One-shot and asymptotic classical capacity in general physical theories

Shintaro Minagawa [1,*] and Hayato Arai [2,†]

[1] *Graduate School of Informatics, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan*

[2] *Mathematical Quantum Information RIKEN Hakubi Research Team, RIKEN, Center for Quantum Computing, 2-1 Hirosawa, Wako 351-0198, Japan*

(Received 9 February 2024; accepted 22 May 2024; published 11 June 2024)

With the recent development of quantum information theory, some attempts have been made to construct information theory beyond quantum theory. Here, we consider hypothesis-testing relative entropy and one-shot classical capacity, that is, the optimal rate of classical information transmitted by using a single channel under the constraint of a certain error probability, in general physical theories where states and measurements are operationally defined. Then we obtain the upper bound of the one-shot classical capacity by generalizing the method given by Wang and Renner [Phys. Rev. Lett. **108**, 200501 (2012)]. Also, we derive the lower bound of the capacity by showing the existence of a good code that can transmit classical information with a certain error probability. Applying the above two bounds, we prove the asymptotic equivalence between classical capacity and hypothesis-testing relative entropy in any general physical theory.

DOI: 10.1103/PhysRevA.109.062416

I. INTRODUCTION

Since Shannon invented information theory [1], it has been increasingly important (see e.g., Ref. [2]). The goal of information theory is basically to express the optimal efficiency for some tasks, and the optimal efficiencies for different tasks are sometimes equivalent or directly related through some information quantities like mutual information; a typical example is the asymptotic equivalence between the exponential rate of hypothesis testing and the classical information-transmission capacity [3].

Recently, because quantum information theory (see, e.g., Refs. [4,5]) has flourished, similar relations have become known in quantum theory. In particular, the same relationship between hypothesis testing and channel capacity also holds in quantum theory [6–11]. These facts imply that an information theory should possess such relations between the optimal efficiencies for some tasks independently of the mathematical structure of its background physical systems.

However, when we establish an information theory that stands by the operationally minimum principles, possible models of background physical systems are not restricted to classical and quantum theory. Such theories are called *general probabilistic theories* [12–49], or GPTs (for a review, see, e.g., Refs. [12–15]). The framework of GPTs is a kind of generalization of classical and quantum theory whose states and measurements are operationally defined, and studies of GPTs have been widespread recently.

Even in such general models, some properties of information theory also hold like in quantum theory. One such result of preceding studies of GPTs is the no-cloning theorem in GPTs [50]. It clarified that no model except for classical theory can copy any information freely, similar to the no-cloning theorem in quantum theory, which means that quantum theory is not a special theory with no cloning but classical theory is a special theory with cloning.

On the other hand, some properties of information theory are drastically changed in GPTs. A typical example is *entropy*. In quantum theory, there are several methods to characterize von Neumann entropy $S(\rho) := -\mathrm{Tr}[\rho \log_2 \rho]$ [51] based on the classical Shannon entropy, but the von Neumann entropy is well defined without a choice of the method of classicalization [4,5]. On the other hand, it is known that such well-definedness is not valid in GPTs; i.e., there is no simple way to define entropy similar to the von Neumann entropy in GPTs [26–28,44,49]. A certain generalization of von Neumann entropy is not even concave [26,28].

Because entropy is not generalized straightforwardly in GPTs, we cannot easily obtain a similar result for optimal efficiency for certain information tasks. Therefore, whether the relations between optimal efficiencies for different tasks are the same as the relations in classical and quantum theory is a difficult question. If the answer is positive, i.e., relations between different information tasks are independent of the mathematical structure of physical systems even though entropies do not behave the same, we can reach a new foundational perspective on information theory: Efficiencies for information tasks give more robust definitions of information quantities than entropies in GPTs because of the independence of the mathematical structure of physical systems.

In this paper, we discuss hypothesis testing and classical information transmission in GPTs in the same way as in classical and quantum theory following Ref. [52]. Next, we estimate the upper and lower bounds of one-shot classical capacity with hypothesis-testing relative entropy [53] in GPTs. As a result, we obtain upper and lower bounds similar to those of quantum theory. Moreover, due to the construction of the achievable case of our bound, our result for the one-shot case can be applied to the asymptotic case even though the

*Contact author: minagawa.shintaro@nagoya-u.jp
†Contact author: hayato.arai@riken.jp

asymptotic scenario is complicated in GPTs. Consequently, we show the asymptotic equivalence between the above two efficiencies even in GPTs.

The structure of this paper is as follows. In Sec. II, we introduce basic notions of GPTs. In Sec. III, we introduce hypothesis-testing relative entropy in GPTs. In Sec. IV, we derive the one-shot classical capacity theorem in GPTs. In Sec. V, we introduce the asymptotic setting and show the asymptotic equivalence between one-shot classical capacity and hypothesis-testing relative entropy in GPTs.

II. GENERAL PROBABILISTIC THEORIES

Here, we introduce the mathematical basics of GPTs following Refs. [12,14,15,40]. Let V be a finite-dimensional real vector space and the subset $K \subset V$ be a positive cone, i.e., a set satisfying the following three conditions: (1) $\lambda x \in K$ holds for any $x \in K$ and any $\lambda \geqslant 0$. (2) K is convex and has a nonempty interior. (3) $K \cap (-K) = \{0\}$. The dual cone of K, denoted K^*, is defined as follows:

$$K^* := \{y \in V^* \mid \langle y, x\rangle \geqslant 0 \ \forall x \in K\}, \tag{1}$$

where $\langle \cdot, \cdot \rangle$ is the inner product of the vector space V. In addition, an inner point $u \in K^*$, called the *unit effect*, is fixed for a model. Then, a state in this model is defined as an element $\rho \in K$ satisfying $\langle \rho, u\rangle = 1$. The state space, i.e., the set of all states, is denoted as $S(K)$. Due to the convexity of K, the state space $S(K)$ is also convex.

Also, a measurement is defined as a family $\boldsymbol{e} := \{e_j\}_{j\in J}$ satisfying $e_j \in K^*$ for any $j \in J$ and $\sum_{j\in J} e_j = u$. The measurement space, i.e., the set of all measurements with finite outcomes, is denoted as $\mathcal{M}(K)$. Here, $\langle e_j, \rho\rangle$ corresponds to the probability of obtaining an outcome $j \in J$ when we perform a measurement $\mathbf{e}$ to a state $\rho \in \mathcal{S}(K)$. Next, we define an order relation $\geqslant$ on K^*. We say that $f \geqslant e$ if $f - e \in K^*$. This means that for any element $x \in K$, $\langle f, x\rangle \geqslant \langle e, x\rangle$. Here, we remark that a family $\{e, u - e\}$ is a measurement in $\mathcal{M}(K)$ if and only if the element $e \in K^*$ satisfies $0 \leqslant e \leqslant u$ using the above order relation.

Next, we give examples of positive cones and models of GPTs. The simplest example of positive cones is the positive part of the vector space $\mathbb{R}^n$, defined as

$$\mathbb{R}^n_+ := \{(x_i)_{i=1}^n \in \mathbb{R}^n \mid x_i \geqslant 0 \ \forall i\}. \tag{2}$$

Considering the standard inner product on $\mathbb{R}^n$, the dual $\mathbb{R}^{n*}_+$ is equivalent to itself. We fix a unit $\mathbf{u} = (1, 1, \dots, 1)$. Then, the state space $\mathcal{S}(\mathbb{R}^{n*}_+)$ is given as the set of all ensembles $\sum_{j\in J} p_j \mathbf{c}^j$, where $\{p_j\}$ is the probability vector and $\mathbf{c}^j := (c^j_i)_{i=1}^n$ is a vector such that $c^j_i = \delta_{ij}$, with δ_{ij} being the Kronecker delta, which corresponds to a classical bit. In this paper, we denote $\mathbf{c}^j$ as $|j\rangle\langle j|$. Also, a measurement is given as a family $\{\mathbf{e}^j\}_{j\in J}$ of $\mathbf{e}^j \in \mathbb{R}^n_+$ such that $\sum_j \mathbf{e}^j = \mathbf{u}$, which corresponds to a strategy to obtain information from an ensemble of the classical n-level system.

This model corresponds to the theory of classical information and classical operations, i.e., classical theory. Quantum theory is also a model of a GPT in the case $V = L_{\mathrm{H}}(\mathcal{H})$, $\langle x, y\rangle = \mathrm{Tr}xy$, $K = L^+_{\mathrm{H}}(\mathcal{H})$, and $u = \mathbb{1}$. Here, $L_{\mathrm{H}}(\mathcal{H})$ denotes the set of Hermitian matrices on a Hilbert space $\mathcal{H}$, and $L^+_{\mathrm{H}}(\mathcal{H})$ denotes the set of positive-semidefinite matrices on $\mathcal{H}$. Also, $\mathbb{1}$ is the identity matrix on $\mathcal{H}$. In this model, a state is given as a density matrix, and a measurement is given as a positive operator-valued measure (POVM).

