Error exponents for entanglement concentration

2003 J. Phys. A: Math. Gen. 36 527

(http://iopscience.iop.org/0305-4470/36/2/316)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **36** (2003) 527–553 PII: S0305-4470(03)53843-4

Error exponents for entanglement concentration

Masahito Hayashi[1], **Masato Koashi**[2], **Keiji Matsumoto**[3], **Fumiaki Morikoshi**[4] **and Andreas Winter**[5]

[1] Laboratory for Mathematical Neuroscience, Brain Science Institute, RIKEN, 2-1Hirosawa, Wako, Saitama, 351-0198, Japan
[2] CREST Research Team for Interacting Carrier Electronics, School of Advanced Sciences, The Graduate University for Advanced Studies (SOKENDAI), Hayama, Kanagawa, 240-0193, Japan
[3] Quantum Computation and Information Project, ERATO, JST, 5-28-3, Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan
[4] NTT Basic Research Laboratories, NTT Corporation, 3-1 Morinosato-Wakamiya, Atsugi-shi, Kanagawa, 243-0198, Japan
[5] Department of Computer Science, University of Bristol, Merchant Venturers Building, Woodland Road, Bristol BS8 1UB, UK

Received 23 September 2002
Published 17 December 2002
Online at stacks.iop.org/JPhysA/36/527

Abstract
Consider entanglement concentration schemes that convert n identical copies of a pure state into a maximally entangled state of a desired size with success probability being close to one in the asymptotic limit. We give the distillable entanglement, the number of Bell pairs distilled per copy, as a function of an *error exponent*, which represents the rate of decrease in failure probability as n tends to infinity. The formula fills the gap between the least upper bound of distillable entanglement in probabilistic concentration, which is the well-known entropy of entanglement, and the maximum attained in deterministic concentration. The method of types in information theory enables the detailed analysis of the distillable entanglement in terms of the error rate. In addition to the probabilistic argument, we consider another type of entanglement concentration scheme, where the initial state is deterministically transformed into a (possibly mixed) final state whose fidelity to a maximally entangled state of a desired size converges to one in the asymptotic limit. We show that the same formula as in the probabilistic argument is valid for the argument on fidelity by replacing the success probability with the fidelity. Furthermore, we also discuss entanglement yield when optimal success probability or optimal fidelity converges to zero in the asymptotic limit (strong converse), and give the explicit formulae for those cases.

PACS numbers: 03.67.−a, 03.67.Hk, 03.65.Ud

1. Introduction

Quantum entanglement, an indispensable resource for quantum information processing such as superdense coding [1], quantum teleportation [2], quantum cryptography [3] and quantum computation [4], is expected to have a rich mathematical structure behind its weirdness. As in the case of other physical resources, quantification of entanglement is the key to understanding its full potential. The essentials of bipartite pure-state entanglement have already been revealed for both finite regimes and the asymptotic limit. The fundamental results are the intimate connection between the mathematical theory of majorization and entanglement manipulation [5–8], and the existence of a unique measure of entanglement in the asymptotic limit [9, 10].

One way of quantifying entanglement is to estimate the number of Bell pairs,

$$\frac{1}{\sqrt{2}}(|00\rangle_{\mathrm{AB}} + |11\rangle_{\mathrm{AB}}) \tag{1}$$

distilled from a given entangled state by local operations and classical communication (LOCC). Though the above quantity of distillable entanglement can be defined for mixed states, we deal with only pure states here. In order to make use of partially entangled states for quantum teleportation, we need to convert the partially entangled states into maximally entangled states by LOCC. The process is called entanglement concentration, and its efficiency in the asymptotic limit is the focus of this paper.

The unique measure of bipartite pure-state entanglement gives the limitation on the efficiency of entanglement concentration. Suppose we share n identical copies of a partially entangled state

$$|\phi\rangle = \sum_{i=1}^{d} \sqrt{p_i}|i\rangle|i\rangle \tag{2}$$

where the Schmidt coefficients squared are arranged in decreasing order, i.e., $p_1 \geqslant p_2 \geqslant \cdots \geqslant p_d \geqslant 0$, and sum to one. (Schmidt coefficients are arranged in decreasing order throughout this paper.) Bennett *et al* [9] proved that the maximum number of Bell pairs distilled per copy from $|\phi\rangle^{\otimes n}$ is given by

$$E_{\mathrm{entropy}}(\phi) = -\sum_{i=1}^{d} p_i \log_2 p_i \tag{3}$$

in the asymptotic limit, $n \to \infty$. (Logarithms are taken to base two throughout this paper unless stated otherwise.) They imposed the condition that the success probability of entanglement concentration tends to one in the asymptotic limit, i.e.,

$$p_{\mathrm{success}} = 1 - \epsilon \tag{4}$$

where

$$\epsilon \to 0 \qquad \text{as} \quad n \to \infty. \tag{5}$$

With this restriction, the maximum attainable entanglement yield is proved to be equation (3).

On the other hand, much research on entanglement concentration has been undertaken from various viewpoints [6–8, 11–14]. Among other things, the bound on entanglement yield in deterministic concentration [13]

$$E_{\mathrm{det}}(\phi) = -\log p_1 \tag{6}$$

gives another quantification of entanglement. The restriction *deterministic* means that the process succeeds with probability one both in finite regimes and in the asymptotic limit.

Though the quantities E_{entropy} and E_{det} give entanglement yield in the asymptotic limit, where both processes succeed with probability one, the two quantities do not coincide. The main purpose of this paper is to find the reason for the discrepancy. We will see that it is caused by the difference of the rate at which failure probabilities decrease when n tends to infinity in both concentration processes. Roughly speaking, while we obtain E_{entropy} when failure probability decreases slowly, we obtain E_{det} when it decreases rapidly. We will represent the rate by the exponent of failure probability in the asymptotic limit (*error exponent*). This is a common approach in the information sciences, and will allow us to 'tune' between the two extremes just mentioned. The quantum fixed-length pure state source coding in [15] is another example that uses the notion of error exponents in quantum information theory.

In the derivation of E_{entropy}, we use the asymptotic equipartition property [16]. However, a detailed analysis of the asymptotic behaviour requires more powerful mathematical tools; namely, the method of types [16, 17], which makes it possible to calculate the probabilities of rare events and derive stronger results than when we focus only on typical sequences.

The argument via the method of types will give entanglement yield as a function of an error exponent and reveal the missing link between E_{entropy} and E_{det}. In addition, we will also see that the success probability exponentially decreases when we try to distil more entanglement than E_{entropy} (strong converse). This was observed in [11], but here we are able to derive the *exact* error rate.

It was suggested that the two extremes E_{entropy} and E_{det} can also be expressed as some limits of Rényi entropy [13]. We will also see that they can be linked by Rényi-entropy-like functions, which is useful for practical calculation.

We can consider two different ways of analysing entanglement concentration: one is estimating the optimal success probability of obtaining the exact copy of a maximally entangled state of a desired size, where we discard failure cases. The other is estimating the fidelity of the final state, which is generally a mixed state, to a maximally entangled state of a desired size, where we obtain the final mixed state with probability one. We will first derive the entanglement yield as a function of an exponent of failure probability (error exponent) under the condition that the optimal success probability converges to one in the asymptotic limit. Then, we will also derive the entanglement yield as a function of an exponent of one minus fidelity, under the condition that the optimal fidelity converges to one in the asymptotic limit. The argument on fidelity can be reduced to that on probability via some lemmata. Finally, the yield functions will turn out to be in the same form in both cases. Since transformations assumed in the argument on fidelity need not produce the exact copy of a maximally entangled state, the treatment is more natural from a physical perspective than that on optimal success probability. In finite regimes, deterministically transforming a pure state into a maximally entangled state with optimal fidelity has been discussed in [18], while we will consider that in the asymptotic limit. The strong converse will also be analysed in terms of fidelity.

This paper is organized as follows. We begin in section 2 by revisiting the entanglement concentration in finite dimensions. After a brief review of the method of types in section 3, we will move on to the main result of this paper, error exponents and asymptotic entanglement concentration, in section 4. Section 5 discusses the strong converse. In section 6, we will present alternative formulae for the result proven in the preceding sections. Finally, we will discuss interpretations of our result and of some properties of the yield function, then conclude the paper. Some lemmata used in the proofs of our results are presented in appendices.

2. Finite-dimensional entanglement concentration revisited

In this section, we revisit entanglement concentration of finite-dimensional states so that the asymptotic limit will be smoothly derived from it. Suppose we distil a maximally entangled

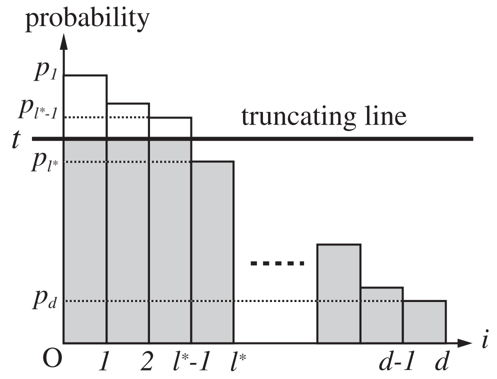

Figure 1. The distribution of the squared Schmidt coefficients of a partially entangled state to be concentrated. Each bar corresponds to a Schmidt coefficient squared. The area of the shaded region below the truncating line t represents the success probability of entanglement concentration.

state with Schmidt number $L(\leqslant d)$,

$$|\Phi_L\rangle = \frac{1}{\sqrt{L}} \sum_{i=1}^{L} |i\rangle|i\rangle \tag{7}$$

from the partially entangled state with Schmidt number d, equation (2). Lo and Popescu [11] derived the optimal probability with which we distil $|\Phi_L\rangle$ from $|\phi\rangle$:

$$P_L = \min_{l\in[1,L]} \frac{L}{L-l+1} \sum_{i=l}^{d} p_i. \tag{8}$$

Note that the Schmidt number of the initial state cannot be expanded by LOCC. Thus, we cannot distil a maximally entangled state with the Schmidt number greater than d.

In the following, we reformulate the optimal probability P_L in a more suitable form for our treatment of asymptotic entanglement concentration. Our strategy for distilling a maximally entangled state consists of two parts:

(i) We perform a two-valued local measurement to change the initial distribution of the Schmidt coefficients (probabilistic part).
(ii) If a desired result is obtained in the above measurement, we distil a maximally entangled state from the resultant state with probability one (deterministic part).

First, we briefly review the second (deterministic) part, which was investigated in [13]. Suppose we wish to distil a maximally entangled state with the greatest possible Schmidt number from $|\phi\rangle$ with probability one. The maximum Schmidt number of the maximally entangled state is $\lfloor 1/p_1 \rfloor$, where $\lfloor x \rfloor$ represents the largest integer equal to or less than x. Thus, if we could make p_1 smaller somehow, the size (Schmidt number) of the resultant maximally entangled state would become greater. Note that, according to Nielsen's theorem [5], the largest Schmidt coefficient cannot be deterministically decreased by LOCC.

In order to distil a maximally entangled state of size $L > \lfloor 1/p_1 \rfloor$, we need to adjust the largest Schmidt coefficient of the initial state before moving on to deterministic entanglement concentration in the second part. So, in the first (probabilistic) part, we perform the measurement presented below to truncate the initial distribution of the Schmidt coefficients.

As shown in figure 1, first we draw a truncating line that represents probability t, which is uniquely determined by the size of the maximally entangled state as proved later.

