Quantum Dichotomies and Coherent Thermodynamics beyond First-Order Asymptotics

Patryk Lipka-Bartosik,[1,†] Christopher T. Chubb,[2,*,†] Joseph M. Renes,[2] Marco Tomamichel,[3,4] and Kamil Korzekwa[5]

[1] *Department of Applied Physics, University of Geneva, Geneva 1211, Switzerland*
[2] *Institute for Theoretical Physics, ETH Zurich, Zürich 8093, Switzerland*
[3] *Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583, Republic of Singapore*
[4] *Centre for Quantum Technologies, National University of Singapore, Singapore 117543, Republic of Singapore*
[5] *Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Kraków 30-348, Poland*

(Received 19 April 2023; accepted 8 March 2024; published 15 May 2024)

We address the problem of exact and approximate transformation of quantum dichotomies in the asymptotic regime, i.e., the existence of a quantum channel $\mathcal{E}$ mapping $\rho_1^{\otimes n}$ into $\rho_2^{\otimes R_n n}$ with an error ϵ_n (measured by trace distance) and $\sigma_1^{\otimes n}$ into $\sigma_2^{\otimes R_n n}$ exactly, for a large number n. We derive second-order asymptotic expressions for the optimal transformation rate R_n in the small-, moderate-, and large-deviation error regimes, as well as the zero-error regime, for an arbitrary pair (ρ_1, σ_1) of initial states and a commuting pair (ρ_2, σ_2) of final states. We also prove that for σ_1 and σ_2 given by thermal Gibbs states, the derived optimal transformation rates in the first three regimes can be attained by thermal operations. This allows us, for the first time, to study the second-order asymptotics of thermodynamic state interconversion with fully general initial states that may have coherence between different energy eigenspaces. Thus, we discuss the optimal performance of thermodynamic protocols with coherent inputs and describe three novel resonance phenomena allowing one to significantly reduce transformation errors induced by finite-size effects. What is more, our result on quantum dichotomies can also be used to obtain, up to second-order asymptotic terms, optimal conversion rates between pure bipartite entangled states under local operations and classical communication.

DOI: 10.1103/PRXQuantum.5.020335

I. INTRODUCTION

A. Statistical inference

Statistical inference is a powerful tool that allows us to explain the inner workings of the physical world by using statistical models based on data that hold crucial information about reality. From scientific discoveries to technological advancements, statistical inference is the backbone of many fields that have shaped our world. This process begins by forming a hypothesis, constructing an appropriate model (often represented by a family of probability distributions), and testing it against observed data. The theoretical foundations of statistical inference provide a solid framework for many essential fields, such as statistical estimation [1–4], metrology [5–8], hypothesis testing [9–12], decision theory [13–16], and machine learning [17–20].

One of the central problems of the theory of statistical inference is to determine which statistical models are more informative, i.e., which probability distributions more accurately reflect reality [21–25]. Given two probability distributions, $\mathbf{p}_1$ and $\mathbf{p}_2$, that describe some property of the physical system (e.g., the probability of observing given energy in the spectrum of a hydrogen atom), we say that $\mathbf{p}_1$ is more informative than $\mathbf{p}_2$ when the latter can be obtained from the former by bistochastic processing. One can also imagine a more general situation in which the physical system depends on some hidden parameter and hence it can be described by multiple models, depending on the value of the hidden parameter (such a parameter can, e.g., specify whether the system is in or out of thermal equilibrium). One is then interested in quantifying how well a given collection of models describes the system in question. In the case in which the hidden parameter is binary, the system can be described with a pair

[†] Corresponding author: paper@christopherchubb.com
[*] These authors contributed equally to this work.

of probability distributions, $(\mathbf{p}, \mathbf{q})$. Now, imagine that we want to decide whether one pair, $(\mathbf{p}_1, \mathbf{q}_1)$, provides a better statistical model, i.e., is more informative, than another pair, $(\mathbf{p}_2, \mathbf{q}_2)$. We say that a pair of probability distributions, or a *dichotomy*, $(\mathbf{p}_1, \mathbf{q}_1)$, is more informative than $(\mathbf{p}_2, \mathbf{q}_2)$ when there exists stochastic processing that maps $\mathbf{p}_1$ into $\mathbf{p}_2$, while also mapping $\mathbf{q}_1$ into $\mathbf{q}_2$. When such processing exists, then the first dichotomy *relatively majorizes* the second [22], a property that can be characterized using the techniques of hypothesis testing [21].

Since the processes that underlie our physical observations are fundamentally quantum and given the recent rapid development of quantum technologies, it is natural to ask how the techniques of statistical inference translate into the quantum realm. This is the main focus of quantum statistical inference [26–30], a theoretical framework that forms the bedrock of quantum estimation theory [26,31–33], quantum sensing and metrology [34–39], quantum statistical mechanics [40–42], and quantum computing [43–45]. The main conceptual difference between the classical and quantum statistical inference is the fact that statistical models in quantum theory must be described by density operators rather than probability distributions. Therefore, the objects to be compared are *quantum dichotomies* denoted by (ρ, σ) for density operators ρ and σ. We say that the dichotomy (ρ_1, σ_1) is more informative than (ρ_2, σ_2) if there exists a quantum channel that jointly transforms ρ_1 into ρ_2 and σ_1 into σ_2. If such a channel exists, then the first dichotomy precedes the second one in the so-called Blackwell order [46,47]. Importantly, when the two density operators forming a quantum dichotomy commute, they can be simultaneously diagonalized and can thus be treated classically. This is not the case for noncommuting quantum dichotomies, in which case the inference task becomes genuinely quantum. This regime naturally leads to much richer behavior but is notoriously harder to characterize.

B. Quantum thermodynamics

Perhaps one of the most impressive applications of statistical inference is in the field of thermodynamics. Indeed, modern thermodynamics started from the realization that statistical models can effectively describe macroscopic processes such as flows of heat and its fluctuations [48,49], phase transitions [50,51], or the dynamics of chemical reactions [40,52]. These processes generally involve unfathomable numbers of degrees of freedom and therefore finding their complete description by solving the corresponding equations of motion is usually beyond reach. It is nowadays widely accepted that when the numbers of particles are large enough, one can use the techniques of statistical inference to build statistical models describing the physical system with an accuracy (or error) that increases (decreases) with the number of particles [53–55]. In the limit when the system of interest is composed of infinitely many particles (the so-called thermodynamic limit), the approximation errors vanish and all relevant macroscopic observables can be fully characterized using only few relevant quantities known as thermodynamic potentials, e.g., the (equilibrium) free energy [40].

The thermodynamic limit is a convenient mathematical idealization but it cannot be justified in many experimentally and theoretically relevant situations. More specifically, when one is interested in the evolution of finite-size systems, fluctuations of thermodynamic variables cannot be neglected and the behavior of the system depends on more than a single thermodynamic potential. This regime is hardly discussed in thermodynamic textbooks, as it often requires rather advanced mathematical techniques of asymptotic analysis. Interestingly, this regime is surprisingly rich and allows us to investigate, i.e., the fundamental irreversibility of thermodynamic transformations [56], which cannot be observed when working solely in the thermodynamic limit.

Some of the techniques developed within the framework of quantum statistical inference have recently been adapted to study (quantum) thermodynamic processes. This has led to the realization that, in an idealized model of thermodynamics known as the resource theory of thermal operations [57–66], a single quantity—the (quantum) nonequilibrium free energy—completely characterizes the optimal rates of all thermodynamic transformations [59]. This interpretation, however, is only valid in the thermodynamic limit of infinitely many copies of quantum systems. Despite many significant efforts, characterizing thermodynamic transformations for general quantum states beyond the thermodynamic limit has remained a central problem for the resource theory of quantum thermodynamics. This difficulty can be easily understood once we realize that the techniques of statistical inference become accurate only when the numbers of particles are sufficiently large. On the other hand, it is known that quantum effects generally become less relevant with the increase in the size of systems, meaning that either the coherence per particle vanishes [67] or that the local observables begin to commute approximately when the system is comprised of a sufficient number of copies [68]. Therefore, a natural question arises: Can we use the tools of quantum statistical inference to gain new insights into the thermodynamics of genuinely quantum systems beyond the thermodynamic limit?

C. Summary of results

In this work, we develop a unified mathematical framework that allows us to compare the informativeness of quantum dichotomies up to second-order asymptotics (i.e., when the transformed dichotomies consist of a large number of identical and independent systems) and for various error regimes. Our results are applicable for arbitrary input dichotomies and commuting target dichotomies. This demonstrates, for the first time, how to compare quantum statistical models outside of the idealized limit of infinite repetitions of the experiments. Second, we apply our results on quantum dichotomies to study the fundamental laws governing thermodynamic transformations for large but finite numbers of particles. As a consequence, we characterize thermodynamic transformations of general energy-coherent input states outside of the thermodynamic limit. We observe that, in this regime, quantum systems can be fully characterized using only a few relevant quantities, in complete analogy with the classical case. Importantly, this shows that the second-order analysis is an especially interesting regime where statistical inference remains highly accurate, while the quantum nature of the thermodynamic process still plays a prominent role. To demonstrate this, we study, in full generality, the fundamental thermodynamic protocols such as work extraction, as well as quantifying the minimal free-energy dissipation when transforming quantum systems. We furthermore discover three novel resonance phenomena, the most interesting of which indicates that quantum coherence can be exploited to increase the reversibility of state transformations. Finally, we also discuss how our general results on quantum dichotomies can be used to bring novel and unifying insights into other fields, such as the theory of entanglement or coherence.

The paper is organized as follows. In Sec. II, we summarize the frameworks of quantum dichotomies (Sec. II A), as well as the resource theories of thermodynamics (Sec. II B) and entanglement (Sec. II C), and define some relevant information-theoretic notions used throughout the paper (Sec. II D). In Sec. III, we discuss our main results. In particular, after presenting an auxiliary lemma on sesquinormal distributions (Sec. III A), we outline our main technical results on quantum dichotomies (Sec. III B), quantum thermodynamics (Sec. III C), and entanglement (Sec. III D). In Sec. IV, we discuss some of the applications of our results to the thermodynamic and entanglement scenarios. In particular, we show how our results can be used to determine optimal thermodynamic protocols with coherent inputs (Sec. IV B), we investigate new types of resonance phenomena (Sec. IV C), and we briefly elaborate on the relevance of our results for entanglement theory (Sec. IV D). In Sec. V, we give proofs for the asymptotic results that we have described in previous sections. Specifically, we review and extend the relationship between quantum dichotomies and hypothesis testing (Sec. V A), present some of the results on hypothesis testing (Sec. V B), and prove the asymptotic transformation rates in different error regimes (Sec. V C). Finally, we finish with Sec. VI, which gives a short outlook on the potential further applications and extensions on our results. Technical derivations not required to understand the results are given in the Appendices A–I.

II. FRAMEWORK

We will denote by $\geq$ the Löwner partial order; i.e., for two Hermitian matrices A and B, the relation $A \geq B$ means that $A - B$ is positive semidefinite. To measure distance between two density matrices, ρ and σ, we will use trace distance $T(\rho,\sigma) := \frac{1}{2}\|\rho-\sigma\|_{\mathrm{tr}}$, where $\|X\|_{\mathrm{tr}} := \mathrm{Tr}|X|$ is the Schatten-1 norm. As a slight abuse of notation, we will also interchangeably refer to the total variation distance on classical distributions, $T(\boldsymbol{p},\boldsymbol{q}) := \frac{1}{2}\sum_i |p_i - q_i|$, as the trace distance. The fidelity between ρ and σ is given by $F(\rho,\sigma) := \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_{\mathrm{tr}}^2$. All states that we will consider are finite dimensional and we will denote the local dimension by d when relevant. We take $\exp(\cdot)$ and $\log(\cdot)$ to be in an arbitrary but compatible base and use $\ln(\cdot)$ to denote the natural logarithm.

A. Quantum dichotomies

For two quantum dichotomies, (ρ_1,σ_1) and (ρ_2,σ_2), we will be interested in whether there exists a completely positive trace-preserving map $\mathcal{E}$ such that $\rho_2 = \mathcal{E}(\rho_1)$ and $\sigma_2 = \mathcal{E}(\sigma_1)$. If such a channel exists, then we say that the first dichotomy precedes the second one in the Blackwell order [21], which we denote by $(\rho_1,\sigma_1) \succeq (\rho_2,\sigma_2)$ [69]. We further consider the concept of an approximate Blackwell order by requiring that the two states are only reproduced approximately by the channel. That is, we write $(\rho_1,\sigma_1) \succeq_{(\epsilon_\rho,\epsilon_\sigma)} (\rho_2,\sigma_2)$ if and only if there exists a quantum channel $\mathcal{E}$ such that

$$T(\mathcal{E}(\rho_1),\rho_2) \leq \epsilon_\rho \quad \text{and} \quad T(\mathcal{E}(\sigma_1),\sigma_2) \leq \epsilon_\sigma. \tag{1}$$

It is known that for commuting dichotomies, $[\rho_1,\sigma_1] = [\rho_2,\sigma_2] = 0$, the problem of determining a suitable channel reduces to the classical problem of comparing probability distributions. It has been observed in Ref. [70] that in this case, by employing Blackwell's equivalence theorem [21], one can show that $(\rho_1,\sigma_1) \succeq_{(\epsilon_\rho,\epsilon_\sigma)} (\rho_2,\sigma_2)$ if and only if

$$\beta_x(\rho_1\|\sigma_1) \leq \beta_{x-\epsilon_\rho}(\rho_2\|\sigma_2) + \epsilon_\sigma \quad \forall x \in (\epsilon_\rho, 1). \tag{2}$$

Here, $\beta_x(\rho\|\sigma)$ is the solution of the semidefinite optimization problem

$$\min_{Q} \quad \mathrm{Tr}(\sigma Q), \tag{3a}$$

$$\text{subject to} \quad 0 \leq Q \leq 1, \tag{3b}$$

$$\mathrm{Tr}(\rho Q) \geq 1 - x. \tag{3c}$$

The two quantities, x and $\beta_x(\rho\|\gamma)$, can be interpreted as two errors appearing in a binary hypothesis-testing problem. More specifically, $\beta_x(\rho\|\sigma)$ is the minimum type-II error given that the type-I error is upper bounded by x for a binary hypothesis testing with a null hypothesis ρ and an alternative hypothesis σ [26]. In the fully quantum case, i.e., when $[\rho_1, \sigma_1] \neq 0$ and $[\rho_2, \sigma_2] \neq 0$, the conditions specified by Eq. (2) (and referred to as relative majorization preorder in Ref. [70]) no longer characterize the Blackwell order [30,71–73], beyond the simplest case of two-dimensional density matrices [27]. For attempts to overcome this limitation, see, e.g., Refs. [30,74,75].

B. Resource theory of thermodynamics

In the resource-theoretic approach to thermodynamics, one focuses on a system S with a Hamiltonian $H = \sum_{i=1}^{d} E_i |i\rangle\langle i|$ and a heat bath B at some fixed inverse temperature β with an arbitrary Hamiltonian H_B [57,58]. The heat bath is always assumed to be prepared in a thermal Gibbs state,

$$\gamma_B = \frac{e^{-\beta H_B}}{Z_B}, \quad Z_B = \mathrm{Tr}\left(e^{-\beta H_B}\right). \tag{4}$$

The interaction of the system with the heat bath is mediated by a unitary U that conserves the total energy, i.e., obeys the additive conservation law $[U, H \otimes \mathbb{1}_B + \mathbb{1} \otimes H_B] = 0$. The effective map $\mathcal{E}$ that is obtained by evolving the system and the heat bath using unitary U and discarding part of the joint system is called a *thermal operation* (TO) and can be formally written as

$$\mathcal{E}(\rho) = \mathrm{Tr}_{B'}\left[U(\rho \otimes \gamma_B)\, U^\dagger\right], \tag{5}$$

where the partial trace can be performed over any subsystem B' of the joint system. Note that since we allow for $B' \neq B$, the Hamiltonian of the final system may differ from H and so we will use γ_1 and γ_2 to denote the Gibbs thermal states of the initial and final systems. We say that $\rho_1 \xrightarrow[\mathrm{TO}]{\epsilon} \rho_2$ when there exists a thermal operation $\mathcal{E}$ such that $\mathcal{E}(\rho_1) = \tilde{\rho}_2$, with $\tilde{\rho}_2$ being a final state that is ϵ-close to the target state ρ_2 in trace distance, i.e., $T(\tilde{\rho}_2, \rho_2) = \epsilon$.

Characterizing the set of transitions achievable via thermal operations in full generality remains an open problem. In the semiclassical case, i.e., when ρ_1 and ρ_2 are block diagonal in the energy eigenbasis (or, equivalently, when $[\rho_1, \gamma_1] = [\rho_2, \gamma_2] = 0$), the existence of a thermal operation transforming ρ_1 into ρ_2 while changing the Hamiltonian from H_1 to H_2 is equivalent to the existence of an *arbitrary* quantum channel mapping a quantum dichotomy (ρ_1, γ_1) into (ρ_2, γ_2) [58,62]. As a consequence, Blackwell's theorem in this case fully characterizes the set of states achievable under thermal operations [58]. More specifically, as observed in Ref. [70], for energy-incoherent (block-diagonal) states ρ_1 and ρ_2, we have $\rho_1 \xrightarrow[\mathrm{TO}]{\epsilon} \rho_2$ if and only if

$$\beta_x(\rho_1 \| \gamma_1) \leq \beta_{x-\epsilon}(\rho_2 \| \gamma_2) \quad \text{for all } x \in (\epsilon, 1). \tag{6}$$

The above condition is just a special case of Eq. (2) and thus we see that the problems of transforming quantum dichotomies and the thermodynamic state transformation are very closely related.

C. Resource theory of entanglement

The resource theory of entanglement investigates the scenario in which a bipartite system is distributed between two spatially separated agents [76]. The agents can act locally on their respective parts and can exchange classical information. The resulting set of free operations is called *local operations and classical communication* (LOCC). Free states of this theory, i.e., states that can be prepared using only LOCC, are given by all separable states. While a complete characterization of LOCC transformations for general mixed states remains an open problem, for pure states there exists a relatively simple characterization known as the Nielsen's theorem [77,78]. The theorem states that a pure bipartite state ψ_1 with Schmidt coefficients $\boldsymbol{p_1}$ can be converted into state ψ_2 with Schmidt coefficients $\boldsymbol{p_2}$ by means of LOCC if and only if there exists a bistochastic matrix mapping $\boldsymbol{p_2}$ to $\boldsymbol{p_1}$.

It has then been observed in Ref. [70] that Nielsen's theorem can be formulated in the language of quantum dichotomies when the Schmidt vectors of input and output states, $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$, have equal dimension. More specifically, by denoting with ρ_i diagonal matrices with $\boldsymbol{p}_i$ on the diagonals, the existence of a transformation that (1) maps ρ_1 to ρ_2 with a transformation error ϵ and (2) maps a maximally mixed state into itself, is equivalent to the existence of a bistochastic matrix mapping $\boldsymbol{p}_2$ into $\boldsymbol{p}_1$ with an error ϵ. Now, in Ref. [56] (see the generalization of Lemma 12 in Appendix D therein), it has been shown that the latter is equivalent to the existence of a bistochastic matrix mapping a distribution ϵ-close to $\boldsymbol{p}_2$ into $\boldsymbol{p}_1$. This means that an LOCC map transforming ψ_1 into ψ_2 with a transformation error ϵ exists if and only if a quantum dichotomy $(\rho_2, \mathbb{1}_d/d)$ can be approximately transformed into $(\rho_1, \mathbb{1}_d/d)$.

To deal with the case of systems with different lengths of Schmidt vectors, d_1 (for input) and d_2 (for output), one can extend the input system with a pure bipartite separable state with local dimensions d_2 and the output system with an analogous state with local dimensions d_1. Then, there exists an LOCC map transforming general pure bipartite state ψ_1 into ψ_2' the Schmidt vector $\boldsymbol{p_2'}$ of which is ϵ away in total variation distance from the Schmidt vector $\boldsymbol{p_2}$ of

ψ_2, if and only if

$$\left(\rho_2 \otimes |0\rangle\langle 0|_{d_1}, \frac{\mathbb{1}_{d_1 d_2}}{d_1 d_2}\right) \succeq_{(\epsilon,0)} \left(\rho_1 \otimes |0\rangle\langle 0|_{d_2}, \frac{\mathbb{1}_{d_1 d_2}}{d_1 d_2}\right). \tag{7}$$

Since states appearing in these dichotomies commute, Blackwell's theorem fully characterizes the states achievable under LOCC. More specifically, an LOCC transformation $\psi_1 \xrightarrow[\text{LOCC}]{\epsilon} \psi_2$ exists if and only for all $x \in (\epsilon, 1)$ one has

$$d_2 \beta_x \left(\rho_2 \middle\| \frac{\mathbb{1}_{d_2}}{d_2}\right) \leq d_1 \beta_{x-\epsilon} \left(\rho_1 \middle\| \frac{\mathbb{1}_{d_1}}{d_1}\right). \tag{8}$$

Of course, the above conditions are again a special case of Eq. (2).

D. Information-theoretic and statistical notions

To formulate our results, we will need the following notions. First, the von Neumann entropy and entropy variance are defined as

$$S(\rho) := -\mathrm{Tr}\left(\rho \log \rho\right), \tag{9a}$$
$$V(\rho) := \mathrm{Tr}\left(\rho (\log \rho)^2\right) - S(\rho)^2, \tag{9b}$$

and their relative cousins, the relative entropy [79] and the relative entropy variance [80,81], as

$$D(\rho\|\sigma) := (\mathrm{Tr}\rho\left(\log \rho - \log \sigma\right)), \tag{10a}$$
$$V(\rho\|\sigma) := \mathrm{Tr}\left(\rho\left(\log \rho - \log \sigma\right)^2\right) - D(\rho\|\sigma)^2. \tag{10b}$$

Note that for σ given by the thermal Gibbs state γ, the above quantities can be interpreted as nonequilibrium free energy [59] and free-energy fluctuations [56,82], respectively. We also define two variants of the Rényi relative entropy [83], namely, the Petz relative entropy $\overline{D}_\alpha$ [84] and the minimal relative entropy $\breve{D}_\alpha$ [85–88], which is

$$\overline{D}_\alpha(\rho\|\sigma) := \frac{\log \mathrm{Tr}\left(\rho^\alpha \sigma^{1-\alpha}\right)}{\alpha - 1}, \tag{11a}$$

$$\breve{D}_\alpha(\rho\|\sigma) := \begin{cases} \dfrac{\log \mathrm{Tr}\left(\left(\sqrt{\rho}\,\sigma^{\frac{1-\alpha}{\alpha}}\sqrt{\rho}\right)^\alpha\right)}{\alpha - 1}, & \alpha \geq \dfrac{1}{2}, \\[2ex] \dfrac{\log \mathrm{Tr}\left(\left(\sqrt{\sigma}\,\rho^{\frac{\alpha}{1-\alpha}}\sqrt{\sigma}\right)^{1-\alpha}\right)}{\alpha - 1}, & \alpha \leq \dfrac{1}{2}. \end{cases} \tag{11b}$$

Note that if the states are commuting, $[\rho, \sigma] = 0$, then both relative entropies are identical and in this case we shall denote this without adornment as D_α. Finally, for classical probability distributions, we will also use the Shannon entropy and the related entropy variance,

$$H(\boldsymbol{p}) := -\sum_i p_i \log p_i, \tag{12a}$$
$$V(\boldsymbol{p}) := \sum_i p_i \left(\log p_i - H(\boldsymbol{p})\right)^2, \tag{12b}$$

as well as the Rényi entropies,

$$H_\alpha(\boldsymbol{p}) = \frac{1}{1-\alpha} \log\left(\sum_i p_i^\alpha\right). \tag{13}$$

The probability density function and the cumulative distribution function of a normal distribution with mean μ and variance ν will be denoted by $\phi_{\mu,\nu}(x)$ and $\Phi_{\mu,\nu}(x)$, whereas their standardized versions (with $\mu = 0$ and $\nu = 1$) by $\phi(x)$ and $\Phi(x)$. We also introduce the function

$$S_\nu^{(\delta)}(\mu) := \inf_{A \geq \Phi} \delta(A', \phi_{\mu,\nu}), \tag{14}$$

where $\nu \in \mathbb{R}^+$ is a parameter, $\mu \in \mathbb{R}$, δ is a statistical distance, and the infimum is taken over cumulative distribution functions A (with probability density function A'), which are pointwise greater than Φ. As we shall see in Lemma 1, this function is a cumulative distribution function if δ is chosen to be the trace distance. The introduction of $S_\nu^{(\delta)}$ is inspired by Ref. [89], where the authors have investigated its special case, called the Rayleigh-normal distribution, with δ given by the infidelity distance. The name of the function comes from the fact that, as ν is varied, it interpolates between the normal and the Rayleigh distribution. In this paper, we will mainly focus on another special case, with δ given by the trace distance, and will denote the corresponding cumulative distribution function simply as

$$S_\nu(\mu) := \frac{1}{2} \inf_{A \geq \Phi} \int_{\mathbb{R}} |A'(x) - \phi_{\mu,\nu}(x)| dx. \tag{15}$$

We will refer to the above as the *sesquinormal distribution*, since we will prove that it interpolates between the normal and half-normal distributions for varying ν.

III. RESULTS

The main technical result of this paper consists of a unified approach for capturing the problem of optimal transformations of quantum dichotomies in the small-, moderate-, large-, and extreme-deviation regimes. It not only provides a much simpler and clearer derivation than the previously known results employing infidelity to measure transformation error [56,89,90] but it also extends the formalism to the case of noncommuting input states. This, in turn, leads to the main conceptual result of the paper: the generalization of the second-order asymptotic analysis of thermodynamic state interconversion to the case of general (energy-coherent) input states. Before formally stating all these results, however, we first present auxiliary results that concern the properties of the sesquinormal distribution, which may be of independent interest.

A. Sesquinormal distribution

The sesquinormal distribution has been defined implicitly via an optimization in Eq. (15). We start by giving an explicitly closed-form solution of this optimization problem and specify some relevant properties of the sesquinormal distribution.