Next, we define a *measurement channel* associated with a measurement $\mathbf{e}$ as the following map $\mathcal{E}_{\mathbf{e}}$ from $\mathcal{S}(K)$ to $\mathcal{S}(\mathbb{R}^n_+)$ [14]:

$$\mathcal{E}_{\mathbf{e}}(\rho) := \sum_{j\in J} \langle e_j, \rho\rangle\, |j\rangle\langle j|\,. \tag{3}$$

We also define an adjoint map of a measurement channel $\mathcal{E}_{\mathbf{e}}$ as the following map $\mathcal{E}^\dagger_{\mathbf{e}}$ from $\mathbb{R}^{n*}_+ = \mathbb{R}^n_+$ to K^* for any $f \in \mathbb{R}^n_+$:

$$\mathcal{E}^\dagger_{\mathbf{e}}(f) := \sum_{j\in J} \langle f, |j\rangle\langle j|\rangle e_j. \tag{4}$$

Note that the following equation holds for any $f \in \mathbb{R}^n_+$ and any $\rho \in \mathcal{S}(K)$:

$$\begin{aligned}\langle \mathcal{E}^\dagger_{\mathbf{e}}(f), \rho\rangle &= \left\langle \sum_{j\in J} \langle f, |j\rangle\langle j|\rangle e_j, \rho \right\rangle \\ &\overset{(a)}{=} \sum_{j\in J} \langle e_j, \rho\rangle \langle f, |j\rangle\langle j|\rangle \\ &\overset{(b)}{=} \left\langle f, \sum_{j\in J} \langle e_j, \rho\rangle\, |j\rangle\langle j| \right\rangle = \langle f, \mathcal{E}_{\mathbf{e}}(\rho)\rangle.\end{aligned} \tag{5}$$

Equalities (a) and (b) hold because of the linearity of the inner product. In addition, if $0 \leqslant f \leqslant u$, it holds that $0 \leqslant \mathcal{E}^\dagger_{\mathbf{e}}(f) \leqslant u$ because $0 \leqslant \langle f, |j\rangle\langle j|\rangle \leqslant 1$ holds. As a result, if f is an effect of a measurement $\{f, u - f\}$, then $\mathcal{E}^\dagger_{\mathbf{e}}(f)$ is also an effect of the measurement $\{\mathcal{E}^\dagger_{\mathbf{e}}(f), u - \mathcal{E}^\dagger_{\mathbf{e}}(f)\}$.

Finally, we define a composite system of classical theory and a general model of the positive cone K in GPTs [12,14]. The vector space of the composite system is given by $\mathbb{R}^n \otimes V$. The corresponding positive cone is given as

$$\mathbb{R}^n_+ \otimes K := \mathrm{Conv}(\{|x\rangle\langle x| \otimes \rho \mid x \in \mathbb{R}^n_+, \rho \in K\}), \tag{6}$$

where $\mathrm{Conv}(S)$ is the convex hull of the set S. The unit is given as the tensor product of units in each system. A state in the composite system is given as an ensemble of tensor products $|x\rangle\langle x| \otimes \rho$ of a classical bit $|x\rangle\langle x|$ and a general state $\rho \in \mathcal{S}(K)$. For a bipartite state ρ^{AB}, the marginal states ρ^{A} and ρ^{B} are defined as the unique states satisfying the following relations for any pair of a classical measurement $\{e^{\mathrm{A}}_i\}$ and a general measurement $\{e^{\mathrm{B}}_j\}$, respectively [12]:

$$\sum_j \left\langle e^{\mathrm{A}}_i \otimes e^{\mathrm{B}}_j, \rho^{\mathrm{AB}}\right\rangle = \left\langle e^{\mathrm{A}}_i, \rho^{\mathrm{A}}\right\rangle, \tag{7}$$

$$\sum_i \left\langle e^{\mathrm{A}}_i \otimes e^{\mathrm{B}}_j, \rho^{\mathrm{AB}}\right\rangle = \left\langle e^{\mathrm{B}}_j, \rho^{\mathrm{B}}\right\rangle. \tag{8}$$

Here, we note that the marginal states are given as follows if the bipartite state is given as the ensemble $\sum_{j,j'} p_{j,j'}\, |j\rangle\langle j|^{\mathrm{A}} \otimes \rho^{\mathrm{B}}_{j'}$:

$$\rho^{\mathrm{A}} = \sum_{j,j'} p_{j,j'}\, |j\rangle\langle j|^{\mathrm{A}}\,, \tag{9}$$

$$\rho^{\mathrm{B}} = \sum_{j,j'} p_{j,j'} \rho^{\mathrm{B}}_{j'}. \tag{10}$$

III. HYPOTHESIS-TESTING RELATIVE ENTROPY IN GPTs

Next, we introduce hypothesis-testing relative entropy in general models. In quantum theory, hypothesis-testing relative entropy is defined for $0 \leqslant \epsilon \leqslant 1$ as follows [52,54–56]:

$$D_{\mathrm{H}}^{\epsilon}(\rho||\sigma) := -\log_2 \min_{\substack{E: 0 \leqslant E \leqslant \mathbb{1}, \\ \mathrm{Tr}\{E\rho\} \geqslant 1-\epsilon}} \mathrm{Tr}\{E\sigma\}, \tag{11}$$

where $\mathbb{1}$ is an identity operator. This definition comes from the hypothesis testing of two quantum states (see, e.g., Ref. [57]). We discriminate states ρ and σ by performing a two-valued measurement with the POVM $\{E, \mathbb{1} - E\}$. There are two kinds of error probabilities, type-I error probability $\mathrm{Tr}\{(\mathbb{1} - E)\rho\}$ and type-II error probability $\mathrm{Tr}\{E\sigma\}$. Definition (11) corresponds to the optimization of the type-II error probability $\mathrm{Tr}\{E\sigma\}$ under the constraint that the type-I error probability has an upper bound ϵ, that is, $\mathrm{Tr}\{(\mathbb{1} - E)\rho\} \leqslant \epsilon$.

As a generalization of this definition, we can introduce hypothesis-testing relative entropy in GPTs as follows [58].

Definition 1. Hypothesis-testing relative entropy in GPTs. Let $\rho, \sigma \in \Omega$ be states and q be an effect where $0 \leqslant \langle q, \rho\rangle \leqslant 1$ holds for any state $\rho \in \Omega$. Let $0 \leqslant \epsilon \leqslant 1$ be a real value. We define hypothesis-testing relative entropy as follows:

$$D_{\mathrm{H,G}}^{\epsilon}(\rho||\sigma) := -\log_2 \min_{\substack{q:\, 0 \leqslant q \leqslant u, \\ \langle q, \rho\rangle \geqslant 1-\epsilon}} \langle q, \sigma\rangle. \tag{12}$$

As the following lemma shows, measurement channels do not increase the hypothesis-testing relative entropy, which is important for the following discussion.

Lemma 1. Data-processing inequality for a measurement channel. Let $\mathcal{E}_{\mathbf{e}} : \mathcal{S}(K) \to S$, defined as $\mathcal{E}_{\mathbf{e}}(\cdot) := \sum_{j \in J} \langle e_j, \cdot\rangle\, |j\rangle\langle j|$, be a measurement channel corresponding to the measurement $\mathbf{e} = \{e_j\}_{j \in J}$. We have

$$D_{\mathrm{H,G}}^{\epsilon}(\rho||\sigma) \geqslant D_{\mathrm{H,G}}^{\epsilon}(\mathcal{E}_{\mathbf{e}}(\rho)||\mathcal{E}_{\mathbf{e}}(\sigma)). \tag{13}$$

IV. ONE-SHOT CLASSICAL CAPACITY IN GPTs

Here, we consider one-shot classical capacity in GPTs based on the setup given by Ref. [52]. First, we describe our setup of one-shot classical information transmission from the sender in system A to the receiver in system B.

The sender and receiver share a channel Φ from $\mathcal{X}$ to $\mathcal{S}(K)$, defined as $\Phi(|x\rangle\langle x|) = \sigma_x^{\mathrm{B}}$, where $\mathcal{X}$ is an alphabet. The sender encodes an n-length bit string $j \in \Gamma := \{0, 1, 2, \ldots, 2^n - 1\}$ to $x \in \mathcal{X}$ by using the function $g(j) = x$, called the *encoder*. The set $\mathcal{G} = g(\Gamma)$ and the element $g(j)$ are called the *codebook* and *codeword*, respectively. The receiver performs a measurement $\mathbf{m}^{\mathrm{B}} := \{m_j^{\mathrm{B}}\}_{j \in \Gamma}$ on the arrived state $\sigma_{g(j)}^{\mathrm{B}}$, where $m_j^{\mathrm{B}} \geqslant 0$ and $\sum_{j \in \Gamma} m_j^{\mathrm{B}} = u$. The error probability for a given message $j \in \Gamma$, encoder g, and measurement $\mathbf{m}^{\mathrm{B}}$ is defined as

$$\Pr(\mathrm{error}|j, g, \mathbf{m}^{\mathrm{B}}) = \langle u - m_j^{\mathrm{B}}, \sigma_{g(j)}^{\mathrm{B}}\rangle. \tag{14}$$

This setting is illustrated in Fig. 1.

The sender and receiver aim to maximize the size of the bit strings under the condition that the average error is small enough. In order to define the rate of this task and the capacity of the channel, we define a code that fulfills this aim.

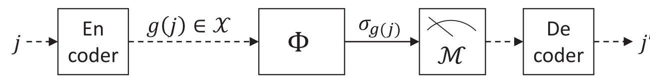

FIG. 1. The setup for sending classical information. The dashed arrows indicate transmissions of classical information. The solid line indicates a transmission of a state in the general theory. The sender chooses the message $j \in \Gamma$ and encodes it using the function $g : \Gamma \to \mathcal{X}$. Classical information $g(j)$ is transformed into state $\sigma_{g(j)}$ by a channel Φ whose input is the classical information and whose output is the state of the GPT of the receiver's system. Then the receiver performs a measurement $\mathcal{M}$ to $\sigma_{g(j)}$ and decodes that result to obtain classical information j'. We say that the measurement decodes the message j correctly when $j' = j$.