Then we perform a two-valued local measurement on either side of the entangled pair by using measurement operators,

$$M_1 = \sum_{i=1}^{l^*-1} \sqrt{\frac{t}{p_i}} |i\rangle\langle i| + \sum_{i=l^*}^{d} |i\rangle\langle i| \tag{9}$$

with $p_1 \geqslant p_2 \geqslant \cdots \geqslant p_{l^*-1} > t \geqslant p_{l^*} \geqslant \cdots \geqslant p_d$, and M_2 such that $M_1^\dagger M_1 + M_2^\dagger M_2 = I$. Measurement outcome 1 corresponds to the success of the entanglement concentration. It is easily seen that it occurs with probability

$$P_{\text{success}} = \|(M_1 \otimes I)|\phi\rangle\|^2 = \sum_{i=1}^{d} \min\{t, p_i\} \tag{10}$$

which is shown schematically as the area of the shaded region below the truncating line t in figure 1. Measurement outcome 2 corresponds to the failure of the concentration, and the failure probability, $1 - P_{\text{success}}$, is equal to the area of the region above the truncating line in figure 1.

In the case of measurement outcome 1, the post-measurement state becomes

$$|\phi'\rangle = \frac{(M_1 \otimes I)|\phi\rangle}{\sqrt{P_{\text{success}}}} = \frac{1}{\sqrt{P_{\text{success}}}} \left(\sum_{i=1}^{l^*-1} \sqrt{t}|i\rangle|i\rangle + \sum_{i=l^*}^{d} \sqrt{p_i}|i\rangle|i\rangle \right) \tag{11}$$

whose Schmidt coefficients squared are represented as the shaded region in figure 1 with an appropriate renormalization; each bar (Schmidt coefficient) is replaced with that divided by P_{success}. Since the largest Schmidt coefficient of the post-measurement state is $\sqrt{t/P_{\text{success}}}$, we can distil a maximally entangled state of size

$$L = \frac{P_{\text{success}}}{t}. \tag{12}$$

In other words, our scheme distils a maximally entangled state of size L with probability P_{success}. The truncating probability t is uniquely determined by a given integer L via an implicit function,

$$L = \sum_{i=1}^{d} \min\left\{1, \frac{p_i}{t}\right\} \tag{13}$$

which is given by equations (10) and (12). Since the right-hand side of equation (13) is strictly monotone decreasing in $[1/p_1, d]$ for $t \in [p_d, p_1]$, t is uniquely determined for a given integer $L \in [1/p_1, d]$. (When $L < 1/p_1$, $|\Phi_L\rangle$ can be distilled with probability one. Thus, we do not need the probabilistic part.)

The rest of this section gives the proof that the probability P_{success} coincides with the optimal probability P_L (equation (8)). The truncating probability t uniquely determines an integer l^* satisfying $p_{l^*-1} > t \geqslant p_{l^*}$. (See figure 1.) From equations (10) and (12), we have

$$P_{\text{success}} = t(l^* - 1) + \sum_{i=l^*}^{d} p_i = tL. \tag{14}$$

Thus,

$$t = \frac{1}{L - l^* + 1} \sum_{i=l^*}^{d} p_i. \tag{15}$$

Substituting this into equation (14), we obtain an alternative expression of P_{success} with l^* and L:

$$P_{\text{success}} = \frac{L}{L - l^* + 1} \sum_{i=l^*}^{d} p_i. \tag{16}$$

On the other hand, equation (14) gives $t(l^* - 1) < tL$, i.e., $l^* \in [1, L]$. Thus, we have $P_L \leqslant P_{\text{success}}$ due to the right-hand side of equation (8), which means the success probability of our scheme is optimal because P_L has already been proven optimal.

Therefore, we obtain the following equation that connects the size of a maximally entangled state L, the truncating probability t and the optimal success probability of concentration P_L:

$$P_L = tL. \tag{17}$$

3. The method of types

In order to analyse entanglement concentration in the asymptotic limit, we will employ the method of types in the following sections. In this section, we briefly summarize relevant definitions and lemmata on the method of types without giving the proofs. For detailed discussions and proofs, see chapter 12 in [16] and chapter 1 in [17].

Let $X_1, X_2, \ldots, X_n$ be a sequence of n symbols from an alphabet $A = \{a_1, a_2, \ldots, a_d\}$, where d is the number of symbols in the alphabet A. A sequence $x_1 x_2 \cdots x_n$ will be denoted by $\mathbf{x}$.

Definition 1. *The type $p_{\mathbf{x}}$ of a sequence $x_1 x_2 \cdots x_n$ is the relative proportion of occurrences of each symbol of A, i.e.,*

$$p_{\mathbf{x}}(a) = \frac{N(a|\mathbf{x})}{n} \qquad \textit{for all} \quad a \in A \tag{18}$$

where $N(a|\mathbf{x})$ is the number of times the symbol a occurs in the sequence $\mathbf{x} \in A^n$.

A type $p_{\mathbf{x}}$ is a map from symbols $a \in A$ to their frequencies in the sequence $\mathbf{x}$ (empirical probability distribution). We denote the set of types with denominator n by $\mathcal{P}_n$. If $p \in \mathcal{P}_n$, then the set of sequences of length n and type p is called the type class of p, denoted by T_p^n, i.e.,

$$T_p^n = \{\mathbf{x} \in A^n | p_{\mathbf{x}} = p\}. \tag{19}$$

In other words, n-letter sequences that coincide when rearranged in alphabetical order are in the same type class.

Lemma 2.

$$|\mathcal{P}_n| \leqslant (n+1)^d. \tag{20}$$

Though the number of sequences in A^n is exponential in n, the number of *types* grows at most polynomially in n, which means that at least one type has exponentially many sequences in its type class.

If each letter in sequences is drawn i.i.d. according to some probability distribution, then we can estimate the probability with which a sequence occurs by using the Shannon entropy $H(p)$ and the relative entropy $D(p\|q)$, i.e.,

$$H(p) = -\sum_{i=1}^{d} p_i \log p_i \tag{21}$$

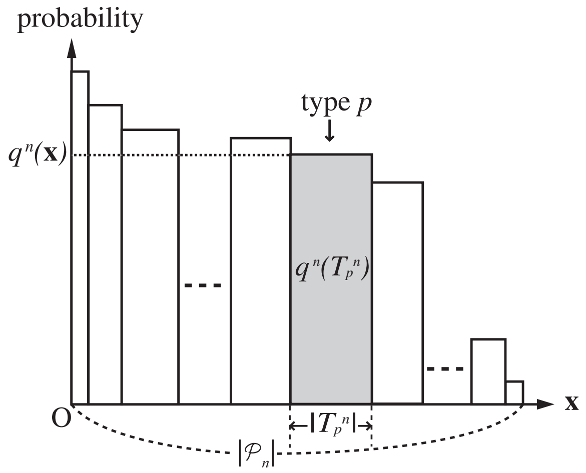

Figure 2. A schematic summary of the method of types. This bar graph represents the probability distribution on n-letter sequences that are drawn i.i.d. according to $q(x)$. Each bar corresponds to a type class, and consists of the sequences of the type. The number of 'stairs' is the number of types $\mathcal{P}_n : |\mathcal{P}_n| \leqslant (n+1)^d$. The height of each bar is the probability with which each sequence in the type class occurs: $q^n(\mathbf{x}) = 2^{-n\{H(p_\mathbf{x})+D(p_\mathbf{x}\|q)\}}$. The width of each bar is the number of sequences in the type class: $(n+1)^{-d}2^{nH(p)} \leqslant |T_p^n| \leqslant 2^{nH(p)}$. The area of each bar is the probability of the type class: $(n+1)^{-d}2^{-nD(p\|q)} \leqslant q^n(T_p^n) \leqslant 2^{-nD(p\|q)}$. (Note that if $q(x)$ *degenerates*, i.e., $q(x_i) = q(x_j)$ for some x_i and x_j such that $i \neq j$, then sequences in different type classes can occur with the same probability. Thus, one bar can consist of different type classes.)

and

$$D(p\|q) = \sum_{i=1}^{d} p_i \log \frac{p_i}{q_i} \tag{22}$$

where p and q are probability distributions.

Lemma 3. *If $X_1, X_2, \ldots, X_n$ are drawn i.i.d. according to $q(x)$, then the probability of* $\mathbf{x}$ *depends only on its type and is given by*

$$q^n(\mathbf{x}) = 2^{-n\{H(p_\mathbf{x})+D(p_\mathbf{x}\|q)\}}. \tag{23}$$

Furthermore, we can also estimate the size of a type class T_p^n; the number of the sequences in T_p^n is bounded as follows:

Lemma 4. *For arbitrary type $p \in \mathcal{P}_n$,*

$$\frac{1}{(n+1)^d} 2^{nH(p)} \leqslant \left|T_p^n\right| \leqslant 2^{nH(p)}. \tag{24}$$

From lemmata 3 and 4, we obtain the bound of the probability of a type class.

Lemma 5. *For arbitrary type $p \in \mathcal{P}_n$ and arbitrary probability distribution q, the probability of the type class T_p^n under q^n is bounded as*

$$\frac{1}{(n+1)^d} 2^{-nD(p\|q)} \leqslant q^n\left(T_p^n\right) \leqslant 2^{-nD(p\|q)}. \tag{25}$$

The lemmata summarized above are also shown schematically in figure 2 for easy reference in the proofs of our results, where the method of types will be heavily used.

4. Asymptotic entanglement concentration

This section presents the main result of this paper, asymptotic entanglement concentration from the viewpoint of error exponents. Suppose we wish to distil a maximally entangled state of size L_n from n identical copies of $|\phi\rangle$, i.e., $|\phi\rangle^{\otimes n} = \sum_{\mathbf{i}} \sqrt{p^n(\mathbf{i})}|\mathbf{i}\rangle|\mathbf{i}\rangle$, where $p^n(\mathbf{i})$ is the n-i.i.d. extension of p_i. Applying the results in section 2 to the n-i.i.d. case, we also have the optimal success probability

$$P_{L_n} = \sum_{\mathbf{i}} \min\{t_n, p^n(\mathbf{i})\} \tag{26}$$

and the relation between P_{L_n}, L_n, and the truncating line t_n,

$$P_{L_n} = t_n L_n. \tag{27}$$

In the following, we consider the case where the optimal success probability P_{L_n} converges to one as the number of entangled pairs n increases. The rate of the convergence is represented by an *error exponent* r, the first-order coefficient in the exponent of the failure probability in the asymptotic limit, which is defined as

$$r = \lim_{n\to\infty} \left\{ -\frac{1}{n} \log\left(1 - P_{L_n}\right) \right\}. \tag{28}$$

Intuitively, this means that the error probability behaves as 2^{-nr}.

We will derive the maximum number of Bell pairs distilled per copy in the asymptotic limit, E, as a function of the error exponent r. First, we prove a theorem that relates entanglement yield and an error exponent via a monotone function, from which we will derive a formula for entanglement yield $E(r)$.

