input string can have outcome $\vec{o}$ and produce outcome string $\vec{\imath}_{h;\vec{o}}$ or a different output string.

The probability that the outcome is accepted, given input string $\vec{\imath}$ and parity-check sequence h, is

$$\begin{aligned}p(\mathrm{accept}|h,\vec{\imath})&=\sum_{\vec{o}\in A_h}p(\vec{o}|h,\vec{\imath})\\ &=\sum_{\vec{o}\in A_h}\delta_{\vec{o},\vec{o}_{h;\vec{\imath}}}\\ &=\begin{cases}1, & \text{if } \vec{\imath} \text{ leads to an accepted outcome,}\\ 0, & \text{if } \vec{\imath} \text{ does not lead to an accepted outcome.}\end{cases}\end{aligned} \tag{31}$$

Notice that this conditional acceptance probability can be nonzero for atypical input strings. The complementary probability, that the outcome is not accepted, given $\vec{\imath}$ and h, is given by

$$\begin{aligned}p(\overline{\mathrm{accept}}|h,\vec{\imath})&=\sum_{\vec{o}\notin A_h}p(\vec{o}|h,\vec{\imath})\\ &=\sum_{\vec{o}\notin A_h}\delta_{\vec{o},\vec{o}_{h;\vec{\imath}}}\\ &=\begin{cases}0, & \text{if } \vec{\imath} \text{ leads to an accepted outcome,}\\ 1, & \text{if } \vec{\imath} \text{ does not lead to an accepted outcome.}\end{cases}\end{aligned} \tag{32}$$

If the input string is a typical string, the conditional acceptance probability can also be written as

$$p[\mathrm{accept}|h,\vec{\imath},\vec{\imath}\in\mathcal{S}_{\mathrm{CK}}(N)]=\sum_{\vec{o}\in A_h}\delta_{\vec{o},\vec{o}_{h;\vec{\imath}}}\delta_{\vec{\imath}_h(\vec{\imath}),\vec{\imath}_{h;\vec{o}}} \tag{33}$$

[see Eq. (30)].

What we are interested in for the present is the probability to have an outcome that is accepted, given a typical input string, but averaged over all parity-check sequences:

$$\begin{aligned}p[\mathrm{accept}|\vec{\imath},\vec{\imath}\in\mathcal{S}_{\mathrm{CK}}(N)]&=\sum_h p[\mathrm{accept}|h,\vec{\imath},\\ &\qquad \vec{\imath}\in\mathcal{S}_{\mathrm{CK}}(N)]p(h)\\ &=\sum_h p(h)\sum_{\vec{o}\in A_h}\delta_{\vec{o},\vec{o}_{h;\vec{\imath}}}\delta_{\vec{\imath}_h(\vec{\imath}),\vec{\imath}_{h;\vec{o}}}.\end{aligned} \tag{34}$$

The complementary probability,

$$\begin{aligned}p[\overline{\mathrm{accept}}|\vec{\imath},\vec{\imath}\in\mathcal{S}_{\mathrm{CK}}(N)]&=\sum_h p[\overline{\mathrm{accept}}|h,\vec{\imath},\\ &\qquad \vec{\imath}\in\mathcal{S}_{\mathrm{CK}}(N)]p(h)\\ &=\sum_h p(h)\sum_{\vec{o}\notin A_h}\delta_{\vec{o},\vec{o}_{h;\vec{\imath}}},\end{aligned} \tag{35}$$

is the average probability not to have an outcome that is accepted, given the typical input string $\vec{\imath}$. This probability is the probability that for a random parity-check sequence, the typical input string $\vec{\imath}$ leads to an outcome that does not pick out a unique output string $\vec{\imath}_{h;\vec{o}}$, i.e., leads to an output string that could have been produced by more than one typical input string. We can bound this probability in the following way. The number of typical sequences satisfies $|\mathcal{S}_{\mathrm{CK}}(N)|\leqslant 2^{N(S_0+\delta)}$. For parity subsets chosen randomly, the probability that two typical input strings, $\vec{\imath}$ and $\vec{\jmath}$, agree on all r