Definition 2. $(2^n, \epsilon)$ *code.* Let $\Gamma = \{0, 1, \ldots, 2^n - 1\}$ be an n-length bit string. A $(2^n, \epsilon)$ code for a map $\Phi : |x\rangle\langle x| \mapsto \sigma_x^{\mathrm{B}}$ consists of an encoder $g : \Gamma \to \mathcal{X}$ and decoding measurement $\mathbf{m}^{\mathrm{B}} := \{m_j^{\mathrm{B}}\}_{j \in \Gamma}$, whose average error probability when the message $j \in \Gamma$ is chosen uniformly at random is bounded from the above by ϵ, with the formula

$$\Pr(\mathrm{error}|g, \mathbf{m}^{\mathrm{B}}) := \frac{1}{2^n} \sum_{j \in \Gamma} \Pr(\mathrm{error}|j, g, \mathbf{m}^{\mathrm{B}}) \leqslant \epsilon. \tag{15}$$

Then, we define the rate and capacity in the same way as they are defined in the quantum case [52,59].

Definition 3. One-shot ϵ-achievable rate. A real number $R \geqslant 0$ is a one-shot ϵ-achievable rate for one-shot classical information transmission through Φ if there is a $(2^R, \epsilon)$ code.

Definition 4. One-shot ϵ-classical capacity. The one-shot ϵ-classical capacity of a map Φ, $\mathrm{C}^{\epsilon}(\Phi)$, is defined as

$$\mathrm{C}^{\epsilon}(\Phi) := \sup\{R \mid R \text{ is a one-shot } \epsilon\text{-achievable rate}\}. \tag{16}$$

Now, we define the following ensemble:

$$\pi_{P_X}^{\mathrm{AB}} := \sum_{x \in \mathcal{X}} P_X(x)\, |x\rangle\langle x|^{\mathrm{A}} \otimes \sigma_x^{\mathrm{B}}, \tag{17}$$

where $P_X(x)$ is the probability distribution of a random variable associated with the alphabet $\mathcal{X}$. The marginal states with respect to A and B are

$$\pi_{P_X}^{\mathrm{A}} = \sum_{x \in \mathcal{X}} P_X(x)\, |x\rangle\langle x|^{\mathrm{A}}, \tag{18}$$

$$\pi_{P_X}^{\mathrm{B}} = \sum_{x \in \mathcal{X}} P_X(x) \sigma_x^{\mathrm{B}}, \tag{19}$$

respectively.

In quantum theory, the ϵ-one-shot classical capacity is asymptotically equivalent to the optimal hypothesis-testing relative entropy between the above ensemble π^{AB} and the product of its marginal states. This paper shows that the equivalence also holds even in GPTs.

First, we show the converse part, i.e., the upper bound of $\mathrm{C}^{\epsilon}(\Phi)$, by applying Lemma 1.

Theorem 1. The ϵ-one-shot classical capacity of a map $\Phi : |x\rangle\langle x| \mapsto \sigma_x^{\mathrm{B}}$ is bounded as follows:

$$\mathrm{C}^{\epsilon}(\Phi) \leqslant \sup_{P_X} D_{\mathrm{H,G}}^{\epsilon}\big(\pi_{P_X}^{\mathrm{AB}} \big|\big| \pi_{P_X}^{\mathrm{A}} \otimes \pi_{P_X}^{\mathrm{B}}\big), \tag{20}$$

where the supremum is taken over all probability distributions P_X and $\pi^{\mathrm{A}}_{P_X}$ and $\pi^{\mathrm{B}}_{P_X}$ are marginal states of $\pi^{\mathrm{AB}}_{P_X}$ with regard to systems A and B, respectively.

We give a proof of Theorem 1 in the Appendix.

Next, we show the achievable part, which gives the lower bound of $\mathrm{C}^{\epsilon}(\Phi)$.

Theorem 2. The ϵ-one-shot classical capacity of a map $\Phi : x \in \mathcal{X} \to \sigma^{\mathrm{B}}_x$ satisfies the following inequality for any $\epsilon' \in (0, \epsilon)$, any $s > 1$, and any $t > s$ satisfying $\epsilon > s\epsilon'$:

$$\mathrm{C}^{\epsilon}(\Phi) \geqslant \sup_{P_X} D^{\epsilon'}_{\mathrm{H,G}}\left(\pi^{\mathrm{AB}}_{P_X} \middle\| \pi^{\mathrm{A}}_{P_X} \otimes \pi^{\mathrm{B}}_{P_X}\right) - \log_2 \frac{t}{\epsilon - s\epsilon'}. \tag{21}$$

To show Theorem 2, we give the following lemma.

Lemma 2. Let K be a positive cone, and let $\mathcal{Y}$ be a finite set. Let $\{A_y\}_{y\in\mathcal{Y}}$ be a family of effects in K^* satisfying $0 \leqslant A_y \leqslant u$ for all $y \in \mathcal{Y}$. Then, for any real numbers $s > 1$ and $t > s$, there is a measurement $\{E_y\}_{y\in\mathcal{Y}} \in \mathcal{M}(K)$ that satisfies the following inequalities for each E_y:

$$u - E_y \leqslant s(u - A_y) + t \sum_{\substack{z\in\mathcal{Y}\\ z\neq y}} A_z. \tag{22}$$

We give proofs of Theorem 2 and Lemma 2 in the Appendix. Simply speaking, an achievable code is obtained with Lemma 2 that is associated with the achievable measurement of hypothesis-testing relative entropy.

Here, we remark on the construction of the measurement in the proofs of Lemma 2 and Theorem 2. In classical-quantum channel coding [9,10,52], the measurement is chosen to be a *pretty good measurement* [60] determined by the square roots of the original family $\{A_y\}_{y\in\mathcal{Y}}$. On the other hand, our construction is trivial, but it needs information about the unique index $y_0 \in \mathcal{Y}$ satisfying a certain property. The construction in Refs. [9,10,52] is valid without such information, and therefore, we can apply the construction in [9,10,52] to actual information tasks, in contrast to our construction. However, this paper aims to estimate the value $\mathrm{C}^{\epsilon}(\Phi)$. For this aim, we have to show only the existence of an optimal measurement, and thus, our construction is sufficient even though our method is not helpful for actual information tasks.

V. ASYMPTOTIC INDEPENDENT AND IDENTICAL DISTRIBUTION CASE

In this section, we consider how the capacity is expressed when a channel is used m (m is a positive integer) times in an independent and identical distribution.

We express an m-length sequence consisting of alphabet $\mathcal{X}$ as $x_1 \cdots x_m$. Let us fix the probability of occurrence for each symbol as $P_X(x)$. Then, when a channel Φ is used m times, the sender and receiver can share the following state:

$$\pi^{AB}_{P_{X^m}} := \sum_{x_1\cdots x_m\in\mathcal{X}^m} P_{X^m}(x_1\cdots x_m)\, |x_1\cdots x_m\rangle\langle x_1\cdots x_m|^{\mathrm{A}} \otimes \sigma^{\mathrm{B}}_{x_1\cdots x_m}. \tag{23}$$

Here, $\mathcal{X}^m$ indicates that the set of all m-length sequences consists of the alphabet $\mathcal{X}$, and $\sigma^{\mathrm{B}}_{x_1\cdots x_m}$ is the abbreviation of $\sigma^{\mathrm{B}}_{x_1} \otimes \sigma^{\mathrm{B}}_{x_2} \otimes \cdots \otimes \sigma^{\mathrm{B}}_{x_m}$. We denote the above map from $|x_1\cdots x_m\rangle\langle x_1\cdots x_m|^{\mathrm{A}}$ to $\sigma^{\mathrm{B}}_{x_1\cdots x_m}$ as $\Phi^{\otimes m}$.

Here, we remark on the composition of the model of GPTs. In the standard setting of GPTs, i.e., the case when we assume no signaling and local tomography [23,61], an n-composite model of a model defined by K is defined by a positive cone K_n satisfying

$$\bigotimes_{i=1}^{n} K_i \subset K_n \subset \left(\bigotimes_{i=1}^{n} K_i^*\right)^*, \tag{24}$$

where the set $\bigotimes_{i=1}^{n} K_i$ is defined as

$$\bigotimes_{i=1}^{n} K_i := \mathrm{Conv}\{\otimes\rho_i | \rho_i \in K_i\}. \tag{25}$$

As shown in Refs. [45,46], an n-composite model of a nonclassical single model is not uniquely determined in GPTs. Therefore, we need to be more careful in GPTs when we consider an asymptotic scenario. However, in the above asymptotic scenario, we need to consider only m uses of a channel Φ, which is a channel from a classical m-length bit to an m-partite product state $\sigma^{\mathrm{B}}_{x_1\cdots x_m}$. Due to the inclusion relation (24), an m-partite product state can be regarded as a state in any composite model of a single system, and therefore, the map $\Phi^{\otimes n}$ is well defined even in GPTs. Hence, we can apply the results in the single-shot scenario to the asymptotic scenario.

Now we consider the situation where we encode a message $j \in \Gamma$ to an m-length sequence $x_1 \cdots x_m$ by an encoder g_m, that is, $g_m(j) = x_1 \cdots x_m$. Here, notice that the size of the set of all messages Γ depends on m, and we denote it as $|g_m|$. Also, let us denote the decoding error ϵ_m when the message j appears uniformly at random. Then, like in the single-shot scenario, we define an ϵ-asymptotic achievable rate as a real number $R \geqslant 0$ if a sequence of $(m, |g_m|, \epsilon)$ codes satisfying $\liminf_{m\to\infty} \frac{1}{m} \log_2 |g_m| = R$ exists. Finally, we define ϵ-asymptotic classical capacity [3].