Lemma 1 (Sesquinormal distribution). The function S_ν is a cumulative distribution function (cdf) for any $\nu \in [0,\infty)$. Moreover, for $\nu \notin \{0,1,\infty\}$ the cdf has the closed form

$$S_\nu(\mu) = \Phi\left(\frac{\mu - \sqrt{\nu}\sqrt{\mu^2 + (\nu-1)\ln\nu}}{1-\nu}\right) - \Phi\left(\frac{\sqrt{\nu}\mu - \sqrt{\mu^2 + (\nu-1)\ln\nu}}{1-\nu}\right) \tag{16}$$

and for $0 < \nu < \infty$, the inverse cdf can be expressed as

$$S_\nu^{-1}(\epsilon) = \min_{x\in(\epsilon,1)} \sqrt{\nu}\Phi^{-1}(x) - \Phi^{-1}(x-\epsilon). \tag{17}$$

The extreme cases $\nu = 0$ and $\nu \to \infty$ reduce to the normal distribution

$$S_0(\mu) = \lim_{\nu\to\infty} S_\nu(\sqrt{\nu}\mu) = \Phi(\mu) \tag{18}$$

and the $\nu = 1$ reduces to the half-normal distribution

$$S_1(\mu) = \max\{2\Phi(\mu/2) - 1, 0\}. \tag{19}$$

Finally, the family of sesquinormal distributions has a duality under reciprocating the parameter

$$S_\nu(\mu) = S_{1/\nu}(\mu/\sqrt{\nu}) \quad \text{or} \quad S_\nu^{-1}(\epsilon) = \sqrt{\nu}S_{1/\nu}^{-1}(\epsilon). \tag{20}$$

Proof. See Appendix A. ■

B. Noncommuting quantum dichotomies

We now turn to our central results on the second-order asymptotic analyzes of transformation rates between quantum dichotomies in all error regimes. Specifically, let $R_n^*(\epsilon_n)$ denote the largest rate R_n such that

$$\left(\rho_1^{\otimes n}, \sigma_1^{\otimes n}\right) \succeq_{(\epsilon_n,0)} \left(\rho_2^{\otimes R_n n}, \sigma_2^{\otimes R_n n}\right). \tag{21}$$

Theorems 2–7 will all concern the asymptotic scaling of $R_n^*(\epsilon_n)$, split by the scaling of the error ϵ_n measured by trace distance. We note that one could also consider a two-sided error variant of this problem with a pair of error sequences, $\epsilon_n^{(\rho)}$ and $\epsilon_n^{(\sigma)}$. We shall neglect this more general problem in the body of this paper but cover the extension of our results to this regime in Appendix C. We do this partially because these two-sided results are not applicable to the resource-theoretic problems on which we are mostly focused and partially because this two-sided problem is in fact no more rich, with the optimal transformation simply diverging to infinity in many regimes.

Before we move on to the second-order analysis, we start with the previously studied [59,91–94] first-order case, which states that the asymptotic transformation rate is controlled by the relative entropy.

Theorem 1 (First-order rate). For constant $\epsilon \in (0,1)$ and $[\rho_2,\sigma_2] = 0$, the optimal rate converges:

$$\lim_{n\to\infty} R_n^*(\epsilon) = \frac{D(\rho_1\|\sigma_1)}{D(\rho_2\|\sigma_2)}. \tag{22}$$

Furthermore, if we consider more general target dichotomies, $[\rho_2,\sigma_2] \neq 0$, then we still have the upper bound:

$$\limsup_{n\to\infty} R_n^*(\epsilon) \leq \frac{D(\rho_1\|\sigma_1)}{D(\rho_2\|\sigma_2)}. \tag{23}$$

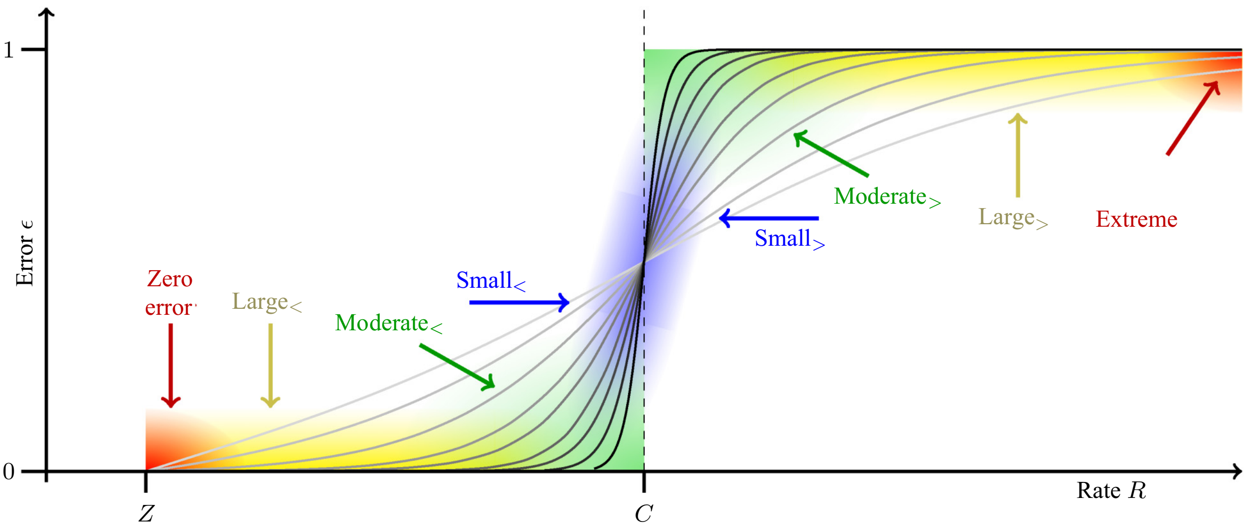
Proof. See Sec. V C. ■

Second-order asymptotics form refinements of Theorem 1 that quantify the rate of convergence to this first-order behavior. A diagram of the different second-order regimes is presented in Fig. 1. Below, we will state all of our second-order theorems, with their proofs left to Sec. V. This analysis is divided up based on the scaling of the error.

The first regime that we consider is that of *small deviations*, in which the errors considered are constants other than 0 or 1. In this regime, we find that the rate approaches the first-order rate as $O(1/\sqrt{n})$, quantified by the relative entropy variance $V(\cdot\|\cdot)$ as well as the sesquinormal distribution $S_{1/\xi}$, where ξ is the *reversibility parameter* [56,89], given by

$$\xi := \frac{V(\rho_1\|\sigma_1)}{D(\rho_1\|\sigma_1)} \bigg/ \frac{V(\rho_2\|\sigma_2)}{D(\rho_2\|\sigma_2)}. \tag{24}$$

Given these, the scaling of the rate in the small-deviation regime is as follows.



Remige		Erorr (ϵ_n)	Rate (R_n)	Tight?
Zero error	Theorem 6	Zero	Z	No
Large$_<$	Theorem 4	Approaching 0 exponentially	$[Z, C)$	No
Moderate$_<$	Theorem 3	Approaching 0 subexponentially	$C - \omega(1/\sqrt{n}) \cap o(1)$	Yes
Small$_<$	Theorem 2	Constant, < 0.5	$C - \Theta(1/\sqrt{n})$	Yes
Small$_>$	Theorem 2	Constant, > 0.5	$C + \Theta(1/\sqrt{n})$	Yes
Moderate$_>$	Theorem 3	Approaching 1 subexponentially	$C + \omega(1/\sqrt{n}) \cap o(1)$	Yes
Large$_>$	Theorem 5	Approaching 1 exponentialy	(C, ∞)	Yes
Extreme	Theorem 7	Approaching 1 superexponentially	∞	Yes

FIG. 1. A summary of our main results. The asymptotics of the transformation rates between quantum dichotomies $(\rho_1, \sigma_1) \to (\rho_2, \sigma_2)$ with an error of at most ϵ_n allowed on the first state. The table summarizes the different error regimes, i.e., the different manners in which the error ϵ_n and rate R_n can scale. In the above, the first-order rate is $C := D(\rho_1\|\sigma_1)/D(\rho_2\|\sigma_2)$ and the zero-error rate is Z. For each result, we just have upper bounds for general target dichotomies but for commuting targets, $[\rho_2, \sigma_2] = 0$, we have upper *and* lower bounds. The final column denotes whether these bounds coincide, which they do in all but one regime.

Theorem 2 (Small-deviation rate). Let $\lesssim/\simeq$ denote (in)equality up to $o(1/\sqrt{n})$. For constant $\epsilon \in (0, 1)$ and for $[\rho_2, \sigma_2] = 0$, the optimal rate is

$$R_n^*(\epsilon) \simeq \frac{D(\rho_1\|\sigma_1) + \sqrt{V(\rho_1\|\sigma_1)/n} \times S_{1/\xi}^{-1}(\epsilon)}{D(\rho_2\|\sigma_2)}. \quad (25)$$

Furthermore, if we consider general output dichotomies, $[\rho_2, \sigma_2] \neq 0$, then we still have the upper bound

$$R_n^*(\epsilon) \lesssim \frac{D(\rho_1\|\sigma_1) + \sqrt{V(\rho_1\|\sigma_1)/n} \times S_{1/\xi}^{-1}(\epsilon)}{D(\rho_2\|\sigma_2)}. \quad (26)$$

Proof. See Sec. V C 1. ■

The second regime that we consider is that of *moderate deviations*, in which errors are tending toward either 0 or 1, but only doing so subexponentially. This causes the rate to approach the first-order rate more slowly than $O(1/\sqrt{n})$; specifically, as follows.

Theorem 3 (Moderate-deviation rate). Consider an $a \in (0, 1)$ and let $\lesssim / \simeq$ denote (in)equality up to $o\left(\sqrt{n^{a-1}}\right)$. Let $\epsilon_n := \exp(-\lambda n^a)$ for some $\lambda > 0$. For $[\rho_2, \sigma_2] = 0$, the optimal rate is

$$R_n^*(\epsilon_n) \simeq \frac{D(\rho_1\|\sigma_1) - |1 - \xi^{-1/2}|\sqrt{2\lambda V(\rho_1\|\sigma_1) n^{a-1}}}{D(\rho_2\|\sigma_2)}, \quad (27a)$$

$$R_n^*(1 - \epsilon_n) \simeq \frac{D(\rho_1\|\sigma_1) + \left[1 + \xi^{-1/2}\right]\sqrt{2\lambda V(\rho_1\|\sigma_1) n^{a-1}}}{D(\rho_2\|\sigma_2)}. \quad (27b)$$

Furthermore, if we consider general output dichotomies, $[\rho_2, \sigma_2] \neq 0$, then we still have the upper bounds

$$R_n^*(\epsilon_n) \lesssim \frac{D(\rho_1\|\sigma_1) - |1-\xi^{-1/2}|\sqrt{2\lambda V(\rho_1\|\sigma_1)n^{a-1}}}{D(\rho_2\|\sigma_2)}, \tag{28a}$$

$$R_n^*(1-\epsilon_n) \lesssim \frac{D(\rho_1\|\sigma_1) + \left[1+\xi^{-1/2}\right]\sqrt{2\lambda V(\rho_1\|\sigma_1)n^{a-1}}}{D(\rho_2\|\sigma_2)}. \tag{28b}$$

Proof. See Sec. V C 3. ■

The third is the *large-deviations* regime, in which the error is either exponentially approaching 0 (large deviation, low error) or exponentially approaching 1 (large deviation, high error). In this case, the error is small or large enough so that the asymptotic rate shifts away from the first-order rate and now depends not just on the relative entropy but also on the Rényi relative entropies; specifically, as follows.

Theorem 4 (Large-deviation rate, low error). For any error of the form $\epsilon_n = \exp(-\lambda n)$ with constant $\lambda > 0$, if $[\rho_2, \sigma_2] = 0$, then the optimal rate is lower bounded by

$$\liminf_{n\to\infty} R_n^*(\epsilon_n) \geq \min_{-\lambda\leq\mu\leq\lambda} \check{r}(\mu). \tag{29}$$

Furthermore, if we consider general output dichotomies, $[\rho_2, \sigma_2] \neq 0$, then the optimal rate is upper bounded by

$$\limsup_{n\to\infty} R_n^*(\epsilon_n) \leq \min_{-\lambda\leq\mu\leq\lambda} \bar{r}(\mu). \tag{30}$$

In the above, $\bar{r}$ and $\check{r}$ are defined in terms of Rényi relative entropies in Sec. V C 2 and they coincide when $[\rho_1, \sigma_1] = [\rho_2, \sigma_2] = 0$.

Proof. See Sec. V C 2. ■

Theorem 5 (Large-deviation rate, high error). For any error of the form $\epsilon_n = 1 - \exp(-\lambda n)$ with constant $\lambda > 0$, if $[\rho_2, \sigma_2] = 0$, then the optimal rate is

$$\lim_{n\to\infty} R_n^*(\epsilon_n) = \inf_{\substack{t_1>1\\0<t_2<1}} \frac{\check{D}_{t_1}(\rho_1\|\sigma_1) + \left(\frac{t_1}{t_1-1} + \frac{t_2}{1-t_2}\right)\lambda}{D_{t_2}(\rho_2\|\sigma_2)}. \tag{31}$$

Furthermore, if we consider general output dichotomies, $[\rho_2, \sigma_2] \neq 0$, then we still have the upper bound

$$\limsup_{n\to\infty} R_n^*(\epsilon_n) \leq \inf_{\substack{t_1>1\\0<t_2<1}} \frac{\check{D}_{t_1}(\rho_1\|\sigma_1) + \left(\frac{t_1}{t_1-1} + \frac{t_2}{1-t_2}\right)\lambda}{\overline{D}_{t_2}(\rho_2\|\sigma_2)}. \tag{32}$$

Proof. See Sec. V C 2. ■

Finally, the fourth regime we consider is that of *extreme deviations*, i.e., with superexponentially decaying errors. The first is the low-error case of errors superexponentially approaching zero, including exactly zero error. In this case, we obtain an expression for the asymptotic rate that is quite similar to the first-order expression but involves a minimization over the minimal relative entropies instead of just *the* relative entropy. It gives an additional operational interpretation of the minimal Rényi entropy [95] and it is a noncommutative generalization of Refs. [96,97]. Specifically, the zero-error rate is as follows.

Theorem 6 (Zero-error rate). For $[\rho_2, \sigma_2] = 0$ the optimal zero-error rate is lower bounded,

$$\liminf_{n\to\infty} R_n^*(0) \geq \max\left\{\inf_{\alpha\in\mathbb{R}} \frac{\overleftarrow{D}_\alpha(\rho_1\|\sigma_1)}{D_\alpha(\rho_2\|\sigma_2)}, \inf_{\alpha\in\mathbb{R}} \frac{\overrightarrow{D}_\alpha(\rho_1\|\sigma_1)}{D_\alpha(\rho_2\|\sigma_2)}\right\}, \tag{33}$$

where the divergences $\overleftarrow{D}_\alpha$ and $\overrightarrow{D}_\alpha$ are defined in Eqs. (81a) and (81b). More generally, if $[\rho_2, \sigma_2] \neq 0$, then the optimal transformation rate for all n is upper bounded:

$$R_n^*(0) \leq \min_{\alpha\in\mathbb{R}} \frac{\check{D}_\alpha(\rho_1\|\sigma_1)}{\check{D}_\alpha(\rho_2\|\sigma_2)}. \tag{34}$$

Proof. See Sec. V C 4. ■

Lastly, we are left with the final case of errors exponentially approaching 1, wherein the rate diverges to infinity.

Theorem 7 (Extremely high error rate). For $[\rho_2, \sigma_2] = 0$, if the error is allowed to be superexponentially close to 1, then the optimal rate is unbounded:

$$\lim_{n\to\infty} R_n^*\big(1 - \exp(-\omega(n))\big) = \infty. \tag{35}$$

Proof. See Sec. V C 4. ■

Given these theorems, we also make two conjectures. First, we note that the form of the small-deviation result (Theorem 2) is identical to the infidelity-based results in Refs. [56,89]. Our first conjecture is that this extends more generally to other distance measures.

Conjecture 1. For any fixed and nonmaximal transformation error $\epsilon > 0$ measured by a (quantum) statistical distance δ (perhaps subject to some additional "niceness" constraints), the optimal rate for transforming quantum dichotomies with commuting target states in the small-deviation regime is given by Eq. (25) with the sesquinormal distribution, $S_{1/\xi}$, replaced by the generalized Rayleigh-normal distributions $S_{1/\xi}^{(\delta)}$, defined in Eq. (14).

Second, all of the achievability bounds rely on connections to hypothesis testing that only apply for commuting targets $[\rho_2,\sigma_2]=0$, while all of our optimality bounds apply for general states. We conjecture that there might exist alternative protocols capable of saturating these bounds.

Conjecture 2. All of the optimality bounds in Theorems 2–7 are achievable, for general states, $[\rho_1,\sigma_1]\neq 0$ and $[\rho_2,\sigma_2]\neq 0$.

C. Coherent quantum thermodynamics

Our technical results find applications in quantum thermodynamics because of the following result, the proof of which can be found in Appendix D.

Theorem 8. For $\sigma_1=\gamma_1$ and $\sigma_2=\gamma_2$, both being thermal Gibbs states, the optimal transformation rates, captured by Theorems 2, 3, 5, and 7, can be attained by thermal operations. Moreover, for energy-incoherent input and output states, ρ_1 and ρ_2, this extends to all error regimes, i.e., Theorems 2–7 characterize optimal transformation rates under thermal operations.

Thus, Theorems 2, 3, 5, and 7 describe optimal rates R_n^* for state transformations under thermal operations between n copies of generic quantum states ρ_1 and $R_n^* n$ copies of energy-incoherent states ρ_2, in most error regimes. This is not true for Theorems 4 and 6 since, as we shall see in Sec. V C, these proofs explicitly leverage nonthermal operations when dealing with energy coherent states. Nevertheless, in Appendix D, we show how we can extract not-necessarily-tight bounds on the achievable rates under thermal operations in these regimes.

Moreover, one can relatively straightforwardly generalize these results to obtain work-assisted optimal transformation rates. In this case, work is either invested to increase the rate of transformation or extracted for the price of decreasing the rate. More precisely, consider an ancillary battery system W with energy levels $|0\rangle_W$ and $|1\rangle_W$ separated by an energy gap w [58–60,98,99]. Then, we say that there exists a w-assisted thermal operation transforming ρ_1 into a state ϵ-close to ρ_2 if

$$\rho_1\otimes|0\rangle\langle 0|_W \xrightarrow[\text{TO}]{\epsilon} \rho_2\otimes|1\rangle\langle 1|_W, \tag{36}$$

where $w>0$ corresponds to work extraction, whereas $w<0$ means work investment. As we show in Appendix E, one can modify the proof of Theorem 2 and arrive at the following result (note that analogous modifications of Theorems 3 and 4 are also possible).

Theorem 9 (Optimal work-assisted rate in the small-deviation regime). Consider a battery system with an energy gap

$$w=w_1 n+w_2\sqrt{n}, \tag{37}$$

with constant w_1 and w_2. Then, for any fixed transformation error $\epsilon\in(0,1)$, the optimal rate R_n^* for w-assisted thermodynamic transformation between n copies of a generic state ρ_1 and $R_n^* n$ copies of an energy-incoherent state ρ_2 is given by

$$R_n^*(\epsilon)\simeq\frac{D(\rho_1\|\gamma_1)-\beta w_1}{D(\rho_2\|\gamma_2)}+\frac{\sqrt{V(\rho_1\|\gamma_1)}S^{-1}_{1/\xi'}(\epsilon)-\beta w_2}{\sqrt{n}D(\rho_2\|\gamma_2)}, \tag{38}$$

where

$$\xi':=\frac{V(\rho_1\|\gamma_1)}{D(\rho_1\|\gamma_1)-\beta w_1}\Big/\frac{V(\rho_2\|\gamma_2)}{D(\rho_2\|\gamma_2)} \tag{39}$$

and $\simeq$ denotes an equality up to terms of order $o(1/\sqrt{n})$. Moreover, when $\rho_2=\gamma_2$, any positive transformation rate R_n^* is possible as long as

$$\frac{\beta w}{n}\lesssim D(\rho_1\|\gamma_1)+\sqrt{\frac{V(\rho_1\|\gamma_1)}{n}}\Phi^{-1}(\epsilon). \tag{40}$$

D. Entanglement transformations

Due to the relation between transforming commuting quantum dichotomies and LOCC transformations discussed in Sec. II C, our technical results also find applications in the resource theory of entanglement.

Theorem 10. For pure bipartite states ψ_1 and ψ_2 characterized by Schmidt vectors $\boldsymbol{p_1}$ and $\boldsymbol{p_2}$, the optimal transformation rates between ψ_1 and ψ_2 under LOCC are captured by Theorems 2–7 with the following substitutions (including the substitutions in the expression for ξ):

$$D(\rho_i\|\sigma_i)\to H(\boldsymbol{p_i}), \tag{41a}$$

$$V(\rho_i\|\sigma_i)\to V(\boldsymbol{p_i}), \tag{41b}$$

$$\overline{D}_t(\rho_i\|\sigma_i),\ \check{D}_t(\rho_i\|\sigma_i)\to H_t(\boldsymbol{p_i}). \tag{41c}$$

The details of the necessary manipulations to arrive at the above result can be found in Appendix F.

IV. DISCUSSION AND APPLICATIONS

A. Phenomenological model

We start the discussion by giving an intuitive but completely nonrigorous "derivation" of the thermodynamic small-deviation rate (rates in other regimes can potentially also be "derived" in a similar fashion). It is based on three assumptions. First, assume that the thermodynamic resource content of a given state ρ is a random variable $\log\rho - \log\gamma$ (a difference between log-likelihoods for the state and the thermal state), so that its mean and variance are given by the nonequilibrium free energy $D(\rho\|\gamma)$ and its fluctuations $V(\rho\|\gamma)$. Second, the distribution of the thermodynamic resource content of $\rho^{\otimes n}$ for large n is a Gaussian with mean $nD(\rho\|\gamma)$ and variance $nV(\rho\|\gamma)$. And third, assume that every transformation that does not increase the resource content, even probabilistically, is allowed.

Using these three assumptions, let us now find the smallest transformation error ϵ for which a thermodynamic transformation with a rate R_n^* from the initial state $\rho_1^{\otimes n}$ to the target state $\rho_2^{\otimes R_n^* n}$ is possible. Cumulative distribution functions of the resource content of the initial and target states are given by Φ_{μ_1,ν_1} and Φ_{μ_2,ν_2}, where

$$\mu_1 = nD(\rho_1\|\gamma), \quad \mu_2 = R_n^* nD(\rho_2\|\gamma), \tag{42a}$$

$$\nu_1 = nV(\rho_1\|\gamma), \quad \nu_2 = R_n^* nV(\rho_2\|\gamma). \tag{42b}$$

Let us also denote the cumulative distribution of the resource content of the final state by A. Then, the condition for not increasing the resource content is given by $A \geq \Phi_{\mu_1,\nu_1}$ (i.e., there is always more probability mass with lower resource content for the final state as compared to the initial state). The minimal transformation error is then given by

$$\epsilon = \inf_{A\geq\Phi_{\mu_1,\nu_1}} \delta(A,\Phi_{\mu_2,\nu_2}) = \inf_{A\geq\Phi} \delta(A,\Phi_{\mu,\nu}) = S_\nu(\mu), \tag{43}$$

where

$$\mu = \frac{\mu_2-\mu_1}{\sqrt{\nu_1}}, \quad \nu = \frac{\nu_2}{\nu_1}. \tag{44}$$

Finally, by applying S_ν^{-1} to both sides of Eq. (43), using the expressions for μ and ν, and keeping only the leading terms in n, we end up recovering the thermodynamic small-deviation rate:

$$R_n^*(\epsilon) = \frac{D(\rho_1\|\gamma) + \sqrt{V(\rho_1\|\gamma)/n} \times S_{1/\xi}^{-1}(\epsilon)}{D(\rho_2\|\gamma)}, \tag{45}$$

where ξ is given by Eq. (24) with $\sigma_1 = \sigma_2 = \gamma$.

B. Optimal thermodynamic protocols with coherent inputs

The obtained results can be straightforwardly applied to study the optimal performance of thermodynamic protocols, where the processed systems may be initially prepared in coherent superpositions of different energy eigenstates. In what follows, we will briefly discuss how this can be done and what it means for work extraction, information erasure, and thermodynamically free encoding of information. We note that in all these protocols, the final states are energy incoherent and thus our results allow one to study them in full generality.

In the work-extraction protocol, one uses a thermal bath and n copies of a system in a state ρ to excite the battery system W over the energy gap w. The aim is to find the largest possible w as a function of the allowed transformation error ϵ. In other words, one wants to find the largest w for which the following thermodynamic transformation exists:

$$\rho^{\otimes n} \otimes |0\rangle\langle 0|_W \xrightarrow[\text{TO}]{\epsilon} |1\rangle\langle 1|_W. \tag{46}$$

This problem can be directly addressed by employing Theorem 9 with the target state of the system being thermal, which results in

$$\frac{w}{n} \leq \frac{1}{\beta}\left(D(\rho\|\gamma) + \sqrt{\frac{V(\rho\|\gamma)}{n}}\Phi^{-1}(\epsilon)\right). \tag{47}$$

The above yields the optimal amount of ϵ-deterministic work that can be extracted per one copy of the processed system and generalizes the previous small-deviation results on work extraction from incoherent states [56] and pure states [82] to general quantum states.

In the information-erasure protocol, one aims at using a thermal bath and the deexcitation of a battery system W with minimal possible energy gap $|w|$ to reset n copies of a system with a trivial Hamiltonian and in a state ρ into a pure state $|0\rangle\langle 0|$. This corresponds to finding the smallest $|w|$ (note that since we deexcite the battery, we have $w < 0$) for which the following thermodynamic transformation exists:

$$\rho^{\otimes n} \otimes |0\rangle\langle 0|_W \xrightarrow[\text{TO}]{\epsilon} |0\rangle\langle 0|^{\otimes n} \otimes |1\rangle\langle 1|_W. \tag{48}$$

Employing Theorem 9 and solving for the w, which allows us to achieve rate 1, we arrive at the thermodynamic cost of information erasure per one copy of the system:

$$\frac{|w|}{n} \simeq \frac{1}{\beta}\left(S(\rho) - \sqrt{\frac{V(\rho)}{n}}\Phi^{-1}(\epsilon)\right), \tag{49}$$

which again generalizes the previously known results for erasing incoherent states.