Theorem 6. *Consider a sequence of entanglement concentration schemes converting n identical copies of $|\phi\rangle = \sum_{i=1}^{d} \sqrt{p_i}|i\rangle|i\rangle$, i.e., $|\phi\rangle^{\otimes n}$, into a maximally entangled state of size L_n, which attain the optimal success probability P_{L_n}. Suppose*

$$\limsup_{n\to\infty} \left(\frac{1}{n}\log L_n\right) < H(p) \tag{29}$$

and

$$\frac{1}{n}\log L_n > -\log p_1 \tag{30}$$

where $p = (p_1, \ldots, p_d)$. Then,

$$\limsup_{n\to\infty} \left(\frac{1}{n}\log L_n\right) = f\left(\liminf_{n\to\infty} \left\{-\frac{1}{n}\log\left(1 - P_{L_n}\right)\right\}\right) \tag{31}$$

and

$$\liminf_{n\to\infty} \left(\frac{1}{n}\log L_n\right) = f\left(\limsup_{n\to\infty} \left\{-\frac{1}{n}\log\left(1 - P_{L_n}\right)\right\}\right) \tag{32}$$

where

$$f(r) \equiv \min_{q:D(q\|p)\leqslant r} \{D(q\|p) + H(q)\}. \tag{33}$$

Proof. Let

$$R_n \equiv -\frac{1}{n}\log t_n. \tag{34}$$

Then, equation (27) gives

$$R_n = \frac{1}{n}\log L_n - \frac{1}{n}\log P_{L_n}. \tag{35}$$

In what follows, we only consider a convergent sub-sequence of $\{R_n\}$, that is, take an infinite subset $\mathcal{N} \subset \{1, 2, \ldots\}$ such that $R' \equiv \lim_{n\to\infty, n\in\mathcal{N}} R_n$ exists. For simplicity, we denote the sub-sequence $\{R_n\}_{n\in\mathcal{N}}$ as $\{R'_n\}$, and omit $n \in \mathcal{N}$. Since equation (29) implies $\lim_{n\to\infty} P_{L_n} = 1$, equation (35) gives

$$\lim_{n\to\infty}\left(\frac{1}{n}\log L'_n\right) = R' \tag{36}$$

where $\{L'_n\}$ denotes the sub-sequence that corresponds to $\{R'_n\}$. Then, equations (29) and (30) imply

$$-\log p_1 \leqslant R' < H(p). \tag{37}$$

Equation (26) gives

$$-\frac{1}{n}\log\left(1 - P_{L'_n}\right) = -\frac{1}{n}\log\left(1 - \sum_{\mathbf{i}}\min\{t'_n, p^n(\mathbf{i})\}\right) \tag{38}$$

where $\{t'_n\}$ is a sub-sequence such that $R'_n = -n^{-1}\log t'_n$.

In the following, we estimate the right-hand side of equation (38) by the method of types. Rewriting the area above the truncating probability t'_n in terms of the type theory, we obtain

$$1 - \sum_{\mathbf{i}}\min\{t'_n, p^n(\mathbf{i})\} = \sum_{q\in\mathcal{P}_n : p^n(q)\geqslant t'_n}\left(p^n\left(T^n_q\right) - \left|T^n_q\right| t'_n\right). \tag{39}$$

Invoking equations (20), (23) and (25), we have the inequalities

$$\begin{aligned}\sum_{q\in\mathcal{P}_n : p^n(q)\geqslant t'_n}\left(p^n\left(T^n_q\right) - \left|T^n_q\right| t'_n\right) &\leqslant \sum_{q\in\mathcal{P}_n : p^n(q)\geqslant t'_n} p^n\left(T^n_q\right)\\ &\leqslant (n+1)^d \max_{q\in\mathcal{P}_n : D(q\|p)+H(q)\leqslant R'_n} 2^{-nD(q\|p)}.\end{aligned} \tag{40}$$

Note that $p^n(q) \geqslant t'_n$ is equivalent to $D(q\|p) + H(q) \leqslant R'_n$ due to equations (23) and (34). Thus, together with equations (38) and (39), we have

$$-\frac{1}{n}\log\left(1 - P_{L'_n}\right) \geqslant -\frac{d}{n}\log(n+1) + \min_{q\in\mathcal{P}_n : D(q\|p)+H(q)\leqslant R'_n} D(q\|p). \tag{41}$$

Therefore, for $-\log p_1 \leqslant R' < H(p)$

$$\liminf_{n\to\infty}\left\{-\frac{1}{n}\log\left(1 - P_{L'_n}\right)\right\} \geqslant \min_{q : D(q\|p)+H(q)\leqslant R'} D(q\|p). \tag{42}$$

Next, first we consider the case $-\log p_1 < R'$. Then, for any q satisfying

$$-\log p_1 \leqslant D(q\|p) + H(q) < R' \tag{43}$$

there exists a sequence of types $q'_n \in \mathcal{P}_n$ such that

$$D(q'_n\|p) + H(q'_n) \leqslant R'_n \qquad \text{with} \qquad \lim_{n\to\infty} q'_n = q. \tag{44}$$

Invoking equations (24) and (25), we also have

$$\begin{aligned}\sum_{q\in\mathcal{P}_n : p^n(q)\geqslant t'_n}\left(p^n\left(T^n_q\right) - \left|T^n_q\right| t'_n\right) &\geqslant p^n\left(T^n_{q'_n}\right) - \left|T^n_{q'_n}\right| t'_n\\ &\geqslant \frac{1}{(n+1)^d}2^{-nD(q'_n\|p)} - 2^{nH(q'_n)}2^{-nR'_n}.\end{aligned} \tag{45}$$

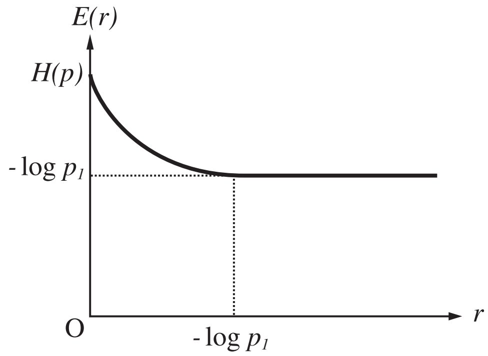

Figure 3. Entanglement yield in asymptotic entanglement concentration with an error exponent r. The horizontal axis represents the error exponent. The vertical axis represents the number of Bell pairs distilled per copy in the asymptotic limit: $E(r) = \min_{q:D(q\|p)\leqslant r}\{D(q\|p) + H(q)\}$.

Thus, together with equations (38) and (39), we have

$$-\frac{1}{n}\log\left(1 - P_{L'_n}\right) \leqslant -\frac{1}{n}\log\left\{2^{-n\left(D(q'_n\|p)+\frac{d}{n}\log(n+1)\right)} - 2^{-n(R'_n - H(q'_n))}\right\}. \tag{46}$$

Equations (43) and (44) imply

$$\lim_{n\to\infty}\left\{D(q'_n\|p) + \frac{d}{n}\log(n+1)\right\} < \lim_{n\to\infty}(R'_n - H(q'_n)). \tag{47}$$

Applying lemma 14 in appendix A to equation (46), we obtain

$$\limsup_{n\to\infty}\left\{-\frac{1}{n}\log\left(1 - P_{L'_n}\right)\right\} \leqslant D(q\|p) \tag{48}$$

which holds for any q satisfying equation (43). Therefore, for $-\log p_1 < R' < H(p)$,

$$\begin{aligned}\limsup_{n\to\infty}\left\{-\frac{1}{n}\log\left(1 - P_{L'_n}\right)\right\} &\leqslant \inf_{q:D(q\|p)+H(q)<R'} D(q\|p)\\ &= \min_{q:D(q\|p)+H(q)\leqslant R'} D(q\|p).\end{aligned} \tag{49}$$

Equations (42) and (49) show that when $-\log p_1 < R' < H(p)$, the sub-sequence $\left\{-n^{-1}\log\left(1 - P_{L'_n}\right)\right\}$ is also convergent and

$$r' \equiv \lim_{n\to\infty}\left\{-\frac{1}{n}\log\left(1 - P_{L'_n}\right)\right\} = \min_{q:D(q\|p)+H(q)\leqslant R'} D(q\|p). \tag{50}$$

According to corollary 16 in appendix B, setting $S(q) \equiv D(q\|p) + H(q)$, $U(q) \equiv D(q\|p)$, $x \equiv R$, and noting that $x_1 = -\log p_1$, and $x_2 = H(p)$, we see that the function $R \mapsto \min_{q:D(q\|p)+H(q)=R} D(q\|p)$ is continuous and strictly monotone decreasing in $(0, -\log p_1)$ for $R \in (-\log p_1, H(p))$. Thus, so is the function $R \mapsto \min_{q:D(q\|p)+H(q)\leqslant R} D(q\|p)$, which is the inverse function of $f(r)$ (equation (33)) (see figure 3). Therefore, equations (36) and (50) provide

$$\lim_{n\to\infty}\left\{-\frac{1}{n}\log\left(1 - P_{L'_n}\right)\right\} = f^{-1}\left(\lim_{n\to\infty}\left(\frac{1}{n}\log L'_n\right)\right) \tag{51}$$

i.e.,

$$\lim_{n\to\infty}\left(\frac{1}{n}\log L'_n\right) = f\left(\lim_{n\to\infty}\left\{-\frac{1}{n}\log\left(1 - P_{L'_n}\right)\right\}\right) \tag{52}$$

for

$$0 < \lim_{n\to\infty} \left\{ -\frac{1}{n} \log\left(1 - P_{L'_n}\right) \right\} < -\log p_1 \tag{53}$$

and

$$-\log p_1 < \lim_{n\to\infty} \left(\frac{1}{n} \log L'_n \right) < H(p). \tag{54}$$

On the other hand, when $R' = -\log p_1$, equations (36) and (42) give

$$\lim_{n\to\infty} \left(\frac{1}{n} \log L'_n \right) = -\log p_1 \tag{55}$$

and

$$\liminf_{n\to\infty} \left\{ -\frac{1}{n} \log\left(1 - P_{L'_n}\right) \right\} \geqslant -\log p_1. \tag{56}$$

In addition, for $r \geqslant -\log p_1$,

$$f(r) = \min_{q: D(q\|p)\leqslant r} \{D(q\|p) + H(q)\} = -\log p_1. \tag{57}$$

Note that $D(q\|p) + H(q) = -\sum_{i=1}^{d} q_i \log p_i \geqslant -\log p_1$ with equality when $q = (1, 0, \ldots, 0)$, i.e., $D(q\|p) = -\log p_1$. Hence,

$$\begin{aligned} -\log p_1 = \lim_{n\to\infty} \left(\frac{1}{n} \log L'_n \right) &= f\left(\liminf_{n\to\infty} \left\{ -\frac{1}{n} \log\left(1 - P_{L'_n}\right) \right\} \right) \\ &= f\left(\limsup_{n\to\infty} \left\{ -\frac{1}{n} \log\left(1 - P_{L'_n}\right) \right\} \right) \end{aligned} \tag{58}$$

for

$$\liminf_{n\to\infty} \left\{ -\frac{1}{n} \log\left(1 - P_{L'_n}\right) \right\} \geqslant -\log p_1. \tag{59}$$

The above argument holds for any convergent sub-sequences $\{R'_n\}$, and $f(r)$ is monotone decreasing. Therefore, equations (52), (53), (54), (58) and (59) provide equations (31) and (32) (see figure 3). □

Theorem 6 leads to the following corollary, which gives the maximum asymptotic entanglement yield $E(r)$ under the requirement that the failure probability decreases as rapidly as 2^{-nr}:

Corollary 7. *Consider a sequence of entanglement concentration schemes converting $|\phi\rangle^{\otimes n}$ into a maximally entangled state of size L_n with success probability $P^{(n)}_{\text{success}}$, such that*

$$r \leqslant \liminf_{n\to\infty} \left\{ -\frac{1}{n} \log\left(1 - P^{(n)}_{\text{success}}\right) \right\}. \tag{60}$$

Let us denote the class of all such sequences by $\mathcal{C}(r)$. Then, for $r > 0$,

$$\begin{aligned} E(r) &\equiv \max_{\mathcal{C}(r)} \limsup_{n\to\infty} \left(\frac{1}{n} \log L_n \right) = \max_{\mathcal{C}(r)} \liminf_{n\to\infty} \left(\frac{1}{n} \log L_n \right) \\ &= \min_{q: D(q\|p)\leqslant r} \{D(q\|p) + H(q)\}. \end{aligned} \tag{61}$$

This corollary connects the following two facts on the distillable entanglement of bipartite pure states E:

(i) If we allow error probability that vanishes in the asymptotic limit, E cannot exceed $H(p)$ [9].
(ii) If we stick to deterministic strategy even in the finite regimes (i.e., no error is allowed), E is equal to $-\log p_1$ [13].