Definition 5. ϵ-asymptotic classical capacity. The ϵ-asymptotic classical capacity of Φ is defined as follows:

$$\tilde{\mathrm{C}}^{\epsilon}(\Phi) := \sup\{R \mid R \text{ is an } \epsilon\text{-achievable rate for } \Phi\}. \tag{26}$$

By the definitions of one-shot ϵ-classical capacity and ϵ-classical capacity, we have

$$\tilde{\mathrm{C}}^{\epsilon}(\Phi) = \liminf_{m\to\infty} \frac{1}{m} \mathrm{C}^{\epsilon}(\Phi^{\otimes m}). \tag{27}$$

Because the map $\Phi^{\otimes m}$ can be regarded as a channel in a model of GPTs, we can apply Theorem 1. Also, in the proof of Theorem 2, the decoding measurement is a trivial measurement with information about a given channel. A trivial measurement is also well defined in any composite system of the single system, and therefore, we can also apply Theorem 2. As a result, the ϵ-asymptotic classical capacity of Φ satisfies

$$\begin{aligned}
&\lim_{m\to\infty} \frac{1}{m} \sup_{P_{X^m}} D^{\epsilon'}_{\mathrm{H,G}}\left(\pi^{\mathrm{AB}}_{P_{X^m}} \middle\| \pi^{\mathrm{A}}_{P_{X^m}} \otimes \pi^{\mathrm{B}}_{P_{X^m}}\right)\\
&\quad\leqslant \tilde{\mathrm{C}}^{\epsilon}(\Phi)\\
&\quad\leqslant \lim_{m\to\infty} \frac{1}{m} \sup_{P_{X^m}} D^{\epsilon}_{\mathrm{H,G}}\left(\pi^{\mathrm{AB}}_{P_{X^m}} \middle\| \pi^{\mathrm{A}}_{P_{X^m}} \otimes \pi^{\mathrm{B}}_{P_{X^m}}\right),
\end{aligned} \tag{28}$$

where $\epsilon' \in (0, \epsilon)$. Here, the hypothesis-testing relative entropy $D^{\epsilon}(\rho||\sigma)$ is left continuous over ϵ (see Lemma 3 in the Appendix). As a result, we obtain the following theorem.

Theorem 3. In any model of GPTs and any $\epsilon \in (0, 1)$, a channel Φ satisfies

$$\tilde{\mathrm{C}}^{\epsilon}(\Phi) = \lim_{m\to\infty} \frac{1}{m} \sup_{P_{X^m}} D^{\epsilon}_{\mathrm{H,G}}\big(\pi^{\mathrm{AB}}_{P_{X^m}} \big\| \pi^{\mathrm{A}}_{P_{X^m}} \otimes \pi^{\mathrm{B}}_{P_{X^m}}\big). \tag{29}$$

In other words, the classical capacity and the hypothesis-testing relative entropy are asymptotically equivalent even in GPTs.

Here, we remark on the dependence of ϵ. In quantum theory, because of the quantum Stein's lemma [6,7], we have the following equality [52]:

$$\forall \epsilon \in (0, 1), \quad \lim_{n\to\infty} \frac{1}{n} D^{\epsilon}_{\mathrm{H}}(\rho^{\otimes n}||\sigma^{\otimes n}) = D(\rho||\sigma).$$

In other words, there is no ϵ dependence in quantum theory, and the rates are equal to Umegaki's relative entropy $D(\rho||\sigma)$ [62,63]. Thus, both the asymptotic classical capacity and asymptotic hypothesis-testing relative entropy are determined independently from ϵ. Therefore, we often consider the case $\epsilon \to 0$, which provides the Holevo-Schumacher-Westmoreland (HSW) theorem [9,10] as stated in Ref. [52]. In contrast to quantum theory, it is not yet known whether they are independent of ϵ in GPTs, which is still an important open problem. However, whether they are dependent or independent, this paper has clarified the asymptotic equivalence between classical capacity and hypothesis-testing relative entropy for each ϵ even in GPTs.

VI. CONCLUSIONS

We introduced classical information transmission in GPTs, and we showed the lower and upper bounds of the one-shot classical capacity theorem in any physical theory given by hypothesis-testing relative entropy. In addition, we showed that the lower and upper bounds are asymptotically equivalent. In other words, we showed the equivalence between asymptotic classical capacity and asymptotic hypothesis-testing relative entropy in any physical theory.

In classical and quantum theory, the above rates of different information tasks are connected by entropies, but in general, entropies do not possess properties similar to classical and quantum theory in GPTs. Our contribution is to clarify the equivalence of two rates of different information tasks without entropies in any physical theory.

Note that the asymptotic equivalence of the capacity and hypothesis-testing relative entropy given by Theorem 3 depends on ϵ. The open problem is then the HSW theorem in GPTs. By virtue of Theorem 3, this problem can be reduced to the following two problems: (1) Are asymptotic classical capacity and asymptotic hypothesis-testing relative entropy independent of ϵ even in GPTs? (2) If so, are asymptotic classical capacity and asymptotic hypothesis-testing relative entropy related to standard relative entropy even in GPTs? The answers to both problems should give an important new operational perspective of entropies and information rates.

ACKNOWLEDGMENTS

The authors thank M. M. Wilde for pointing out errors in the previous version and for helpful comments and discussions. Helpful comments from and discussions with F. Buscemi are gratefully acknowledged. Comments from T. V. Vu on the previous version of the manuscript are gratefully acknowledged. S.M. would like to take this opportunity to acknowledge the Nagoya University Interdisciplinary Frontier Fellowship supported by Nagoya University and JST, a university fellowship established for the creation of science technology innovation, Grant No. JPMJFS2120, and "THERS Make New Standards Program for the Next Generation Researchers" supported by JST SPRING, Grant No. JPMJSP2125.

APPENDIX: LEMMAS AND PROOFS

1. Proof of Lemma 1

Proof. Let $0 \leqslant \overline{q} \leqslant u$ be a classical effect that attains the minimization in the definition of classical hypothesis-testing relative entropy $D^{\epsilon}_{\mathrm{H,G}}(\mathcal{E}_{\mathbf{e}}(\rho)||\mathcal{E}_{\mathbf{e}}(\sigma))$. In other words, $\overline{q}$ satisfies the following relations:

$$\langle \overline{q}, \mathcal{E}_{\mathbf{e}}(\rho)\rangle \geqslant 1 - \epsilon, \tag{A1}$$

$$-\log_2\langle \overline{q}, \mathcal{E}_{\mathbf{e}}(\sigma)\rangle = D^{\epsilon}_{\mathrm{H,G}}(\mathcal{E}_{\mathbf{e}}(\rho)||\mathcal{E}_{\mathbf{e}}(\sigma)). \tag{A2}$$

Because of Eq. (5), we have $\langle \mathcal{E}^{\dagger}_{\mathbf{e}}(\overline{q}), \rho\rangle = \langle \overline{q}, \mathcal{E}_{\mathbf{e}}(\rho)\rangle$. The combination of this with Eqs. (A1) and (A2) leads to the following relations, respectively:

$$\langle \mathcal{E}^{\dagger}_{\mathbf{e}}(\overline{q}), \rho\rangle \geqslant 1 - \epsilon, \tag{A3}$$

$$-\log_2\langle \mathcal{E}^{\dagger}_{\mathbf{e}}(\overline{q}), \sigma\rangle = D^{\epsilon}_{\mathrm{H,G}}(\mathcal{E}_{\mathbf{e}}(\rho)||\mathcal{E}_{\mathbf{e}}(\sigma)). \tag{A4}$$

Equation (A3) implies that the effect $\mathcal{E}^{\dagger}_{\mathbf{e}}(\overline{q})$ satisfies the condition of the minimization of $D^{\epsilon}_{\mathrm{H,G}}(\rho||\sigma)$ [but not necessarily the optimal effect that achieves $D^{\epsilon}_{\mathrm{H,G}}(\rho||\sigma)$], which implies the desired inequality (13) [64]. ■

2. Proof of Theorem 1

Proof. Let $R := \mathrm{C}^{\epsilon}(\Phi)$ be the maximum ϵ-achievable rate. Take an $(2^R, \epsilon)$ code, i.e., an encoder $g : \Gamma \to \mathcal{X}$ and a measurement $\mathbf{m}^{\mathrm{B}} := \{m^{\mathrm{B}}_j\}_{j\in\Gamma}$ satisfying

$$\frac{1}{2^R} \sum_{j\in\Gamma} \big\langle u - m^{\mathrm{B}}_j, \sigma^{\mathrm{B}}_{g(j)} \big\rangle \leqslant \epsilon. \tag{A5}$$

To show inequality (20), we consider the concrete probability distribution P_X as a uniform distribution, and we define the bipartite state $\pi^{\mathrm{AB}}_{\mathrm{uni}}$ as

$$\pi^{\mathrm{AB}}_{\mathrm{uni}} := \frac{1}{2^R} \sum_{j\in\Gamma} |g(j)\rangle\langle g(j)|^{\mathrm{A}} \otimes \sigma^{\mathrm{B}}_{g(j)}. \tag{A6}$$

The marginal states of $\pi^{\mathrm{AB}}_{\mathrm{uni}}$ are given as

$$\pi^{\mathrm{A}}_{\mathrm{uni}} = \frac{1}{2^R} \sum_{j\in\Gamma} |g(j)\rangle\langle g(j)|^{\mathrm{A}}, \tag{A7}$$

$$\pi^{\mathrm{B}}_{\mathrm{uni}} = \frac{1}{2^R} \sum_{j\in\Gamma} \sigma^{\mathrm{B}}_{g(j)}. \tag{A8}$$

We will show inequality (20) by showing the inequality

$$R \leqslant D^{\epsilon}_{\mathrm{H,G}}\big(\pi^{\mathrm{AB}}_{\mathrm{uni}}\big|\big|\pi^{\mathrm{A}}_{\mathrm{uni}} \otimes \pi^{\mathrm{B}}_{\mathrm{uni}}\big) \tag{A9}$$

as discussed later.