parity checks—i.e., have the same outcome—is $\leqslant 2^{-r}$; thus the probability that $\vec{\imath}$ and $\vec{\jmath}$ agree on all r parity checks *and* produce *different* output strings, $\vec{\imath}_h(\vec{\imath})$ and $\vec{\imath}_h(\vec{\jmath})$, is $\leqslant 2^{-r}$. Hence, the probability of not producing a unique output, given a typical input $\vec{\imath}$, is bounded by

$$p[\overline{\text{accept}}|\vec{\imath},\vec{\imath}\in\mathcal{S}_{\text{CK}}(N)]\leqslant 2^{-r}\times 2^{N(S_0+\delta)}=\zeta. \tag{36}$$

This implies that the conditional acceptance probability (34) satisfies

$$p[\text{accept}|\vec{\imath},\vec{\imath}\in\mathcal{S}_{\text{CK}}(N)]\geqslant 1-\zeta. \tag{37}$$

The Bayes rule tells us that the posterior probability for output string $\vec{\imath}_h$, given h and $\vec{o}$, is

$$\begin{aligned} p(\vec{\imath}_h|h,\vec{o}) &= \sum_{\vec{\imath}\in I_h(\vec{o},\vec{\imath}_h)} p(\vec{\imath}|h,\vec{o}) \\ &= \sum_{\vec{\imath}\in I_h(\vec{o},\vec{\imath}_h)} \frac{p(\vec{o}|h,\vec{\imath})p(h)p(\vec{\imath})}{p(\vec{o}|h)p(h)} \\ &= \sum_{\vec{\imath}\in I_h(\vec{o},\vec{\imath}_h)} \frac{p(\vec{o}|h,\vec{\imath})p(\vec{\imath})}{p(\vec{o}|h)}, \end{aligned} \tag{38}$$

where

$$p(\vec{o}|h)=\sum_{\vec{\imath}} p(\vec{o}|h,\vec{\imath})p(\vec{\imath}) \tag{39}$$

is the probability for outcome string $\vec{o}$, given parity-check sequence h.

Given a parity-check sequence h and an accepted outcome $\vec{o}\in A_h$ for that sequence, we judge the ''success'' of the accepted output string $\vec{\imath}_{h;\vec{o}}$ by the posterior probability, i.e,

$$\begin{aligned} p(\text{success}|h,\vec{o}) &= p(\vec{\imath}_{h;\vec{o}}|h,\vec{o}) \\ &= \sum_{\vec{\imath}\in I_h(\vec{o},\vec{\imath}_{h;\vec{o}})} p(\vec{\imath}|h,\vec{o}) \quad \text{for } \vec{o}\in A_h. \end{aligned} \tag{40}$$

The total probability of success, $p(\text{success})$, is obtained by averaging over all parity-check sequences h and over all accepted outcomes $\vec{o}\in A_h$. This probability can be manipulated in the following ways:

$$\begin{aligned} p(\text{success}) &= \sum_h \sum_{\vec{o}\in A_h} p(\text{success}|h,\vec{o})p(\vec{o}|h)p(h) \\ &= \sum_h \sum_{\vec{o}\in A_h} \sum_{\vec{\imath}\in I_h(\vec{o},\vec{\imath}_{h;\vec{o}})} p(\vec{\imath}|h,\vec{o})p(\vec{o}|h)p(h) \\ &= \sum_h \sum_{\vec{o}\in A_h} \sum_{\vec{\imath}\in I_h(\vec{o},\vec{\imath}_{h;\vec{o}})} p(\vec{o}|h,\vec{\imath})p(h)p(\vec{\imath}) \\ &\geqslant \sum_h \sum_{\vec{o}\in A_h} \sum_{\substack{\vec{\imath}\in I_h(\vec{o},\vec{\imath}_{h;\vec{o}}) \\ \vec{\imath}\in\mathcal{S}_{\text{CK}}(N)}} p(\vec{o}|h,\vec{\imath})p(h)p(\vec{\imath}) \\ &= \sum_h \sum_{\vec{o}\in A_h} \sum_{\vec{\imath}\in\mathcal{S}_{\text{CK}}(N)} \delta_{\vec{o},\vec{o}_{h;\vec{\imath}}}\delta_{\vec{\imath}_h(\vec{\imath}),\vec{\imath}_{h;\vec{o}}}p(h)p(\vec{\imath}). \end{aligned} \tag{41}$$