Equation (61) provides the missing link between them, i.e., $E_{\text{entropy}} = H(p) = \lim_{r\to 0} E(r)$ and $E_{\text{det}} = -\log p_1 = \lim_{r\to\infty} E(r)$. The distillable entanglement $E(r)$ monotonically decreases and reaches $-\log p_1$ when $r = -\log p_1$, which means that probabilistic concentration schemes with error exponents greater than $-\log p_1$ effectively give the same result as deterministic ones.

Next, we move to a discussion about entanglement yield and fidelity. Suppose we wish to transform the initial state $|\phi\rangle^{\otimes n}$ deterministically into some final (possibly mixed) state that is as close to a maximally entangled state $|\Phi_{L_n}\rangle$ of size L_n as possible. Instead of the success probability $P^{(n)}_{\text{success}}$ in the previous argument, we here require that the fidelity F_n between the final state and the maximally entangled state $|\Phi_{L_n}\rangle$ approaches unity as rapidly as $1 - 2^{-nr}$, namely,

$$r \leqslant \liminf_{n\to\infty} \left\{ -\frac{1}{n} \log(1 - F_n) \right\}. \tag{62}$$

Let us denote the class of all such sequences by $\mathcal{C}_F(r)$. The maximum asymptotic entanglement yield $E_F(r)$ over $\mathcal{C}_F(r)$ can be obtained by reducing the problem to that of probabilistic schemes via the following two lemmata.

Lemma 8. *If the transformation* $|\phi\rangle \longrightarrow |\Phi_L\rangle$ *is possible with probability* $1 - \epsilon$*, then there exists a deterministic transformation* $|\phi\rangle \longrightarrow \rho$ *with fidelity* $\langle\Phi_L|\rho|\Phi_L\rangle \geqslant 1 - \epsilon$.

The proof is straightforward by considering the case $\rho = (1-\epsilon)|\Phi_L\rangle\langle\Phi_L| + \epsilon\rho'$. This lemma implies that for any sequence of probabilistic schemes belonging to $\mathcal{C}(r)$ with size $\{L_n\}$, there exists a sequence belonging to $\mathcal{C}_F(r)$ with the same size $\{L_n\}$. Hence,

$$\max_{\mathcal{C}_F(r)} \liminf_{n\to\infty} \left(\frac{1}{n} \log L_n \right) \geqslant E(r). \tag{63}$$

Lemma 9. *If there exists a deterministic transformation* $|\phi\rangle \longrightarrow \rho$ *with fidelity* $\langle\Phi_T|\rho|\Phi_T\rangle \geqslant 1 - \epsilon$*, the transformation* $|\phi\rangle \longrightarrow |\Phi_L\rangle$ *is possible with probability* $1 - 6\epsilon$*, for* $L = \lfloor T(1 - 6\epsilon)/6 \rfloor$.

Proof. See appendix C. □

This lemma implies that for any sequence of deterministic schemes belonging to $\mathcal{C}_F(r)$ with size $\{L_n\}$, there exists a sequence belonging to $\mathcal{C}(r)$ with the size $\{L'_n\}$, where $L'_n \geqslant L_n/7$ for large n. Hence,

$$\max_{\mathcal{C}_F(r)} \limsup_{n\to\infty} \left(\frac{1}{n} \log L_n \right) \leqslant \max_{\mathcal{C}(r)} \limsup_{n\to\infty} \left\{ \frac{1}{n} \log(7L'_n) \right\} = E(r). \tag{64}$$

From equations (63) and (64), we have

$$E_F(r) \equiv \max_{\mathcal{C}_F(r)} \limsup_{n\to\infty} \left(\frac{1}{n} \log L_n \right) = \max_{\mathcal{C}_F(r)} \liminf_{n\to\infty} \left(\frac{1}{n} \log L_n \right) = E(r) \tag{65}$$

which shows that the entanglement yield E_F can be expressed as the same function as that for probabilistic cases. Though the probabilistic argument assumes that we can transform the initial state to the exact copy of a maximally entangled state with probability close to one, the fidelity argument is much more natural in that it imposes weaker restrictions on manipulation where the fidelity close to one, which is common to more general entanglement manipulations, such as entanglement dilution.

5. Strong converse

We have investigated how the entanglement yield behaves when the failure probability exponentially decreases. In this section, conversely, we discuss asymptotic entanglement concentration with exponentially decreasing *success* probability, which will finally lead to the strong converse of asymptotic entanglement concentration.

The following discussion is parallel to that in the previous section. So, when we distil a maximally entangled state of size L_n^* from $|\phi\rangle^{\otimes n}$, the following relations also hold:

$$P_{L_n^*} = \sum_{\mathbf{i}} \min\{t_n, p^n(\mathbf{i})\} \tag{66}$$

and

$$P_{L_n^*} = t_n L_n^*. \tag{67}$$

We assume that the success probability converges to zero as the number of entangled pairs n increases. Then, the first-order coefficient in the exponent of the success probability in the asymptotic limit becomes

$$r = \lim_{n\to\infty} \left(-\frac{1}{n}\log P_{L_n^*}\right). \tag{68}$$

Intuitively, this means that the success probability behaves as 2^{-nr}. We will derive the maximum number of Bell pairs distilled per copy in the asymptotic limit, E^*, as a function of the exponent r. First, we prove a theorem that relates entanglement yield and an exponent via a monotone function, from which we will derive a formula for entanglement yield $E^*(r)$.

Theorem 10. *Consider a sequence of entanglement concentration schemes that convert n identical copies of $|\phi\rangle = \sum_{i=1}^d \sqrt{p_i}|i\rangle|i\rangle$, i.e., $|\phi\rangle^{\otimes n}$, into a maximally entangled state of size L_n^*, which attain the optimal success probability $P_{L_n^*}$. Suppose*

$$\liminf_{n\to\infty}\left(\frac{1}{n}\log L_n^*\right) > H(p) \tag{69}$$

and

$$\frac{1}{n}\log L_n^* < \log d \tag{70}$$

where $p = (p_1, \ldots, p_d)$. Then,

$$\limsup_{n\to\infty}\left(\frac{1}{n}\log L_n^*\right) = g\left(\limsup_{n\to\infty}\left(-\frac{1}{n}\log P_{L_n^*}\right)\right) \tag{71}$$

and

$$\liminf_{n\to\infty}\left(\frac{1}{n}\log L_n^*\right) = g\left(\liminf_{n\to\infty}\left(-\frac{1}{n}\log P_{L_n^*}\right)\right) \tag{72}$$

where

$$g(r) \equiv \max_{q:D(q\|p)\leqslant r} H(q). \tag{73}$$

Proof. Let

$$R_n \equiv -\frac{1}{n}\log t_n. \tag{74}$$

Then, equation (67) gives

$$-\frac{1}{n}\log P_{L_n^*} = R_n - \frac{1}{n}\log L_n^*. \tag{75}$$

In what follows, we only consider a convergent sub-sequence of $\{R_n\}$, that is, take an infinite subset $\mathcal{N} \subset \{1, 2, \ldots\}$ such that $R' \equiv \lim_{n\to\infty, n\in\mathcal{N}} R_n$ exists. For simplicity, we denote the sub-sequence $\{R_n\}_{n\in\mathcal{N}}$ as $\{R'_n\}$, and omit $n \in \mathcal{N}$. Equations (69) and (75) imply

$$H(p) < R'. \tag{76}$$

First, we consider the case

$$R' < -\frac{1}{d}\sum_{i=1}^{d} \log p_i. \tag{77}$$

Then, for sufficiently large n, we have

$$R'_n < -\frac{1}{d}\sum_{i=1}^{d} \log p_i. \tag{78}$$

Equation (66) gives

$$-\frac{1}{n}\log P_{L_n^{*\prime}} = -\frac{1}{n}\log \sum_{\mathbf{i}} \min\{t'_n, p^n(\mathbf{i})\} \tag{79}$$

where $\left\{P_{L_n^{*\prime}}\right\}$ and $\{t'_n\}$ are sub-sequences related to $\{R'_n\}$ by equations (74) and (75).

In the following, we estimate the right-hand side of equation (79) by the method of types. Rewriting the area below the truncating probability t'_n in terms of the type theory, we obtain

$$\sum_{\mathbf{i}} \min\{t'_n, p^n(\mathbf{i})\} = \sum_{q\in\mathcal{P}_n : p^n(q) > t'_n} \left|T_q^n\right| t'_n + \sum_{q\in\mathcal{P}_n : p^n(q)\leqslant t'_n} p^n\left(T_q^n\right). \tag{80}$$

Invoking equations (20), (23), (24) and (25), we have the inequalities

$$\begin{aligned} &\sum_{q\in\mathcal{P}_n : p^n(q) > t'_n} \left|T_q^n\right| t'_n + \sum_{q\in\mathcal{P}_n : p^n(q)\leqslant t'_n} p^n\left(T_q^n\right) \\ &\quad\leqslant \sum_{q\in\mathcal{P}_n : D(q\|p)+H(q) < R'_n} 2^{n(H(q)-R'_n)} + \sum_{q\in\mathcal{P}_n : D(q\|p)+H(q)\geqslant R'_n} 2^{-nD(q\|p)} \qquad (81) \\ &\quad\leqslant (n+1)^d \left[\max_{q\in\mathcal{P}_n : D(q\|p)+H(q)\leqslant R'_n} 2^{n(H(q)-R'_n)} + \max_{q\in\mathcal{P}_n : D(q\|p)+H(q)\geqslant R'_n} 2^{-nD(q\|p)}\right]. \qquad (82) \end{aligned}$$

According to lemma 15 in appendix B, setting $S(q) \equiv D(q\|p) + H(q), T(q) \equiv H(q), x \equiv R$, and noting that $x_1 = -\log p_1$, and $x_2 = -d^{-1}\sum_{i=1}^{d}\log p_i$, we see that the function $R \mapsto \max_{q:D(q\|p)+H(q)=R} H(q)$ is continuous and strictly monotone increasing in $[0, \log d]$ for $R \in \left[-\log p_1, -d^{-1}\sum_{i=1}^{d}\log p_i\right]$. Thus, $\max_{q\in\mathcal{P}_n:D(q\|p)+H(q)\leqslant R'_n} H(q) = \max_{q\in\mathcal{P}_n:D(q\|p)+H(q)=R'_n} H(q) = \max_{q\in\mathcal{P}_n:D(q\|p)+H(q)=R'_n}[R'_n - D(q\|p)]$.