Next, take a classical measurement $\{\lambda^{\mathrm{A}}_j\}_{j\in\Gamma}$ such that $\langle\lambda^{\mathrm{A}}_j, |g(j')\rangle\langle g(j')|\rangle = \delta_{j,j'}$. Then, we define a product measurement $\mathbf{m}' := \{\lambda_j \otimes m_{j'}\}_{j,j'\in\Gamma}$, and the measurement channel $\mathcal{E}_{\mathbf{m}'}$ associated with the product measurement, i.e., $\mathcal{E}_{\mathbf{m}'}$, is given as

$$\mathcal{E}_{\mathbf{m}'}(\rho) = \sum_{j,j'\in\Gamma} \langle \lambda_j \otimes m_{j'}, \rho\rangle\, |j\rangle\langle j| \otimes |j'\rangle\langle j'|\,. \tag{A10}$$

From Lemma 1, we obtain the following inequality:

$$\begin{aligned} &D^{\epsilon}_{\mathrm{H,G}}\big(\mathcal{E}_{\mathbf{m}'}\big(\pi^{\mathrm{AB}}_{\mathrm{uni}}\big)\big|\big|\mathcal{E}_{\mathbf{m}'}\big(\pi^{\mathrm{A}}_{\mathrm{uni}} \otimes \pi^{\mathrm{B}}_{\mathrm{uni}}\big)\big)\\ &\quad\leqslant D^{\epsilon}_{\mathrm{H,G}}\big(\pi^{\mathrm{AB}}_{\mathrm{uni}}\big|\big|\pi^{\mathrm{A}}_{\mathrm{uni}} \otimes \pi^{\mathrm{B}}_{\mathrm{uni}}\big). \end{aligned} \tag{A11}$$

To calculate the left-hand side of (A11), we define the classical state $P^{\mathrm{AB}} := \mathcal{E}_{\mathbf{m}'}(\pi^{\mathrm{AB}}_{\mathrm{uni}})$. Then, the marginal states of P^{AB} are given as

$$P^{\mathrm{A}} = \sum_{j,j'\in\Gamma} \big\langle \lambda_j \otimes m_{j'}, \pi^{\mathrm{AB}}_{\mathrm{uni}}\big\rangle\, |j\rangle\langle j|^{\mathrm{A}} \tag{A12}$$

$$= \sum_{j,\in\Gamma} \big\langle \lambda_j \otimes u, \pi^{\mathrm{AB}}_{\mathrm{uni}}\big\rangle\, |j\rangle\langle j|^{\mathrm{A}} \tag{A13}$$

$$= \sum_{j,\in\Gamma} \big\langle \lambda_j, \pi^{\mathrm{A}}_{\mathrm{uni}}\big\rangle\, |j\rangle\langle j|^{\mathrm{A}}\,, \tag{A14}$$

$$P^{\mathrm{B}} = \sum_{j,j'\in\Gamma} \big\langle \lambda_j \otimes m_{j'}, \pi^{\mathrm{AB}}_{\mathrm{uni}}\big\rangle\, |j'\rangle\langle j'|^{\mathrm{B}} \tag{A15}$$

$$= \sum_{j',\in\Gamma} \big\langle u \otimes m_{j'}, \pi^{\mathrm{AB}}_{\mathrm{uni}}\big\rangle\, |j'\rangle\langle j'|^{\mathrm{B}} \tag{A16}$$

$$= \sum_{j',\in\Gamma} \big\langle m_{j'}, \pi^{\mathrm{B}}_{\mathrm{uni}}\big\rangle\, |j'\rangle\langle j'|^{\mathrm{B}}\,. \tag{A17}$$

With these forms of marginal states P^{A} and P^{B}, we obtain the following relation:

$$P^{\mathrm{A}} \otimes P^{\mathrm{B}} = \sum_{j,j'\in\Gamma} \big\langle \lambda_j, \pi^{\mathrm{A}}_{\mathrm{uni}}\big\rangle\big\langle m_{j'}, \pi^{\mathrm{B}}_{\mathrm{uni}}\big\rangle\, |j\rangle\langle j|^{\mathrm{A}} \otimes |j'\rangle\langle j'|^{\mathrm{B}} \tag{A18}$$

$$= \sum_{j,j'\in\Gamma} \big\langle \lambda_j \otimes m_{j'}, \pi^{\mathrm{A}}_{\mathrm{uni}} \otimes \pi^{\mathrm{B}}_{\mathrm{uni}}\big\rangle\, |j\rangle\langle j|^{\mathrm{A}} \otimes |j'\rangle\langle j'|^{\mathrm{B}} \tag{A19}$$

$$= \mathcal{E}_{\mathbf{m}'}\big(\pi^{\mathrm{A}}_{\mathrm{uni}} \otimes \pi^{\mathrm{B}}_{\mathrm{uni}}\big). \tag{A20}$$

Therefore, to show the desirable inequality (A9), we need to show the inequality

$$R \leqslant D^{\epsilon}_{\mathrm{H,G}}(P^{\mathrm{AB}}||P^{\mathrm{A}} \otimes P^{\mathrm{B}}), \tag{A21}$$

which will be shown below.

To estimate the value $D^{\epsilon}_{\mathrm{H,G}}(P^{\mathrm{AB}}||P^{\mathrm{A}} \otimes P^{\mathrm{B}})$, take a concrete classical two-outcome measurement $\{q, u-q\}$ to distinguish classical states P^{AB} and $P^{\mathrm{A}} \otimes P^{\mathrm{B}}$ such that $q = (q_i)_{i\in\Gamma^2} \in \mathbb{R}^{\Gamma^2}$ and $q_i = \delta_{j,j'}$ for $i = (j, j')$. Then, the type-I error probability $\langle q, P^{\mathrm{AB}}\rangle$ is given as

$$\langle q, P^{\mathrm{AB}}\rangle = \sum_{j=j'} \big\langle \lambda_j \otimes m_{j'}, \pi^{\mathrm{AB}}_{\mathrm{uni}}\big\rangle \tag{A22}$$

$$= \sum_{j\in\Gamma}\sum_{k\in\Gamma} \frac{1}{2^R} \langle \lambda_j, |g(k)\rangle\langle g(k)|^{\mathrm{A}}\rangle\big\langle m_j, \sigma^{\mathrm{B}}_{g(k)}\big\rangle \tag{A23}$$

$$\overset{(a)}{=} \frac{1}{2^R}\sum_{j\in\Gamma}\langle m_j, \sigma_{g(j)}\rangle \tag{A24}$$

$$= 1 - \frac{1}{2^R}\sum_{j\in\Gamma}\langle u - m_j, \sigma_{g(j)}\rangle \tag{A25}$$

$$\overset{(b)}{\geqslant} 1 - \epsilon. \tag{A26}$$

The equality (a) holds because $\langle\lambda_j, |g(k)\rangle\langle g(k)| = \delta_{j,k}$. The inequality (b) holds because of inequality (A5). Also, the type-II error probability $\langle q, P^{\mathrm{A}} \otimes P^{\mathrm{B}}\rangle$ is given as

$$\langle q, P^{\mathrm{A}} \otimes P^{\mathrm{B}}\rangle = \sum_{j=j'} \big\langle \lambda_j \otimes m_{j'}, \pi^{\mathrm{A}}_{\mathrm{uni}} \otimes \pi^{\mathrm{A}}_{\mathrm{uni}}\big\rangle \tag{A27}$$

$$= \sum_{j\in\Gamma} \big\langle \lambda_j \pi^{\mathrm{A}}_{\mathrm{uni}}\big\rangle\big\langle m_j \pi^{\mathrm{B}}_{\mathrm{uni}}\big\rangle \tag{A28}$$

$$\begin{aligned} &= \sum_{j\in\Gamma}\left(\frac{1}{2^R}\sum_{k\in\Gamma}\langle\lambda_j\, |g(k)\rangle\langle g(k)|^{\mathrm{A}}\rangle\right)\\ &\quad\times\left(\frac{1}{2^R}\sum_{k'\in\Gamma}\big\langle m_j, \sigma^{\mathrm{B}}_{k'}\big\rangle\right) \end{aligned} \tag{A29}$$

$$\begin{aligned} &\overset{(a)}{=} \frac{1}{2^{2R}}\sum_{j\in\Gamma}\sum_{k'\in\Gamma}\big\langle m_j, \sigma^{\mathrm{B}}_{k'}\big\rangle\\ &= \frac{1}{2^{2R}}\sum_{k'\in\Gamma}\big\langle u, \sigma^{\mathrm{B}}_{k'}\big\rangle = \frac{1}{2^R}. \end{aligned} \tag{A30}$$

The equality (a) holds because $\langle\lambda_j, |g(k)\rangle\langle g(k)| = \delta_{j,k}$. Relations (A22) and (A27) imply inequality (A21). As a result, we obtain inequality (20) as follows:

$$\mathrm{C}^{\epsilon}(\Phi) = R \leqslant D^{\epsilon}_{\mathrm{H,G}}(P^{\mathrm{AB}}||P^{\mathrm{A}} \otimes P^{\mathrm{B}}) \tag{A31}$$

$$= D^{\epsilon}_{\mathrm{H,G}}\big(\mathcal{E}_{\mathbf{m}'}\big(\pi^{\mathrm{AB}}_{\mathrm{uni}}\big)\big|\big|\mathcal{E}_{\mathbf{m}'}\big(\pi^{\mathrm{A}}_{\mathrm{uni}} \otimes \pi^{\mathrm{B}}_{\mathrm{uni}}\big)\big) \tag{A32}$$

$$\leqslant D^{\epsilon}_{\mathrm{H,G}}\big(\pi^{\mathrm{AB}}_{\mathrm{uni}}\big|\big|\pi^{\mathrm{A}}_{\mathrm{uni}} \otimes \pi^{\mathrm{B}}_{\mathrm{uni}}\big) \tag{A33}$$

$$\leqslant \sup_{P_X} D^{\epsilon}_{\mathrm{H,G}}\big(\pi^{\mathrm{AB}}_{P_X}\big|\big|\pi^{\mathrm{A}}_{P_X} \otimes \pi^{\mathrm{B}}_{P_X}\big). \tag{A34}$$

Here, we remark that this proof is an operational generalization of the proof of the converse part given in Ref. [52].