The inequality here follows from restricting the sum over input strings to typical strings and reflects the fact that an atypical string might lead to an accepted outcome *and* to the accepted output string $\vec{\imath}_{h;\vec{o}}$, thereby contributing to the success probability. The final equality comes from using Eq. (30) for $p(\vec{o}|h,\vec{\imath})$. Using Eqs. (27), (34), and (37), we can now bound the probability of success:

$$\begin{aligned} p(\text{success}) &\geqslant \sum_{\vec{\imath}\in\mathcal{S}_{\text{CK}}(N)} p(\vec{\imath})\sum_h p(h)\sum_{\vec{o}\in A_h} \delta_{\vec{o},\vec{o}_{h;\vec{\imath}}}\delta_{\vec{\imath}_h(\vec{\imath}),\vec{\imath}_{h;\vec{o}}} \\ &= \sum_{\vec{\imath}\in\mathcal{S}_{\text{CK}}(N)} p(\vec{\imath})p[\text{accept}|\vec{\imath},\vec{\imath}\in\mathcal{S}_{\text{CK}}(N)] \\ &\geqslant (1-\zeta)\sum_{\vec{\imath}\in\mathcal{S}_{\text{CK}}(N)} p(\vec{\imath}) \\ &= (1-\zeta)p[\mathcal{S}_{\text{CK}}(N)] \\ &\geqslant (1-\zeta)(1-\eta-\epsilon) \\ &\geqslant 1-\zeta-\eta-\epsilon. \end{aligned} \tag{42}$$

This is the desired result. Assuming we can choose arbitrary positive constants ϵ and η and have sufficiently large N, the probability (42) can be made arbitrarily close to one.

Except for certain singular distributions $p(\vec{w})$, given an exchangeable state of the form (11), it is always possible to make η in Eq. (26) arbitrarily small by choosing the entropy S_0 sufficiently large ($0\leqslant S_0<2$); if $S_0\geqslant 1$, however, then the number of hashing rounds $r\geqslant N$, which means there is no yield since $N-r\leqslant 0$. To decrease the value of S_0 and thereby make the yield positive or increase an already positive yield, one can perform quantum-state tomography on some of the pairs to obtain more data about the state, generally producing a narrower posterior distribution $p'(\vec{w})$ (see Sec. II). The width of the posterior distribution depends on the number of pairs sacrificed for the tomographic measurements, but not on the total number of pairs N. The number of pairs needed for tomography can therefore be neglected in the asymptotic limit of large N.

Asymptotically, the prior probability of obtaining a posterior $p'(\vec{w})$ concentrated at $\vec{w}=\vec{w}_0$ with an entropy $S_{\vec{w}_0}<S_0$ is given by the expression

$$p(S<S_0)\equiv\int_{S_{\vec{w}}<S_0} d\vec{w}\,p(\vec{w}), \tag{43}$$

where $p(\vec{w})$ is the prior distribution (11) defining the initial state. Putting everything together we see that, for $S_0<1$, $p(S<S_0)$ is the probability of obtaining an asymptotic yield of $N(1-S_0)$ using a combination of quantum-state tomography and one-way hashing.