Finally, the problem of thermodynamically free encoding of information, introduced in Ref. [100] and studied for incoherent and pure states in Ref. [82], is stated as follows. A sender is given n copies of a quantum system ρ that acts as an information carrier and wants to encode one of M messages into these systems without using any thermodynamic resources, hence employing only thermal operations. The aim is to find the maximal number M of messages that can be encoded in a way that allows for decoding them with error probability at most ϵ. In Ref. [100], it has been shown that, in the small-deviation regime, M is upper bounded by

$$\frac{\log M(\rho^{\otimes n}, \epsilon)}{n} \lesssim D(\rho\|\gamma) + \sqrt{\frac{V(\rho\|\gamma)}{n}}\Phi^{-1}(\epsilon) \tag{50}$$

and in Ref. [82], it has been proved that the above bound can be achieved for states ρ that are either energy incoherent or pure. Using the results that we obtained have here, this can be generalized to arbitrary quantum states ρ in the following way. Consider the following thermodynamic transformation:

$$\rho^{\otimes n} \xrightarrow[\text{TO}]{\epsilon} |0\rangle\langle 0|_A^{\otimes Rn}, \tag{51}$$

where the final system A consists of Rn two-level subsystems with trivial Hamiltonians. Note that since all energy levels of the final systems are degenerate, the sender can map the state $|0\rangle\langle 0|_A^{\otimes Rn}$ to any of 2^{Rn} basis states using thermal operations. Thus, the sender can encode $M = 2^{Rn}$ messages that, moreover, can be decoded with probability of error ϵ simply through a computational-basis measurement. It is then straightforward to employ Theorems 2 and 8 to obtain Eq. (50) with $\lesssim$ replaced by $\simeq$.

C. Resonance phenomena

One of the fundamental observations in the resource theory of thermodynamics is that all state transformations become reversible in the asymptotic limit [59]. Indeed, Theorem 1 clearly states that for $n \to \infty$, the conversion rates R and R' for transformations $\rho_1^{\otimes n} \to \rho_2^{\otimes Rn}$ and $\rho_2^{\otimes n} \to \rho_1^{\otimes R'n}$ become inversely proportional to each other, $R = 1/R'$. This is generally no longer true when we move outside of the idealized asymptotic scenario with $n \to \infty$. For example, in the thermodynamic protocols analyzed in Sec. IV B, we have seen the deteriorating effect of finite-size transformations, i.e., due to the finite number of thermodynamically processed systems, the transformations are irreversible and lead to free-energy dissipation that is related to the free-energy fluctuations measured by $V(\rho\|\gamma)$ [82]. As a result, the performance of small quantum thermal machines may be seriously limited. Similar behavior can be observed in the resource theory of pure-state entanglement or coherence.

Interestingly, it has recently been found that these finite-size effects can be significantly mitigated by carefully engineering the resource-conversion process [101]. More specifically, by appropriately tuning the initial and final states, so that the reversibility parameter $\xi = 1$, the second-order correction to the optimal rate may vanish in the limit of zero transformation error. Thus, up to higher-order terms, reversibility is restored. This intriguing phenomenon, termed resource resonance, was first predicted in Ref. [89] for pure-state entanglement transformations and then generalized to thermodynamic transformations between energy-incoherent states in Ref. [101]. The results that we have presented in this paper allow us to extend the resource resonance phenomenon in three novel ways that we will now discuss.

1. Coherent resonance

For simplicity, let us focus on thermodynamic transformations between n copies of a two-level system in a general state ρ_1 and Rn copies of a two-level system in an energy-incoherent state ρ_2, assuming that the thermal Gibbs state γ is the same for the initial and final systems. Using Theorem 2 together with Theorem 8, we obtain that the optimal transformation error ϵ for the asymptotic rate $R = D(\rho_1\|\gamma)/D(\rho_2\|\gamma)$ (i.e., avoiding dissipation) is given by

$$\epsilon = S_{1/\xi}(0), \tag{52}$$

which vanishes for $\xi = 1$ and increases from 0 to 1/2 for $\xi > 1$ and $\xi < 1$. Without loss of generality, let us parametrize the initial and final states in the energy eigenbasis by

$$\rho_1(x) = \begin{pmatrix} p & x\sqrt{p(1-p)} \\ x\sqrt{p(1-p)} & 1-p \end{pmatrix},$$
$$\rho_2 = \begin{pmatrix} q & 0 \\ 0 & 1-q \end{pmatrix}, \tag{53}$$

with $p, q, x \in [0,1]$. Then, for a fixed p and q (and given γ), we can consider a family of initial states parametrized by x [see Fig. 2(a)]. This corresponds to probabilistic mixtures of an energy-incoherent state $p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|$ and a pure coherent superposition of energy eigenstates $\sqrt{p}|0\rangle + \sqrt{1-p}|1\rangle$, so that $x \in [0,1]$ smoothly connects between completely incoherent and completely coherent initial states. In Fig. 2(b), we present the nontrivial dependence of the transformation error ϵ on the coherence level x, where we can observe two resonant values of x for which error-free and dissipationless transformations (up to second-order asymptotics) are possible. This clearly illustrates that quantum coherence can play an important role in avoiding free-energy dissipation in thermodynamic transformations of quantum states.

2. Work-assisted resonance

Looking at the optimal work-assisted rate from Theorem 9, we see that the whenever $\xi' = 1$, one can choose $w_2 = 0$ to make the second-order asymptotic correction vanish for

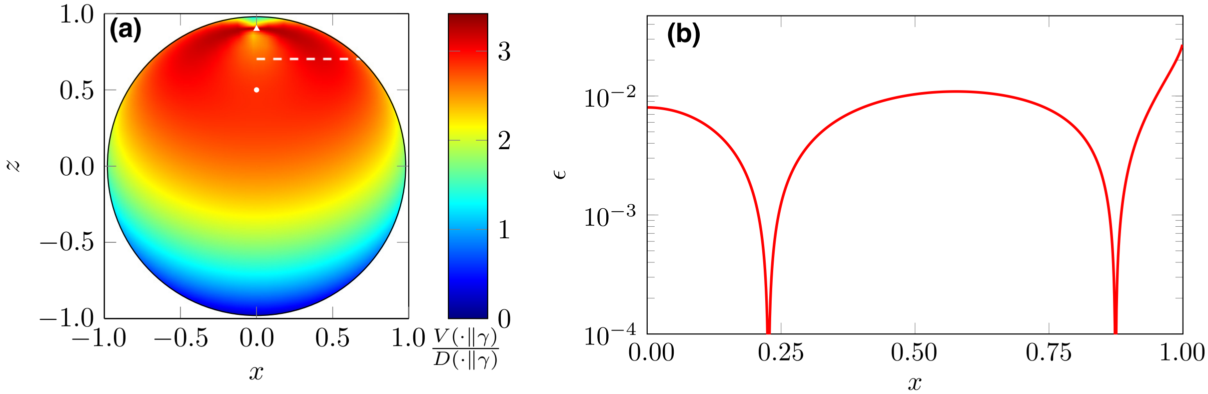


FIG. 2. Coherent resonance in thermodynamic transformations of two-level systems. (a) The ratio $V(\rho\|\gamma)/D(\rho\|\gamma)$ (encoding the resonance condition) for qubit states lying in the x-z plane of the Bloch sphere for a thermal state $\gamma = \mathrm{diag}(0.95, 0.05)$ (indicated by a white triangle). The white disk corresponds to the final state $\rho_2 = \mathrm{diag}(0.75, 0.25)$, while the dashed white line indicates a family of initial states $\rho_1(x)$ with diagonal (0.85, 0.15) and off-diagonal elements equal to $\sqrt{0.85 \times 0.15} \times x$ for $x \in [0, 1]$. (b) The threshold transformation error ϵ required to achieve the asymptotic transformation rate $D(\rho_1(x)\|\gamma)/D(\rho_2\|\gamma)$ for a finite number n of transformed systems (i.e., ϵ such that the second-order correction term in Eq. (25) disappears). Resonance is obtained when the relative free-energy fluctuations V/D are the same for the initial state $\rho_1(x)$ and the final state ρ_2, i.e., when $\xi = 1$.

zero transformation error ϵ. Crucially, the value of ξ' can be controlled by the amount w_1 of invested (or extracted) work per one copy of the system. By choosing

$$w_1 = \frac{1}{\beta}\left(D(\rho_1\|\gamma_1) - \frac{V(\rho_1\|\gamma_1)}{V(\rho_2\|\gamma_2)}D(\rho_2\|\gamma_2)\right), \tag{54}$$

which results in the optimal rate given by

$$R = \frac{V(\rho_1\|\gamma_1)}{V(\rho_2\|\gamma_2)}, \tag{55}$$

one can perform an error-free and dissipationless transformation. In other words, the total initial state of the system and battery gets transformed to the total final state with zero error and equal free-energy content (up to second-order terms). This thus opens a way for bringing two states into resonance by investing or extracting work.

The work-assisted resonance can be understood by first noting that the resonance condition can be seen as requiring the total fluctuations of the initial system to be equal to the total fluctuations of the final system, up to first order in n. Without work assistance, this means that

$$V(\rho_1^{\otimes n}\|\gamma_1^{\otimes n}) = V(\rho_2^{\otimes Rn}\|\gamma_2^{\otimes Rn}) \tag{56}$$

and given the asymptotic value of R, it yields

$$V(\rho_1\|\gamma_1) = \frac{D(\rho_1\|\gamma_1)}{D(\rho_2\|\gamma_2)}V(\rho_2\|\gamma_2), \tag{57}$$

which is exactly the original resonance condition $\xi = 1$. Now, introducing the battery system does not change the fluctuations (since at the initial and final time, the battery is in a pure energy eigenstate with zero fluctuations) but it affects the rate R. The work-assisted resonance condition is achieved by increasing or decreasing R through an appropriate choice of w, so that Eq. (56) is satisfied, which happens for R given by Eq. (55).

3. Strong resonance

In Ref. [101], a resonance phenomenon has been observed for transformations operating at the first-order asymptotic rate. Specifically, such transformations generically incur a constant error but it has been shown that if a resonance condition is met, these errors are in fact exponentially suppressed. That result was built upon the small- and moderate-deviation results of Refs. [56,89,90] but large- and extreme-deviation analyses had not been performed at the time that would allow for exponentially small errors to be analyzed. By extending to large- and extreme-deviation analyses in this paper, it can in fact be seen that there exists an even stronger notion of resonance, which we term *strong resonance*, in contrast to the *weak resonance* of Ref. [101], in which errors are not just exponentially suppressed but eliminated entirely.

An illustration of weak and strong resonance is presented in Fig. 3. Weak resonance corresponds to the second-order corrections in the small- and moderate-deviation rates (Theorems 2 and 3) vanishing and occurs when

$$\frac{V(\rho_1\|\sigma_1)}{D(\rho_1\|\sigma_1)} = \frac{V(\rho_2\|\sigma_2)}{D(\rho_2\|\sigma_2)}. \tag{58}$$

Strong resonance corresponds to the situation in which the large- and extreme-deviation rates (Theorems 4 and 6) also

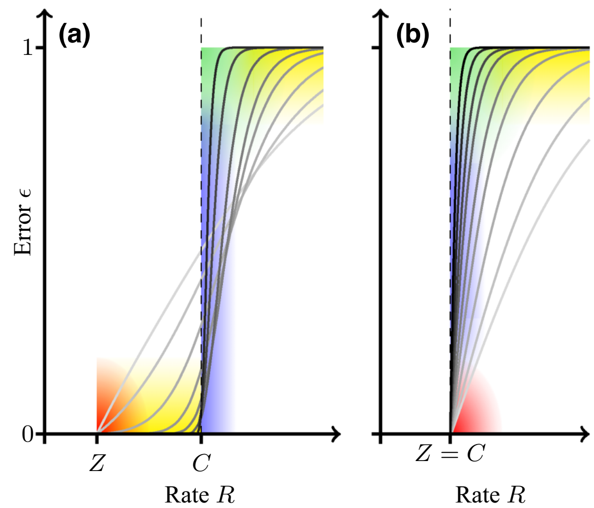


FIG. 3. Weak- and strong-resonance phenomena. (a) Weak resonance, in which the small and moderate regimes at rates $R < C$ collapse but the large and extreme regimes persist, i.e., $Z < C$. (b) Strong resonance, in which *all* error regimes at rates below C collapse, i.e., $Z = C$. For an explanation of the various error regimes indicated, as well as the definitions of Z and C, see Fig. 1.

collapse down to the first-order rate; in other words, when

$$\arg\min_{\alpha\in\overline{\mathbb{R}}} \frac{\breve{D}_\alpha(\rho_1\|\sigma_1)}{D_\alpha(\rho_2\|\sigma_2)} = 1. \tag{59}$$

We present a numerical example of a set of states that exhibit both strong and weak resonance in Appendix G and discuss the relationship between weak and strong resonance in Appendix H.

D. Entanglement transformations

Let us now make a few brief comments on Theorem 10. It is very important to note that related results have previously appeared in the literature. First, in Ref. [89], the authors have derived the optimal second-order rates for pure-state bipartite-entanglement transformations in the small-deviation regime using infidelity to measure transformation error. Later, in Ref. [90], the authors have extended these results to the moderate-deviation regime. Finally, in Ref. [102], the author has investigated exact asymptotic transformations and derived optimal rates in the zero-error regime.

Our work differs from these results in three ways. First, we extend the analysis to the previously unaddressed large-deviation regime. This allows us to, e.g., predict a strong-resonance phenomenon for entanglement transformations. Second, our results hold for a different error measure (trace distance instead of infidelity). And third, probably most importantly, we propose a novel methodology employing dichotomies and hypothesis testing that allows us to easily characterize asymptotic rates in a unified manner across various error regimes and avoid many arduous subtleties along the way. We believe that this approach brings a significant simplification and clarity as compared to the previous techniques.

On the flip side, we need to mention that while infidelity measure between the Schmidt vectors $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ has a clear operational meaning (since it is precisely the infidelity between the corresponding entangled states ψ_1 and ψ_2), the use of the trace distance may be less useful. Still, one can directly relate the trace distance δ between $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ to the probability P of distinguishing bipartite entangled states ψ_1 and ψ_2 locally by one party: $P = (1+\delta)/2$.

Finally, we recall that it has been proven in Ref. [103] that the pure-state transformation laws in the resource theory of coherence [104], i.e., conditions under which pure superpositions of distinguished basis states can be mapped to each other under incoherent operations, are also characterized by the majorization relation. Thus, Theorem 10 can be straightforwardly applied to describe optimal rates for pure-state coherence transformations (put simply, $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ need to represent occupations of the initial and target states in the distinguished basis).

V. DERIVATIONS

In this section, we will give proofs of our results on the asymptotic analysis of the transformation rates between quantum dichotomies in several different error regimes. We will break this analysis down into three stages. In Sec. V A, we will review the relationship between Blackwell ordering and hypothesis testing, generalizing the existing analysis beyond the fully commuting case to allow for results where the input dichotomy is noncommuting and partial results when the target dichotomy is also noncommuting. Critically, once established, this connection allows us to rather straightforwardly extend the existing asymptotic analyses of hypothesis testing to transformation rates between quantum dichotomies. In Sec. V B, we review the existing results around hypothesis testing, with some necessary technical extensions. Finally, in Sec. V C, we put everything together, giving the final proofs of transformation rates in each error regime.

A. Hypothesis testing and pinched hypothesis testing

The data-processing inequality ensures that the approximate Blackwell ordering $(\rho_1,\sigma_1) \succeq_{(\epsilon_\rho,\epsilon_\sigma)} (\rho_2,\sigma_2)$ implies

$$\beta_x(\rho_1\|\sigma_1) \le \beta_{x-\epsilon_\rho}(\rho_2\|\sigma_2) + \epsilon_\sigma \quad \forall x \in (\epsilon_\rho, 1). \tag{60}$$

Extending this, it is shown in Ref. [70] that the two conditions are in fact equivalent for commuting states $[\rho_1,\sigma_1] = [\rho_2,\sigma_2] = 0$. Thus, for such states, the analysis of transformation rates can be entirely reduced to the analysis of hypothesis testing. Unfortunately, the situation for noncommuting quantum states is not so straightforward: it is known that such a hypothesis-testing condition is *not* generally a sufficient condition for Blackwell ordering [30,71–73].

While there is no known sufficient condition that can be phrased in terms of regular hypothesis testing, we instead consider a modified task that we call *pinched hypothesis testing*, which does provide such a sufficient condition for noncommuting input states. This condition does not, however, extend to noncommuting target states and we leave this for future work.

We will use $\mathcal{P}_\tau(\cdot)$ to denote the pinching with respect to the eigenspaces of τ. Specifically, it is defined by

$$\mathcal{P}_\tau(X) := \sum_\lambda \Pi_\lambda X \Pi_\lambda, \tag{61}$$

where Π_λ are the eigenspace projectors of τ, i.e., $\tau = \sum_\lambda \lambda \Pi_\lambda$. The task of *pinched hypothesis testing* is to distinguish between the states $\mathcal{P}_\sigma(\rho)$ and σ or between ρ and $\mathcal{P}_\rho(\sigma)$. Correspondingly, we define the *left-pinched* and *right-pinched* type-II hypothesis-testing error as

$$\overleftarrow{\beta}_x(\rho\|\sigma) := \beta_x(\mathcal{P}_\sigma(\rho)\|\sigma), \tag{62a}$$

$$\overrightarrow{\beta}_x(\rho\|\sigma) := \beta_x\big(\rho\,\big\|\,\mathcal{P}_\rho(\sigma)\big). \tag{62b}$$

By the data-processing inequality, we know that pinching cannot make states easier to distinguish and thus the pinched error is at least the nonpinched error,

$$\overleftarrow{\beta}_x(\rho\|\sigma), \overrightarrow{\beta}_x(\rho\|\sigma) \geq \beta_x(\rho\|\sigma), \tag{63}$$

with equality if $[\rho,\sigma]=0$. Having defined the pinched error, we now show how it can be used to construct a *sufficient* condition for noncommuting Blackwell ordering.

Lemma 2 (Conditions for approximate Blackwell ordering for commuting second dichotomy). Consider the approximate Blackwell ordering $(\rho_1,\sigma_1) \succeq_{(\epsilon_\rho,\epsilon_\sigma)} (\rho_2,\sigma_2)$. A *necessary* condition for this ordering is given by

$$\beta_x(\rho_1\|\sigma_1) \leq \beta_{x-\epsilon_\rho}(\rho_2\|\sigma_2) + \epsilon_\sigma \quad \forall x \in (\epsilon_\rho, 1). \tag{64}$$

If the second dichotomy is commuting, $[\rho_2,\sigma_2]=0$, then a sufficient condition for this ordering is given by either

$$\overleftarrow{\beta}_x(\rho_1\|\sigma_1) \leq \beta_{x-\epsilon_\rho}(\rho_2\|\sigma_2) + \epsilon_\sigma \quad \forall x \in (\epsilon_\rho, 1), \tag{65a}$$

or

$$\overrightarrow{\beta}_x(\rho_1\|\sigma_1) \leq \beta_{x-\epsilon_\rho}(\rho_2\|\sigma_2) + \epsilon_\sigma \quad \forall x \in (\epsilon_\rho, 1). \tag{65b}$$

Proof. As noted above, the necessary condition simply follows from the data-processing inequality [70], so we need only prove the sufficient condition. We start by assuming that the pinched hypothesis-testing inequality, Eq. (65a), holds. Expanding out the definition of $\overleftarrow{\beta}_x(\cdot\|\cdot)$, this is equivalent to

$$\beta_x\big(\mathcal{P}_{\sigma_1}(\rho_1)\big\|\sigma_1\big) \leq \beta_{x-\epsilon_\rho}(\rho_2\|\sigma_2) + \epsilon_\sigma \quad \forall x \in (\epsilon_\rho, 1). \tag{66}$$

Pinching a state causes it to commute, in the sense that $[\mathcal{P}_\sigma(\cdot),\sigma] \equiv 0$. As such, the first dichotomy $(\mathcal{P}_{\sigma_1}(\rho_1),\sigma_1)$ is commuting and the second dichotomy (ρ_2,σ_2) is also commuting by assumption. Applying Ref. [70, Theorem 2], this in turn implies the Blackwell ordering on the pinched states,

$$\big(\mathcal{P}_{\sigma_1}(\rho_1),\sigma_1\big) \succeq_{(\epsilon_\rho,\epsilon_\sigma)} (\rho_2,\sigma_2). \tag{67}$$

Next, we want to argue that approximate Blackwell ordering has a data-processing property. By definition, this ordering implies the existence of a channel $\mathcal{E}$ such that

$$T\Big(\mathcal{E}\big(\mathcal{P}_{\sigma_1}(\rho_1)\big),\rho_2\Big) \leq \epsilon_\rho \quad \text{and} \quad T(\mathcal{E}(\sigma_1),\sigma_2) \leq \epsilon_\sigma. \tag{68}$$

If we define $\overleftarrow{\mathcal{E}} := \mathcal{E}\circ\mathcal{P}_{\sigma_1}$ and recall that $\mathcal{P}_{\sigma_1}(\sigma_1)=\sigma_1$, then these expressions can be rewritten as

$$T\Big(\overleftarrow{\mathcal{E}}(\rho_1),\rho_2\Big) \leq \epsilon_\rho \quad \text{and} \quad T\Big(\overleftarrow{\mathcal{E}}(\sigma_1),\sigma_2\Big) \leq \epsilon_\sigma, \tag{69}$$

which in turn implies the required Blackwell ordering of the two dichotomies. A similar argument can be given for Eq. (65b), with $\overrightarrow{\mathcal{E}} := \mathcal{E}\circ\mathcal{P}_{\rho_1}$. ■

We now have both necessary and sufficient conditions for approximate Blackwell ordering of quantum dichotomies that are of the same form. Unlike the commuting case, however, these two conditions are no longer identical, involving the pinched and nonpinched variants of hypothesis testing. As such, this will generally open up a gap between the upper and lower bounds that can be placed upon transformation rates using this technique, which makes this approach unsuitable in the single-shot setting. However, as we will see later in this section, in the asymptotic setting the pinched and nonpinched variants of hypothesis testing have identical asymptotic behavior in most error regimes, closing these gaps and allowing us to give optimal expressions of transformation rates beyond the first-order asymptotics.

B. Asymptotic analyses of hypothesis testing

In this subsection, we want to review the relevant asymptotic analyses of hypothesis testing, putting these results into a common notation for easier use later, as well as extending these analyses to the pinched variant of the task where necessary. The cornerstone of asymptotic analysis of hypothesis testing is Stein's lemma. While sufficient to give a first-order analysis of transformation rates between quantum dichotomies, we will see that we require refinements upon Stein's lemma to go beyond first order. We will start this section by describing Stein's lemma and giving some intuition for the different regimes of refinements thereto. We will then go through each error regime reviewing the refined asymptotic analysis in each, extending these analyses to the pinched variant of the problem as necessary.

Consider the task of distinguishing between two states $\rho^{\otimes n}$ and $\sigma^{\otimes n}$. To avoid technical issues, we will assume that σ is of full support. Intuitively, each additional copy of the states should give us a constant amount of new information, allowing us to multiplicatively reduce the chance of failing to distinguish the two states, leading to exponentially decreasing hypothesis-testing errors. In general, there is a trade-off between the type-I and -II errors. A natural simplification of this more general question would be the following: if we constrain one of our errors to be constant, how does the other error decay? The answer is that the error decays exponentially, with that exponent being given by the relative entropy. This fact is known as Stein's lemma and it will form the backbone of this subsection.

Lemma 3 (Quantum Stein's lemma [105,106]). For any $\epsilon \in (0,1)$

$$\lim_{n\to\infty} -\frac{1}{n}\log \beta_\epsilon\left(\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right) = D(\rho\|\sigma). \tag{70}$$

As mentioned above, Stein's lemma alone will only be sufficient to give first-order rates and we will require more refined asymptotic analysis to go beyond this. In Fig. 4, we present a sketch of the various error regimes that we will consider. The idea is that as n increases, Stein's lemma states that the trade-off between the type-I error and type-II error exponent becomes a step at the relative entropy and each of the refinements seeks to quantify the rate of that convergence in different ways. Specifically, as shown in the table in Fig. 4, our analysis will be divided up based on the scaling of the type-I error considered. The two most important regimes will be the small- and large-deviation regimes, in which the type-I error is a constant bounded away from 0/1, or exponentially approaching 0/1, respectively. This leaves us with two edge cases: the intermediate regime of subexponential decay is termed "moderate deviation" and for completeness we also consider the regime in which the type-I error superexponentially approaches 1, which will be required for our analysis of transformation rates in the zero-error setting.

In Appendix H, we will nonrigorously discuss the interplay between these regimes and the consistency between these results and in Appendix I, we will show how all of the analyses below can be strengthened to have a uniformity property that will be necessary in some of the proofs of transformation rates given in Sec. V C.

1. Small deviation

The first regime that we consider is the small deviation. As stated in Fig. 4, in this regime, the type-I error is a fixed constant between 0 and 1 and we want to know the asymptotic behavior of the type-II error exponent. Stein's lemma tells us that this exponent must approach the relative entropy and the small-deviation analysis, also known as the *second-order expansion*, states that this convergence happens as $\Theta(1/\sqrt{n})$.