3. Proof of Theorem 2

First, we prove Lemma 2.

Proof of Lemma 2. Let us denote $u' := \sum_{y\in\mathcal{Y}} A_y$. Also, let us define

$$B_y := -(s-1)u - tu' + (s+t)A_y \tag{A35}$$

for each $y \in \mathcal{Y}$. Then the desired inequality (22) is rewritten as $B_y \leqslant E_y$.

Next, we show that at most one element $y_0 \in \mathcal{Y}$ satisfying $B_{y_0} \geqslant 0$ by contradiction exists. Assume that two different elements $y_0, y_1 \in \mathcal{Y}$ satisfy $B_{y_0} \geqslant 0$ and $B_{y_1} \geqslant 0$, respectively. Then we have the following two inequalities:

$$(s+t)A_{y_0} \geqslant (s-1)u + tu', \tag{A36}$$

$$(s+t)A_{y_1} \geqslant (s-1)u + tu'. \tag{A37}$$

Therefore, we have

$$(s+t)u' \geqslant (s+t)\big(A_{y_0} + A_{y_1}\big) \tag{A38}$$
$$\geqslant 2(s-1)u + 2tu' \tag{A39}$$
$$> (s+t)u'. \tag{A40}$$

The first inequality is from $A_{y_0} + A_{y_1} \leqslant u'$. The second inequality is from inequalities (A36) and (A37). The final inequality holds because $s > 1$ and $t > s$. However, this is a contradicting relation, and thus, at most one y_0 such that $B_{y_0} \geqslant 0$ exits.

Finally, we define a measurement. Let $y_0 \in \mathcal{Y}$ be a unique element satisfying $B_{y_0} \geqslant 0$ if it exists. If such a y_0 does not exist, we choose a fixed element $y_0 \in \mathcal{Y}$. Then, we define a measurement $\{E_y\}_{y\in\mathcal{Y}}$ as

$$E_y := 0 \quad (y \neq y_0),$$
$$E_y := u \quad (y = y_0). \tag{A41}$$

As discussed above, $B_i < 0$ holds for all i except y_0, and therefore, $B_i < E_i$ holds for any $i \neq y_0$. Therefore, what we need to show is $B_{y_0} \leqslant E_{y_0}$, which is shown as follows:

$$E_{y_0} - B_{y_0} = u - [-(s-1)u - tu' + (s+t)A_{y_0}] \tag{A42}$$
$$= s\big(u - A_{y_0}\big) + t\big(u' - A_{y_0}\big) \geqslant 0. \tag{A43}$$

As a result, the measurement defined as (A41) satisfies $B_y \leqslant E_y$ for all $y \in \mathcal{Y}$, and thus, the desired inequality (22) holds. ■

Now, we prove Theorem 2 by applying Lemma 2.

Proof of Theorem 2. The structure of this proof is also essentially the same as the method given in Ref. [52].

To show inequality (21), we fix an arbitrary parameter $\epsilon' \in (0, \epsilon)$ and an arbitrary probability distribution $P_X(x)$. Then, we take an effect e that achieves the optimization $D^{\epsilon'}_{\mathrm{H,G}}(\pi^{\mathrm{AB}}_{P_X} || \pi^{\mathrm{A}}_{P_X} \otimes \pi^{\mathrm{B}}_{P_X})$. Therefore, the effect e satisfies $0 \leqslant e \leqslant u$ and $\langle e, \pi^{\mathrm{AB}}_{P_X}\rangle \geqslant 1 - \epsilon'$. Then, we need to show the existence of a $(2^R, \epsilon)$ code satisfying

$$R \geqslant -\log_2 \big\langle e, \pi^{\mathrm{A}}_{P_X} \otimes \pi^{\mathrm{B}}_{P_X}\big\rangle - \log_2 \frac{t}{\epsilon - s\epsilon'}, \tag{A44}$$

which can be rewritten as

$$\epsilon \leqslant s\epsilon' + 2^R t\big\langle e, \pi^{\mathrm{A}}_{P_X} \otimes \pi^{\mathrm{B}}_{P_X}\big\rangle. \tag{A45}$$

To show this, we need to show the existence of the code with size R, i.e., the existence of a measurement $\mathbf{m}^B$ such that

$$\Pr(\mathrm{error}|\mathbf{m}^B) \leqslant s\epsilon' + 2^R t\big\langle e, \pi^{\mathrm{A}}_{P_X} \otimes \pi^{\mathrm{B}}_{P_X}\big\rangle. \tag{A46}$$

Now, we consider the situation in which we send an R-length bit string $j \in \Gamma$. We generate an encoder $g : \Gamma \to \mathcal{X}$ by choosing a codeword $g(j) = x_j \in \mathcal{X}$ at random, where each x_j is chosen according to the distribution P_X independently. Also, let us define $0 \leqslant A_x \leqslant u$ as an effect such that

$$\langle A_x, \rho^{\mathrm{B}}\rangle = \langle e, |x\rangle\langle x|^{\mathrm{A}} \otimes \rho^{\mathrm{B}}\rangle \tag{A47}$$

for any $\rho^{\mathrm{B}} \in \mathcal{S}(K^{\mathrm{B}})$. By applying Lemma 2, we choose a decoding measurement $\mathbf{m}^{\mathrm{B}} := \{m^{\mathrm{B}}_j\}_{j\in\Gamma}$ for which the following inequality holds for any $s > 1$ and $t > s$:

$$u - m^{\mathrm{B}}_j \leqslant s\big(u - A_{g(j)}\big) + t\sum_{\substack{i\in\Gamma\\ i\neq j}} A_{g(i)}. \tag{A48}$$

Therefore, an upper bound of the error probability with respect to a message $j \in \Gamma$ with an encoder g is bounded as follows:

$$\Pr(\mathrm{error}|j, g, \mathbf{m}^{\mathrm{B}}) \leqslant s\big\langle u - A_{g(j)}, \sigma^{\mathrm{B}}_{g(j)}\big\rangle + t\sum_{\substack{i\in\Gamma\\ i\neq j}} \big\langle A_{g(i)}, \sigma^{\mathrm{B}}_{g(j)}\big\rangle. \tag{A49}$$

Next, for an arbitrary fixed bit string j, we consider the average of the error probability $\Pr(\mathrm{error}|j, g, \mathbf{m}^{\mathrm{B}})$ over all the encoders g. We denote the probability to generate g and the average value as $P(g)$ and $\Pr(\mathrm{error}|j, \overline{g}, \mathbf{m}^{\mathrm{B}})$, respectively. Also, we denote the set of all encoders and the set of encoders satisfying $g(j) = x$ as G and G_x, respectively. Then, the value $\Pr(\mathrm{error}|j, \overline{g}, \mathbf{m}^{\mathrm{B}})$ is bounded as follows due to the method to generate g:

$$\Pr(\mathrm{error}|j, \overline{g}, \mathbf{m}^{\mathrm{B}}) := \sum_{g\in G} P(g)\Pr(\mathrm{error}|j, g, \mathbf{m}^{\mathrm{B}}) \tag{A50}$$

$$\overset{(a)}{\leqslant} \sum_{g\in G} P(g)\left(s\big\langle u - A_{g(j)}, \sigma^{\mathrm{B}}_{g(j)}\big\rangle + t\sum_{\substack{i\in\Gamma\\ i\neq j}} \big\langle A_{g(i)}, \sigma^{\mathrm{B}}_{g(j)}\big\rangle\right) \tag{A51}$$

$$= \sum_{x\in\mathcal{X}}\sum_{g\in G_x} P(g)\left(s\big\langle u - A_{g(j)}, \sigma^{\mathrm{B}}_{g(j)}\big\rangle + t\sum_{\substack{i\in\Gamma\\ i\neq j}} \big\langle A_{g(i)}, \sigma^{\mathrm{B}}_{g(j)}\big\rangle\right) \tag{A52}$$

$$\overset{(b)}{=} \sum_{x\in\mathcal{X}}\sum_{g\in G_x} P(g)s\big\langle u - A_x, \sigma^{\mathrm{B}}_x\big\rangle + \sum_{x\in\mathcal{X}}\sum_{g\in G_x} P(g)t\sum_{\substack{i\in\Gamma\\ i\neq j}} \big\langle A_{g(i)}, \sigma^{\mathrm{B}}_x\big\rangle \tag{A53}$$