Similarly, according to corollary 16 in appendix B, setting $S(q) \equiv -\{D(q\|p) + H(q)\}, U(q) \equiv D(q\|p), x \equiv -R$, and noting that $x_1 = \log p_d$, and $x_2 = -H(p)$, we see that the function $-R \mapsto \min_{q:D(q\|p)+H(q)=R} D(q\|p)$ is continuous and strictly monotone decreasing in $[0, -\log p_d]$ for $-R \in [\log p_d, -H(p)]$. This means that the function $R \mapsto \min_{q:D(q\|p)+H(q)=R} D(q\|p)$ is continuous and strictly monotone increasing in $[0, -\log p_d]$ for $R \in [H(p), -\log p_d]$. Thus, $\min_{q\in\mathcal{P}_n:D(q\|p)+H(q)\geqslant R'_n} D(q\|p) = \min_{q\in\mathcal{P}_n:D(q\|p)+H(q)=R'_n} D(q\|p)$.

Then, combining equations (80) and (82) and using equation (20), we have

$$\sum_{\mathbf{i}} \min\{t'_n, p^n(\mathbf{i})\} \leqslant 2(n+1)^d \max_{q\in\mathcal{P}_n:D(q\|p)+H(q)=R'_n} 2^{-nD(q\|p)}. \tag{83}$$

Together with equation (79), we obtain

$$\liminf_{n\to\infty}\left(-\frac{1}{n}\log P_{L_n^{*'}}\right)\geqslant \min_{q:D(q\|p)+H(q)=R'} D(q\|p). \tag{84}$$

Next, we derive a lower bound of equation (80). For any q satisfying $D(q\|p)+H(q)=R'$, there exists a sequence of types $q'_n\in\mathcal{P}_n$ such that $D(q'_n\|p)+H(q'_n)\geqslant R'_n$ and $\lim_{n\to\infty} q'_n=q$, due to the condition $R'_n<-d^{-1}\sum_{i=1}^d \log p_i\leqslant -\log p_d$.

Invoking equation (25), we have

$$\begin{aligned}\sum_{q\in\mathcal{P}_n:p^n(q)>t'_n}\left|T_q^n\right| t'_n+\sum_{q\in\mathcal{P}_n:p^n(q)\leqslant t'_n} p^n\left(T_q^n\right) &\geqslant \sum_{q\in\mathcal{P}_n:p^n(q)\leqslant t'_n} p^n\left(T_q^n\right)\\ &\geqslant p^n\left(T_{q'_n}^n\right)\\ &\geqslant \frac{1}{(n+1)^d}2^{-nD(q'_n\|p)}.\end{aligned} \tag{85}$$

Together with equations (79) and (80), we obtain

$$\limsup_{n\to\infty}\left(-\frac{1}{n}\log P_{L_n^{*'}}\right)\leqslant D(q\|p) \tag{86}$$

which holds for any q satisfying $D(q\|p)+H(q)=R'$. Therefore,

$$\limsup_{n\to\infty}\left(-\frac{1}{n}\log P_{L_n^{*'}}\right)\leqslant \min_{q:D(q\|p)+H(q)=R'} D(q\|p). \tag{87}$$

Equations (84) and (87) imply that the sub-sequence $\left\{-n^{-1}\log P_{L_n^{*'}}\right\}$ is also convergent, i.e.,

$$r'\equiv\lim_{n\to\infty}\left(-\frac{1}{n}\log P_{L_n^{*'}}\right)=\min_{q:D(q\|p)+H(q)=R'} D(q\|p). \tag{88}$$

As stated above, the function $R\mapsto \min_{q:D(q\|p)+H(q)=R} D(q\|p)$ is continuous and strictly monotone increasing in $[0,-\log p_d]$ for $R\in[H(p),-\log p_d]$. Hence, the inverse function of equation (88) clearly exists for $R\in(H(p),-\log p_d)$. However, now we consider the range restricted to equation (77), we obtain, for $r\in(0,c)$ with $c\equiv D(u\|p)$ being the relative entropy between the uniform distribution $u=(1/d,\ldots,1/d)$ and p,

$$R'(r')=\max_{q:D(q\|p)=r'}\{D(q\|p)+H(q)\}. \tag{89}$$

Furthermore, from equation (75), we see that if R' and r' exist, then the sub-sequence $\left\{n^{-1}\log L_n^{*'}\right\}$ is also convergent, i.e., $E^{*'}\equiv\lim_{n\to\infty}\left(n^{-1}\log L_n^{*'}\right)$ exists, and

$$r'=R'-E^{*'}. \tag{90}$$

Hence,

$$\begin{aligned}E^{*'}=R'(r')-r'&=\max_{q:D(q\|p)=r'}\{D(q\|p)+H(q)\}-r'\\ &=\max_{q:D(q\|p)=r'} H(q).\end{aligned} \tag{91}$$

According to lemma 15 in appendix B, setting $S(q)\equiv D(q\|p)$, $T(q)\equiv H(q)$, $x\equiv r$, and noting that $x_1=0$, and $x_2=c$, we see that the function $r\mapsto \max_{q:D(q\|p)=r} H(q)$ is continuous and strictly monotone increasing in $(H(p),\log d)$ for $r\in(0,c)$, thus

$$\max_{q:D(q\|p)=r} H(q)=\max_{q:D(q\|p)\leqslant r} H(q). \tag{92}$$

Therefore, in the case of equation (77), we have $0 < r' < c$, $H(p) < E^{*'} < \log d$, and

$$E^{*'}(r') = g(r'). \tag{93}$$

Next, we consider the case $R' \geqslant -d^{-1}\sum_{i=1}^{d}\log p_i$. Equation (70) provides

$$\limsup_{n\to\infty}\left(\frac{1}{n}\log L_n^{*'}\right) \leqslant \log d. \tag{94}$$

Together with equation (75), we have

$$\liminf_{n\to\infty}\left(-\frac{1}{n}\log P_{L_n^{*'}}\right) = R' - \limsup_{n\to\infty}\left(\frac{1}{n}\log L_n^{*'}\right) \geqslant R' - \log d. \tag{95}$$

Hence,

$$\liminf_{n\to\infty}\left(-\frac{1}{n}\log P_{L_n^{*'}}\right) \geqslant -\frac{1}{d}\sum_{i=1}^{d}\log p_i - \log d = c. \tag{96}$$

Then, since the function $E^{*'}(r')$ is monotone increasing, we have $\liminf_{n\to\infty}\left(n^{-1}\log L_n^{*'}\right) \geqslant \lim_{r'\uparrow c} E^{*'}(r') = g(c) = \log d$. Note that for $r \geqslant c = D(u\|p)$, $g(r) = \max_{q:D(q\|p)\leqslant r} H(q) = \log d$. Hence, together with equation (94) we have

$$\begin{aligned}\log d &= \lim_{n\to\infty}\left(\frac{1}{n}\log L_n^{*'}\right)\\ &= g\left(\liminf_{n\to\infty}\left(-\frac{1}{n}\log P_{L_n^{*'}}\right)\right)\\ &= g\left(\limsup_{n\to\infty}\left(-\frac{1}{n}\log P_{L_n^{*'}}\right)\right)\end{aligned} \tag{97}$$

for

$$\liminf_{n\to\infty}\left(-\frac{1}{n}\log P_{L_n^{*'}}\right) \geqslant c. \tag{98}$$

The above argument holds for any convergent sub-sequence $\{R_n'\}$. Since r is an exponent of success probability, the function $E^*(r)$ is monotone increasing. And, so is $g(r)$. Therefore, equations (93), (97) and (98) provide equations (71) and (72). □

Theorem 10 leads to the following corollary that gives the maximum asymptotic entanglement yield $E^*(r)$ under the requirement that the success probability decreases as slowly as 2^{-nr}:

Corollary 11. *Consider a sequence of entanglement concentration schemes converting $|\phi\rangle^{\otimes n}$ into a maximally entangled state of size L_n^* with success probability $P_{\text{success}}^{(n)}$, such that*

$$r \geqslant \limsup_{n\to\infty}\left(-\frac{1}{n}\log P_{\text{success}}^{(n)}\right). \tag{99}$$

Let us denote the class of all such sequences by $\mathcal{C}^*(r)$. Then, for $r > 0$,

$$\begin{aligned}E^*(r) &\equiv \max_{\mathcal{C}^*(r)}\limsup_{n\to\infty}\left(\frac{1}{n}\log L_n^*\right)\\ &= \max_{\mathcal{C}^*(r)}\liminf_{n\to\infty}\left(\frac{1}{n}\log L_n^*\right)\\ &= \max_{q:D(q\|p)\leqslant r} H(q).\end{aligned} \tag{100}$$

Entanglement yield can be related to fidelity also in the high-entanglement regime, where the aim of the task is to convert $|\phi\rangle^{\otimes n}$ deterministically into some final (possibly mixed) state that is as close to a maximally entangled state $\left|\Phi_{T_n^*}\right\rangle$ of size T_n^* as possible. We here require that the fidelity F_n between the final state and the maximally entangled state $\left|\Phi_{T_n^*}\right\rangle$ decrease as slowly as 2^{-nr}, namely,

$$r \geqslant \limsup_{n\to\infty}\left(-\frac{1}{n}\log F_n\right). \tag{101}$$

Let us denote the class of all such sequences by $\mathcal{C}_F^*(r)$. The maximum asymptotic entanglement yield $E_F^*(r)$ over $\mathcal{C}_F^*(r)$ can be obtained by reducing the problem to that of probabilistic schemes via the following two lemmata.

Lemma 12. *If the transformation $|\phi\rangle \longrightarrow |\Phi_L\rangle$ is possible with probability ϵ_1, then, for any $T \geqslant L$, there exists a deterministic transformation $|\phi\rangle \longrightarrow \rho$ with fidelity $\langle\Phi_T|\rho|\Phi_T\rangle \geqslant \epsilon_1\epsilon_2$ with $\epsilon_2 = L/T$.*

The proof is straightforward by considering the case $\rho = \epsilon_1|\Phi_L\rangle\langle\Phi_L| + (1-\epsilon_1)\rho'$, by noting that $|\langle\Phi_T|\Phi_L\rangle|^2 = L/T$.