Lemma 4 (Small-deviation analysis of hypothesis testing). For any constant $\epsilon \in (0,1)$, the hypothesis-testing type-II errors scale as

$$-\frac{1}{n}\log \beta_\epsilon\left(\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right) \simeq D(\rho\|\sigma) + \sqrt{\frac{V(\rho\|\sigma)}{n}}\Phi^{-1}(\epsilon), \tag{71a}$$

$$-\frac{1}{n}\log \overleftarrow{\beta}_\epsilon\left(\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right) \simeq D(\rho\|\sigma) + \sqrt{\frac{V(\rho\|\sigma)}{n}}\Phi^{-1}(\epsilon), \tag{71b}$$

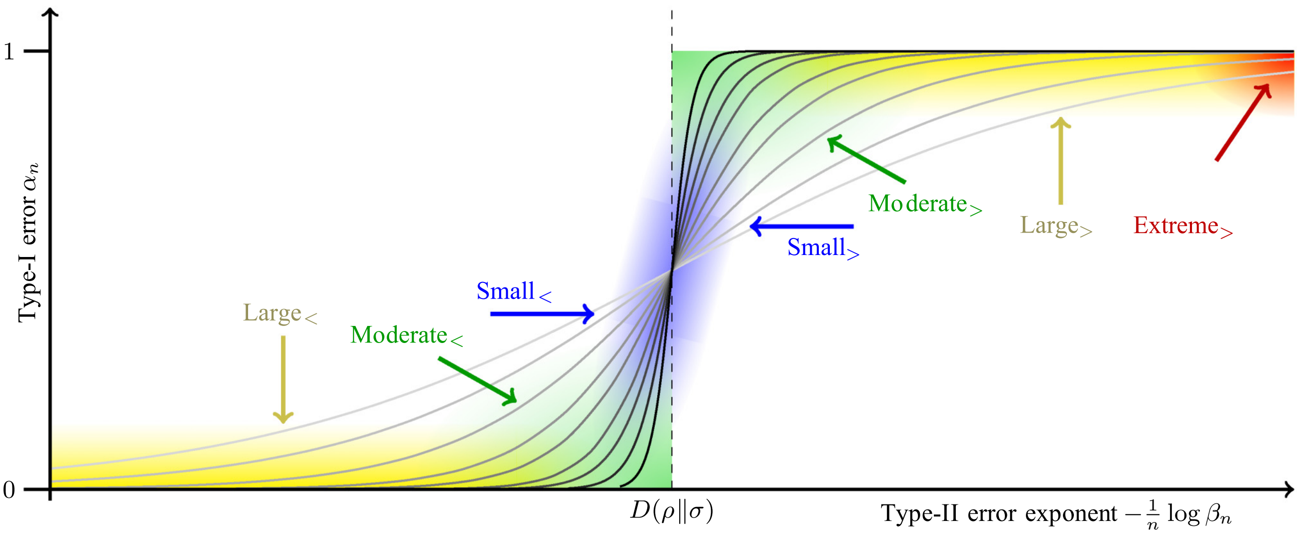
where $\simeq$ denotes equality up to terms $o(1/\sqrt{n})$.

Proof. The scaling of $\beta_\epsilon\left(\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right)$ is directly a restatement of Ref. [107, Proposition 16], a result that originates in Refs. [81,108], so we are just left with showing that the pinched variant $\overleftarrow{\beta}_\epsilon\left(\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right)$ has the same scaling up to second order.

First, we note that $\overleftarrow{\beta}_\epsilon(\cdot\|\cdot) \geq \beta_\epsilon(\cdot\|\cdot)$, so the upper bound holds straightforwardly, leaving only the lower bound left to be proved. For this, we turn to Ref. [108], specifically combining Eqs. (14), (20), and (27) to give that, for any $0 < \delta < \epsilon/3$,

$$\overleftarrow{\beta}_\epsilon(\rho\|\sigma) \leq \beta_{\epsilon-2\delta}(\rho\|\sigma) \times \frac{2^8(\epsilon-\delta)\nu(\sigma)^2}{\delta^5(1-\epsilon+\delta)}, \tag{72}$$

where $\nu(\sigma)$ denotes the number of unique eigenvalues of σ. Note that for any finite-dimensional σ, the number of eigenvalues of the tensor power $\sigma^{\otimes n}$ only scales polynomially, $\nu\left(\sigma^{\otimes n}\right) \leq n^{\nu(\sigma)} = n^{O(1)}$. Using this, we can now



Regime		Type-I (α_n)	Type-II exponent $(-\frac{1}{n}\log\beta_n)$	Equal?
Large$_<$	Lemma 5	Approaching 0 exponentially	$[0, D(\rho\|\sigma))$	No
Moderate$_<$	Lemma 6	Approaching 0 subexponentially	$D(\rho\|\sigma) - \omega(1/\sqrt{n}) \cap o(1)$	Yes
Small$_<$	Lemma 4	Constant, < 0.5	$D(\rho\|\sigma) - \Theta(1/\sqrt{n})$	Yes
Small$_>$	Lemma 4	Constant, > 0.5	$D(\rho\|\sigma) + \Theta(1/\sqrt{n})$	Yes
Moderate$_>$	Lemma 6	Approaching 1 subexponentially	$D(\rho\|\sigma) + \omega(1/\sqrt{n}) \cap o(1)$	Yes
Large$_>$	Lemma 5	Approaching 1 exponentialy	$(D(\rho\|\sigma), \infty)$	Yes
Extreme$_>$	Lemma 8	Approaching 1 superexponentially	∞	Yes

FIG. 4. The trade-off between the optimal type-I and -II errors. An illustrative sketch of the trade-off between the optimal type-I and -II errors of the hypothesis test between two states, $\rho^{\otimes n}$ and $\sigma^{\otimes n}$, as n grows. Here, α_n is the optimal type-I error, β_n the optimal type-II error, and $-(1/n)\log\beta_n$ the type-II error exponent. Each of the gray curves corresponds to a trade-off (α_n, β_n), for a given n, with darker curves corresponding to growing n. The fact that these curves approach a step at the relative entropy is equivalent to Stein's lemma (see Lemma 3). Each of the colored regions corresponds to a deviation regime in which we will consider refinements to Stein's lemma in this subsection. In the table, we present the scaling in each regime. For the details and explicit expressions for all of the scaling constants, see the corresponding lemmas, (Lemmas 4–6 and Lemma 8). The final column denotes whether the asymptotics of the pinched and nonpinched variants of hypothesis testing are identical, which they are in all regimes in which both errors are exponentially decreasing.

substitute $\rho \to \rho^{\otimes n}$ and $\sigma \to \sigma^{\otimes n}$, giving

$$\log \overleftarrow{\beta}_\epsilon\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big) \leq \log \beta_{\epsilon-2\delta}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big) + O(\log n). \tag{73}$$

Importantly, this logarithmic error is $o(\sqrt{n})$ and can therefore be neglected to second order. As such, we obtain the bound

$$-\frac{1}{n}\log \overleftarrow{\beta}_\epsilon\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big) \gtrsim -\frac{1}{n}\log \beta_{\epsilon-2\delta}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big), \tag{74a}$$

$$\gtrsim D(\rho\|\sigma) + \sqrt{\frac{V(\rho\|\sigma)}{n}}\Phi^{-1}(\epsilon - 2\delta). \tag{74b}$$

As this holds for any $\delta \in (0, \epsilon/3)$ and Φ^{-1} is continuous on $(0,1)$, we can take $\delta \to 0^+$, giving

$$-\frac{1}{n}\log \overleftarrow{\beta}_\epsilon\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big) \gtrsim D(\rho\|\sigma) + \sqrt{\frac{V(\rho\|\sigma)}{n}}\Phi^{-1}(\epsilon), \tag{75}$$

as required. ■

2. Large deviation

The next most important error regime is that of large deviations, which is the regime in which both errors are exponentially approaching either 0 or 1. Stein's lemma suggests that as long as the type-II error exponent is less than the relative entropy, then the type-I error will also be exponentially decreasing with n; but if it exceeds the relative entropy, then we expect the type-I error to be exponentially increasing toward 1.

The existing expressions of these results in the literature are all phrased in terms of these exponents directly. But, using this notation, the large-deviation regime would need to be divided up into several different forms based on whether the errors are approaching 0 or 1, usually termed the *error-exponent* and *strong-converse-exponent* regimes. Instead, we will combine all of these results in a single unified notation by concerning ourselves not with error *probabilities* but with the error *log odds*. This unified notation dramatically simplifies our later proofs, which rely upon these bounds, and to our knowledge this formulation has not appeared elsewhere in the literature.

The idea to unify these regimes is to consider a "signed exponent." For a quantity p_n that is exponentially approaching 0, the exponent is given by $-(1/n)\log p_n$ and, similarly, if p_n is exponentially approaching 1, then the exponent is given by $-(1/n)\log(1-p_n)$. The idea is to combine these two functions to give a single expression that can yield both exponents. Specifically, we will use the *logit function*, which is simply the difference between $\log p$ and $\log(1-p)$,

$$L[p] := \log\frac{p}{1-p}. \tag{76}$$

As required, now we can think of $(1/n)L[p_n]$ as an exponent that covers both cases in which p_n is approaching 0 or 1 in the sign of this exponent. Specifically, for any $\lambda > 0$,

$$\frac{1}{n}L[p_n]\to-\lambda \quad\Longleftrightarrow\quad -\frac{1}{n}\log p_n\to\lambda, \tag{77a}$$

$$\frac{1}{n}L[p_n]\to+\lambda \quad\Longleftrightarrow\quad -\frac{1}{n}\log(1-p_n)\to\lambda. \tag{77b}$$

While there are other functions that have this property, one thing to note about the logit function specifically is that if p is a probability, then $L[p]$ is the associated *log odds*. As we will see below, it turns out that the standard large-deviation results can be more succinctly expressed in terms of the type-I and -II error *log odds* instead of error *probabilities*.

To allow us to express the large- (and moderate-) deviation results in terms of the *log odds*, we will define the optimal type-II log odds $\gamma_x(\rho\|\sigma)$—in analogy to the optimal type-II error probability $\beta_x(\rho\|\sigma)$—as the solution to the optimization

$$\min_{Q}\quad L\big[\mathrm{Tr}(\sigma Q)\big], \tag{78a}$$

$$\text{subject to}\quad 0\le Q\le 1, \tag{78b}$$

$$L\big[1-\mathrm{Tr}(\rho Q)\big]\le x. \tag{78c}$$

And as with the error probability, we will also require the pinched variants, defined as

$$\overleftarrow{\gamma}_x(\rho\|\sigma) := \gamma_x(\mathcal{P}_\sigma(\rho)\|\sigma), \tag{79a}$$

$$\overrightarrow{\gamma}_x(\rho\|\sigma) := \gamma_x\big(\rho\,\big\|\,\mathcal{P}_\rho(\sigma)\big). \tag{79b}$$

In terms of error probabilities, these definitions are equivalent to

$$\gamma_x(\rho\|\sigma) = L\left[\beta_{L^{-1}[x]}(\rho\|\sigma)\right], \tag{80a}$$

$$\overleftarrow{\gamma}_x(\rho\|\sigma) = L\left[\overleftarrow{\beta}_{L^{-1}[x]}(\rho\|\sigma)\right], \tag{80b}$$

$$\overrightarrow{\gamma}_x(\rho\|\sigma) = L\left[\overrightarrow{\beta}_{L^{-1}[x]}(\rho\|\sigma)\right]. \tag{80c}$$

As argued above, a nice feature of this formulation is that we can describe all large-deviation results in a single unified way. There are three different regimes of large-deviation results, only two of which are captured in Fig. 4. Applying Stein's lemma (Lemma 3) alongside a dual version where we swap the states, we can see that there is a regime in which both error probabilities decay exponentially, albeit with exponents no greater than the respective relative entropies. If, however, one of the errors decays with an exponent greater than the relative entropy, then the other error will in fact exponentially approach 1. This is illustrated in Fig. 5, using our log-odds formulation.

Before we give the large-deviation bound, we need several additional definitions that will be critical for the pinched case. We define the *left-pinched* and *right-pinched* Rényi relative entropies as

$$\overleftarrow{D}_\alpha(\rho\|\sigma) := \lim_{n\to\infty}\frac{1}{n}D_\alpha\big(\mathcal{P}_{\sigma^{\otimes n}}\big(\rho^{\otimes n}\big)\,\big\|\,\sigma^{\otimes n}\big), \tag{81a}$$

$$\overrightarrow{D}_\alpha(\rho\|\sigma) := \lim_{n\to\infty}\frac{1}{n}D_\alpha\big(\rho^{\otimes n}\,\big\|\,\mathcal{P}_{\rho^{\otimes n}}\big(\sigma^{\otimes n}\big)\big). \tag{81b}$$

We note that due to the duality property of the classical relative entropy $(1-\alpha)D_\alpha(p\|q) = \alpha D_{1-\alpha}(q\|p)$, we straightforwardly have

$$(1-\alpha)\overleftarrow{D}_\alpha(\rho\|\sigma) = \alpha\overrightarrow{D}_{1-\alpha}(\sigma\|\rho). \tag{82}$$

We also note that for $\alpha\ge 0$, the left-pinched relative entropy coincides with the sandwiched relative entropy and for $\alpha\le 1$, the right-pinched relative entropy coincides with the reverse-sandwiched relative entropy [85,87], i.e.,

$$\overleftarrow{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1}\log\mathrm{Tr}\left(\left(\sqrt{\rho}\,\sigma^{\frac{1-\alpha}{\alpha}}\sqrt{\rho}\right)^{\alpha}\right)\quad\forall\alpha\ge 0, \tag{83a}$$

$$\overrightarrow{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1}\log\mathrm{Tr}\left(\left(\sqrt{\sigma}\,\rho^{\frac{\alpha}{1-\alpha}}\sqrt{\sigma}\right)^{1-\alpha}\right)\quad\forall\alpha\le 1, \tag{83b}$$

but we know of no closed-form solution for either outside of these ranges. In Appendix B, we show the existence and some important properties of these relative entropies. A consequence of these expressions is that the regular relative entropy can be recovered by taking the limits of α going to 1 and 0, respectively:

$$D(\rho\|\sigma) = \lim_{\alpha\to 1} \overleftarrow{D}_\alpha(\rho\|\sigma) = \lim_{\alpha\to 0} \frac{1-\alpha}{\alpha} \overrightarrow{D}_\alpha(\sigma\|\rho). \tag{84}$$

In our results, we will also need the counterpart quantity found when exchanging these limits, which we will denote by $D^\star(\rho\|\sigma)$, defined as

$$D^\star(\rho\|\sigma) := \lim_{\alpha\to 1} \overrightarrow{D}_\alpha(\rho\|\sigma) = \lim_{\alpha\to 0} \frac{1-\alpha}{\alpha} \overleftarrow{D}_\alpha(\sigma\|\rho). \tag{85}$$

With regard to this, we note that the data-processing inequality gives $D^\star(\rho\|\sigma) \le D(\rho\|\sigma)$, with equality if $[\rho,\sigma]=0$.

With these definitions in hand, we can now present the large-deviation bound on hypothesis testing.

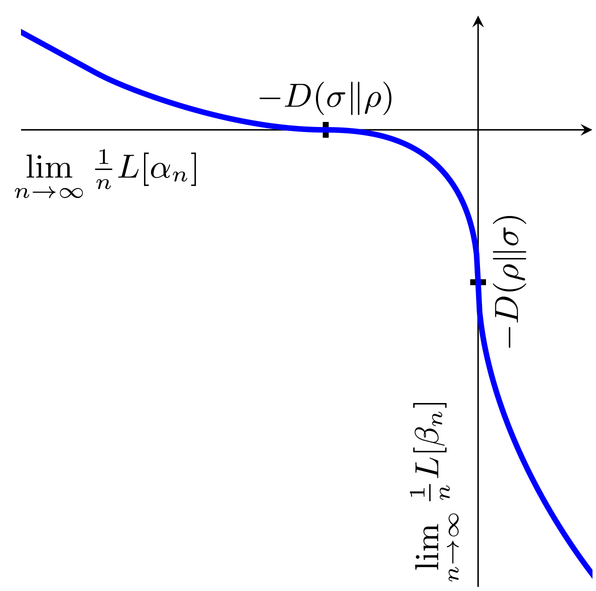


FIG. 5. The trade-off using log odds. The trade-off between the type-I and -II error log odds per copy—$\lim_{n\to\infty}(1/n)L[\alpha_n]$ and $\lim_{n\to\infty}(1/n)L[\beta_n]$, respectively—for the hypothesis test between $\rho^{\otimes n}$ and $\sigma^{\otimes n}$, in the limit of growing n. The bottom-left quadrant corresponds to the regime in which both errors are decaying exponentially, with exponents bound by the relative entropies $D(\sigma\|\rho)$ and $D(\rho\|\sigma)$, respectively. The top-left regime corresponds to a type-I error that is decaying even more rapidly, causing the type-II error to instead increase toward 1 and the bottom-right regime shows the converse of this. This curve is generated by plotting $\Gamma_\lambda(\rho\|\sigma)$ from Lemma 5 for two randomly generated $d=5$ qudit states.

Lemma 5 (Large-deviation analysis of hypothesis testing). For any $\lambda \in \mathbb{R}$, define the asymptotic nonpinched and pinched log-odds error per copy as

$$\Gamma_\lambda(\rho\|\sigma) := \lim_{n\to\infty} \frac{1}{n} \gamma_{\lambda n}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big), \tag{86a}$$

$$\overleftarrow{\Gamma}_\lambda(\rho\|\sigma) := \lim_{n\to\infty} \frac{1}{n} \overleftarrow{\gamma}_{\lambda n}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big), \tag{86b}$$

$$\overrightarrow{\Gamma}_\lambda(\rho\|\sigma) := \lim_{n\to\infty} \frac{1}{n} \overrightarrow{\gamma}_{\lambda n}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big). \tag{86c}$$

Then, all of these limits exist, and each is given by

$$\Gamma_\lambda(\rho\|\sigma) = \begin{cases} \sup_{t<0} \breve{D}_t(\rho\|\sigma) + \dfrac{t}{1-t}\lambda, & \lambda < -D(\sigma\|\rho), \\ \inf_{0<t<1} -\overline{D}_t(\rho\|\sigma) - \dfrac{t}{1-t}\lambda, & -D(\sigma\|\rho) < \lambda < 0, \\ \sup_{t>1} -\breve{D}_t(\rho\|\sigma) + \dfrac{t}{1-t}\lambda, & \lambda > 0, \end{cases} \tag{87a}$$

$$\overleftarrow{\Gamma}_\lambda(\rho\|\sigma) = \begin{cases} \sup_{t<0} \overleftarrow{D}_t(\rho\|\sigma) + \dfrac{t}{1-t}\lambda, & \lambda < -D^\star(\sigma\|\rho), \\ \inf_{0<t<1} -\overleftarrow{D}_t(\rho\|\sigma) - \dfrac{t}{1-t}\lambda, & -D^\star(\sigma\|\rho) < \lambda < 0, \\ \sup_{t>1} -\overleftarrow{D}_t(\rho\|\sigma) + \dfrac{t}{1-t}\lambda, & \lambda > 0, \end{cases} \tag{87b}$$

$$\overrightarrow{\Gamma}_\lambda(\rho\|\sigma) = \begin{cases} \sup_{t<0} \overrightarrow{D}_t(\rho\|\sigma) + \dfrac{t}{1-t}\lambda, & \lambda < -D(\sigma\|\rho), \\ \inf_{0<t<1} -\overrightarrow{D}_t(\rho\|\sigma) - \dfrac{t}{1-t}\lambda, & -D(\sigma\|\rho) < \lambda < 0, \\ \sup_{t>1} -\overrightarrow{D}_t(\rho\|\sigma) + \dfrac{t}{1-t}\lambda, & \lambda > 0, \end{cases} \tag{87c}$$

including the edge cases

$$\Gamma_{-D(\sigma\|\rho)}(\rho\|\sigma) = 0, \quad \Gamma_0(\rho\|\sigma) = -D(\rho\|\sigma), \tag{88a}$$

$$\overleftarrow{\Gamma}_{-D^\star(\sigma\|\rho)}(\rho\|\sigma) = 0, \quad \overleftarrow{\Gamma}_0(\rho\|\sigma) = -D(\rho\|\sigma), \tag{88b}$$

$$\overrightarrow{\Gamma}_{-D(\sigma\|\rho)}(\rho\|\sigma) = 0, \quad \overrightarrow{\Gamma}_0(\rho\|\sigma) = -D^\star(\rho\|\sigma) \tag{88c}$$

and the limits

$$\Gamma_{\pm\infty}(\rho\|\sigma) = \overleftarrow{\Gamma}_{\pm\infty}(\rho\|\sigma) = \overrightarrow{\Gamma}_{\pm\infty}(\rho\|\sigma) = \mp\infty. \tag{89}$$

Proof. We start by dividing the range of λ into three parameter regions, corresponding to the quadrants of

Fig. 5:

$$\mathcal{R}_L := \big(-\infty, -D(\sigma\|\rho)\big), \tag{90a}$$
$$\mathcal{R}_M := \big(-D(\sigma\|\rho), 0\big), \tag{90b}$$
$$\mathcal{R}_R := (0, \infty). \tag{90c}$$

We note that the region $\mathcal{R}_M$ corresponds to the so-called "error-exponent" regime, in which both the type-I and type-II errors are exponentially decreasing, whereas $\mathcal{R}_L/\mathcal{R}_R$ are referred to as the "strong-converse" regime, in which one error is exponentially decaying but the other is exponentially approaching 1. Regions $\mathcal{R}_M$ and $\mathcal{R}_L$ have been previously studied: in fact, the expressions for Γ_λ on $\mathcal{R}_M$ can be derived from Refs. [109,110] and the expressions for Γ_λ on $\mathcal{R}_L$ can be derived from Refs. [111–113].

By swapping the states, we can extend the result in $\mathcal{R}_L$ to $\mathcal{R}_R$. Recall that $\Gamma_\gamma(\rho\|\sigma)$ quantifies the optimal type-II error possible for a given type-I error. As swapping the states corresponds to exchanging the two error types, this means that $\Gamma_\gamma(\sigma\|\rho)$ must therefore quantify the optimal type-I error given for a given type-II error. As such, we can think of the two functions

$$\lambda \mapsto \Gamma_\lambda(\rho\|\sigma) \quad \text{and} \quad \lambda \mapsto \Gamma_\lambda(\sigma\|\rho), \tag{91}$$

as both describing the trade-off between two types of error, as functions of the type-I and -II errors, respectively, and they are therefore inverses [114]. So, to find an expression for Γ_λ on $\mathcal{R}_R$, we simply need to find the inverse of the flipped version on $\lambda \le D(\rho\|\sigma)$. To this end, we define

$$f(\lambda) := \sup_{t>1} \frac{1-t}{t}\Big[\lambda + \breve{D}_t(\rho\|\sigma)\Big], \tag{92a}$$
$$g(\lambda) := \sup_{s>1} \frac{s}{1-s}\lambda - \breve{D}_s(\rho\|\sigma). \tag{92b}$$

By utilizing the identity

$$\breve{D}_\alpha(\rho\|\sigma) = \frac{\alpha}{1-\alpha}\breve{D}_{1-\alpha}(\sigma\|\rho), \tag{93}$$

we can see that $f(\lambda) = \Gamma_\lambda(\sigma\|\rho)$ for $\lambda \le -D(\rho\|\sigma)$ and we now want to argue that g is its inverse. Consider a composition of f and g, which gives

$$(f\circ g)(\lambda) = \sup_{t>1}\inf_{s>1} \frac{1-t}{t}\left[\breve{D}_t(\rho\|\sigma) - \breve{D}_s(\rho\|\sigma) + \frac{s}{1-s}\lambda\right], \tag{94a}$$

$$(g\circ f)(\lambda) = \sup_{s>1}\inf_{t>1} \frac{s}{1-s}\frac{1-t}{t}\Big[\lambda + \breve{D}_t(\rho\|\sigma)\Big] - \breve{D}_s(\rho\|\sigma). \tag{94b}$$

If we let $t = s$ and $s = t$ in the inner optimizations in each of these, we obtain $(f\circ g)(\lambda) \ge \lambda$ and $(g\circ f)(\lambda) \ge \lambda$, respectively. Using these, together with the fact that f is monotonically nonincreasing, we have

$$f\circ g\circ f = (f\circ g)\circ f \ge f, \tag{95a}$$
$$f\circ g\circ f = f\circ(g\circ f) \le f. \tag{95b}$$

Thus, we have that $f\circ g\circ f = f$, and so f and g are quasiinverses. As f is $\Gamma_\lambda(\sigma\|\rho)$ on $\lambda \le -D(\rho\|\sigma)$, then g must correspond to $\Gamma_\lambda(\rho\|\sigma)$ for $\lambda \in \mathcal{R}_R$, as required.

Next, we turn to the pinched variants $\overleftarrow{\Gamma}_\lambda/\overrightarrow{\Gamma}_\lambda$—we will start with $\overleftarrow{\Gamma}_\lambda$. Here, we want to consider the hypothesis testing between the pinched state $\mathcal{P}_{\sigma^{\otimes n}}\big(\rho^{\otimes n}\big)$ and $\sigma^{\otimes n}$. As these states are no longer independent identically distributed (IID), this can be considered as a hypothesis-testing problem between two correlated states. Thankfully, the problem of extending the previously mentioned hypothesis-testing analyses to the case of correlated states has been considered. Specifically, Ref. [115, Theorem 4.8] considers the error-exponent regime ($\mathcal{R}_M$) and Ref. [112, Corollary IV.6] the strong-converse regime ($\mathcal{R}_L/\mathcal{R}_R$). In both cases, it has been shown that if the regularized relative entropy exists and is differentiable (see Appendix B), then the standard IID results can be extended, where the single-copy relative entropy is replaced with this regularized quantity. For our case, looking at pinched states, this means that the change when going from nonpinched to pinched hypothesis testing is

$$\overline{D}(\rho\|\sigma), \quad \breve{D}(\rho\|\sigma) \to \overleftarrow{D}(\rho\|\sigma). \tag{96}$$

Making this substitution, we obtain the required expression for $\overleftarrow{\Gamma}_\lambda$ on $\mathcal{R}_L$, $\mathcal{R}_M$, and $\mathcal{R}_R$. An analogous argument can also be made for $\overrightarrow{\Gamma}_\lambda$.