$$\overset{(c)}{=} \sum_{x\in\mathcal{X}} P_X(x)s\langle u - A_x, \sigma_x^{\mathrm{B}}\rangle + \sum_{x\in\mathcal{X}}\sum_{g\in G_x} P(g)t \sum_{\substack{i\in\Gamma\\ i\neq j}} \langle A_{g(i)}, \sigma_x^{\mathrm{B}}\rangle \tag{A54}$$

$$\overset{(d)}{=} \sum_{x\in\mathcal{X}} P_X(x)s\langle u - A_x, \sigma_x^{\mathrm{B}}\rangle + \sum_{x\in\mathcal{X}} t \sum_{\substack{i\in\Gamma\\ i\neq j}} \left\langle P_X(x)\sum_{x'\in\mathcal{X}} P_X(x')A_{x'}, \sigma_x^{\mathrm{B}}\right\rangle \tag{A55}$$

$$\overset{(e)}{=} s\left(1 - \sum_{x\in\mathcal{X}} P_X(x)\langle A_x, \sigma_x^{\mathrm{B}}\rangle\right) + t(2^R - 1)\left\langle \sum_{x'\in\mathcal{X}} P_X(x')A_{x'}, \sum_{x\in\mathcal{X}} P_X(x)\sigma_x^{\mathrm{B}}\right\rangle \tag{A56}$$

$$\overset{(f)}{\leqslant} s\left(1 - \sum_{x\in\mathcal{X}} P_X(x)\langle A_x, \sigma_x^{\mathrm{B}}\rangle\right) + t2^R\left\langle \sum_{x'\in\mathcal{X}} P_X(x')A_{x'}, \sum_{x\in\mathcal{X}} P_X(x)\sigma_x^{\mathrm{B}}\right\rangle. \tag{A57}$$

The inequality (a) is derived from (A49). The equality (b) is derived from the definition of G_x. The equality (c) holds because the first term is independent of g and the average over $g \in G_x$ is given by $P_X(x)$ due to the method to generate g. Because each codeword $g(i)$ is chosen at random with probability P_X independent of the choice of $g(j) = x$ in the second term, the average of $A_{g(i)}$ over $g \in G_x$ is given as $P_X(x)\sum_{x'\in\mathcal{X}} P_X(x')A_{x'}$. Therefore, the equality (d) holds. The equality (e) holds because the second term is independent of j. The inequality (f) is shown from the trivial inequality $2^R - 1 \leqslant 2^R$ and the positivity of the inner product.

By applying the above upper bound, the average of $\Pr(\text{error}|g, \mathbf{m}^{\mathrm{B}})$ over all of the encoders g is bounded as follows:

$$\begin{aligned}\Pr(\text{error}|\overline{g}, \mathbf{m}^{\mathrm{B}}) &:= \sum_{g\in G} P(g) \sum_j \frac{1}{2^R}\Pr(\text{error}|j, g, \mathbf{m}^{\mathrm{B}})\\ &= \sum_j \frac{1}{2^R}\sum_{g\in G} P(g)\Pr(\text{error}|j, g, \mathbf{m}^{\mathrm{B}})\\ &\leqslant \sum_j \frac{1}{2^R}\left[s\left(1 - \sum_{x\in\mathcal{X}} P_X(x)\langle A_x, \sigma_x^{\mathrm{B}}\rangle\right)\right.\\ &\quad \left. + t2^R\left\langle \sum_{x'\in\mathcal{X}} P_X(x')A_{x'}, \sum_{x\in\mathcal{X}} P_X(x)\sigma_x^{\mathrm{B}}\right\rangle\right]\\ &\overset{(a)}{=} s\left(1 - \sum_{x\in\mathcal{X}} P_X(x)\langle A_x, \sigma_x^{\mathrm{B}}\rangle\right)\\ &\quad + t2^R\left\langle \sum_{x'\in\mathcal{X}} P_X(x')A_{x'}, \sum_{x\in\mathcal{X}} P_X(x)\sigma_x^{\mathrm{B}}\right\rangle.\end{aligned} \tag{A58}$$

The equality (a) holds because the summation part is independent of j.

Finally, by using Eq. (A47) and the linearity of the inner product, we have

$$\begin{aligned}\sum_{x\in\mathcal{X}} P_X(x)\langle A_x, \sigma_x^{\mathrm{B}}\rangle &= \sum_{x\in\mathcal{X}} P_X(x)\langle e, |x\rangle\langle x|^{\mathrm{A}} \otimes \sigma_x^{\mathrm{B}}\rangle\\ &= \left\langle e, \sum_{x\in\mathcal{X}} P_X(x)\,|x\rangle\langle x|^{\mathrm{A}} \otimes \sigma_x^{\mathrm{B}}\right\rangle\\ &= \langle e, \pi_{P_X}^{\mathrm{AB}}\rangle \geqslant 1 - \epsilon'.\end{aligned} \tag{A59}$$

Also, we have

$$\begin{aligned}&\left\langle \sum_{x'} P_X(x')A_{x'}, \sum_x P_X(x)\sigma_x^{\mathrm{B}}\right\rangle\\ &= \sum_{x'} P_X(x')\left\langle A_{x'} \sum_x P_X(x)\sigma_x^{\mathrm{B}}\right\rangle\\ &\overset{(a)}{=} \sum_{x'} P_X(x')\left\langle e, |x'\rangle\langle x'|^{\mathrm{A}} \otimes \sum_x P_X(x)\sigma_x^{\mathrm{B}}\right\rangle\\ &= \left\langle e, \sum_{x'} P_X(x')\,|x'\rangle\langle x'|^{\mathrm{A}} \otimes \sum_x P_X(x)\sigma_x^{\mathrm{B}}\right\rangle\\ &= \langle e, \pi_{P_X}^{\mathrm{A}} \otimes \pi_{P_X}^{\mathrm{B}}\rangle.\end{aligned} \tag{A60}$$

The equality (a) is derived directly from (A47). Substituting (A59) and (A60) for (A58), we obtain the inequality

$$\Pr(\text{error}|\overline{g}, \mathbf{m}^B) \leqslant s\epsilon' + 2^R t\langle e, \pi_{P_X}^{\mathrm{A}} \otimes \pi_{P_X}^{\mathrm{B}}\rangle. \tag{A61}$$

Due to the same logic as Shannon's random encoding, at least one encoder and decoder satisfying the desired inequality (A46) exist. As a result, we show the existence of the $(2^R, \epsilon)$ code satisfying (A44), which implies the statement of Theorem 2. ■

4. Proof of Theorem 3

The remainder of the proof of Theorem 3 is the following lemma.

Lemma 3. For any states $\rho, \sigma \in \mathcal{S}(K)$ and $\epsilon > 0$, the ϵ-hypothesis-testing relative entropy is left continuous over ϵ.

Proof of Lemma 3 Due to the definition (12), we need to show only the continuity of the function $D(\epsilon)$, defined as

$$D(\epsilon) := \min_{\substack{q:\, 0\leqslant q\leqslant u,\\ \langle q,\rho\rangle\geqslant 1-\epsilon}} \langle q, \sigma\rangle \tag{A62}$$

for any $\rho, \sigma \in \mathcal{S}(K)$. Because the effect $(1-\epsilon)u$ satisfies the condition of the minimization in (A62), $0 \leqslant D(\epsilon) \leqslant 1 - \epsilon$ holds. Let $\eta > 0$ be an arbitrary parameter for the so-called $\epsilon - \delta$ discussion. We need to show only the existence of δ for any $0 < \eta < 1 - D(\epsilon)$ such that any $\epsilon' < \epsilon$ with $\epsilon - \epsilon' < \delta$ satisfies $|D(\epsilon) - D(\epsilon')| < \eta$. We take an argument-minimum effect $q_0 := \operatorname{argmin} D(\epsilon)$, and we choose δ to be $\delta := \frac{\epsilon\eta}{1-D(\epsilon)}$.

In addition, we take an effect q_1 as

$$q_1 := \frac{\delta}{\epsilon}u + \left(1 - \frac{\delta}{\epsilon}\right)q_0. \tag{A63}$$

Because $0 < \frac{\delta}{\epsilon} \leqslant 1$, the effect q_1 satisfies $0 \leqslant q_1 \leqslant u$ and the following inequality:

$$\begin{aligned}\langle q_1, \rho\rangle =&\left(1 - \frac{\delta}{\epsilon}\right)\langle q_0, \rho\rangle + \frac{\delta}{\epsilon}\langle u, \rho\rangle\\ \geqslant&\left(1 - \frac{\delta}{\epsilon}\right)(1-\epsilon) + \frac{\delta}{\epsilon}\\ =&1 - \epsilon + \delta \geqslant 1 - \epsilon'. \end{aligned}\tag{A64}$$

Then, we obtain the following inequality:

$$|D(\epsilon) - D(\epsilon')| \overset{(a)}{=} D(\epsilon') - D(\epsilon) \tag{A65}$$

$$= \min_{\substack{q' : 0 \leqslant q' \leqslant u,\\ \langle q', \rho\rangle \geqslant 1-\epsilon'}} \langle q', \sigma\rangle - D(\epsilon) \tag{A66}$$

$$\overset{(b)}{\leqslant} \langle q_1, \sigma\rangle - D(\epsilon) \tag{A67}$$

$$= \frac{\delta}{\epsilon}[1 - D(\epsilon)] = \eta. \tag{A68}$$

The equality (a) holds because $\epsilon' < \epsilon$. The inequality (b) holds because of inequality (A64). As a result, $D(\epsilon)$ is left continuous, and therefore, ϵ-hypothesis-testing relative entropy is also left continuous over ϵ. ■

[1] C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. **27**, 379 (1948).