Applying this lemma to asymptotic sequences by setting $\epsilon_2 = 2^{-nx}$ where x is an arbitrary nonnegative number, we have the following: for any sequence of probabilistic schemes belonging to $\mathcal{C}^*(r)$ with size $\{L_n^*\}$, there exists a sequence belonging to $\mathcal{C}_F^*(r+x)$ with the size $\{T_n^* = L_n^* 2^{nx}\}$. Hence,

$$\max_{\mathcal{C}_F^*(r+x)}\liminf_{n\to\infty}\left(\frac{1}{n}\log T_n^*\right) \geqslant E^*(r) + x \qquad \forall x \geqslant 0 \tag{102}$$

and therefore

$$\max_{\mathcal{C}_F^*(r)}\liminf_{n\to\infty}\left(\frac{1}{n}\log T_n^*\right) \geqslant \sup_{0\leqslant x<r}\{E^*(r-x)+x\}. \tag{103}$$

Lemma 13. *Suppose that there exists a deterministic transformation $|\phi\rangle \longrightarrow \rho$ with fidelity $F = \langle\Phi_T|\rho|\Phi_T\rangle$. Then, there exist an integer $L \leqslant T$ and a transformation $|\phi\rangle \longrightarrow |\Phi_L\rangle$ with success probability P, satisfying*

$$\sqrt{PL} \geqslant \frac{\sqrt{TF}-1}{\ln(T)}. \tag{104}$$

Proof. See appendix C. □

Using this lemma, we derive an upper bound of $\max_{\mathcal{C}_F^*(r)}\limsup_{n\to\infty}(n^{-1}\log T_n^*)$. Consider a sequence of deterministic schemes where the fidelity between the final state of the nth scheme and the maximally entangled state $\left|\Phi_{T_n^*}\right\rangle$ is F_n. Suppose that

$$r \geqslant \limsup_{n\to\infty}\left(-\frac{1}{n}\log F_n\right) \tag{105}$$

and

$$\beta \equiv \limsup_{n\to\infty}\left(\frac{1}{n}\log T_n^*\right) > r. \tag{106}$$

Let us take an infinite set of integers $\mathcal{N}$ such that $\lim_{n\to\infty;n\in\mathcal{N}}(n^{-1}\log T_n^*) = \beta$. According to lemma 13, there exists a sequence of transformations $|\phi\rangle^{\otimes n} \longrightarrow \left|\Phi_{L_n^*}\right\rangle (L_n^* \leqslant T_n^*)$ with success probability $P_{\text{success}}^{(n)}$ satisfying

$$\sqrt{P_{\text{success}}^{(n)}L_n^*} \geqslant \frac{\sqrt{T_n^*F_n}-1}{\ln(T_n^*)}. \tag{107}$$

Since $\beta > r$, for any $\delta > 0$, there exists n_1 such that for all $n \geqslant n_1$,

$$\frac{1}{n}\left(\log P_{\text{success}}^{(n)} + \log L_n^*\right) \geqslant \frac{1}{n}(\log F_n + \log T_n^*) - \delta. \tag{108}$$

Further, by equations (105) and (106), there exists $n_2 \geqslant n_1$ such that for all $n \geqslant n_2$ with $n \in \mathcal{N}$,

$$\frac{1}{n}\left(\log P_{\text{success}}^{(n)} + \log L_n^*\right) \geqslant \beta - r - \delta. \tag{109}$$

Let us define $r_1 \equiv -\liminf_{n\to\infty;n\in\mathcal{N}}\left(n^{-1}\log P_{\text{success}}^{(n)}\right)$ (or, r_1 can be any accumulation value). Note that $L_n^* \leqslant T_n^*$ and the above inequalities imply $r_1 \leqslant r$. Then, for any $\delta_1, \delta_2 > 0$, there are infinite values of n satisfying $-\frac{1}{n}\log P_{\text{success}}^{(n)} \leqslant r_1 + \delta_1$ and $n^{-1}\log L_n^* \geqslant \beta - r + r_1 - \delta_2$. This means that $E^*(r_1 + \delta_1) \geqslant \beta + r_1 - r - \delta_2$ holds for any $\delta_1, \delta_2 > 0$. Noting that $E^*(r)$ is monotone increasing, we have

$$\lim_{x\downarrow r_1} E^*(x) \geqslant \beta - r + r_1. \tag{110}$$

Note that what we have shown here is that for any sequence belonging to $\mathcal{C}_F^*(r)$ and having $\beta > r$, there exists a value r_1 ($\leqslant r$) satisfying equation (110). The same statement also holds for sequences with $\beta \leqslant r$, since the left-hand side of equation (110) is nonnegative and the right-hand side is nonpositive for $r_1 = 0$. Hence,

$$\max_{\mathcal{C}_F^*(r)} \limsup_{n\to\infty}\left(\frac{1}{n}\log T_n^*\right) \leqslant r + \sup_{0\leqslant r_1\leqslant r}\left\{\lim_{x\downarrow r_1} E^*(x) - r_1\right\}. \tag{111}$$

Since $E^*(r)$ is continuous in $(0, \infty)$ and monotone increasing in $[0, \infty)$, equations (103) and (111) determine $E_F^*(r)$ as

$$\begin{aligned} E_F^*(r) &\equiv \max_{\mathcal{C}_F^*(r)} \limsup_{n\to\infty}\left(\frac{1}{n}\log T_n^*\right) \\ &= \max_{\mathcal{C}_F^*(r)} \liminf_{n\to\infty}\left(\frac{1}{n}\log T_n^*\right) \\ &= \sup_{0<x\leqslant r}\{E^*(x) + r - x\}. \end{aligned} \tag{112}$$

The function $E_F^*(r)$ is shown by the broken curve in figure 4. Note that $\frac{\mathrm{d}}{\mathrm{d}r}E^*$ goes to ∞ at $r \to 0$ and that it tends to 0 as $r \to \infty$. The value r' in the figure satisfies $\frac{\mathrm{d}}{\mathrm{d}r}E^* = 1$ and from this point $E_F^*(r)$ starts to deviate from $E^*(r)$ and increases linearly.

In contrast to $E^*(r)$, $E_F^*(r)$ can reach any large value by making r large enough. This is because we are here allowing an exponentially small fidelity. Indeed, for a separable initial state in which $E^*(r) = 0$ everywhere, we still have $E_F^*(r) = r$, which corresponds to the fact that separable states can attain fidelity 2^{-N} to N Bell pairs. In the region of large r, the entanglement in the initial state contributes to the vertical offset of $E_F^*(r)$, which is given by $E_F^*(r') - r'$.

6. Alternative formulae

Corollaries 7 and 11 give the entanglement yields $E(r)$ and $E^*(r)$ as a minimum or a maximum over a probability distribution q. In practical calculation, however, it is not easy to deal with such minimization or maximization over probability distributions. We therefore present alternative formulae for entanglement yield, which contain only minimization or maximization over a variable s.

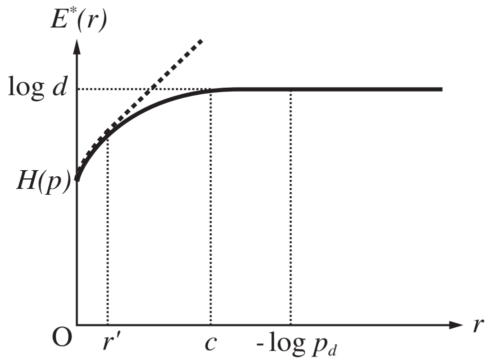

Figure 4. Entanglement yield in entanglement concentration whose success probability exponentially decreases (strong converse). The horizontal axis represents the exponent of the success probability. The vertical axis represents the number of Bell pairs distilled per copy in the asymptotic limit: $E^*(r) = \max_{q:D(q\|p)\leqslant r} H(q)$. $E^*(r)$ reaches the maximum value, $\log d$, at $r = c = -\log d - d^{-1}\sum_i \log p_i$, which is the relative entropy between the uniform distribution $u \equiv (1/d, \ldots, 1/d)$ and $\{p_i\}$. The broken line represents the entanglement yield $E_F^*(r)$ as a function of the exponent of the fidelity. At the value $r = r'$, where $\frac{\mathrm{d}}{\mathrm{d}r}E^*(r) = 1$, $E_F^*(r)$ starts to deviate from $E^*(r)$.

The entanglement yields in corollaries 7 and 11 are also expressed as

$$E(r) = \sup_{s\geqslant 1} \frac{r + \psi(s)}{1 - s} \tag{113}$$

and

$$E^*(r) = \min_{0\leqslant s\leqslant 1} \frac{rs + \psi(s)}{1 - s} \tag{114}$$

where

$$\psi(s) = \log \sum_i p_i^s. \tag{115}$$

The proof of equations (113) and (114) is given in appendix D. The entanglement yields in the fidelity arguments can also be expressed in the same manner:

$$E_F(r) = \sup_{s\geqslant 1} \frac{r + \psi(s)}{1 - s} \tag{116}$$

and

$$E_F^*(r) = \begin{cases} \min_{0\leqslant s\leqslant 1} \frac{rs+\psi(s)}{1-s} & (0 < r \leqslant r') \\ r - r' + E^*(r') & (r' < r) \end{cases} \tag{117}$$

where $r = r'$ is a value at which $\frac{\mathrm{d}}{\mathrm{d}r}E^*(r) = 1$. Equations (116) and (117) are easily obtained from equations (65) and (112).

As an example derived by using the above formulae, we show non-additivity of the entanglement yield $E(r)$. Additivity is an important property of $E_{\text{entropy}}(\rho) = H(p)$ and $E_{\text{det}}(\rho) = -\log p_1$: entanglement of a composite system is the sum of the contributions from each system, e.g., $E_{\text{entropy}}(\rho \otimes \sigma) = E_{\text{entropy}}(\rho) + E_{\text{entropy}}(\sigma)$. (Note that ρ and σ represent pure states.) This property, however, does not hold for the general expression of entanglement yield, $E(r)$, as seen below.

We can prove the following inequality of E_r:

$$E_{r_1+r_2}(\rho \otimes \sigma) \leqslant E_{r_1}(\rho) + E_{r_2}(\sigma). \tag{118}$$

From equation (113), we have

$$\begin{aligned}E_{r_1+r_2}(\rho\otimes\sigma) &= \sup_{s\geqslant 1}\frac{(r_1+r_2)+\psi_{\rho\otimes\sigma}(s)}{1-s}\\ &= \sup_{s\geqslant 1}\left\{\frac{r_1+\psi_\rho(s)}{1-s}+\frac{r_2+\psi_\rho(s)}{1-s}\right\}\\ &\leqslant \sup_{s\geqslant 1}\frac{r_1+\psi_\rho(s)}{1-s}+\sup_{s\geqslant 1}\frac{r_2+\psi_\sigma(s)}{1-s}\\ &= E_{r_1}(\rho)+E_{r_2}(\sigma).\end{aligned} \tag{119}$$

Equality holds if and only if $s_+(r_1)_\rho = s_+(r_2)_\sigma$, where $s_+(r)\geqslant 1$ is a unique solution of $r=-\psi(s)-(1-s)\psi'(s)$ (see appendix B). Thus we obtain equation (118).

When $r_1=r_2=r/2$, equation (118) provides

$$E_r(\rho\otimes\sigma)\leqslant E_{\frac{r}{2}}(\rho)+E_{\frac{r}{2}}(\sigma). \tag{120}$$

Let the largest squared Schmidt coefficients of ρ and σ be p_1 and q_1, respectively. Equality holds if and only if ρ and σ satisfy $s_+(r/2)_\rho = s_+(r/2)_\sigma$. Note that this condition is satisfied when $r=0$ or $r\geqslant -2\log(\min\{p_1,q_1\})$, which recovers the additivity of E_{entropy} or E_{det}, respectively.

Hence,

$$E_{\frac{r}{2}}(\rho)=\tfrac{1}{2}E_r(\rho\otimes\rho). \tag{121}$$

Substituting this into equation (120), we obtain

$$E_r(\rho\otimes\sigma)\leqslant \tfrac{1}{2}\left\{E_r(\rho\otimes\rho)+E_r(\sigma\otimes\sigma)\right\}. \tag{122}$$

This means that the entanglement yield in collective distillation of different states does not exceed the average of those in separate distillations of each state. For ρ and σ such that $s_+(r/2)_\rho \neq s_+(r/2)_\sigma$, the left-hand side of equation (122) is strictly less than the right-hand side.