If most of the prior distribution $p(\vec{w})$ is concentrated on states with an entropy exceeding one bit, i.e., if $p(S<1)$ is small, then it will normally be a better strategy to precede the hashing procedure by a few iterations of the recurrence method. This is the content of the next section.

IV. ENTANGLEMENT PURIFICATION BY RECURRENCE

If the generating function $p(\vec{w})$ has no significant support on weights $\vec{w}$ with $S_{\vec{w}}<1$, then hashing cannot be used for entanglement purification, at least initially. It might still be possible, however, to distill some entanglement by using the more robust (but far more wasteful) technique of *recurrence* [2,4].

In the recurrence algorithm, an initial set of $2N$ entangled qubit pairs is grouped into N sets of two pairs each. In each set, one pair is designated the *target* pair, and the other the *control* pair. Alice and Bob thus have N target qubits and N control qubits each. Alice now rotates all her qubits by $\pi/2$ about the x axis, while Bob rotates all his qubits by $-\pi/2$ about the x axis. Each of them then performs a controlled-NOT operation from each control qubit onto the corresponding target qubit and measures his or her target qubits in the z basis ($|0\rangle$ and $|1\rangle$). The target qubits are then discarded. If Alice and Bob both get the same result for a given target pair (i.e., both 0 or both 1), the procedure has succeeded, and the control pair can be shown to have increased entanglement. If their results differ, the procedure has failed, and the control qubits must also be discarded.

If the state of both target and control pairs is of form (10), the probability of success is

$$p_s=p_s(\vec{w})=(w_1+w_4)^2+(w_2+w_3)^2, \tag{44}$$

and the new state of the control pair after the measurement has weights [4]

$$\begin{aligned} w_1'&=2w_2w_3/p_s,\\ w_2'&=(w_2^2+w_3^2)/p_s,\\ w_3'&=2w_1w_4/p_s,\\ w_4'&=(w_1^2+w_4^2)/p_s. \end{aligned} \tag{45}$$

If initially $w_4>1/2$, then this procedure converges towards $w_4=1$. The convergence is slow, however, and since more than half of all the pairs is discarded each time, the yield is generally low.

Suppose that instead of a product state we have an exchangeable state of the form (11), perhaps arising from a Jaynes-type state assignment. We can carry out the procedure exactly as before, grouping the pairs into sets of two, with a target and control bit. If there are initially $2N$ pairs in the state

$$\rho^{(2N)}=\int d\vec{w}\, p(\vec{w})\rho_{\vec{w}}^{\otimes 2N}, \tag{46}$$

then after performing the measurements, Alice and Bob will get the same result N_s times and different results $N-N_s$ times, leaving them with a new state of the form (46) for N_s pairs. For large N, the posterior distribution $p(\vec{w}|N_s)$ will generally be sharply peaked about those $\vec{w}$ that give a value of p_s close to N_s/N. Unlike hashing, the recurrence algorithm produces a posterior state $\rho^{(N_s)}$ which is exchangeable. We now turn to how we find this state in light of the measurement results.

Compared with the hashing algorithm, where precisely one bit of information is obtained in each round of the procedure, in the recurrence method much more information is obtained, namely the value of N_s. We can therefore deduce the posterior distribution

$$p(\vec{w}|N_s)=\frac{p(N_s|\vec{w})p(\vec{w})}{p(N_s)}, \tag{47}$$

where

$$p(N_s|\vec{w})=\binom{N}{N_s}[p_s(\vec{w})]^{N_s}[1-p_s(\vec{w})]^{N-N_s}, \tag{48}$$

and

$$p(N_s)=\int d\vec{w}\, p(N_s|\vec{w})p(\vec{w}). \tag{49}$$

Because the remaining states have been transformed according to Eq. (45), we must also change to the new variables $\vec{w}'$. So the new state is

$$\tilde{\rho}^{(N_s)}=\int d\vec{w}'\, p'(\vec{w}')\rho(\vec{w}')^{\otimes N_s}, \tag{50}$$