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Wiley Series in Telecommunications and Signal Processing (Wiley, New York, 2006).

[3] S. Verdú and T. S. Han, A general formula for channel capacity, IEEE Trans. Inf. Theory **40**, 1147 (1994).

[4] M. M. Wilde, *Quantum Information Theory*, 2nd ed. (Cambridge University Press, Cambridge, 2017).

[5] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, 2018).

[6] F. Hiai and D. Petz, The proper formula for relative entropy and its asymptotics in quantum probability, Commun. Math. Phys. **143**, 99 (1991).

[7] T. Ogawa and H. Nagaoka, Strong converse and Stein's lemma in quantum hypothesis testing, in *Asymptotic Theory of Quantum Statistical Inference: Selected Papers* (World Scientific, Singapore, 2005), pp. 28–42.

[8] B. Schumacher, Quantum coding, Phys. Rev. A **51**, 2738 (1995).

[9] B. Schumacher and M. D. Westmoreland, Sending classical information via noisy quantum channels, Phys. Rev. A **56**, 131 (1997).

[10] A. S. Holevo, The capacity of the quantum channel with general signal states, IEEE Trans. Inf. Theory **44**, 269 (1998).

[11] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem, IEEE Trans. Inf. Theory **48**, 2637 (2002).

[12] P. Janotta and H. Hinrichsen, Generalized probability theories: What determines the structure of quantum theory? J. Phys. A **47**, 323001 (2014).

[13] G. M. D'Ariano, G. Chiribella, and P. Perinotti, *Quantum Theory from First Principles: An Informational Approach* (Cambridge University Press, Cambridge, 2017).

[14] M. Plávala, General probabilistic theories: An introduction, Phys. Rep. **1033**, 1 (2023).

[15] M. P. Müller, Probabilistic theories and reconstructions of quantum theory, SciPost Phys. Lect. Notes, 28 (2021).

[16] G. Ludwig, Versuch einer axiomatischen grundlegung der quantenmechanik und allgemeinerer physikalischer theorien, Z. Angew. Phys. **181**, 233 (1964).

[17] G. Ludwig, Attempt of an axiomatic foundation of quantum mechanics and more general theories, II, Commun. Math. Phys. **4**, 331 (1967).

[18] E. B. Davies and J. T. Lewis, An operational approach to quantum probability, Commun. Math. Phys. **17**, 239 (1970).

[19] S. Gudder, Convex structures and operational quantum mechanics, Commun. Math. Phys. **29**, 249 (1973).

[20] M. Ozawa, Optimal measurements for general quantum systems, Rep. Math. Phys. **18**, 11 (1980).

[21] S. Popescu and D. Rohrlich, Quantum nonlocality as an axiom, Found. Phys. **24**, 379 (1994).

[22] L. Hardy, Quantum theory from five reasonable axioms, arXiv:quant-ph/0101012.

[23] J. Barrett, Information processing in generalized probabilistic theories, Phys. Rev. A **75**, 032304 (2007).

[24] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski, Information causality as a physical principle, Nature (London) **461**, 1101 (2009).

[25] G. Kimura, T. Miyadera, and H. Imai, Optimal state discrimination in general probabilistic theories, Phys. Rev. A **79**, 062306 (2009).

[26] A. J. Short and S. Wehner, Entropy in general physical theories, New J. Phys. **12**, 033023 (2010).

[27] H. Barnum, J. Barrett, L. O. Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, and R. Wilke, Entropy and information causality in general probabilistic theories, New J. Phys. **12**, 033024 (2010).

[28] G. Kimura, K. Nuida, and H. Imai, Distinguishability measures and entropies for general probabilistic theories, Rep. Math. Phys. **66**, 175 (2010).

[29] K. Nuida, G. Kimura, and T. Miyadera, Optimal observables for minimum-error state discrimination in general probabilistic theories, J. Math. Phys. **51**, 093505 (2010).

[30] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Probabilistic theories with purification, Phys. Rev. A **81**, 062348 (2010).

[31] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Informational derivation of quantum theory, Phys. Rev. A **84**, 012311 (2011).

[32] M. P. Müller and C. Ududec, Structure of reversible computation determines the self-duality of quantum theory, Phys. Rev. Lett. **108**, 130401 (2012).

[33] H. Barnum, M. P. Müller, and C. Ududec, Higher-order interference and single-system postulates characterizing quantum theory, New J. Phys. **16**, 123029 (2014).
[34] G. Chiribella and C. M. Scandolo, Entanglement and thermodynamics in general probabilistic theories, New J. Phys. **17**, 103027 (2015).
[35] G. Chiribella and C. M. Scandolo, Operational axioms for diagonalizing states, Electr. Proc. Theor. Comput. Sci. **195**, 96 (2015).
[36] G. Kimura, J. Ishiguro, and M. Fukui, Entropies in general probabilistic theories and their application to the Holevo bound, Phys. Rev. A **94**, 042113 (2016).
[37] M. Krumm, H. Barnum, J. Barrett, and M. P. Müller, Thermodynamics and the structure of quantum theory, New J. Phys. **19**, 043025 (2017).
[38] A. Jenčová, Incompatible measurements in a class of general probabilistic theories, Phys. Rev. A **98**, 012133 (2018).
[39] H. Arai, Y. Yoshida, and M. Hayashi, Perfect discrimination of non-orthogonal separable pure states on bipartite system in general probabilistic theory, J. Phys. A **52**, 465304 (2019).
[40] Y. Yoshida, H. Arai, and M. Hayashi, Perfect discrimination in approximate quantum theory of general probabilistic theories, Phys. Rev. Lett. **125**, 150402 (2020).
[41] F. Shahandeh, Contextuality of general probabilistic theories, PRX Quantum **2**, 010330 (2021).
[42] J. H. Selby, D. Schmid, E. Wolfe, A. B. Sainz, R. Kunjwal, and R. W. Spekkens, Accessible fragments of generalized probabilistic theories, cone equivalence, and applications to witnessing nonclassicality, Phys. Rev. A **107**, 062203 (2023).
[43] E. Wakakuwa, Gentle measurement as a principle of quantum theory, arXiv:2103.15110.
[44] P. Perinotti, A. Tosini, and L. Vaglini, Shannon theory beyond quantum: Information content of a source, Phys. Rev. A **105**, 052222 (2022).
[45] G. Aubrun, L. Lami, C. Palazuelos, and M. Plávala, Entangleability of cones, Geom. Funct. Anal. **31**, 181 (2021).
[46] G. Aubrun, L. Lami, C. Palazuelos, and M. Plávala, Entanglement and superposition are equivalent concepts in any physical theory, Phys. Rev. Lett. **128**, 160402 (2022).
[47] A. Jenčová, Assemblages and steering in general probabilistic theories, J. Phys. A **55**, 434001 (2022).
[48] S. Minagawa, H. Arai, and F. Buscemi, Von Neumann's information engine without the spectral theorem, Phys. Rev. Res. **4**, 033091 (2022).
[49] P. Perinotti, A. Tosini, and L. Vaglini, Which entropy for general physical theories? arXiv:2302.01651.
[50] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, Cloning and broadcasting in generic probabilistic theories, arXiv:quant-ph/0611295.
[51] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, NJ, 1955).
[52] L. Wang and R. Renner, One-shot classical-quantum capacity and hypothesis testing, Phys. Rev. Lett. **108**, 200501 (2012).
[53] This quantity was first introduced in Ref. [54], and the relationship to one-shot classical capacity was derived in Ref. [52], but we generalize it to GPTs in this paper.
[54] F. Buscemi and N. Datta, The quantum capacity of channels with arbitrarily correlated noise, IEEE Trans. Inf. Theory **56**, 1447 (2010).
[55] F. Buscemi and N. Datta, Distilling entanglement from arbitrary resources, J. Math. Phys. **51**, 102201 (2010).
[56] F. Dupuis, L. Kraemer, P. Faist, J. M. Renes, and R. Renner, Generalized entropies, in *XVIIth International Congress on Mathematical Physics* (World Scientific, Singapore, 2014), pp. 134–153.
[57] M. Hayashi, *Quantum Information Theory Mathematical Foundation* (Springer, Berlin, 2017).
[58] The previous work [26] does not explicitly define hypothesis-testing relative entropy in GPTs like (12). However, our definition follows the argument in Ref. [26]. Indeed, if we take the logarithm of Eq. (31) in Ref. [26] and subtract, we get hypothesis-testing relative entropy with $\epsilon = 1/2$. Our main focuses are the properties and applications of the explicit definition of hypothesis-testing relative entropy in GPTs.
[59] N. Datta, M. Mosonyi, M.-H. Hsieh, and F. G. Brandao, A smooth entropy approach to quantum hypothesis testing and the classical capacity of quantum channels, IEEE Trans. Inf. Theory **59**, 8014 (2013).
[60] P. Hausladen and W. K. Wootters, A 'pretty good' measurement for distinguishing quantum states, J. Mod. Opt. **41**, 2385 (1994).
[61] P. Janotta and R. Lal, Generalized probabilistic theories without the no-restriction hypothesis, Phys. Rev. A **87**, 052131 (2013).
[62] H. Umegaki, On information in operator algebras, Proc. Jpn. Acad. **37**, 459 (1961).
[63] H. Umegaki, Conditional expectation in an operator algebra, IV (entropy and information), Kodai Math. J. **14**, 59 (1962).
[64] The argument for this proof of the monotonicity is similar to that proposed in Ref. [52], where monotonicity in general completely positive trace-preserving maps was proven in quantum theory. In GPTs, however, we prove the monotonicity only in measurement channels, which is sufficient for this paper's discussion.