The only values of λ left to consider are the edge cases and limits. In each case, these follow from the monotonicity of $\Gamma_\lambda/\overleftarrow{\Gamma}_\lambda/\overrightarrow{\Gamma}_\lambda$ in λ. The edge cases

$$\Gamma_{-D(\sigma\|\rho)}(\rho\|\sigma) = 0, \quad \Gamma_0(\rho\|\sigma) = -D(\rho\|\sigma) \tag{97}$$

correspond to Lemma 3 (and its state-reversed analogue) and, similarly,

$$\overleftarrow{\Gamma}_{-D^\star(\sigma\|\rho)}(\rho\|\sigma) = 0, \quad \overleftarrow{\Gamma}_0(\rho\|\sigma) = -D(\rho\|\sigma), \tag{98a}$$
$$\overrightarrow{\Gamma}_{-D(\sigma\|\rho)}(\rho\|\sigma) = 0, \quad \overrightarrow{\Gamma}_0(\rho\|\sigma) = -D^\star(\rho\|\sigma) \tag{98b}$$

to the pinched variants of Stein's lemma. Lastly, the limits

$$\Gamma_{\pm\infty}(\rho\|\sigma) = \overleftarrow{\Gamma}_{\pm\infty}(\rho\|\sigma) = \overrightarrow{\Gamma}_{\pm\infty}(\rho\|\sigma) = \mp\infty \tag{99}$$

follow from the fact that $\beta_0(\cdot\|\cdot) = \overleftarrow{\beta}_0(\cdot\|\cdot) = \overrightarrow{\beta}_0(\cdot\|\cdot) = 1$ and $\beta_1(\cdot\|\cdot) = \overleftarrow{\beta}_1(\cdot\|\cdot) = \overrightarrow{\beta}_1(\cdot\|\cdot) = 0$ generically for any states of full support. ■

3. Moderate deviation

In the small-deviation regime, we have considered type-I errors that are constant and in the large-deviation regime, we have considered type-I errors that are exponentially approaching 0/1. This leaves a gap for errors that are approaching 0 or 1 but doing so subexponentially. This is referred to as the *moderate-deviation regime* [116–118]. As with large deviations, it will be advantageous to express these results in terms of type-I and/or type-II log odds.

Whereas "small deviation" corresponds to type-I log odds that are constant in n and "large deviation" to any log odds that scale as $\pm\lambda n$, "moderate" will refer to any log odds that scale as $\pm\lambda n^a$, where $\lambda > 0$ and $a \in (0, 1)$. We note that in some other papers considering moderate deviations—such as Refs. [117,118]—these results are considered more generally for any sequence x_n such that $\lim_{n\to\infty} x_n = 0$ and $\lim_{n\to\infty} nx_n = \pm\infty$. We will restrict to this polynomial subset of such sequences primarily for notational convenience but note that all of the results below can be extended to these more general moderate sequences.

Lemma 6 (Moderate-deviation analysis of hypothesis testing). For any $\lambda > 0$ and $a \in (0, 1)$, the type-II log odds of error scale as

$$\frac{1}{n}\gamma_{\pm\lambda n^a}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big) \simeq -D(\rho\|\sigma) \mp \sqrt{2V(\rho\|\sigma)\lambda n^{a-1}}, \tag{100a}$$

$$\frac{1}{n}\overleftarrow{\gamma}_{\pm\lambda n^a}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big) \simeq -D(\rho\|\sigma) \mp \sqrt{2V(\rho\|\sigma)\lambda n^{a-1}}, \tag{100b}$$

where $\simeq$ denotes equality up to terms scaling as $o\left(\sqrt{n^{a-1}}\right)$.

Proof. The nonpinched result is just a restatement of the hypothesis-testing result from Ref. [117, Theorem 1]. For the pinched quantity, we will use a similar argument to that used in the small-deviation case of Lemma 4.

The data-processing inequality $\overleftarrow{\gamma}_x(\rho\|\sigma) \geq \gamma_x(\rho\|\sigma)$ trivially gives us a lower bound on the scaling of $\overleftarrow{\gamma}$. For the upper bound, we return to the inequality

$$\overleftarrow{\beta}_\epsilon(\rho\|\sigma) \leq \beta_{\epsilon-2\delta}(\rho\|\sigma) \times \frac{2^8(\epsilon-\delta)\nu(\sigma)^2}{\delta^5(1-\epsilon+\delta)} \tag{101}$$

considered previously in the proof of Lemma 4. We want to use this bound in the two moderate regimes, in which the type-I error is approaching 0/1. To do this, consider the two equalities when we substitute $\delta = \epsilon/4$ for the case of ϵ approaching 0 and $\delta = 1-\epsilon$ for the case of ϵ approaching 1 (with the added requirement that $\epsilon > 3/4$), giving

$$\overleftarrow{\beta}_\epsilon(\rho\|\sigma) \leq \beta_{\epsilon/2}(\rho\|\sigma) \times \frac{3\times 2^{16}\nu(\sigma)^2}{\epsilon^4(1-3\epsilon/4)}, \tag{102a}$$

$$\overleftarrow{\beta}_\epsilon(\rho\|\sigma) \leq \beta_{3\epsilon-2}(\rho\|\sigma) \times \frac{2^7(2\epsilon-1)\nu(\sigma)^2}{(1-\epsilon)^6}. \tag{102b}$$

Now, we want to make the substitutions $\rho \to \rho^{\otimes n}$ and $\sigma \to \sigma^{\otimes n}$. In the former case, we will substitute $\epsilon \to L^{-1}[-\lambda n^a]$ and in the latter $\epsilon \to L^{-1}[+\lambda n^a]$. Taking logarithms and recalling that the number of unique eigenvalues $\nu\left(\sigma^{\otimes n}\right)$ scales only polynomially with n, this gives us the upper bounds expressed in terms of log odds per copy of

$$\frac{1}{n}\overleftarrow{\gamma}_{\pm\lambda n^a}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big) \leq \frac{1}{n}\gamma_{\pm\lambda n^a}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big) + O(n^{a-1}). \tag{103}$$

Lastly, as $a < 1$, we have that $n^{a-1} = o\left(\sqrt{n^{a-1}}\right)$, allowing us to neglect this error term, and thus we conclude that the pinched and nonpinched log odds per copy must scale identically up to $\simeq$, as required. ■

4. Extreme deviation

Now that we have dealt with type-I errors that do not approach 0 or 1, approach them subexponentially, and approach them exponentially, we are left with one final case: when the error approaches 0 or 1 *super*exponentially. We will see that if we consider superexponentially scaling errors, the problem of hypothesis testing becomes "boring," in the sense that we obtain a simple linear trade-off between the two types of error, as the error constraints are too strict for any meaningful trade-off. While boring in and of itself, the analysis of hypothesis testing in this regime is needed for technical reasons in the proof of the zero-error transformation rates between quantum dichotomies to come in Sec. V C 4.

One silver lining of the boringness of this regime is that we do not need to consider an asymptotic number of states and we can start with a single-shot statement. The quantum Neyman-Pearson lemma [26,119] states that the trade-off between type-I and type-II hypothesis-testing error can be characterized by Neyman-Pearson tests of the form

$$Q_t := \{\rho - t\sigma > 0\}, \tag{104}$$

for $t \geq 0$, where $\{M > 0\}$ denotes the projector onto the eigenspaces of M corresponding to positive eigenvalues. Specifically, the claim is that the optimal trade-off between the errors is either given by a test of the form Q_t or, when t corresponds to a value at which Q_t changes rank, a convex combination of $\lim_{s\to t^-} Q_s$ and Q_t.

Recall that the $\alpha \to \pm\infty$ limiting cases of the minimal divergence are given by the max-divergence [87, Sec. 4.2],

$$\breve{D}_{+\infty}(\rho\|\sigma) = \log \lambda_{\max}\left(\sigma^{-1/2}\rho\sigma^{-1/2}\right), \tag{105a}$$

$$\breve{D}_{-\infty}(\rho\|\sigma) = -\log \lambda_{\max}\left(\rho^{-1/2}\sigma\rho^{-1/2}\right). \tag{105b}$$

Lemma 7 (Single-shot extreme-deviation analysis of hypothesis testing). For any $x \le \lambda_{\min}(\rho)$,

$$\beta_{1-x}(\rho\|\sigma) = x \times \exp\left(-\breve{D}_{+\infty}(\rho\|\sigma)\right), \tag{106a}$$

$$1 - \beta_x(\rho\|\sigma) = x \times \exp\left(-\breve{D}_{-\infty}(\rho\|\sigma)\right) \tag{106b}$$

and

$$\overrightarrow{\beta}_{1-x}(\rho\|\sigma) = x \times \exp\Big(-D_{+\infty}\big(\rho\big\|\mathcal{P}_\rho(\sigma)\big)\Big), \tag{107a}$$

$$1 - \overrightarrow{\beta}_x(\rho\|\sigma) = x \times \exp\Big(-D_{-\infty}\big(\rho\big\|\mathcal{P}_\rho(\sigma)\big)\Big). \tag{107b}$$

Also, for any $x \le \lambda_{\min}\left(\mathcal{P}_\sigma(\rho)\right)$,

$$\overleftarrow{\beta}_{1-x}(\rho\|\sigma) = x \times \exp\Big(-D_{+\infty}(\mathcal{P}_\sigma(\rho)\|\sigma)\Big), \tag{108a}$$

$$1 - \overleftarrow{\beta}_x(\rho\|\sigma) = x \times \exp\Big(-D_{-\infty}(\mathcal{P}_\sigma(\rho)\|\sigma)\Big). \tag{108b}$$

Proof. First, we note that we can rewrite the Neyman-Pearson test Q_t as

$$Q_t = \left\{\rho^{1/2}\left(I - t\rho^{-1/2}\sigma\rho^{-1/2}\right)\rho^{1/2} > 0\right\}. \tag{109}$$

Clearly, $Q_0 = I$ but from this we can also see that $Q_t = I$ for any $t < t^*$, where

$$t^* := 1/\lambda_{\max}(\rho^{-1/2}\sigma\rho^{-1/2}) = \exp\left(\breve{D}_{-\infty}(\rho\|\sigma)\right). \tag{110}$$

As such, we can see that the first nontrivial projective Neyman-Pearson test is given by Q_{t^*}. Since Q_t are necessarily not full rank for any $t > t^*$, they must have a type-I error of at least $\lambda_{\min}(\rho)$. Thus, to obtain a type-I error of $0 < x < \lambda_{\min}(\rho)$, we must consider a test that is a convex combination of $Q_0 = I$ and Q_{t^*}.

Assume for the moment that $\rho^{-1/2}\sigma\rho^{-1/2}$ has a nondegenerate maximal eigenvector (we will return to this below), with eigenvalue $1/t^*$ and eigenvector $|\psi\rangle$. Then, this first nontrivial projective test is

$$Q_{t^*} = I - \frac{\rho^{-1/2}|\psi\rangle\langle\psi|\rho^{-1/2}}{\langle\psi|\rho^{-1}|\psi\rangle}. \tag{111}$$

So a test Q that is a convex combination of Q_0 and Q_{t^*}, and has type-I error of x, takes the form

$$Q := I - x \times \rho^{-1/2}|\psi\rangle\langle\psi|\rho^{-1/2}. \tag{112}$$

As $0 \le x < \lambda_{\min}(\rho)$, we have that this is a valid test and the type-I error is simply given by x as required:

$$1 - \mathrm{Tr}(\rho Q) = x. \tag{113}$$

For the type-II error, we obtain the desired expression:

$$\beta_x(\rho\|\sigma) = \mathrm{Tr}(\sigma Q) \tag{114a}$$

$$= 1 - x \times \mathrm{Tr}(\sigma\rho^{-1/2}|\psi\rangle\langle\psi|\rho^{-1/2}) \tag{114b}$$

$$= 1 - x \times \rangle\psi|\rho^{-1/2}\sigma\rho^{-1/2}|\psi\rangle \tag{114c}$$

$$= 1 - x/t^* \tag{114d}$$

$$= 1 - x \times \exp\left(-\breve{D}_{-\infty}(\rho\|\sigma)\right). \tag{114e}$$

For $x > 1 - \lambda_{\min}(\rho)$, a similar argument can be given *mutatis mutandis* by considering the last nontrivial test, which gives

$$\beta_x(\rho\|\sigma) = (1-x)\exp\left(-\breve{D}_{+\infty}(\rho\|\sigma)\right). \tag{115}$$

Finally, the pinched results trivially follow from the non-pinched variants by making the substitution $\sigma \to \mathcal{P}_\rho(\sigma)$ and $\rho \to \mathcal{P}_\sigma(\rho)$, respectively.

In the above, we have assumed that $\rho^{-1/2}\sigma\rho^{-1/2}$ has a nondegenerate maximal eigenvalue—an assumption to which we now return. The idea now is to show that we can perturb the state σ by an arbitrarily small amount to break any such degeneracy. Specifically, consider letting $|\psi\rangle$ be an arbitrary maximal eigenvector of $\rho^{-1/2}\sigma\rho^{-1/2}$ and define the (unnormalized) state σ_ϵ as

$$\sigma_\epsilon := \sigma + \epsilon\rho^{1/2}|\psi\rangle\langle\psi|\rho^{1/2}. \tag{116}$$

We can see that this breaks the degeneracy as

$$\rho^{-1/2}\sigma_\epsilon\rho^{-1/2} = \rho^{-1/2}\sigma\rho^{-1/2} + \epsilon|\psi\rangle\langle\psi|, \tag{117}$$

allowing us to apply the above proof to give expressions for $\beta_x(\rho\|\sigma_\epsilon)$. Next, we can see that $\beta_x(\rho\|\sigma_\epsilon) \to \beta_x(\rho\|\sigma)$ as $\epsilon \to 0^+$ as the difference is bounded by the trace norm,

$$|\beta_x(\rho\|\sigma_\epsilon) - \beta_x(\rho\|\sigma)| \le \|\sigma_\epsilon - \sigma\|_{\mathrm{Tr}} \tag{118a}$$

$$= \epsilon\langle\psi|\rho|\psi\rangle \tag{118b}$$

$$\le \epsilon. \tag{118c}$$

■

Applying this single-shot analysis to the case of an asymptotically large number of copies of each state, we can obtain asymptotic expressions for the log odds per copy that are comparable with the large-deviation expressions of Lemma 5, for both nonpinched and pinched hypothesis-testing problems.

Lemma 8 (Extreme-deviation analysis of hypothesis testing). For any $\lambda > -\log \lambda_{\min}(\rho)$,

$$\Gamma_{\pm\lambda}(\rho\|\sigma) = \mp\lambda - \check{D}_{\pm\infty}(\rho\|\sigma), \tag{119a}$$

$$\overleftarrow{\Gamma}_{\pm\lambda}(\rho\|\sigma) = \mp\lambda - \overleftarrow{D}_{\pm\infty}(\rho\|\sigma), \tag{119b}$$

$$\overrightarrow{\Gamma}_{\pm\lambda}(\rho\|\sigma) = \mp\lambda - \overrightarrow{D}_{\pm\infty}(\rho\|\sigma), \tag{119c}$$

where we recall that

$$\Gamma_{\lambda}(\rho\|\sigma) := \lim_{n\to\infty} \frac{1}{n}\gamma_{+\lambda n}(\rho\|\sigma), \tag{120a}$$

$$\overleftarrow{\Gamma}_{\lambda}(\rho\|\sigma) := \lim_{n\to\infty} \frac{1}{n}\overleftarrow{\gamma}_{+\lambda n}(\rho\|\sigma), \tag{120b}$$

$$\overrightarrow{\Gamma}_{\lambda}(\rho\|\sigma) := \lim_{n\to\infty} \frac{1}{n}\overrightarrow{\gamma}_{+\lambda n}(\rho\|\sigma). \tag{120c}$$

Proof. For the nonpinched case, we can simply apply the single-shot result (Lemma 7) to the states $\rho^{\otimes n}$ and $\sigma^{\otimes n}$, using the additivity of the $\check{D}_{\pm\infty}$, and expressing the type-I and -II errors in terms of log odds.

We will do similarly for the pinched case. First, we use the pinching inequality

$$\mathcal{P}_{\sigma^{\otimes n}}\left(\rho^{\otimes n}\right) \geq \frac{\rho^{\otimes n}}{|\mathrm{spec}(\sigma^{\otimes n})|} \geq \frac{\rho^{\otimes n}}{n^d} \tag{121}$$

and so

$$\log \lambda_{\min}\left(\mathcal{P}_{\sigma^{\otimes n}}\left(\rho^{\otimes n}\right)\right) \geq n\lambda_{\min}(\rho) - O(\log n). \tag{122}$$

Thus, if we have a *strict* inequality $\lambda > -\log \lambda_{\min}(\rho)$, then

$$n\lambda \overset{\text{ev.}}{>} -\log \lambda_{\min}\left(\mathcal{P}_{\sigma^{\otimes n}}\left(\rho^{\otimes n}\right)\right). \tag{123}$$

Given this, we can now substitute the pinched states into Lemma 7, which gives the desired expressions for $\overleftarrow{\Gamma}_{\pm\lambda}$ and $\overrightarrow{\Gamma}_{\pm\lambda}$. ■

C. Transformation rates

In this section, we will take the asymptotic analysis of hypothesis testing from the previous section and extend it to transformation rates between quantum dichotomies. To be more concrete, for some sequence of errors ϵ_n and fixed states ρ_1, ρ_2, σ_1, and σ_2, recall the definition of $R_n^*(\epsilon_n)$ as the maximum R_n such that

$$(\rho_1^{\otimes n}, \sigma_1^{\otimes n}) \succeq_{(\epsilon_n,0)} (\rho_2^{\otimes R_n n}, \sigma_2^{\otimes R_n n}). \tag{124}$$

We will be studying the scaling of $R_n^*(\epsilon_n)$ for various scaling regimes of ϵ_n. While we will restrict below to just the case of such one-sided errors, we cover how these techniques can be extended, and what the resulting rates are, in the more general regime of two-sided errors in Appendix C.

To spare the reader from being subjected to the phrase "for sufficiently large n" *ad nauseam*, we use the notation $\overset{\text{ev.}}{<}$ and $\overset{\text{ev.}}{>}$ to denote eventual inequalities for the following proofs. In other words, we will use $a_n \overset{\text{ev.}}{<} b_n$ and $b_n \overset{\text{ev.}}{>} a_n$ as shorthand for

$$\exists N:\ a_n < b_n\ \forall n \geq N. \tag{125}$$

We note that if the quantities are functions, then this will just denote pointwise eventual inequality, i.e., $f_n(x) \overset{\text{ev.}}{<} g_n(x)$ is shorthand for

$$\forall x\ \exists N(x):\ f_n(x) < g_n(x)\ \forall n \geq N(x) \tag{126}$$

and not

$$\exists N:\ f_n(x) < g_n(x)\ \forall x, \forall n \geq N. \tag{127}$$

Upgrading from such pointwise inequalities to uniform inequalities will be important in the achievability proofs to come, which will require uniform versions of the lemmas in Sec. V B, which are presented in Appendix I.

We start with the first-order rate, since all of the remaining results in this section are refinements of this first-order rate. Moreover, all of the remaining proofs will follow a general approach that extends the proof below. We will quantify the optimal-transformation-rates regime by providing both upper and lower bounds, referred to as the *optimality* and *achievability* bounds, respectively. In all cases, these bounds will follow from Lemma 2, which provides both necessary and sufficient conditions for the existence of a transformation in terms of hypothesis-testing quantities.

The first-order transformation rate is captured by the following theorem.

Theorem 1 (Restated) (First-order rate). For constant $\epsilon \in (0,1)$ and $[\rho_2, \sigma_2] = 0$, the optimal rate converges:

$$\lim_{n\to\infty} R_n^*(\epsilon) = \frac{D(\rho_1\|\sigma_1)}{D(\rho_2\|\sigma_2)}. \tag{128}$$

Furthermore, if we consider more general target dichotomies, $[\rho_2, \sigma_2] \neq 0$, then we still have the upper bound:

$$\limsup_{n\to\infty} R_n^*(\epsilon) \leq \frac{D(\rho_1\|\sigma_1)}{D(\rho_2\|\sigma_2)}. \tag{129}$$

Proof. We start with optimality. Consider a rate $R > D(\rho_1\|\sigma_1)/D(\rho_2\|\sigma_2)$. As $\epsilon \in (0,1)$, we also have that $(1 \pm \epsilon)/2 \in (0,1)$, allowing us to apply Lemma 3. On the input side, this gives

$$\lim_{n\to\infty} -\frac{1}{n}\log \beta_{\frac{1+\epsilon}{2}}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) = D(\rho_1\|\sigma_1) \tag{130}$$

and on the target side

$$\begin{aligned} &\lim_{n\to\infty} -\frac{1}{n}\log \beta_{\frac{1-\epsilon}{2}}\left(\rho_2^{\otimes Rn}\middle\|\sigma_2^{\otimes Rn}\right) \\ &= \lim_{m\to\infty} -\frac{R}{m}\log \beta_{\frac{1-\epsilon}{2}}\left(\rho_2^{\otimes m}\middle\|\sigma_2^{\otimes m}\right) &(131a)\\ &= RD(\rho_2\|\sigma_2) &(131b)\\ &> D(\rho_1\|\sigma_1). &(131c)\end{aligned}$$

If we let $x = (1+\epsilon)/2$, then

$$\beta_x\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{\text{ev.}}{>} \beta_{x-\epsilon}\left(\rho_2^{\otimes Rn}\middle\|\sigma_2^{\otimes Rn}\right), \tag{132}$$

and so by Lemma 2 this means that transformation at a rate of R is eventually *not* possible. Thus, R provides an upper bound for asymptotic optimal transformation rate. As this argument holds for *any* $R > D(\rho_1\|\sigma_1)/D(\rho_2\|\sigma_2)$ this means that

$$\limsup_{n\to\infty} R_n^*(\epsilon) \le \frac{D(\rho_1\|\sigma_1)}{D(\rho_2\|\sigma_2)}. \tag{133}$$

Next, we proceed to proving the achievability for commuting target dichotomies, $[\rho_2,\sigma_2]=0$. Consider a rate $r < D(\rho_1\|\sigma_1)/D(\rho_2\|\sigma_2)$. To show that a rate is achievable, we need to consider the pinched hypothesis testing of the input dichotomy. Specifically, we will use the limits from Lemma 4:

$$\lim_{n\to\infty} -\frac{1}{n}\log \overleftarrow{\beta}_{\epsilon}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) = D(\rho_1\|\sigma_1), \tag{134a}$$

$$\lim_{n\to\infty} -\frac{1}{n}\log \beta_{1-\epsilon}\left(\rho_2^{\otimes rn}\middle\|\sigma_2^{\otimes rn}\right) = rD(\rho_2\|\sigma_2). \tag{134b}$$

As $r < D(\rho_1\|\sigma_1)/D(\rho_2\|\sigma_2)$, combining these gives

$$\overleftarrow{\beta}_{\epsilon}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{\text{ev.}}{<} \beta_{1-\epsilon}\left(\rho_2^{\otimes rn}\middle\|\sigma_2^{\otimes rn}\right). \tag{135}$$

But this is only true for specific errors, while Lemma 2 requires such an inequality for *all* x in a range. How do we span this gap? For the first-order problem (and high-error cases of moderate and large deviations), this is easily solved by simply using the monotonicity of $\beta_x(\cdot\|\cdot)$, $\overleftarrow{\beta}_x(\cdot\|\cdot)$ and $\overrightarrow{\beta}_x(\cdot\|\cdot)$ as functions of x for fixed states. We can use this to relax the preceding inequality to

$$\overleftarrow{\beta}_x\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{\text{ev.}}{<} \beta_{x-\epsilon}\left(\rho_2^{\otimes rn}\middle\|\sigma_2^{\otimes rn}\right) \quad \forall x \in (\epsilon, 1). \tag{136}$$

We note that because this set of inequalities (parametrized by x) follows from the previous x-independent inequality, there is no issue of uniformity, i.e., there exists an x-independent N such that this holds for $n \ge N$. As such, we can apply Lemma 2, which gives that transformation at a rate of r *is* eventually achievable and thus

$$\liminf_{n\to\infty} R_n^*(\epsilon) \ge \frac{D(\rho_1\|\sigma_1)}{D(\rho_2\|\sigma_2)}. \tag{137}$$

■

As with hypothesis testing in Sec. V B, we will now spend the rest of this subsection giving our refinements on this first-order result for different regimes of the scaling of the error ϵ_n. A summary of these different regimes is given in Fig. 1.

1. Small deviation

We start with the small-deviation regime, in which the error is a constant $\epsilon \in (0,1)$. The proof is broadly similar to that of the first-order rate in Theorem 1 but more care has to be taken to capture the second-order contribution, especially on the achievability side. Also, recall that Eq. (24) defines the reversibility parameter as

$$\xi := \frac{V(\rho_1\|\sigma_1)}{D(\rho_1\|\sigma_1)} \bigg/ \frac{V(\rho_2\|\sigma_2)}{D(\rho_2\|\sigma_2)}. \tag{138}$$

Then, we have the following.