where

$$p'(\vec{w}')d\vec{w}'=p(\vec{w}|N_s)\, d\vec{w}. \tag{51}$$

While this Bayesian procedure is very simple compared to the hashing method, it is still a bit too complicated for simple illustration. There is, however, an even simpler variant of this technique that is easy to analyze. Suppose that, instead of the general Bell-diagonal state (10), we have an initial Werner state,

$$\rho(F)=F\mathbf{\Phi}_+ +\frac{1-F}{3}(\mathbf{\Phi}_- +\mathbf{\Psi}_+ +\mathbf{\Psi}_-). \tag{52}$$

We can carry out the recurrence procedure exactly as above, with the probability of success

$$p_s(F)=(8F^2-4F+5)/9; \tag{53}$$

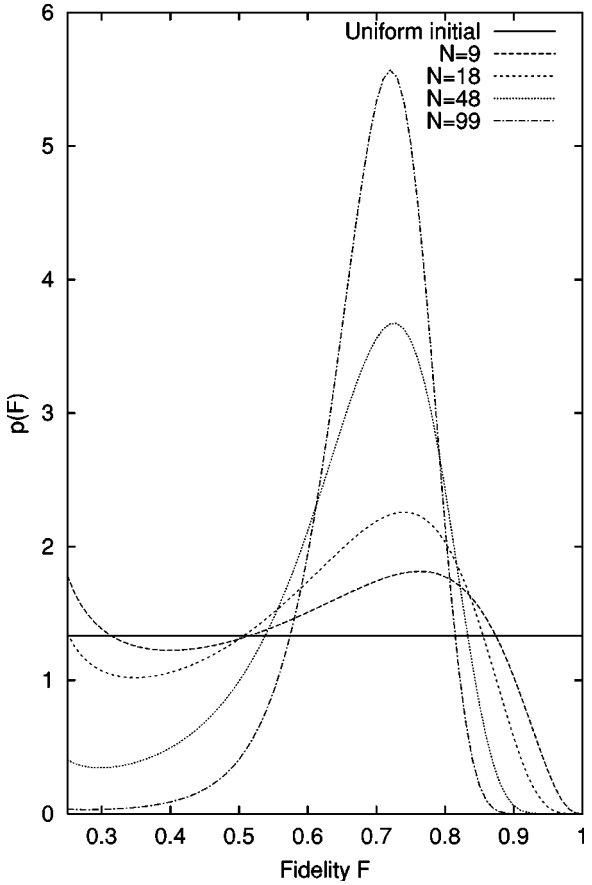

FIG. 1. Plots of an initially uniform distribution for the generalized Werner state for fidelities between $F=1/4$ (maximally mixed) and $F=1$ (maximally entangled) and updated distributions after one round of the simplified recurrence method. Before the round there are $2N$ pairs; we assume the procedure succeeds in $N_s=2N/3$ cases ($p_s=2/3$). The new distribution is plotted for $N=9,18,48,99$. The new distribution is more and more highly peaked for bigger N, and the probability of unentangled states is more and more strongly suppressed.

here F denotes the fidelity of the state with $\mathbf{\Phi}_+$, with $F>1/2$ necessary for distillability. The recurrence procedure does not in general lead to a new state of form (52), but by twirling the state, can be put in this form, at the cost of some increase in entropy. The new state has a fidelity

$$F'=\frac{10F^2-2F+1}{8F^2-4F+5}. \tag{54}$$

Suppose that we have $2N$ entangled pairs, with partial information sufficient to determine that they are all in a state of the form (52), but not to determine the exact fidelity F. The joint state of the pairs is then

$$\rho^{(2N)}=\int dF\, p(F)\rho(F)^{\otimes 2N}. \tag{55}$$

We then group the pairs into sets of two and carry out the recurrence procedure on each set, with N_s successful results. We can then deduce a revised generating function