On the other hand, consider the independent concentration of ρ and σ with the same error exponent r. The failure probability of the whole process is

$$1-(1-2^{-nr})^2 = 2^{-nr+1}-2^{-2nr}. \tag{123}$$

Thus, the error exponent of the whole process is also r (see appendix A). Therefore,

$$E_r(\rho\otimes\sigma)\geqslant E_r(\rho)+E_r(\sigma) \tag{124}$$

which means that E_r is generally non-additive. In fact, equation (121) provides

$$E_r(\rho\otimes\rho)=2E_{\frac{r}{2}}(\rho)>2E_r(\rho)\qquad\text{for}\quad r<-2\log p_1 \tag{125}$$

thus, the inequality in equation (124) is strict in this case.

To sum up, we obtain the following relation on the general entanglement yield of composite pairs:

$$E_r(\rho)+E_r(\sigma)\leqslant E_r(\rho\otimes\sigma)\leqslant \tfrac{1}{2}\{E_r(\rho\otimes\rho)+E_r(\sigma\otimes\sigma)\}. \tag{126}$$

7. Conclusion

We have discussed entanglement concentration with exponentially decreasing failure probability, as well as fidelity exponentially close to one, in the asymptotic limit. By employing the method of types, we derived the entanglement yield $E(r)$ as a function of an error exponent r.

The result fills the gap between the well-known least upper bound of entanglement yield represented by entropy and the maximum attained in deterministic concentration. The explicit dependence on the exponent of the success probability and of the fidelity was also presented, for the large-yield regime, in the form of a strong converse.

In entanglement manipulation as well as other types of quantum information processing, deterministic, probabilistic and high-fidelity transformations are considered. Our results represent a common generalization and refinement of all three approaches. They provide a unified view of probabilistic and deterministic transformations, and show that success probability and fidelity are essentially equivalent concepts; to be precise, high fidelity and high probability result in the same yield function, while the yield function for low fidelity coincides with that for low probability only for small exponents: for large exponents the latter saturates whereas the former becomes a straight line. The power of the error rate approach (i.e., of quantifying 'rare events') in information theory is also demonstrated in quantum information theory.

Acknowledgments

MK was supported by a Grant-in-Aid for Encouragement of Young Scientists (grant no 12740243) and a Grant-in-Aid for Scientific Research (B) (grant no 12440111) by the Japan Society of the Promotion of Science. KM is thankful to Professor Imai for support. FM is grateful to Takaaki Mukai for his support. AW was supported by the UK Engineering and Physical Sciences Research Council, and he acknowledges gratefully the hospitality of the ERATO Quantum Computation and Information Project, Tokyo.

Appendix A

Lemma 14. *If* $\lim_{n\to\infty} a_n \leqslant \lim_{n\to\infty} b_n$, *then*

$$\lim_{n\to\infty} -\frac{1}{n}\log(2^{-na_n} + 2^{-nb_n}) = \lim_{n\to\infty} a_n. \tag{A.1}$$

Proof.

$$\begin{aligned}\lim_{n\to\infty} -\frac{1}{n}\log(2^{-na_n} + 2^{-nb_n}) &= \lim_{n\to\infty} -\frac{1}{n}\log\left\{2^{-na_n}\left(1 + 2^{-n(b_n-a_b)}\right)\right\}\\ &= \lim_{n\to\infty}\left\{a_n - \frac{1}{n}\log\left(1 + 2^{-n(b_n-a_b)}\right)\right\}\\ &= \lim_{n\to\infty} a_n. \end{aligned}\tag{A.2}$$

□

Appendix B

Lemma 15. *Let* $\mathcal{P}$ *be a convex subset of* $\mathbb{R}^n$. *Let* $S : \mathcal{P} \to \mathbb{R}$ *be a continuous function, and let* $T : \mathcal{P} \to \mathbb{R}$ *be a strictly concave function. Suppose that* T *takes its maximum at* $q_2 \in \mathcal{P}$ *and* $S(q_2) = x_2$, *and that* $\min_{q\in\mathcal{P}} S(q) = x_1$. *Then,*

$$R(x) \equiv \max_{q:S(q)=x} T(q) \qquad (x_1 \leqslant x \leqslant x_2) \tag{B.1}$$

is strictly monotone increasing.

Proof. Consider arbitrary x' and x'' such that $x_1 \leqslant x' < x'' \leqslant x_2$. Then, there exists q' such that $R(x') = \max_{q:S(q)=x'} T(q) = T(q')$ and $S(q') = x'$. Due to the continuity of $S(q)$, there exists q'' such that $S(q'') = x''$. Since the domain $\mathcal{P}$ is a convex set, we have

$$q'' = \lambda q_2 + (1-\lambda)q' \qquad (0 < \lambda \leqslant 1). \tag{B.2}$$

Then, the strict concavity of $T(q)$ and our assumption provide

$$T(q'') \geqslant \lambda T(q_2) + (1-\lambda)T(q') > T(q'). \tag{B.3}$$

Note that the strict concavity of T ensures $T(q_2) > T(q')$. Together with $T(q'') \leqslant \max_{q:S(q)=x''} T(q)$, we have

$$R(x') < R(x'') \qquad \text{for} \quad x' < x''. \tag{B.4}$$

□

We can easily derive the following corollary on *convex* functions.

Corollary 16. *Let $\mathcal{P}$ be a convex subset of $\mathbb{R}^n$. Let $S : \mathcal{P} \to \mathbb{R}$ be a continuous fuction, and let $U : \mathcal{P} \to \mathbb{R}$ be a strictly convex function. Suppose that U takes its minimum at $q_2 \in \mathcal{P}$ and $S(q_2) = x_2$, and that* $\min_{q\in\mathcal{P}} S(q) = x_1$. *Then,*

$$Q(x) \equiv \min_{q:S(q)=x} U(q) \qquad (x_1 \leqslant x \leqslant x_2) \tag{B.5}$$

is strictly monotone decreasing.

Appendix C

This appendix gives the proofs of the lemmata used for reducing fidelity arguments to probabilistic ones in sections 4 and 5. Note that these lemmata are not confined to the analysis of i.i.d. entanglement concentration, but apply in the most general situations. In fact, they apply whenever we are only interested in the *rate* of the entanglement produced and in the error probabilities (fidelities).

Proof of lemma 9. According to theorem 3 in [18], the optimum fidelity is always achieved with a deterministic transformation that actually produces a *pure* state. So, our assumption implies that $|\phi\rangle$ can be deterministically transformed into a pure state $|\phi'\rangle$ such that $|\langle\Phi_T|\phi'\rangle|^2 \geqslant 1-\epsilon$. We may assume (after suitable local rotations)

$$|\phi'\rangle = \sum_i \sqrt{p_i}|i\rangle|i\rangle \tag{C.1}$$

$$|\Phi_T\rangle = \sum_{i=1}^{T} \frac{1}{\sqrt{T}}|i\rangle|i\rangle \tag{C.2}$$

because this will only increase the fidelity.

Then our assumption becomes

$$\sum_{i=1}^{T} \sqrt{p_i}\frac{1}{\sqrt{T}} \geqslant \left(\sum_{i=1}^{T} \sqrt{p_i}\frac{1}{\sqrt{T}}\right)^2 \geqslant 1-\epsilon \tag{C.3}$$

from which we directly obtain

$$\sum_{i=1}^{T} \left(\sqrt{p_i} - \frac{1}{\sqrt{T}}\right)^2 \leqslant 2\epsilon. \tag{C.4}$$

Suppose

$$p_1 \geqslant p_2 \geqslant \cdots \geqslant p_K \geqslant \frac{(1+\sqrt{2})^2}{T} > p_{K+1} \geqslant \cdots \geqslant p_T. \tag{C.5}$$

(Note that if $p_1 < (1+\sqrt{2})^2/T < 6/T$, then we can deterministically obtain $|\Phi_L\rangle$ with $L = \lfloor T/6 \rfloor$ by the deterministic entanglement concentration [13] reviewed in section 2.) Then,

$$\frac{2\epsilon}{T} \geqslant \frac{1}{T}\sum_{i=1}^{T}\left(\sqrt{p_i} - \frac{1}{\sqrt{T}}\right)^2 \geqslant \frac{1}{T}\sum_{i=1}^{K}\left(\sqrt{p_i} - \frac{1}{\sqrt{T}}\right)^2 \geqslant \frac{1}{T}\sum_{i=1}^{K}\frac{2}{T} = \frac{2K}{T^2}. \tag{C.6}$$

Introducing $\beta \equiv K/T \leqslant \epsilon$, and defining $\delta \equiv \sum_{i=1}^{K} p_i$, our aim is to bound the latter probability.

Note that with the restriction that the largest K probabilities p_i sum to δ, the fidelity to $|\Phi_T\rangle$ is maximized for

$$\hat{p}_1 = \ldots = \hat{p}_K = \frac{\delta}{K} \tag{C.7}$$

$$\hat{p}_{K+1} = \ldots = \hat{p}_T = \frac{1-\delta}{T-K} \tag{C.8}$$

yielding

$$\begin{aligned} 1-\epsilon \leqslant F(\phi', \Phi_T) &\leqslant F_{\max}(\delta, K) \\ &= \left(K\sqrt{\frac{\delta}{K}}\frac{1}{\sqrt{T}} + (T-K)\sqrt{\frac{1-\delta}{T-K}}\frac{1}{\sqrt{T}}\right)^2 \\ &= (\sqrt{\delta}\sqrt{\beta} + \sqrt{1-\delta}\sqrt{1-\beta})^2 \\ &\leqslant \sqrt{\delta}\sqrt{\beta} + \sqrt{1-\delta}\sqrt{1-\beta}. \end{aligned} \tag{C.9}$$

If we denote the right-hand side by $1-\eta$, we can solve for $\sqrt{\delta}$, and obtain

$$\begin{aligned} \sqrt{\delta} &= \sqrt{\beta}(1-\eta) \pm \sqrt{(1-\beta)\eta(2-\eta)} \\ &\leqslant \sqrt{\beta} + \sqrt{2\epsilon} \\ &\leqslant \left(1+\sqrt{2}\right)\sqrt{\epsilon}. \end{aligned} \tag{C.10}$$

Hence, $\delta \leqslant 6\epsilon$, and we can execute the following probabilistic protocol: first, Alice and Bob observe whether the state is in the subspace spanned by $|1\rangle, \ldots, |K\rangle$ or in the orthogonal complement, using a projective measurement. The first result occurs only with probability $\delta \leqslant 6\epsilon$. In the other case, the post-measurement state has Schmidt coefficients $p'_i = p_i/(1-\delta)$. Hence,

$$\frac{6}{T(1-6\epsilon)} \geqslant p'_{K+1} \geqslant p'_{K+2} \geqslant \ldots \tag{C.11}$$

so by the deterministic entanglement concentration [13], we can obtain $|\Phi_L\rangle$ deterministically, with $L = \lfloor T(1-6\epsilon)/6 \rfloor$. Therefore, a pure state $|\phi'\rangle$, into which $|\phi\rangle$ is deterministically transformed, can be transformed into $|\Phi_L\rangle$ with probability $1-6\epsilon$. □

Proof of lemma 13. According to theorem 3 in [18], the optimum fidelity is always achieved with a deterministic transformation that actually produces a pure state. So, our assumption implies that $|\phi\rangle$ can be deterministically transformed into a pure state $|\phi'\rangle$ such that $|\langle\Phi_T|\phi'\rangle|^2 = F$. Let $|\phi'\rangle = \sum_{i=1}^{M}\sqrt{p_i}|i\rangle|i\rangle$, and introduce the function $p(x)$ defined by $p(x) = p_i\,(i-1 < x \leqslant i)$, $p(x) = 0\,(M < x)$. Then, the fidelity F to the state $|\Phi_T\rangle$ satisfies

$$\sqrt{TF} = \sqrt{p_1} + \int_1^T \sqrt{p(x)}\,\mathrm{d}x. \tag{C.12}$$

Substituting $x = \mathrm{e}^s$, we obtain

$$\sqrt{TF} - \sqrt{p_1} = \int_0^{\ln(T)} x\sqrt{p(x)}\,\mathrm{d}s \leqslant \ln(T)\max_{0\leqslant x\leqslant T}[x\sqrt{p(x)}]. \tag{C.13}$$

On the other hand, for any x, the probabilistic concentration with truncation value $p(x)$ gives a probabilistic transformation into $|\Phi_L\rangle$ with success probability P with $L \geqslant x$ and $P \geqslant xp(x)$. (Note that $P = p(x)L$ due to equation (17) in section 2.) Hence, $\sqrt{PL} \geqslant x\sqrt{p(x)}$.