Theorem 2 (Restated) (Small-deviation rate). Let $\lesssim$ / $\simeq$ denote (in)equality up to $o(1/\sqrt{n})$. For constant $\epsilon \in (0,1)$ and $[\rho_2,\sigma_2]=0$, the optimal rate is

$$R_n^*(\epsilon) \simeq \frac{D(\rho_1\|\sigma_1) + \sqrt{V(\rho_1\|\sigma_1)/n} \times S_{1/\xi}^{-1}(\epsilon)}{D(\rho_2\|\sigma_2)}. \tag{139}$$

Furthermore, if we consider general output dichotomies, $[\rho_2,\sigma_2] \neq 0$, then we still have the upper bound

$$R_n^*(\epsilon) \lesssim \frac{D(\rho_1\|\sigma_1) + \sqrt{V(\rho_1\|\sigma_1)/n} \times S_{1/\xi}^{-1}(\epsilon)}{D(\rho_2\|\sigma_2)}. \tag{140}$$

Proof. Consider a small slack parameter $\delta > 0$ and define the rate R_n as

$$R_n := \frac{D(\rho_1\|\sigma_1) + \sqrt{V(\rho_1\|\sigma_1)/n} \times S^{-1}_{1/\xi}(\epsilon) + \delta/\sqrt{n}}{D(\rho_2\|\sigma_2)}. \tag{141}$$

Recall that by Lemma 1, the term $S^{-1}_{1/\xi}(\epsilon)$ can be expressed as a minimum,

$$S^{-1}_{1/\xi}(\epsilon) = \min_{x\in(\epsilon,1)} \left[\Phi^{-1}(x) - \sqrt{1/\xi} \times \Phi^{-1}(x-\epsilon)\right]. \tag{142}$$

Let $x^* > \epsilon$ denote the value of x at which this minimum is attained, such that

$$R_n = \frac{D(\rho_1\|\sigma_1)}{D(\rho_2\|\sigma_2)} + \frac{\delta}{\sqrt{n}D(\rho_2\|\sigma_2)} + \sqrt{\frac{V(\rho_1\|\sigma_1)}{nD^2(\rho_2\|\sigma_2)}}\Phi^{-1}(x^*) - \sqrt{\frac{V(\rho_2\|\sigma_2)D(\rho_1\|\sigma_1)}{nD^3(\rho_2\|\sigma_2)}}\Phi^{-1}(x^*-\epsilon). \tag{143}$$

Next, we turn to Lemma 4. For the input dichotomy, this gives

$$-\frac{1}{n}\log \beta_{x^*}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \simeq D(\rho_1\|\sigma_1) + \sqrt{\frac{V(\rho_1\|\sigma_1)}{n}}\Phi^{-1}(x^*), \tag{144}$$

and for the target we can substitute in R_n to obtain

$$\begin{aligned} -\frac{1}{n}\log \beta_{x^*-\epsilon}\left(\rho_2^{\otimes R_n n}\middle\|\sigma_2^{\otimes R_n n}\right) &\simeq R_n D(\rho_2\|\sigma_2) + \sqrt{\frac{R_n V(\rho_2\|\sigma_2)}{n}}\Phi^{-1}(x^*-\epsilon) && \text{(145a)}\\ &= D(\rho_1\|\sigma_1) + \sqrt{\frac{V(\rho_1\|\sigma_1)}{n}}\Phi^{-1}(x^*) + \frac{\delta}{\sqrt{n}} && \text{(145b)}\\ &\simeq -\frac{1}{n}\log\beta_{x^*}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) + \frac{\delta}{\sqrt{n}}. && \text{(145c)} \end{aligned}$$

Due to this $\delta > 0$ term, we can therefore conclude that

$$\beta_{x^*}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{\text{ev.}}{>} \beta_{x^*-\epsilon}\left(\rho_2^{\otimes R_n n}\middle\|\sigma_2^{\otimes R_n n}\right) \tag{146}$$

and so by Lemma 2 the transformations at the rate R_n are eventually *not* possible. As this is true for all $\delta > 0$, this then upper bounds the optimal rate

$$R_n^*(\epsilon) \lesssim \frac{D(\rho_1\|\sigma_1) + \sqrt{V(\rho_1\|\sigma_1)/n} \times S^{-1}_{1/\xi}(\epsilon)}{D(\rho_2\|\sigma_2)}. \tag{147}$$

We now turn to achievability. Once again, consider a small slack parameter $0 < \delta < \epsilon/2$ and define the rate r_n as

$$\begin{aligned} r_n := \frac{D(\rho_1\|\sigma_1)}{D(\rho_2\|\sigma_2)} - \frac{\delta}{\sqrt{n}D(\rho_2\|\sigma_2)} &+ \frac{1}{\sqrt{n}}\min_{y\in[\epsilon,1]}\left[\frac{\sqrt{V(\rho_1\|\sigma_1)}}{D(\rho_2\|\sigma_2)}\Phi^{-1}(y-\delta)\right.\\ &\left. - \sqrt{\frac{V(\rho_2\|\sigma_2)D(\rho_1\|\sigma_1)}{D(\rho_2\|\sigma_2)^3}}\Phi^{-1}(y-\epsilon+\delta)\right]. \end{aligned} \tag{148}$$

Again, for convenience, let y^* denote a minimizer in the above optimization over y. If we let $x \in (\epsilon, 1)$ then applying Lemma 4 to the input gives

$$-\frac{1}{n}\log \overleftarrow{\beta}_{x-\delta}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \simeq D(\rho_1\|\sigma_1) + \sqrt{\frac{V(\rho_1\|\sigma_1)}{n}} \times \Phi^{-1}(x-\delta) \tag{149}$$

and to the target gives

$$\begin{aligned} -\frac{1}{n}\log \beta_{x-\epsilon+\delta}\left(\rho_2^{\otimes r_n n}\middle\|\sigma_2^{\otimes r_n n}\right) &\simeq r_n D(\rho_2\|\sigma_2) + \sqrt{\frac{r_n V(\rho_2\|\sigma_2)}{n}} \times \Phi^{-1}(x-\epsilon+\delta), && \text{(150a)}\\ &\simeq D(\rho_1\|\sigma_1) - \delta/\sqrt{n} + \sqrt{\frac{V(\rho_2\|\sigma_2)D(\rho_1\|\sigma_1)}{nD(\rho_2\|\sigma_2)}}\Phi^{-1}(x-\epsilon+\delta)\\ &\quad + \sqrt{\frac{V(\rho_1\|\sigma_1)}{n}}\Phi^{-1}(y^*-\delta) - \sqrt{\frac{V(\rho_2\|\sigma_2)D(\rho_1\|\sigma_1)}{nD(\rho_2\|\sigma_2)}}\Phi^{-1}(y^*-\epsilon+\delta). && \text{(150b)} \end{aligned}$$

Combining these, we have

$$\begin{aligned}
&-\frac{1}{n}\log\beta_{x-\epsilon+\delta}\left(\rho_2^{\otimes r_n n}\middle\|\sigma_2^{\otimes r_n n}\right)\\
&\simeq -\frac{1}{n}\log\overleftarrow{\beta}_{x-\delta}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right)-\delta/\sqrt{n}\\
&\quad-\sqrt{\frac{V(\rho_1\|\sigma_1)}{n}}\Phi^{-1}(x-\delta)\\
&\quad+\sqrt{\frac{V(\rho_2\|\sigma_2)D(\rho_1\|\sigma_1)}{nD(\rho_2\|\sigma_2)}}\Phi^{-1}(x-\epsilon+\delta)\\
&\quad-\sqrt{\frac{V(\rho_2\|\sigma_2)D(\rho_1\|\sigma_1)}{nD(\rho_2\|\sigma_2)}}\Phi^{-1}(y^*-\epsilon+\delta)\\
&\quad+\sqrt{\frac{V(\rho_1\|\sigma_1)}{n}}\Phi^{-1}(y^*-\delta). \qquad (151)
\end{aligned}$$

Recalling that y^* has been defined as the minimizer over just such an expression, we obtain

$$\begin{aligned}
&-\frac{1}{n}\log\beta_{x-\epsilon+\delta}\left(\rho_2^{\otimes r_n n}\middle\|\sigma_2^{\otimes r_n n}\right)\\
&\qquad\lesssim -\frac{1}{n}\log\overleftarrow{\beta}_{x-\delta}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right)-\delta/\sqrt{n}. \qquad (152)
\end{aligned}$$

Because of $\delta>0$, this in turn implies the eventual inequality

$$\overleftarrow{\beta}_{x-\delta}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right)\overset{\text{ev.}}{<}\beta_{x-\epsilon+\delta}\left(\rho_2^{\otimes r_n n}\middle\|\sigma_2^{\otimes r_n n}\right). \qquad (153)$$

Finally, we can relax out the slack parameter, once again using the fact that β_x and $\overleftarrow{\beta}_x$ are monotone decreasing as functions of x, giving

$$\overleftarrow{\beta}_{x}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right)\leq\beta_{x-\epsilon}\left(\rho_2^{\otimes r_n n}\middle\|\sigma_2^{\otimes r_n n}\right). \qquad (154)$$

We note that in the above proof we have skipped over the issue of uniformity. While we did show that Eq. (154) holds for all x eventually, it still remains to be seen that it eventually holds for all x, the latter of which would be required to apply Lemma 2. To put it less confusingly, we have shown that there exists an N such that Eq. (154) holds for all $n>N$ but we have not ruled out the possibility that N depends on x. Such a dependence might mean that there is no x-independent N beyond which this expression holds for all x, which is what would be needed by Lemma 2. However, if we swap out Lemma 4 with its uniform version (Lemma 21, presented and proven in Appendix I), then we do indeed obtain an N that is independent of x (though still dependent on ρ_1, ρ_2, σ_1, σ_2, ϵ, and δ, of course), removing this issue. Given this, we can now conclude that Eq. (154) eventually holds for all $x\in(\epsilon,1)$. Having dealt with this uniformity issue, we can we can return to Lemma 2, which allows us to conclude that transformation at the rate r_n *is* eventually possible. This held true for all small $\delta>0$. Taking the limit $\delta\to 0^+$ and using the continuity of Φ^{-1} on $(0,1)$, we can see that this rate does indeed limit to the desired expression, lower bounding the optimal rate as

$$R_n^*(\epsilon_n)\gtrsim\frac{D(\rho_1\|\sigma_1)+\sqrt{V(\rho_1\|\sigma_1)/n}\times S_{1/\xi}^{-1}(\epsilon)}{D(\rho_2\|\sigma_2)}. \qquad (155)$$

■

2. Large deviation

We now turn to the large-deviation regime, in which we consider errors that are exponentially approaching either 0 or 1, which we refer to as low and high error. The general structure of the proof follows that of the small-deviation case but will be split into two subregimes: high and low error. In high error, we have that ϵ_n is exponentially close to 1 and so the region $x\in(\epsilon_n,1)$ is quite small. This allows us to obtain an optimal expression for the transformation rate with a single application of the large-deviation analysis of hypothesis testing Lemma 5, similar to the proof of the first-order rate (Theorem 1). For the low-error case, however, the region $x\in(\epsilon_n,1)$ is quite large, requiring us to consider hypothesis testing for a whole interval of possible error exponents. Here, the proof will more closely follow that of the small-deviation case (Theorem 2), running into the same uniformity issues. As the high-error proof is simpler, we shall start there.

Theorem 5 (Restated) (Large-deviation rate, high error). For any error of the form $\epsilon_n=1-\exp(-\lambda n)$ with constant $\lambda>0$, if $[\rho_2,\sigma_2]=0$, then the optimal rate is

$$\lim_{n\to\infty}R_n^*(\epsilon_n)=\inf_{\substack{t_1>1\\0<t_2<1}}\frac{\breve{D}_{t_1}(\rho_1\|\sigma_1)+\left(\frac{t_1}{t_1-1}+\frac{t_2}{1-t_2}\right)\lambda}{D_{t_2}(\rho_2\|\sigma_2)}. \qquad (156)$$

Furthermore, if we consider general output dichotomies, $[\rho_2,\sigma_2]\neq 0$, then we still have the upper bound

$$\limsup_{n\to\infty}R_n^*(\epsilon_n)\leq\inf_{\substack{t_1>1\\0<t_2<1}}\frac{\breve{D}_{t_1}(\rho_1\|\sigma_1)+\left(\frac{t_1}{t_1-1}+\frac{t_2}{1-t_2}\right)\lambda}{\overline{D}_{t_2}(\rho_2\|\sigma_2)}. \qquad (157)$$

Proof. Consider a rate R such that

$$R>\inf_{\substack{t_1>1\\0<t_2<1}}\frac{\breve{D}_{t_1}(\rho_1\|\sigma_1)+\left(\frac{t_1}{t_1-1}+\frac{t_2}{1-t_2}\right)\lambda}{\overline{D}_{t_2}(\rho_2\|\sigma_2)}. \qquad (158)$$

Rearranging this gives

$$\sup_{t_1>1} -\check{D}_{t_1}(\rho_1\|\sigma_1) + \frac{t_1}{1-t_1}\lambda > \inf_{0<t_2<1} -R\overline{D}_{t_2}(\rho_2\|\sigma_2) + \frac{t_2}{1-t_2}\lambda. \tag{159}$$

Recalling the definition of Γ from Lemma 5, this is equivalent to

$$\Gamma_{+\lambda}(\rho_1\|\sigma_1) > R\Gamma_{-\lambda/R}(\rho_2\|\sigma_2). \tag{160}$$

The idea now is to connect this rate to the large-deviation hypothesis-testing quantities from Lemma 5. Consider hypothesis testing of the input or target with type-I error log odds of $\pm\lambda$. Lemma 5 gives, for the input dichotomy,

$$\lim_{n\to\infty} \frac{1}{n}\gamma_{+\lambda n}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) = \Gamma_{+\lambda}(\rho_1\|\sigma_1), \tag{161}$$

and for the target dichotomy,

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n}\gamma_{-\lambda n}\left(\rho_2^{\otimes Rn}\middle\|\sigma_1^{\otimes Rn}\right) &= \lim_{m\to\infty} \frac{R}{m}\gamma_{-\lambda m/R}\left(\rho_2^{\otimes m}\middle\|\sigma_1^{\otimes m}\right) &(162a)\\ &= R\Gamma_{-\lambda/R}(\rho_2\|\sigma_2). &(162b)\end{aligned}$$

We can put the above limits back in terms of the type-II error *probability* as

$$\lim_{n\to\infty} \frac{1}{n}L\left[\beta_{L^{-1}[-\lambda n]}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right)\right] = \Gamma_{+\lambda}(\rho_1\|\sigma_1), \tag{163a}$$

$$\lim_{n\to\infty} \frac{1}{n}L\left[\beta_{L^{-1}[+\lambda n]}\left(\rho_2^{\otimes Rn}\middle\|\sigma_2^{\otimes Rn}\right)\right] = R\Gamma_{-\lambda/R}(\rho_2\|\sigma_2), \tag{163b}$$

where we recall that $L[x] := \log(x/(1-x))$. As ϵ_n is exponentially approaching 1 with an exponent of λ, $(1\pm\epsilon_n)/2$ are exponentially approaching 0 and 1 respectively, both also with an exponent of λ. Putting this in terms of log odds as we did in Sec. V B 2, this means that

$$\lim_{n\to\infty} \frac{1}{n}L\left[\frac{1\pm\epsilon_n}{2}\right] = \pm\lambda, \tag{164}$$

Using the uniformity of the large-deviation analysis of hypothesis testing shown in Lemma 22 then implies that Eq. (163) can be extended to

$$\lim_{n\to\infty} \frac{1}{n}L\left[\beta_{\frac{1+\epsilon_n}{2}}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right)\right] = \Gamma_{+\lambda}(\rho_1\|\sigma_1), \tag{165a}$$

$$\lim_{n\to\infty} \frac{1}{n}L\left[\beta_{\frac{1-\epsilon_n}{2}}\left(\rho_2^{\otimes Rn}\middle\|\sigma_2^{\otimes Rn}\right)\right] = R\Gamma_{-\lambda/R}(\rho_2\|\sigma_2). \tag{165b}$$

Recalling Eq. (160) and using the monotonicity of $L[\cdot]$, we then have

$$\beta_{x_n}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{\text{ev.}}{>} \beta_{x_n-\epsilon_n}\left(\rho_2^{\otimes Rn}\middle\|\sigma_2^{\otimes Rn}\right) \tag{166}$$

for $x_n := (1+\epsilon_n)/2$. By Lemma 2, this means that transformation at a rate of R is eventually *not* possible. As this held for *any* R above satisfying Eq. (158), this means that this implies a corresponding upper bound on the optimal rate,

$$\limsup_{n\to\infty} R_n^*(\epsilon_n) \le \inf_{\substack{t_1>1\\0<t_2<1}} \frac{\check{D}_{t_1}(\rho_1\|\sigma_1) + \left(\frac{t_1}{t_1-1}+\frac{t_2}{1-t_2}\right)\lambda}{\overline{D}_{t_2}(\rho_2\|\sigma_2)}, \tag{167}$$

as required.

Next, consider a rate r such that

$$r < \inf_{\substack{t_1>1\\0<t_2<1}} \frac{\overleftarrow{D}_{t_1}(\rho_1\|\sigma_1) + \left(\frac{t_1}{t_1-1}+\frac{t_2}{1-t_2}\right)\lambda}{D_{t_2}(\rho_2\|\sigma_2)}. \tag{168}$$

Similar to optimality, this can be rearranged into the inequality

$$\overleftarrow{\Gamma}_{+\lambda}(\rho_1\|\sigma_1) < r\Gamma_{-\lambda/r}(\rho_2\|\sigma_2) \tag{169}$$

and so

$$\overleftarrow{\gamma}_{+\lambda n}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{\text{ev.}}{<} \gamma_{-\lambda n}\left(\rho_2^{\otimes rn}\middle\|\sigma_2^{\otimes rn}\right) \tag{170}$$

or, equivalently,

$$\overleftarrow{\beta}_{L^{-1}[+\lambda n]}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{\text{ev.}}{<} \beta_{L^{-1}[-\lambda n]}\left(\rho_2^{\otimes rn}\middle\|\sigma_2^{\otimes rn}\right). \tag{171}$$

Recalling that $\epsilon_n = 1-\exp(-\lambda n)$, we have

$$\lim_{n\to\infty} \frac{1}{n}L[1-\epsilon_n] = -\lambda, \tag{172a}$$

$$\lim_{n\to\infty} \frac{1}{n}L[\epsilon_n] = +\lambda \tag{172b}$$

and therefore we eventually have

$$\overleftarrow{\beta}_{\epsilon_n}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{\text{ev.}}{<} \beta_{1-\epsilon_n}\left(\rho_2^{\otimes rn}\middle\|\sigma_2^{\otimes rn}\right). \tag{173}$$

Lastly, we use monotonicity of the type-II error, which allows us to relax this to the broader inequality

$$\overleftarrow{\beta}_x\big(\rho_1^{\otimes n}\big\|\sigma_1^{\otimes n}\big)\overset{\text{ev.}}{<}\beta_{x-\epsilon_n}\big(\rho_2^{\otimes rn}\big\|\sigma_2^{\otimes rn}\big), \tag{174}$$

for all $x \in (\epsilon_n, 1)$. We note that unlike with the proof of the small-deviation rate (see Lemma 4), there is no concern about the uniformity of the asymptotic analysis of hypothesis testing, as we only need to apply the large-deviation analysis in Lemma 5 for a single error exponent λ. This diversion aside, we can now apply Lemma 2, which allows us to conclude that the rate r *is* eventually achievable. As this was true for any rate of the form Eq. (168), this implies a corresponding lower bound on the optimal rate,

$$\liminf_{n\to\infty} R_n^*(\epsilon_n) \geq \inf_{\substack{t_1>1\\0<t_2<1}} \frac{\breve{D}_{t_1}(\rho_1\|\sigma_1)+\left(\frac{t_1}{t_1-1}+\frac{t_2}{1-t_2}\right)\lambda}{D_{t_2}(\rho_2\|\sigma_2)}. \tag{175}$$

■

Next, we turn to the trickier case of low error. Before stating and proving the result, we will need some definitions. As we saw in the high-error case, the proof came down to satisfying inequalities of the form

$$\Gamma_{+\lambda}\big(\rho_1^{\otimes n}\big\|\rho_2^{\otimes n}\big) \leq R\Gamma_{-\lambda/R}\big(\rho_2^{\otimes Rn}\big\|\rho_2^{\otimes Rn}\big), \tag{176a}$$

$$\overleftarrow{\Gamma}_{+\lambda}\big(\rho_1^{\otimes n}\big\|\rho_2^{\otimes n}\big) \leq R\Gamma_{-\lambda/R}\big(\rho_2^{\otimes Rn}\big\|\rho_2^{\otimes Rn}\big), \tag{176b}$$

$$\overrightarrow{\Gamma}_{+\lambda}\big(\rho_1^{\otimes n}\big\|\rho_2^{\otimes n}\big) \leq R\Gamma_{-\lambda/R}\big(\rho_2^{\otimes Rn}\big\|\rho_2^{\otimes Rn}\big). \tag{176c}$$

In the low-error case, we will need inequalities of the form

$$\Gamma_{+\mu}\big(\rho_1^{\otimes n}\big\|\rho_2^{\otimes n}\big) \leq R\Gamma_{+\mu/R}\big(\rho_2^{\otimes Rn}\big\|\rho_2^{\otimes Rn}\big), \tag{177a}$$

$$\overleftarrow{\Gamma}_{+\mu}\big(\rho_1^{\otimes n}\big\|\rho_2^{\otimes n}\big) \leq R\Gamma_{+\mu/R}\big(\rho_2^{\otimes Rn}\big\|\rho_2^{\otimes Rn}\big), \tag{177b}$$

$$\overrightarrow{\Gamma}_{+\mu}\big(\rho_1^{\otimes n}\big\|\rho_2^{\otimes n}\big) \leq R\Gamma_{+\mu/R}\big(\rho_2^{\otimes Rn}\big\|\rho_2^{\otimes Rn}\big). \tag{177c}$$

Importantly, for the low-error case, we will need to satisfy these not just for a single μ but for all $-\lambda \leq \mu \leq \lambda$ (for details, see the proof below). As such, it will be helpful to define the rates that saturate the above inequalities. Specifically, let $\bar{r}(\mu)$, $\overleftarrow{r}(\mu)$, and $\overrightarrow{r}(\mu)$ denote the largest rates satisfying these inequalities, in the nonpinched, left-pinched, and right-pinched cases, respectively. By expanding the definitions of Γ, $\overleftarrow{\Gamma}$, and $\overrightarrow{\Gamma}$, one can come up with explicit formulations of these rates, which share their piecewise structure; specifically,

$$\bar{r}(\mu) := \begin{cases} r_1(\mu), & \mu < -D(\sigma_1\|\rho_1), \\ \bar{r}_2(\mu), & -D(\sigma_1\|\rho_1) < \mu < 0, \\ r_3(\mu), & \mu > 0, \end{cases} \tag{178a}$$

$$\overleftarrow{r}(\mu) := \begin{cases} \overleftarrow{r}_1(\mu), & \mu < -D^\star(\sigma_1\|\rho_1), \\ \overleftarrow{r}_2(\mu), & -D^\star(\sigma_1\|\rho_1) < \mu < 0, \\ r_3(\mu), & \mu > 0, \end{cases} \tag{178b}$$

$$\overrightarrow{r}(\mu) := \begin{cases} r_1(\mu), & \mu < -D(\sigma_1\|\rho_1), \\ \overrightarrow{r}_2(\mu), & -D(\sigma_1\|\rho_1) < \mu < 0, \\ \overrightarrow{r}_3(\mu), & \mu > 0, \end{cases} \tag{178c}$$

where

$$r_1(\mu) := \sup_{t_2<0}\inf_{t_1<0} \frac{-\breve{D}_{t_1}(\rho_1\|\sigma_1)+\left(\frac{t_1}{t_1-1}-\frac{t_2}{t_2-1}\right)\mu}{-\breve{D}_{t_2}(\rho_2\|\sigma_2)}, \tag{179a}$$

$$\bar{r}_2(\mu) := \inf_{0<t_2<1}\sup_{0<t_1<1} \frac{\overline{D}_{t_1}(\rho_1\|\sigma_1)+\left(\frac{t_1}{1-t_1}-\frac{t_2}{1-t_2}\right)\mu}{\overline{D}_{t_2}(\rho_2\|\sigma_2)}, \tag{179b}$$

$$r_3(\mu) := \sup_{t_2>1}\inf_{t_1>1} \frac{\breve{D}_{t_1}(\rho_1\|\sigma_1)+\left(\frac{t_1}{t_1-1}-\frac{t_2}{t_2-1}\right)\mu}{\breve{D}_{t_2}(\rho_2\|\sigma_2)}, \tag{179c}$$

and

$$\overleftarrow{r}_1(\mu) := \sup_{t_2<0}\inf_{t_1<0} \frac{-\overleftarrow{D}_{t_1}(\rho_1\|\sigma_1)+\left(\frac{t_1}{t_1-1}-\frac{t_2}{t_2-1}\right)\mu}{-D_{t_2}(\rho_2\|\sigma_2)}, \tag{180a}$$

$$\overleftarrow{r}_2(\mu) := \inf_{0<t_2<1}\sup_{0<t_1<1} \frac{\overleftarrow{D}_{t_1}(\rho_1\|\sigma_1)+\left(\frac{t_1}{1-t_1}-\frac{t_2}{1-t_2}\right)\mu}{D_{t_2}(\rho_2\|\sigma_2)}, \tag{180b}$$

$$\overrightarrow{r}_2(\mu) := \inf_{0<t_2<1}\sup_{0<t_1<1} \frac{\overrightarrow{D}_{t_1}(\rho_1\|\sigma_1)+\left(\frac{t_1}{1-t_1}-\frac{t_2}{1-t_2}\right)\mu}{D_{t_2}(\rho_2\|\sigma_2)}, \tag{180c}$$

$$\overrightarrow{r}_3(\mu) := \sup_{t_2>1}\inf_{t_1>1} \frac{\overrightarrow{D}_{t_1}(\rho_1\|\sigma_1)+\left(\frac{t_1}{t_1-1}-\frac{t_2}{t_2-1}\right)\mu}{D_{t_2}(\rho_2\|\sigma_2)}. \tag{180d}$$

Lastly, similar to $\bar{r}(\mu)$, we define $\breve{r}(\mu)$ as

$$\breve{r}(\mu) := \begin{cases} r_1(\mu), & \mu < -D(\sigma_1\|\rho_1), \\ \breve{r}_2(\mu), & -D(\sigma_1\|\rho_1) < \mu < 0, \\ r_3(\mu), & \mu > 0, \end{cases} \tag{181}$$

where

$$\breve{r}_2(\mu) := \inf_{0<t_2<1} \sup_{0<t_1<1} \frac{\breve{D}_{t_1}(\rho_1\|\sigma_1) + \left(\frac{t_1}{1-t_1} - \frac{t_2}{1-t_2}\right)\mu}{D_{t_2}(\rho_2\|\sigma_2)}. \tag{182}$$

These definitions in hand, we can turn to the low-error rate.

Theorem 4 (Restated) (Large-deviation rate, low error). For any error of the form $\epsilon_n = \exp(-\lambda n)$ with constant $\lambda > 0$, if $[\rho_2, \sigma_2] = 0$, then the optimal rate is lower bounded by

$$\liminf_{n\to\infty} R_n^*(\epsilon_n) \geq \min_{-\lambda\leq\mu\leq\lambda} \breve{r}(\mu). \tag{183}$$

Furthermore, if we consider general output dichotomies, $[\rho_2, \sigma_2] \neq 0$, then the optimal rate is upper bounded by

$$\limsup_{n\to\infty} R_n^*(\epsilon_n) \leq \min_{-\lambda\leq\mu\leq\lambda} \bar{r}(\mu). \tag{184}$$

In the above, $\bar{r}$ and $\breve{r}$ are defined in terms of Rényi relative entropies in Sec. V C 2 and coincide when $[\rho_1, \sigma_1] = [\rho_2, \sigma_2] = 0$.