$$p(F|N_s)=\frac{p(N_s|F)p(F)}{p(N_s)}, \tag{56}$$

where

$$p(N_s|F)=\binom{N}{N_s}[p_s(F)]^{N_s}[1-p_s(F)]^{N-N_s}, \tag{57}$$

and

$$p(N_s)=\int dF\, p(N_s|F)p(F). \tag{58}$$

The new density operator for the N_s remaining pairs is

$$\rho^{(N_s)}=\int dF'\, p'(F')\rho(F')^{\otimes N_s}, \tag{59}$$

where the posterior distribution is expressed in terms of the new variable F' given by Eq. (54). Working this out explicitly, we get

$$p'(F')=\left(8F(F')-2+\frac{3(3-4F')}{\sqrt{6F'-4F'^2-1}}\right)\frac{p[F(F')|N_s]}{10-8F'}, \tag{60}$$

where $F(F')$ is the inverse of Eq. (54):

$$F(F')=\frac{(1-2F')+3\sqrt{6F'-4F'^2-1}}{10-8F'}. \tag{61}$$

We can see how much information is gained by a single round of the recurrence method using this simplified version as an example. If the initial generating function is a uniform distribution, $p(F)=4/3$ for $1/4<F<1$, then for large N, the posterior distribution is highly peaked after one round. We see this in Fig. 1, where the prior and posterior distributions are shown for different values of N and a typical choice of N_s. Note that states with $1/4<F<1/2$ move towards $F=1/4$ under the procedure, producing a peak about the completely mixed state; for high N and the value of p_s used in our example, this peak is suppressed by the Bayesian updating. States with $F>1/2$ move towards $F=1$. The procedure has fixed points at $F=1/4$, $F=1/2$, and $F=1$.

It should be noted that because of its extremely small yield, the recurrence method should never be used if hashing is possible. An initial state that cannot be distilled by the hashing method, however, might, after one or more rounds of the recurrence method, satisfy the criterion (26) for some value of $S_0<1$. If that is so, then a combination of tomography and hashing should be used thereafter, as described in the last section.

Similarly, if $p(\rho)$ has some support on distillable and some on undistillable states, a few rounds of the recurrence method generally produces convergence on either a distillable or undistillable state, without ambiguity. Under certain circumstances, however, it might be beneficial to supplement this with tomographic measurements on a number of pairs as well. For example, the updating procedure (45) treats the coefficients w_1, w_4 and w_2, w_3 symmetrically. An initially symmetric state thus has this symmetry preserved, and the

distribution $p(\vec{w})$ might become double peaked. In this case, measuring a small number of pairs would suffice to eliminate one of the two peaks.

V. CONCLUSION

In this paper, we have discussed the problems that arise in naïvely applying the Jaynes maximum-entropy construction to multiple copies of a system, about which only partial information is available. Rather than simply assigning N copies of the single-system Jaynes state ρ_J to the N systems, one should instead assign an *exchangeable* state of the form (5). This assumption is the starting point for entanglement manipulation in the case of partially known or unknown states. Given such an exchangeable state, by measuring some of the systems, one can obtain information about the state of the others; the state of the remaining systems is then updated with the quantum version of the Bayes rule.

Using this rule, we have given a Bayesian account of the entanglement purification procedures of one-way hashing and recurrence. The Bayesian formulation allows us to provide a straightforward discussion of the conditions under which maximally entangled states can be distilled from unknown or partially known quantum states. For one-way hashing, we have given the *a priori* probabilities for the possible asymptotic yields of maximally entangled pairs. Our results can be used to decide which combination of quantum-state tomography, recurrence, and hashing to use to obtain the highest expected yield, both asymptotically and in the case of a fixed number of initially given pairs. Although our discussion is entirely in terms of pairs of qubits, the method is general and can be applied to any generalization of hashing or recurrence in Hilbert spaces of higher dimension.