Combining those, we arrive at the conclusion that there exists a probabilistic scheme for some $L \leqslant T$ satisfying

$$\sqrt{PL} \geqslant \frac{\sqrt{TF} - \sqrt{p_1}}{\ln(T)} \geqslant \frac{\sqrt{TF} - 1}{\ln(T)}. \tag{C.14}$$

$\square$

Appendix D

Proof of equations (113) and (114)

Consider the function

$$F(s) = -\psi(s) - (1-s)\psi'(s) \qquad (s \geqslant 0) \tag{D.1}$$

where

$$\psi(s) = \log\sum_{i=1}^{d} p_i^s \tag{D.2}$$

and p_i is a probability distribution. Differentiating these functions, we have the following relations:

$$\psi'(s) = \sum_i h_i(s)\log p_i \tag{D.3}$$

where $h_i(s)$ is a probability distribution such that

$$h_i(s) = \frac{p_i^s}{\sum_j p_j^s} \tag{D.4}$$

$$\psi''(s) = \ln 2\left\{\sum_i h_i(s)(\log p_i)^2 - \left(\sum_i h_i(s)\log p_i\right)^2\right\} > 0 \tag{D.5}$$

and

$$F'(s) = -(1-s)\psi''(s)\begin{cases} >0 & (s > 1) \\ =0 & (s = 1) \\ <0 & (0 \leqslant s < 1). \end{cases} \tag{D.6}$$

(Note that $\ln x$ denotes natural logarithm.) Since $F(1) = 0$, $\lim_{s\to\infty} F(s) = -\log p_1$ and $F(0) = D(u\|p) \equiv c$, where $u = (1/d, \ldots, 1/d)$ is the uniform distribution, there exist unique $s_+(r) > 1$ and $0 < s_-(r) < 1$ such that

$$r = \begin{cases} F(s_+(r)) & r \in (0, -\log p_1) \\ F(s_-(r)) & r \in (0, c). \end{cases} \tag{D.7}$$

Let q and $h(s)$ be probability distributions such that $D(q\|p) = D(h(s)\|p) = r$. Then, we have

$$
\begin{aligned}
H(h(s)) - H(q) &= -D(h(s)\|p) - \sum_i h_i(s)\log p_i + D(q\|p) + \sum_i q_i \log p_i \\
&= \sum_i \{q_i - h_i(s)\}\log p_i
\end{aligned} \tag{D.8}
$$

and

$$
\begin{aligned}
\frac{1}{1-s}D(q\|h(s)) &= \frac{1}{1-s}\{D(q\|h(s)) + D(h(s)\|p) - D(q\|p)\} \\
&= \frac{1}{1-s}\sum_i \{-q_i \log h_i(s) + h_i(s)\log h_i(s) - h_i(s)\log p_i + q_i \log p_i\} \\
&= \frac{1}{1-s}\left\{(1-s)\sum_i (q_i - h_i(s))\log p_i + \sum_i (q_i - h_i(s))\log \sum_j p_j^s\right\} \\
&= \sum_i \{q_i - h_i(s)\}\log p_i .
\end{aligned} \tag{D.9}
$$

Thus,

$$
H(h(s)) - H(q) = \frac{1}{1-s}D(q\|h(s)) \tag{D.10}
$$

which gives

$$
H(h(s_+(r))) \leqslant H(q) \qquad \text{and} \qquad H(h(s_-(r)) \geqslant H(q). \tag{D.11}
$$

On the other hand,

$$
\begin{aligned}
F(s) &= -\log \sum_j p_j^s - (1-s)\sum_i h_i(s)\log p_i \\
&= \sum_i h_i(s)\log h_i(s) - \sum_i h_i(s)\log p_i \\
&= D(h(s)\|p)
\end{aligned} \tag{D.12}
$$

hence,

$$
D(h(s_\pm(r))\|p) = F(s_\pm(r)) = r. \tag{D.13}
$$

Therefore,

$$
\begin{aligned}
\frac{s_\pm r + \psi(s_\pm)}{1 - s_\pm} &= \frac{s_\pm F(s_\pm) + \psi(s_\pm)}{1 - s_\pm} \\
&= \psi(s_\pm) - s_\pm \psi'(s_\pm) \\
&= \log \sum_i p_i^{s_\pm} - s_\pm \sum_i h_i(s_\pm)\log p_i \\
&= -\sum_i h_i(s_\pm)\log h_i(s_\pm) \\
&= H(h(s_\pm)).
\end{aligned} \tag{D.14}
$$

From equations (D.11), (D.13) and (D.14), we obtain, for $r \in (0, -\log p_1)$,

$$
\min_{q:D(q\|p)=r} H(q) = \frac{s_+ r + \psi(s_+)}{1 - s_+} \tag{D.15}
$$

i.e.,

$$\min_{q:D(q\|p)=r} \{D(q\|p) + H(q)\} = \frac{r + \psi(s_+)}{1 - s_+} \tag{D.16}$$

and, for $r \in (0, c)$,

$$\max_{q:D(q\|p)=r} H(q) = \frac{s_- r + \psi(s_-)}{1 - s_-} = \frac{r + \psi(s_-)}{1 - s_-} - r. \tag{D.17}$$

Next, consider the function

$$G(s) \equiv \frac{r + \psi(s)}{1 - s}. \tag{D.18}$$

Then,

$$G'(s) = \frac{r - F(s)}{(1 - s)^2}. \tag{D.19}$$

Since

$$F(s) \begin{cases} \geqslant r & (s \leqslant s_-, s_+ \leqslant s) \\ < r & (s_- < s < s_+) \end{cases} \tag{D.20}$$

we have

$$G'(s) \begin{cases} > 0 & (s_- < s < s_+) \\ = 0 & (s = s_\pm) \\ < 0 & (s < s_-, s_+ < s). \end{cases} \tag{D.21}$$

(Note that we take $s_+ = \infty$ for $r \geqslant -\log p_1$.) Thus, we obtain

$$\max_{s \geqslant 1} G(s) = G(s_+(r)) \qquad \text{for} \quad r \in (0, -\log p_1) \tag{D.22}$$

and

$$\sup_{s \geqslant 1} G(s) = \lim_{s \to \infty} G(s) = -\log p_1 \qquad \text{for} \quad r \geqslant -\log p_1. \tag{D.23}$$

As discussed in section 4, the left-hand side of equation (D.16) is monotone decreasing with respect to r. Thus, together with equation (D.18), we obtain

$$\min_{q:D(q\|p) \leqslant r} \{D(q\|p) + H(q)\} = G(s_+) \qquad \text{for} \quad r \in (0, -\log p_1). \tag{D.24}$$

Therefore, equations (D.22), (D.23) and (D.24) give equation (113).

On the other hand, for $r \in (0, c)$, we have $s_-(r) > 0$, thus

$$\min_{0 \leqslant s \leqslant 1} G(s) = G(s_-(r)) \tag{D.25}$$

hence,

$$\min_{0 \leqslant s \leqslant 1} \frac{rs + \psi(s)}{1 - s} = \min_{0 \leqslant s \leqslant 1} (G(s) - r) = G(s_-(r)) - r. \tag{D.26}$$

In addition, for $r \geqslant c$, we have $s_-(r) \leqslant 0$, thus

$$\min_{0 \leqslant s \leqslant 1} G(s) = G(0) = r + \log d \tag{D.27}$$

hence,

$$\min_{0 \leqslant s \leqslant 1} \frac{rs + \psi(s)}{1 - s} = \min_{0 \leqslant s \leqslant 1} (G(s) - r) = \log d. \tag{D.28}$$

As discussed in section 5, the left-hand side of equation (D.17) is monotone increasing with respect to r. Thus, together with equation (D.18), we obtain

$$\max_{q:D(q\|p) \leqslant r} H(q) = G(s_-) - r \qquad \text{for} \quad r \in (0, c). \tag{D.29}$$

Therefore, equations (D.26), (D.28) and (D.29) give equation (114). □

References

[1] Bennett C H and Wiesner S J 1992 *Phys. Rev. Lett.* **69** 2881
[2] Bennett C H, Brassard G, Crépeau C, Jozsa R, Peres A and Wootters W K 1993 *Phys. Rev. Lett.* **70** 1895
[3] Ekert A K 1991 *Phys. Rev. Lett.* **67** 661
[4] Jozsa R and Linden N 2002 *Preprint* quant-ph/0201143
[5] Nielsen M A 1999 *Phys. Rev. Lett.* **83** 436 (*Preprint* quant-ph/9811053)
[6] Vidal G 1999 *Phys. Rev. Lett.* **83** 1046 (*Preprint* quant-ph/9902033)
[7] Hardy L 1999 *Phys. Rev.* A **60** 1912 (*Preprint* quant-ph/9903001)
[8] Jonathan D and Plenio M B 1999 *Phys. Rev. Lett.* **83** 1455 (*Preprint* quant-ph/9903054)
[9] Bennett C H, Bernstein H J, Popescu S and Schumacher B 1996 *Phys. Rev.* A **53** 2046 (*Preprint* quant-ph/9511030)
[10] Popescu S and Rohrlich D 1997 *Phys. Rev.* A **56** R3319 (*Preprint* quant-ph/9610044)
[11] Lo H-K and Popescu S 2001 *Phys. Rev.* A **63** 022301 (*Preprint* quant-ph/9707038)
[12] Morikoshi F 2000 *Phys. Rev. Lett.* **84** 3189 (*Preprint* quant-ph/9911019)
[13] Morikoshi F and Koashi M 2001 *Phys. Rev.* A **64** 022316 (*Preprint* quant-ph/0107120)
[14] Hayashi M and Matsumoto K 2001 *Preprint* quant-ph/0109028
[15] Hayashi M 2002 *Phys. Rev.* A **66** 032321 (*Preprint* quant-ph/0202002)
[16] Cover T M and Thomas J A 1991 *Elements of Information Theory* (New York: Wiley)
[17] Csiszár I and Körner J 1981 *Information Theory: Coding Theorems for Discrete Memoryless Systems* (New York: Academic)
[18] Vidal G, Jonathan D and Nielsen M A 2000 *Phys. Rev.* A **62** 012304 (*Preprint* quant-ph/9910099)