Proof. We once again start with optimality. Let $0 < \delta < \lambda$ be a small constant and consider a rate R such that

$$R > \min_{-\lambda+\delta\leq\mu\leq\lambda-\delta} \bar{r}(\mu). \tag{185}$$

This means that there exists a $-\lambda < \mu^* < \lambda$ such that $R > \bar{r}(\mu^*)$ and therefore that

$$\Gamma_{\mu^*}(\rho_1\|\sigma_1) > R\Gamma_{\mu^*/R}(\rho_2\|\sigma_2). \tag{186}$$

By a similar chain of reasoning to that used in the proof of Theorem 5, this implies that

$$\beta_{L^{-1}[\mu^* n]}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{\text{ev.}}{>} \beta_{L^{-1}[\mu^* n]}\left(\rho_2^{\otimes Rn}\middle\|\sigma_2^{\otimes Rn}\right). \tag{187}$$

Let $x_n := L^{-1}[\mu^* n]$. As $\mu^* > -\lambda$, we have that x_n dominates over ϵ_n and so the monotonicity of β_x allows us to relax this to

$$\beta_{x_n}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{\text{ev.}}{>} \beta_{x_n-\epsilon_n}\left(\rho_2^{\otimes Rn}\middle\|\sigma_2^{\otimes Rn}\right). \tag{188}$$

By Lemma 2, this means that transformation at a rate of R is eventually *not* possible. If we now take $\delta \to 0^+$, this gives an upper bound on the optimal rate of

$$\limsup_{n\to\infty} R_n^*(\epsilon_n) \leq \min_{-\lambda\leq\mu\leq\lambda} \bar{r}(\mu). \tag{189}$$

We now turn to achievability. Let $\delta > 0$ be a small constant and consider a rate r such that

$$r < \min_{-\lambda-\delta\leq\mu\leq\lambda+\delta} \overleftarrow{r}(\mu). \tag{190}$$

This means that

$$\overleftarrow{\Gamma}_\mu(\rho_1\|\sigma_1) < r\Gamma_{\mu/r}(\rho_2\|\sigma_2) \tag{191}$$

and thus, by Lemma 5,

$$\overleftarrow{\beta}_{L^{-1}[\mu n]}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{\text{ev.}}{<} \beta_{L^{-1}[\mu n]}\left(\rho_2^{\otimes Rn}\middle\|\sigma_2^{\otimes Rn}\right), \tag{192}$$

for any $-\lambda - \delta \leq \mu \leq \lambda + \delta$. Note that applying Lemma 5 only gives this convergence pointwise but if we swap this out for the uniform version (Lemma 22, given in Appendix I), then this can be strengthened to a uniform statement. Specifically, we obtain that for sufficiently large n, this holds for all μ such that $|\mu| \leq \lambda + \delta$ in that range. Recalling that $\epsilon_n := \exp(-\lambda n)$ and therefore corresponds to log odds per copy of $-\lambda$, we can see that for any probability $y_n \in (\epsilon_n/2, 1 - \epsilon_n/2)$, we have

$$\frac{1}{n} L(y_n) \in [-\lambda, \lambda] \subset (-\lambda - \delta, \lambda + \delta) \tag{193}$$

for sufficiently large n. As such, we have

$$\overleftarrow{\beta}_{y_n}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{\text{ev.}}{<} \beta_{y_n}\left(\rho_2^{\otimes Rn}\middle\|\sigma_2^{\otimes Rn}\right), \tag{194}$$

for all such y_n. Use of the monotonicity of β_x and $\overleftarrow{\beta}_x$ allows us to relax this to

$$\overleftarrow{\beta}_{y_n-\epsilon_n/2}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{\text{ev.}}{<} \beta_{y_n+\epsilon_n/2}\left(\rho_2^{\otimes Rn}\middle\|\sigma_2^{\otimes Rn}\right). \tag{195}$$

Lastly, we shift this by $x_n := y_n - \epsilon/2$, which yields

$$\overleftarrow{\beta}_{y_n-\epsilon_n/2}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{\text{ev.}}{<} \beta_{y_n+\epsilon_n/2}\left(\rho_2^{\otimes Rn}\middle\|\sigma_2^{\otimes Rn}\right), \tag{196}$$

for $x_n \in (\epsilon_n, 1)$. As in deriving this inequality we have employed not just the pointwise Lemma 5 but also the uniform Lemma 22, we therefore have it uniformly, which allows us to utilize Lemma 2. This in turn tells us that transformation at rate r *is* eventually possible. Taking

$\delta \to 0^+$, this yields the corresponding lower bound on the optimal rate of

$$\liminf_{n\to\infty} R_n^*(\epsilon_n) \geq \min_{-\lambda\leq\mu\leq\lambda} \overleftarrow{r}(\mu). \tag{197}$$

Repeating the above argument for $\overrightarrow{r}(\mu)$ also gives an achievability bound

$$\liminf_{n\to\infty} R_n^*(\epsilon_n) \geq \min_{-\lambda\leq\mu\leq\lambda} \overrightarrow{r}(\mu). \tag{198}$$

Combining these gives

$$\liminf_{n\to\infty} R_n^*(\epsilon_n) \geq \min_{-\lambda\leq\mu\leq\lambda} \max\left\{\overleftarrow{r}(\mu), \overrightarrow{r}(\mu)\right\}. \tag{199}$$

By applying Lemma 13, it can be shown that

$$\max\{\overleftarrow{r}(\mu), \overrightarrow{r}(\mu)\} = \check{r}(\mu), \tag{200}$$

giving the final achievability bound

$$\liminf_{n\to\infty} R_n^*(\epsilon_n) \geq \min_{-\lambda\leq\mu\leq\lambda} \check{r}(\mu). \tag{201}$$

■

3. Moderate deviation

So far, we have considered constant error and exponentially decaying error, which leaves a gap of errors that decay subexponentially, known as the moderate-deviation regime. Much like the large-deviation case, this will contain a slightly easier high-error case and a slightly trickier low-error case and the proof will follow as a streamlined version of the proof used for Theorems 4 and 5. Recall from Eq. (24) that the reversibility parameter is defined as

$$\xi := \frac{V(\rho_1\|\sigma_1)}{D(\rho_1\|\sigma_1)} \Bigg/ \frac{V(\rho_2\|\sigma_2)}{D(\rho_2\|\sigma_2)}. \tag{202}$$

Then, we have the following.

Theorem 3 (Restated) (Moderate-deviation rate). Consider an $a \in (0,1)$, and let $\lesssim / \simeq$ denote (in)equality up to $o\left(\sqrt{n^{a-1}}\right)$. Let $\epsilon_n := \exp(-\lambda n^a)$ for some $\lambda > 0$. For $[\rho_2, \sigma_2] = 0$, the optimal rate is

$$R_n^*(\epsilon_n) \simeq \frac{D(\rho_1\|\sigma_1) - |1-\xi^{-1/2}|\sqrt{2\lambda V(\rho_1\|\sigma_1) n^{a-1}}}{D(\rho_2\|\sigma_2)}, \tag{203a}$$

$$R_n^*(1-\epsilon_n) \simeq \frac{D(\rho_1\|\sigma_1) + \left[1+\xi^{-1/2}\right]\sqrt{2\lambda V(\rho_1\|\sigma_1) n^{a-1}}}{D(\rho_2\|\sigma_2)}. \tag{203b}$$

Furthermore, if we consider general output dichotomies, $[\rho_2, \sigma_2] \neq 0$, then we still have the upper bounds

$$R_n^*(\epsilon_n) \lesssim \frac{D(\rho_1\|\sigma_1) - |1-\xi^{-1/2}|\sqrt{2\lambda V(\rho_1\|\sigma_1) n^{a-1}}}{D(\rho_2\|\sigma_2)}, \tag{204a}$$

$$R_n^*(1-\epsilon_n) \lesssim \frac{D(\rho_1\|\sigma_1) + \left[1+\xi^{-1/2}\right]\sqrt{2\lambda V(\rho_1\|\sigma_1) n^{a-1}}}{D(\rho_2\|\sigma_2)}. \tag{204b}$$

Proof. We begin with the more involved low-error case of $R_n^*(\epsilon_n)$, returning to the high-error case of $R_n^*(1-\epsilon_n)$ at the end of the proof. As is customary, we start with optimality. Let $0 < \lambda' < \lambda$ be a constant and consider a rate R_n defined as

$$R_n := \frac{D(\rho_1\|\sigma_1) - |1-\xi^{-1/2}|\sqrt{2\lambda' V(\rho_1\|\sigma_1) n^{a-1}}}{D(\rho_2\|\sigma_2)}. \tag{205}$$

Applying Lemma 6 to the input state, we obtain

$$\frac{1}{n}\gamma_{\pm\lambda n^a}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \simeq -D(\rho_1\|\sigma_1) \mp \sqrt{2\lambda V(\rho_1\|\sigma_1) n^{a-1}}, \tag{206}$$

and for the target state, we have

$$\begin{aligned}&\frac{1}{n}\gamma_{\pm\lambda n^a}\left(\rho_2^{\otimes R_n n}\middle\|\sigma_2^{\otimes R_n n}\right)\\ &\quad\simeq -R_n D(\rho_1\|\sigma_1) \mp \sqrt{2\lambda R_n V(\rho_2\|\sigma_2) n^{a-1}}.\end{aligned} \tag{207a}$$

Taking the difference of these and expanding out R_n gives

$$\begin{aligned}&\frac{1}{n}\gamma_{\pm\lambda n^a}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) - \frac{1}{n}\gamma_{\pm\lambda n^a}\left(\rho_2^{\otimes R_n n}\middle\|\sigma_2^{\otimes R_n n}\right)\\ &\quad\simeq -\sqrt{2\lambda' V(\rho_1\|\sigma_1) n^{a-1}}|1-\xi^{-1/2}|\\ &\quad\mp \sqrt{2\lambda V(\rho_1\|\sigma_1) n^{a-1}}\left[1-\xi^{-1/2}\right].\end{aligned} \tag{208a}$$

So if we take $s := \operatorname{sgn}(\xi^{-1/2}-1)$ to be the sign that makes this second term positive, then $\lambda' < \lambda$ tells us that this term must asymptotically dominate and, as such, we can conclude that

$$\frac{1}{n}\gamma_{s\lambda n^a}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{\text{ev.}}{>} \frac{1}{n}\gamma_{s\lambda n^a}\left(\rho_2^{\otimes R_n n}\middle\|\sigma_2^{\otimes R_n n}\right). \tag{209}$$

Recalling that $\epsilon_n := \exp(-\lambda n^a)$, this means that ϵ_n has asymptotic log odds per copy of $-\lambda n^a$, as does $2\epsilon_n$. Using this, we can reexpress the above in terms of the type-II error probabilities as

$$\xi > 1: \quad \beta_{2\epsilon_n}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{\text{ev.}}{>} \beta_{\epsilon_n}\left(\rho_2^{\otimes R_n n}\middle\|\sigma_2^{\otimes R_n n}\right), \quad (210a)$$

$$\xi < 1: \quad \beta_{1-\epsilon_n}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{\text{ev.}}{>} \beta_{1-2\epsilon_n}\left(\rho_2^{\otimes R_n n}\middle\|\sigma_2^{\otimes R_n n}\right). \quad (210b)$$

By Lemma 2, this means that transformation at a rate of R_n is asymptotically not possible and thus that

$$R_n^*(\epsilon_n) \lesssim \frac{D(\rho_1\|\sigma_1) - |1-\xi^{-1/2}|\sqrt{2\lambda V(\rho_1\|\sigma_1)n^{a-1}}}{D(\rho_2\|\sigma_2)}. \quad (211)$$

The achievability proof follows similarly. The idea is that the bottleneck will appear once at log odds of $\pm\lambda n^{a-1}$ and to satisfy both, the rate will require an absolute value around the term $1-\xi^{1/2}$. For this to give achievability, we will need to use the uniform version of the moderate-deviation analysis of hypothesis testing (Lemma 23, presented in Appendix I).

As for the high-error case, this sign issue does not arise. In this case, we can follow an approach similar to the achievability proof of Theorem 1. By using Lemma 6, we can show that

$$\beta_{\frac{1+\epsilon_n}{2}}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{\text{ev.}}{>} \beta_{\frac{1-\epsilon_n}{2}}\left(\rho_2^{\otimes R_n n}\middle\|\sigma_1^{\otimes R_n n}\right) \quad (212)$$

for an appropriately chosen rate R_n that will yield the optimality bound and for achievability we first show that

$$\overleftarrow{\beta}_{\epsilon_n}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{\text{ev.}}{<} \beta_{1-\epsilon_n}\left(\rho_2^{\otimes r_n n}\middle\|\sigma_1^{\otimes r_n n}\right) \quad (213)$$

and then use monotonicity to extend this to the ordering required by Lemma 2. ∎

4. Extreme deviation

The argument for the zero-error case follows similarly to the low-error large-deviation case.

Theorem 6 (Restated) (Zero-error rate). For $[\rho_2,\sigma_2] = 0$, the optimal zero-error rate is lower bounded:

$$\liminf_{n\to\infty} R_n^*(0) \geq \max\left\{\inf_{\alpha\in\mathbb{R}} \frac{\overleftarrow{D}_\alpha(\rho_1\|\sigma_1)}{D_\alpha(\rho_2\|\sigma_2)}, \inf_{\alpha\in\mathbb{R}} \frac{\overrightarrow{D}_\alpha(\rho_1\|\sigma_1)}{D_\alpha(\rho_2\|\sigma_2)}\right\}, \quad (214)$$

where the divergences $\overleftarrow{D}_\alpha$ and $\overrightarrow{D}_\alpha$ are defined in Eqs. (81a) and (81b). More generally, if $[\rho_2,\sigma_2] \neq 0$, then the optimal transformation rate for all n is upper bounded:

$$R_n^*(0) \leq \min_{\alpha\in\mathbb{R}} \frac{\breve{D}_\alpha(\rho_1\|\sigma_1)}{\breve{D}_\alpha(\rho_2\|\sigma_2)}. \quad (215)$$

Proof. As this is the zero-error case, the optimality side is pretty straightforward. Any additive and data-processing quantity $Q(\cdot\|\cdot)$ puts a single-shot bound on the largest possible transformation rate for all n of the form

$$R_n^*(\epsilon) \leq \frac{Q(\rho_1\|\sigma_1)}{Q(\rho_2\|\sigma_2)}. \quad (216)$$

If we consider the minimal relative entropies $\breve{D}_\alpha$, this gives

$$R_n^*(0) \leq \min_{\alpha\in\mathbb{R}} \frac{\breve{D}_\alpha(\rho_1\|\sigma_1)}{\breve{D}_\alpha(\rho_2\|\sigma_2)}. \quad (217)$$

In the case of coherent outputs, one could include other possible monotones Q, which could constrain the zero-error rate further.

Now, we turn to the tricky part, achievability. Consider a rate constant r such that

$$r < \inf_{\alpha\in\mathbb{R}} \frac{\overleftarrow{D}_\alpha(\rho_1\|\sigma_1)}{D_\alpha(\rho_2\|\sigma_2)}. \quad (218)$$

We note that this rate is almost of the form that we want but involves the pinched relative entropy and not the minimal and is therefore suboptimal—we will return to this. We want to prove that

$$\overleftarrow{\beta}_x\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{!}{<} \beta_x\left(\rho_2^{\otimes rn}\middle\|\sigma_2^{\otimes rn}\right) \quad (219)$$

eventually holds for all x. To do this, we will need to combine both the extreme- and large-deviation analyses.

First, we start with high errors. Noting that $r < \overleftarrow{D}_{+\infty}(\rho_1\|\sigma_1)/D_{+\infty}(\rho_2\|\sigma_2)$ and recalling that $\overleftarrow{D}_\alpha$ is defined as the pinched-and-regularized relative entropy, this means that

$$r \overset{\text{ev.}}{<} \frac{D_{+\infty}\left(\mathcal{P}_{\sigma_1^{\otimes n}}\left(\rho_1^{\otimes n}\right)\middle\|\sigma_1^{\otimes n}\right)}{nD_{+\infty}(\rho_2\|\sigma_2)}. \quad (220)$$

The pinching inequality gives

$$\lambda_{\min}\left(\mathcal{P}_{\sigma_1^{\otimes n}}\left(\rho_1^{\otimes n}\right)\right) \geq \frac{\lambda_{\min}(\rho)^n}{|\text{spec}(\sigma^{\otimes n})|} \geq \lambda_{\min}^{dn}(\rho), \quad (221)$$

so any $x \leq \min\{\lambda_{\min}^d(\rho_1), \lambda_{\min}^r(\rho_2)\}^n$ satisfies

$$x \leq \lambda_{\min}\left(\mathcal{P}_{\sigma_1^{\otimes n}}\left(\rho_1^{\otimes n}\right)\right) \quad \text{and} \quad x \leq \lambda_{\min}\left(\sigma_1^{\otimes rn}\right) \quad (222)$$

and so we can apply Lemma 7 to both states, giving

$$\overleftarrow{\beta}_{1-x}(\rho_1^{\otimes n}\|\sigma_1^{\otimes n}) = x\exp\left(-D_{+\infty}\left(\mathcal{P}_{\sigma_1^{\otimes n}}(\rho_1^{\otimes n})\middle\|\sigma_1^{\otimes n}\right)\right) \tag{223a}$$

$$\overset{\text{ev.}}{<} x\exp\left(-rnD_{+\infty}(\rho_2\|\sigma_2)\right) \tag{223b}$$

$$= x\exp\left(-\breve{D}_{+\infty}(\rho_2^{\otimes rn}\|\sigma_2^{\otimes rn})\right) \tag{223c}$$

$$= \beta_{1-x}(\rho_2^{\otimes rn}\|\sigma_2^{\otimes rn}). \tag{223d}$$

Similarly, for the low-error case, we can use $r < \overleftarrow{D}_{-\infty}(\rho_1\|\sigma_1)/D_{-\infty}(\rho_2\|\sigma_2)$, which gives, for sufficiently large n, that

$$1-\overleftarrow{\beta}_x(\rho_1^{\otimes n}\|\sigma_1^{\otimes n}) = x\exp\left(-D_{-\infty}\left(\mathcal{P}_{\sigma_1^{\otimes n}}(\rho_1^{\otimes n})\middle\|\sigma_1^{\otimes n}\right)\right) \tag{224a}$$

$$\overset{\text{ev.}}{>} x\exp\left(-rn\breve{D}_{-\infty}(\rho_2\|\sigma_2)\right) \tag{224b}$$

$$= 1-\beta_x(\rho_2^{\otimes rn}\|\sigma_2^{\otimes rn}). \tag{224c}$$

For the remaining range of x, we resort to the method used in the large-deviation regime. As $r < \overleftarrow{D}_\alpha(\rho_1\|\sigma_1)/D_\alpha(\rho_2\|\sigma_2)$, we have

$$\overleftarrow{\Gamma}_\lambda(\rho_1\|\sigma_1) < r\Gamma_{\lambda/r}(\rho_2\|\sigma_2) \tag{225}$$

for all λ. Using Lemma 5, this means that

$$\overleftarrow{\beta}_{L^{-1}[\lambda n]}(\rho_1^{\otimes n}\|\sigma_1^{\otimes n}) \overset{\text{ev.}}{<} \beta_{L^{-1}[\lambda n]}(\rho_2^{\otimes rn}\|\sigma_2^{\otimes rn}). \tag{226}$$

This is only a pointwise convergence, which is insufficient for achievability but, similarly to the proof of Theorem 4, we can leverage the uniform analysis of Lemma 22 to show that this inequality must eventually hold *uniformly* for λ on a closed interval. If we specifically consider the interval

$$|\lambda \leq \max\{-d\log\lambda_{\min}(\rho_1), -r\log\lambda_{\min}(\rho_2)\}| + 1, \tag{227}$$

then this overlaps with the extreme-deviation cases and thus we have that, for sufficiently large n,

$$\overleftarrow{\beta}_x(\rho_1^{\otimes n}\|\sigma_1^{\otimes n}) < \beta_x(\rho_2^{\otimes rn}\|\sigma_2^{\otimes rn}) \tag{228}$$

holds for all $x \in (0,1)$. Application of Lemma 2 gives that transformation at rate r is eventually possible and so

$$\liminf_{n\to\infty} R_n^*(0) \geq \inf_{\alpha\in\mathbb{R}} \frac{\overleftarrow{D}_\alpha(\rho_1\|\sigma_1)}{D_\alpha(\rho_2\|\sigma_2)}. \tag{229}$$

Similarly, if we consider the right pinching, we also obtain

$$\liminf_{n\to\infty} R_n^*(0) \geq \inf_{\alpha\in\mathbb{R}} \frac{\overrightarrow{D}_\alpha(\rho_1\|\sigma_1)}{D_\alpha(\rho_2\|\sigma_2)}. \tag{230}$$

Now combining both achievability results for left and right pinching and recalling Lemma 13, gives

$$\liminf_{n\to\infty} R_n^*(0) \geq \max\left\{\inf_{\alpha\in\mathbb{R}} \frac{\overleftarrow{D}_\alpha(\rho_1\|\sigma_1)}{D_\alpha(\rho_2\|\sigma_2)}, \inf_{\alpha\in\mathbb{R}} \frac{\overrightarrow{D}_\alpha(\rho_1\|\sigma_1)}{D_\alpha(\rho_2\|\sigma_2)}\right\}, \tag{231}$$

as required. Note that this is quite close to the achievability which, due to Lemma 13, can be rewritten as

$$\limsup_{n\to\infty} R_n^*(0) \leq \inf_{\alpha\in\mathbb{R}} \max\left\{\frac{\overleftarrow{D}_\alpha(\rho_1\|\sigma_1)}{D_\alpha(\rho_2\|\sigma_2)}, \frac{\overrightarrow{D}_\alpha(\rho_1\|\sigma_1)}{D_\alpha(\rho_2\|\sigma_2)}\right\}. \tag{232}$$

■

In Theorem 8, it has been noted that all of the achievability results in this paper are, in the thermodynamic setting, achievable with only thermal operations, *except* for Theorem 6, which requires Gibbs-preserving maps. While all the achievability results in this paper leverage pinching, which is itself a thermal operation (see Appendix D), the problem has arisen in this final step involving pinching either the first or second state. In the case where switching the pinching is unnecessary, then this is a thermal operation but that is not generally the case.

Instead of a rate-based statement, we can also phrase this zero-error statement in terms of *eventual* Blackwell ordering [96,97,102], in line with some of the existing papers looking at similar zero-error transformation questions. For a pair of dichotomies, we define a notion of *eventual* Blackwell ordering as an ordering that appears for a sufficiently large number of copies, i.e., $(\rho_1^{\otimes n},\sigma_1^{\otimes n}) \overset{\text{ev.}}{\succeq} (\rho_2^{\otimes n},\sigma_2^{\otimes n})$ is a shorthand for

$$\exists N: \ (\rho_1^{\otimes n},\sigma_1^{\otimes n}) \succeq (\rho_2^{\otimes n},\sigma_2^{\otimes n}) \quad \forall n \geq N. \tag{233}$$

Corollary 1 (Eventual Blackwell ordering). Consider a pair of dichotomies (ρ_1,σ_1) and (ρ_2,σ_2). If the target is commuting, $[\rho_2,\sigma_2]=0$, and

$$\overleftarrow{D}_\alpha(\rho_1\|\sigma_1) > D_\alpha(\rho_2\|\sigma_2) \quad \forall\alpha\in\overline{\mathbb{R}}, \tag{234a}$$

or

$$\overrightarrow{D}_\alpha(\rho_1\|\sigma_1) > D_\alpha(\rho_2\|\sigma_2) \quad \forall\alpha\in\overline{\mathbb{R}}, \tag{234b}$$

then $(\rho_1^{\otimes n},\sigma_1^{\otimes n}) \overset{\text{ev.}}{\succeq} (\rho_2^{\otimes n},\sigma_2^{\otimes n})$. Moreover, if $(\rho_1^{\otimes n},\sigma_1^{\otimes n}) \overset{\text{ev.}}{\succeq} (\rho_2^{\otimes n},\sigma_2^{\otimes n})$, then this implies the inequalities

$$\breve{D}_\alpha(\rho_1\|\sigma_1) \geq \breve{D}_\alpha(\rho_2\|\sigma_2) \quad \forall\alpha\in\overline{\mathbb{R}}, \tag{235}$$

even for noncommuting targets $[\rho_2,\sigma_2]\neq 0$.

Proof. This follows directly from considering the $R = 1$ cases of Theorem 6. The inequalities

$$\overleftarrow{D}_\alpha(\rho_1\|\sigma_1) > D_\alpha(\rho_2\|\sigma_2) \quad \forall\alpha \in \overline{\mathbb{R}}, \tag{236a}$$

or

$$\overrightarrow{D}_\alpha(\rho_1\|\sigma_1) > D_\alpha(\rho_2\|\sigma_2) \quad \forall\alpha \in \overline{\mathbb{R}}, \tag{236b}$$

give that the zero-error rate is strictly greater than unity, $R_n^*(0) \overset{\text{ev.}}{>} 1$, and the inequalities

$$\breve{D}_\alpha(\rho_1\|\sigma_1) \geq \breve{D}_\alpha(\rho_2\|\sigma_2) \quad \forall\alpha \in \overline{\mathbb{R}} \tag{237}$$

all follow from the data-processing inequality of the minimal Rényi relative entropy. ■

Lastly, for completeness, we consider the case of a superexponentially high error, wherein the asymptotic transformation rate is unbounded.