ACKNOWLEDGMENTS

We would like to thank Howard Barnum, Oliver Cohen, Chris Fuchs, and Bob Griffiths for helpful conversations. T.A.B. was supported in part by NSF Grant No. PHY-9900755 and DOE Grant No. DE-FG02-90ER40542, R.S. was supported by the UK Engineering and Physical Sciences Research Council, and C.M.C. was supported in part by ONR Grant No. N00014-00-1-0578. Some of this work was done at the workshop on ''Quantum Information Processing'' at the Benasque Center for Science in Benasque, Spain.

[1] C. H. Bennett *et al.*, Phys. Rev. Lett. **76**, 722 (1996).
[2] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).
[3] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. **78**, 574 (1997).
[4] D. Deutsch *et al.*, Phys. Rev. Lett. **80**, 2022 (1998).
[5] G. Giedke, H.-J. Briegel, J. I. Cirac, and P. Zoller, Phys. Rev. A **59**, 2641 (1999).
[6] J. Eisert *et al.*, Phys. Rev. Lett. **84**, 1611 (2000).
[7] R. Horodecki, M. Horodecki, and P. Horodecki, Phys. Rev. A **59**, 1799 (1999).
[8] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).
[9] E. T. Jaynes, Phys. Rev. **108**, 171 (1957).
[10] A. K. Rajagopal, Phys. Rev. A **60**, 4338 (1999).
[11] A. Rigo, A. R. Plastino, A. Plastino, and M. Casas, Phys. Lett. A **270**, 1 (2000).
[12] N. Linden, S. Massar, and S. Popescu, Phys. Rev. Lett. **81**, 3279 (1998).
[13] E. T. Jaynes, in *Maximum Entropy and Bayesian Methods in Applied Statistics*, edited by J. H. Justice (Cambridge University Press, Cambridge, 1986), pp. 26–58.
[14] B. de Finetti, *Theory of Probability* (Wiley, New York, 1990).
[15] R. L. Hudson and G. R. Moody, Z. Wahrscheinlichkeitstheor. Verwandte Geb. **33**, 343 (1976).
[16] E. Størmer, J. Funct. Anal. **3**, 48 (1969).
[17] C. M. Caves, C. A. Fuchs, and R. Schack (unpublished).
[18] D. Bures, Trans. Am. Math. Soc. **135**, 199 (1969).
[19] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994).
[20] P. B. Slater, J. Math. Phys. **38**, 2274 (1997).
[21] K. Życzkowski, P. Horodecki, A. Sanpera, and M. Lewenstein, Phys. Rev. A **58**, 883 (1998).
[22] One possibility, whose analog for classical probabilities was proposed by Skilling, [J. Skilling, in *Maximum Entropy and Bayesian Methods*, edited by J. Skilling (Kluwer, Dordrecht, 1989), pp. 45–52] is to use $f(\rho)=e^{\alpha S(\rho)}$, along with one of the proposed unbiased measures for $d\rho$ [18–21]. Here, $\alpha \gg 1$ is a parameter that characterizes one's confidence in the single-copy maximum-entropy assignment ρ_J: for $N \ll \alpha$, the exchangeable state (5) becomes effectively the product state $\rho_J^{\otimes N}$, but for $N \gg \alpha$, Eq. (5) predicts measurement statistics different from the product state.
[23] K. Kraus, *States, Effects, and Operations. Fundamental Notions of Quantum Theory*, Lecture Notes in Physics Vol. 190 (Springer, Berlin, 1983).
[24] R. Schack, T. A. Brun, and C. M. Caves, e-print quant-ph/0008113.
[25] K. Vogel and H. Risken, Phys. Rev. A **40**, 2847 (1989).
[26] D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani, Phys. Rev. Lett. **70**, 1244 (1993).
[27] U. Leonhardt, Phys. Rev. Lett. **74**, 4101 (1995).
[28] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
[29] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).
[30] I. Cziszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems* (Academic Press, New York, 1981).
[31] R. Jozsa, M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. **81**, 1714 (1998).