Theorem 7 (Restated) (Extremely high error rate). For $[\rho_2, \sigma_2] = 0$, if the error is allowed to be superexponentially close to 1, then the optimal rate is unbounded,

$$\lim_{n\to\infty} R_n^*\big(1 - \exp(-\omega(n))\big) = \infty. \tag{238}$$

Proof. Consider any constant rate r. As ϵ_n is superexponentially approaching 1, it must dominate any other expression approaching 1 exponentially, specifically

$$\epsilon_n \overset{\text{ev.}}{>} 1 - (\lambda_{\min}(\rho_1)/2)^n \quad \text{and} \quad \epsilon_n \overset{\text{ev.}}{>} 1 - \lambda_{\min}(\rho_2)^{rn}. \tag{239}$$

Thus, we have that $1 - \epsilon_n \overset{\text{ev.}}{<} (\lambda_{\min}(\rho_1)/2)^n$ and $1 - \epsilon_n \overset{\text{ev.}}{<} \lambda_{\min}^{rn}(\rho_2)$. Application of Lemma 7 to the input gives

$$\frac{\overleftarrow{\beta}_{\epsilon_n}\big(\rho_1^{\otimes n}\big\|\sigma_1^{\otimes n}\big)}{1 - \epsilon_n} = \exp\left(-D_{+\infty}\big(\mathcal{P}_{\sigma^{\otimes n}}\big(\rho^{\otimes n}\big)\big\|\sigma^{\otimes n}\big)\right), \tag{240}$$

and to the target gives

$$\frac{1 - \beta_{1-\epsilon_n}\big(\rho_2^{\otimes rn}\big\|\sigma_2^{\otimes rn}\big)}{1 - \epsilon_n} = \exp\Big(-rn\breve{D}_{-\infty}(\rho_2\|\sigma_2)\Big). \tag{241}$$

As $n \to \infty$, these type-II errors approach 0 and 1, respectively, and so

$$\overleftarrow{\beta}_{\epsilon_n}\big(\rho_1^{\otimes n}\big\|\sigma_1^{\otimes n}\big) \overset{\text{ev.}}{<} \beta_{1-\epsilon_n}\big(\rho_2^{\otimes rn}\big\|\sigma_2^{\otimes rn}\big). \tag{242}$$

Use of the monotonicity of $x \mapsto \beta_x(\cdot\|\cdot)$ allows us to relax this to

$$\overleftarrow{\beta}_x\big(\rho_1^{\otimes n}\big\|\sigma_1^{\otimes n}\big) < \beta_{x-\epsilon_n}\big(\rho_2^{\otimes rn}\big\|\sigma_2^{\otimes rn}\big) \tag{243}$$

for $x \in (\epsilon_n, 1)$. So, by Lemma 2, this means that transformation at the rate r is eventually achievable. As this entire argument has worked for any constant r, this therefore means that the optimal rate must diverge:

$$\liminf_{n\to\infty} R_n^*(\epsilon_n) = \infty. \tag{244}$$

■

VI. CONCLUSIONS AND OUTLOOK

In this work, we have analyzed one of the central problems of the theory of quantum statistical inference, namely, that of comparing informativeness of two quantum dichotomies (which is directly related to transforming the first dichotomy into the second one). By focusing on the asymptotic version of the problem, we have been able to solve it in various error regimes under the assumption that the second dichotomy is commutative. More precisely, we have found optimal transformation rates between many copies of pairs of quantum states in the small-, moderate-, large-, and zero-error regimes. We have then employed the obtained results to derive new thermodynamic laws for quantum systems prepared in coherent superpositions of energy eigenstates. Thus, for the first time, we have been able to analyze the optimal performance of thermodynamic protocols with coherent inputs beyond the thermodynamic limit and we have discussed new resonance phenomena that allow one to mitigate thermodynamic dissipation by, e.g., employing quantum coherence.

We believe that the success of employing quantum statistical inference techniques to accurately describe quantum thermodynamic transformations strongly motivates further exploration of the connections between the two frameworks. We propose the following three avenues. First, one of the problems within the resource-theoretic approach to quantum thermodynamics is the lack of techniques for addressing the regimes of nonindependent systems. Interestingly, in Refs. [112,115], it is suggested that the hypothesis-testing approach can be effective for studying ensembles composed of weakly correlated states. Potentially, such techniques can be adapted to study the thermodynamic state transformation problem outside of the usual uncorrelated setting. Second, one could use quantum statistical inference techniques to develop explicit thermodynamic protocols. Indeed, one of the criticisms of the resource-theoretic approach is that many of its consequences are implicit, i.e., one often shows the existence of protocols but their explicit form is usually not possible to infer. However, as observed in Ref. [70], the hypothesis-testing approach allows us to construct explicit thermal operations starting from the optimal measurement in the related hypothesis-testing problem. We expect that investigating the explicit form of optimal thermodynamic protocols in different asymptotic regimes can lead to interesting new insights on the nature of fundamental limitations imposed by thermodynamic laws on dissipation, reversibility, work processes, etc. And third, it is a long-standing problem to connect the resource-theoretic approach to thermodynamics with more standard approaches [120]. We believe that an especially interesting connection might exist between the resource-theoretic approach and so-called slow-driving protocols [121]. Here, we note that both approaches use similar statistical and geometric techniques. For example, the optimal thermodynamic protocols in the slow-driving regime can be quantified using the so-called Kubo-Mori metric, which is also related to the problem of hypothesis testing [122]. Exploring these intrinsic similarities might improve our understanding of quantum thermodynamics.

On a more technical side, we think that a very interesting avenue for further research is to try to generalize our results so that they also apply to noncommutative output dichotomies [123]. This would open a way to study fully quantum laws of thermodynamics, where both initial and final states could be given by superpositions of different energy eigenstates. While we think that Conjecture 2 may be true, this does not necessarily mean that the transformation rates in the fully coherent regime would be simple generalizations of the current results. The reason for that is that proving Conjecture 2 would only guarantee such a simple generalization of the rate under transformations with the so-called Gibbs-preserving operations [124] and not under thermal operations. In fact, we believe that in a fully quantum regime, there may be a gap between the rates achievable with these two sets of free operations (especially for the zero-error case).

Another technical generalization of our result that we find highly interesting is to study transformations between *multichotomies*, i.e., multipartite transformations from m states to m states, with dichotomies being the special case of $m = 2$. The classical zero-error case of this has recently been analyzed in Ref. [97] and the quantum and/or nonzero-error cases are natural generalizations worthy of study. Physically, such a result could help us to understand transformations of quantum systems under the constraint of the symmetry. This is because for a symmetry group G, the existence of a G-covariant quantum channel mapping the initial state to the final one is equivalent to the existence of an unconstrained channel mapping the orbit of the initial state to the orbit of the final state, with the orbit being generated by the symmetry elements of G [125].

Finally, there are two aspects of the resonance phenomenon described in this paper that we believe deserve more attention. First, we think it would be very interesting to find the equivalent of the resonance phenomenon in more traditional approaches to thermodynamics, beyond the resource-theoretic treatment. In other words, we would like to investigate whether such a potential reduction of free-energy dissipation may appear in actual physical processes when the parameters are tuned appropriately. Second, one could look for similar resonance effects in other resource theories. In particular, we note that pure state interconversion conditions in the resource theory of $U(1)$-asymmetry [126] are ruled by a generalization of the majorization partial order (called cyclic majorization in Ref. [127]). Since the resonance appeared for standard majorization (as we have seen for pure bipartite-entanglement transformations in this paper), the resource theory of $U(1)$-asymmetry seems to be a good candidate to look for novel resource resonance effects.

ACKNOWLEDGMENTS

We would like to thank Francesco Buscemi for helpful pointers to the literature. K.K. acknowledges financial support from the Foundation for Polish Science through the TEAM-NET project (Grant No. POIR.04.04.00-00-17C1/18-00). P.L.B., C.T.C., and J.M.R. acknowledge the Swiss National Science Foundation for financial support through the National Centres of Competence in Research (NCCR) SwissMAP and “Quantum Science and Technology” (QSIT) programs and C.T.C. and J.M.R. acknowledge their support through Sinergia Grant No. CRSII5_186364. M.T. acknowledges the hospitality of the Pauli Center, which supported a long-term visit to ETH Zürich. He is also supported by National University of Singapore (NUS) startup Grants No. R-263-000-E32-133 and No. R-263-000-E32-731.

APPENDIX A: PROOF OF LEMMA 1

Lemma 1 (Restated) (Sesquinormal distribution). The function S_ν is a cumulative distribution function (cdf) for any $\nu \in [0, \infty)$. Moreover, for $\nu \notin \{0, 1, \infty\}$, the cdf has the closed form

$$S_\nu(\mu) = \Phi\left(\frac{\mu - \sqrt{\nu}\sqrt{\mu^2 + (\nu - 1)\ln\nu}}{1 - \nu}\right) - \Phi\left(\frac{\sqrt{\nu}\mu - \sqrt{\mu^2 + (\nu - 1)\ln\nu}}{1 - \nu}\right), \quad \text{(A1)}$$

and for $0 < \nu < \infty$, the inverse cdf can be expressed as

$$S_\nu^{-1}(\epsilon) = \min_{x \in (\epsilon, 1)} \sqrt{\nu}\Phi^{-1}(x) - \Phi^{-1}(x - \epsilon). \quad \text{(A2)}$$

The extreme cases $\nu = 0$ and $\nu \to \infty$ reduce to the normal distribution

$$S_0(\mu) = \lim_{\nu \to \infty} S_\nu(\sqrt{\nu}\mu) = \Phi(\mu), \quad \text{(A3)}$$

and the $\nu = 1$ reduces to the half-normal distribution

$$S_1(\mu) = \max\{2\Phi(\mu/2) - 1, 0\}. \quad \text{(A4)}$$

Finally, the family of sesquinormal distributions has a duality under reciprocating the parameter

$$S_\nu(\mu) = S_{1/\nu}(\mu/\sqrt{\nu}) \quad \text{or} \quad S_\nu^{-1}(\epsilon) = \sqrt{\nu} S_{1/\nu}^{-1}(\epsilon). \quad \text{(A5)}$$

Proof. To prove that S_ν is a valid cdf, we need to prove that it is continuous, monotone nondecreasing, and has the limits

$$\lim_{\mu\to-\infty} S_\nu(\mu) = 0 \quad \text{and} \quad \lim_{\mu\to+\infty} S_\nu(\mu) = 1. \quad \text{(A6)}$$

We will see that continuity and the limits both follow from the closed form below. For monotonicity, we can use the fact that the total variation distance of two distributions is unchanged if we shift them along $\mathbb{R}$:

$$\begin{aligned} S_\nu(\mu+\epsilon) &= \inf_{A\geq\Phi_{0,1}} T\big(A, \Phi_{\mu+\epsilon,\nu}\big) \\ &= \inf_{A\geq\Phi_{-\epsilon,1}} T\big(A, \Phi_{\mu,\nu}\big) \\ &\geq \inf_{A\geq\Phi_{0,1}} T\big(A, \Phi_{\mu,\nu}\big) = S_\nu(\mu). \end{aligned} \quad \text{(A7)}$$

Next, we turn to a closed form of S_ν. Recall the definition

$$S_\nu(\mu) := \frac{1}{2}\inf_{A\geq\Phi}\int_{\mathbb{R}} |A'(x) - \phi_{\mu,\nu}(x)|\, dx. \quad \text{(A8)}$$

To find a closed form, we will suggest a candidate A, evaluate its total variation distance, and then construct a lower bound to show that this is optimal among distributions with $A\geq\Phi$. We split into two cases based on whether $\nu \lesseqgtr 1$. ∎

a. Case $0 < \nu < 1$. Start by recalling that the total variation distance between two measures is the largest possible difference in probability that they assign to an event, i.e.,

$$S_\nu(\mu) = \inf_{A\geq\Phi}\sup_{R\subseteq\mathbb{R}}\int_R \big(A'(x) - \phi_{\mu,\nu}(x)\big)\, dx. \quad \text{(A9)}$$

Next, consider the set of x such that $\Phi(x) \geq \Phi_{\mu,\nu}(x)$ *and* $\phi(x) \geq \phi_{\mu,\nu}(x)$. For $\nu < 1$, this region is given precisely by $x \leq X$, where

$$X = \frac{\mu - \sqrt{\nu}\sqrt{\mu^2 + (\nu-1)\ln\nu}}{1-\nu}. \quad \text{(A10)}$$

Thus, we can lower bound the total variation distance by considering the region $R = (-\infty, X]$, which gives

$$\begin{aligned} S_\nu(\mu) &\geq \inf_{A\geq\Phi}\int_{-\infty}^{X} \big(A'(x) - \phi_{\mu,\nu}(x)\big)\, dx && \text{(A11a)} \\ &= \inf_{A\geq\Phi} A(X) - \Phi_{\mu,\nu}(X) && \text{(A11b)} \\ &\geq \Phi(X) - \Phi_{\mu,\nu}(X) && \text{(A11c)} \\ &= \Phi\left(\frac{\mu - \sqrt{\nu}\sqrt{\mu^2 + (\nu-1)\ln\nu}}{1-\nu}\right) \\ &\quad - \Phi_{\mu,\nu}\left(\frac{\mu - \sqrt{\nu}\sqrt{\mu^2 + (\nu-1)\ln\nu}}{1-\nu}\right) && \text{(A11d)} \\ &= \Phi\left(\frac{\mu - \sqrt{\nu}\sqrt{\mu^2 + (\nu-1)\ln\nu}}{1-\nu}\right) \\ &\quad - \Phi\left(\frac{\sqrt{\nu}\mu - \sqrt{\mu^2 + (\nu-1)\ln\nu}}{1-\nu}\right). && \text{(A11e)} \end{aligned}$$

Moreover, it can be seen that by taking $A(x) := \max\{\Phi(x), \Phi_{\mu,\nu}(x)\}$ we can saturate this lower bound, proving it to be optimal among all cdfs such that $A \geq \Phi$.

b. Case $\nu > 1$ For $\nu > 1$, we can do a similar proof to that for $\nu < 1$. Here, we are interested in the region in which $\Phi(x) \geq \Phi_{\mu,\nu}(x)$ *and* $\phi(x) \leq \phi_{\mu,\nu}(x)$, which is now given by $x \geq X$, with X defined as before. Now, looking at the lower bound given by $R = [X, \infty)$, we obtain the same as previously:

$$\begin{aligned} S_\nu(\mu) &\geq \inf_{A\geq\Phi}\int_{X}^{\infty} \big(\phi_{\mu,\nu}(x) - A'(x)\big)\, dx && \text{(A12a)} \\ &= \inf_{A\geq\Phi}\big(1 - \Phi_{\mu,\nu}(X)\big) - \big(1 - A(X)\big) && \text{(A12b)} \\ &= \inf_{A\geq\Phi} A(X) - \Phi_{\mu,\nu}(X) && \text{(A12c)} \\ &\geq \Phi(X) - \Phi_{\mu,\nu}(X) && \text{(A12d)} \\ &= \Phi\left(\frac{\mu - \sqrt{\nu}\sqrt{\mu^2 + (\nu-1)\ln\nu}}{1-\nu}\right) \\ &\quad - \Phi\left(\frac{\sqrt{\nu}\mu - \sqrt{\mu^2 + (\nu-1)\ln\nu}}{1-\nu}\right). && \text{(A12e)} \end{aligned}$$

Once again, optimality of this bound is implied by the fact it is still saturated by $A(x) := \max\{\Phi(x), \Phi_{\mu,\nu}(x)\}$.

Now, by taking limits of the closed form, we can see that

$$\begin{aligned} &\lim_{\nu\to0^+} S_\nu(\mu) = \lim_{\nu\to\infty} S_\nu(\sqrt{\nu}\mu) = \Phi(\mu) \quad \text{and} \\ &\lim_{\nu\to1} S_\nu(\mu) = \max\{2\Phi(\mu/2) - 1, 0\} \end{aligned} \quad \text{(A13)}$$

and by substituting $\nu \to 1/\nu$ it can be straightforwardly seen that this expression has the duality property

$$S_\nu(\mu) = S_{1/\nu}(\mu/\sqrt{\nu}). \quad \text{(A14)}$$

Having a closed form of the cdf, we turn to the inverse cdf. To start with, consider an arbitrary $\epsilon \in (0,1)$ and define

$$\mu := \inf_{x\in(\epsilon,1)} \sqrt{\nu}\Phi^{-1}(x) - \Phi^{-1}(x-\epsilon). \quad \text{(A15)}$$

To prove the form of the inverse cdf $S_\nu^{-1}(\epsilon) = \mu$, it suffices to show the inverse expression $S_\nu(\mu) = \epsilon$. We start by noting that the function $f(x) := \sqrt{\nu}\Phi^{-1}(x) - \Phi^{-1}(x-\epsilon)$ is bounded for any $x \in (\epsilon, 1)$ and diverges to $+\infty$ for either $x \to \epsilon^+$ and $x \to 1^-$ and, as such, the infimum is in fact a minimum, so

$$\mu = \min_{x\in(\epsilon,1)} \sqrt{\nu}\Phi^{-1}(x) - \Phi^{-1}(x-\epsilon). \tag{A16}$$

Thus, there must exist a $y \in (\epsilon, 1)$ at which the infimum is attained, i.e., $f(y) = \mu$. By the interior extremum theorem, we must have that this is a stationary point, $f'(y) = 0$. The inverse function rule allows us to evaluate the derivative of f to be

$$f'(x) = \frac{\sqrt{\nu}}{\phi\left(\Phi^{-1}(x)\right)} - \frac{1}{\phi\left(\Phi^{-1}(x-\epsilon)\right)} \tag{A17}$$

and thus $f'(y) = 0$ reduces to

$$\left[\Phi^{-1}(y)\right]^2 + \ln\nu = \left[\Phi^{-1}(y-\epsilon)\right]^2. \tag{A18}$$

To get rid of this shifted Gaussian term, we can use a substitution

$$\Phi^{-1}(y-\epsilon) = \sqrt{\nu}\Phi^{-1}(y) - \mu, \tag{A19}$$

which gives us a quadratic expression for $\Phi^{-1}(y)$,

$$\left[\Phi^{-1}(y)\right]^2 + \ln\nu = \left[\sqrt{\nu}\Phi^{-1}(y) - \mu\right]^2, \tag{A20}$$

with a pair of solutions

$$\Phi^{-1}(y) = -\frac{\mu\sqrt{\nu} \pm \sqrt{\mu^2 + (\nu-1)\ln\nu}}{1-\nu}. \tag{A21}$$

This, however, still has a lingering $\pm$ ambiguity. If we rearrange Eq. (A19) to make ϵ the subject, we obtain

$$\epsilon = y - \Phi\left(\sqrt{\nu}\Phi^{-1}(y) - \mu\right). \tag{A22}$$

Substituting our pair of solutions into this expression gives

$$\epsilon = \Phi\left(\Phi^{-1}(y)\right) - \Phi\left(\sqrt{\nu}\Phi^{-1}(y) - \mu\right) \tag{A23a}$$

$$= \Phi\left(-\frac{\mu\sqrt{\nu} \pm \sqrt{\mu^2 + (\nu-1)\ln\nu}}{1-\nu}\right) - \Phi\left(-\sqrt{\nu}\frac{\mu\sqrt{\nu} \pm \sqrt{\mu^2 + (\nu-1)\ln\nu}}{1-\nu} - \mu\right) \tag{A23b}$$

$$= \Phi\left(\frac{\mu \mp \sqrt{\nu}\sqrt{\mu^2 + (\nu-1)\ln\nu}}{1-\nu}\right) - \Phi\left(\frac{\sqrt{\nu}\mu \mp \sqrt{\mu^2 + (\nu-1)\ln\nu}}{1-\nu}\right). \tag{A23c}$$

Now, we note that $\sqrt{\mu^2 + (\nu-1)\ln\nu} \geq \mu$. Using this, we can see that the positive solution for $\Phi^{-1}(x)$ will correspond to $\epsilon \leq 0$ and thus the minimizing x must correspond to the negative solution, i.e.,

$$\epsilon = \Phi\left(\frac{\mu - \sqrt{\nu}\sqrt{\mu^2 + (\nu-1)\ln\nu}}{1-\nu}\right) - \Phi\left(\frac{\sqrt{\nu}\mu - \sqrt{\mu^2 + (\nu-1)\ln\nu}}{1-\nu}\right). \tag{A24}$$

Finally, we now have $\epsilon = S_\nu(\mu)$, as required.

Lemma 9 (Asymptotic expansions). For $\mu \to \infty$, the sesquinormal cdf can be expanded as

$$\ln\left[S_\nu(-\mu)\right] \approx -\frac{1}{2}\left(\frac{\mu}{1-\sqrt{\nu}}\right)^2, \tag{A25a}$$

$$\ln\left[1 - S_\nu(\mu)\right] \approx -\frac{1}{2}\left(\frac{\mu}{1+\sqrt{\nu}}\right)^2. \tag{A25b}$$

Similarly, for $\epsilon \to 0^+$, the sesquinormal inverse cdf can be expanded as

$$S_\nu^{-1}(\epsilon) \approx |1-\sqrt{\nu}|\sqrt{2\log 1/\epsilon}, \tag{A26a}$$

$$S_\nu^{-1}(1-\epsilon) \approx (1+\sqrt{\nu})\sqrt{2\log 1/\epsilon}. \tag{A26b}$$

Proof. The expansions of the cdf can be found simply by expanding the closed-form expression in Lemma 1 to leading order in μ; specifically, using the approximation

$$\sqrt{\mu^2 + (\nu-1)\ln\nu} \approx |\mu|. \tag{A27}$$

Use of this approximation and the $x \to \infty$ expansion $\ln[1-\Phi(x)] \approx -x^2/2$ gives the cdf expansions. By inverting this, we can equivalently obtain the inverse cdf expansions as well. ■

APPENDIX B: PINCHED RELATIVE ENTROPY

In this appendix, we will show the existence and properties of the pinched Rényi relative entropies. Suppose that the states ρ, σ are fixed and full rank and define

$$f_n(\alpha) := \frac{1}{n}D_\alpha\big(\mathcal{P}_{\sigma^{\otimes n}}\left(\rho^{\otimes n}\right)\big\|\sigma^{\otimes n}\big). \tag{B1}$$

The left-pinched Rényi relative entropy is, as we shall see below, defined as $\overleftarrow{D}_\alpha := \lim_{n\to\infty} f_n(\alpha)$, with $\overrightarrow{D}_\alpha$ defined similarly. Its existence and properties will be given below in Theorem 11. The first thing we will note is that while we do not know of a closed-form solution for $\overleftarrow{D}_\alpha$ and $\overrightarrow{D}_\alpha$

in general, they are known to reduce to the sandwiched and reverse sandwiched entropies for $\alpha \geq 0$ and $\alpha \leq 1$ respectively [87, Proposition 4.12],

$$\forall\alpha \geq 0 \quad \overleftarrow{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1}\log\mathrm{Tr}\left(\left(\sqrt{\rho}\sigma^{\frac{1-\alpha}{\alpha}}\sqrt{\rho}\right)^{\alpha}\right), \tag{B2a}$$

$$\forall\alpha \leq 1 \quad \overrightarrow{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1}\log\mathrm{Tr}\left(\left(\sqrt{\sigma}\rho^{\frac{\alpha}{1-\alpha}}\sqrt{\sigma}\right)^{1-\alpha}\right), \tag{B2b}$$

and so they inherit the desired properties within these ranges. As such, we will focus on showing that these properties extend beyond these ranges where we lack closed-form expressions. We start by showing that f_n, f_n', and f_n'' are uniformly bounded.

Lemma 10. For all n and $\alpha \leq 0, f_n(\alpha)$ is nonpositive and bounded by the minimal entropy

$$0 \geq f_n(\alpha) \geq \breve{D}_\alpha(\rho\|\sigma). \tag{B3}$$

Moreover, there exist uniform (i.e., independent of n and α) bounds on the value and first two derivatives of f_n,

$$|f_n(\alpha)| \leq C_0, \quad |f_n'(\alpha)| \leq C_1, \quad |f_n''(\alpha)| \leq C_2. \tag{B4}$$

Proof. First, the nonpositivity of f_n follows from the fact that D_α is nonpositive for $\alpha \leq 0$. The lower bound on f_n follows from the data-processing inequality (recalling that the data-processing inequality (DPI) is reversed for $\alpha \leq 0$) and additivity of the minimal relative entropy:

$$f_n = \frac{1}{n}\breve{D}_\alpha\big(\mathcal{P}_{\sigma^{\otimes n}}\big(\rho^{\otimes n}\big)\big\|\sigma^{\otimes n}\big) \geq \frac{1}{n}\breve{D}_\alpha\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big) = \breve{D}_\alpha(\rho\|\sigma). \tag{B5}$$

Furthermore, given that $\breve{D}_\alpha(\rho\|\sigma) \geq \breve{D}_{-\infty}(\rho\|\sigma)$ for all α, we have a uniform lower bound on f_n, i.e., $C_0 := \breve{D}_{-\infty}(\rho\|\sigma)$.

Next, we turn to the derivatives. Before applying it to our states, we start by looking at what form the derivatives of the (classical) Rènyi relative entropy take in the abstract, say, for two classical distributions, p and q. For notational simplicity, we are going to assume that all logarithms below are natural, to avoid factors of $\ln b$. Given that we are only concerned with nonpositive α and are not concerned with $\alpha = 1$, we can switch to looking at the unnormalized variant of the Rényi relative entropy of the form $(\alpha-1)D_\alpha(p\|q)$. Taking derivatives of this gives

$$(\alpha-1)D_\alpha(p\|q) = \log\sum_i p_i^\alpha q_i^{1-\alpha}, \tag{B6a}$$

$$((\alpha-1)D_\alpha(p\|q))' = \frac{\sum_i \ln\frac{p_i}{q_i}\times p_i^\alpha q_i^{1-\alpha}}{\sum_i p_i^\alpha q_i^{1-\alpha}}, \tag{B6b}$$

$$((\alpha-1)D_\alpha(p\|q))'' = \frac{\left(\sum_i \ln^2\frac{p_i}{q_i}\times p_i^\alpha q_i^{1-\alpha}\right)\times\left(\sum_i p_i^\alpha q_i^{1-\alpha}\right)^2 - \left(\sum_i \ln\frac{p_i}{q_i}\times p_i^\alpha q_i^{1-\alpha}\right)^2}{\left(\sum_i p_i^\alpha q_i^{1-\alpha}\right)^2}. \tag{B6c}$$

Conveniently, the first and second derivatives take the form of moments. Specifically, if we consider the distribution

$$w_i := \frac{p_i^\alpha q_i^{1-\alpha}}{\sum_j p_j^\alpha q_j^{1-\alpha}}, \tag{B7}$$

then the derivatives become the mean and variance of $\ln\frac{p}{q}$ with respect to w,

$$((\alpha-1)D_\alpha(p\|q))' = \sum_i w_i \ln\frac{p_i}{q_i}, \tag{B8a}$$

$$((\alpha-1)D_\alpha(p\|q))'' = \sum_i w_i \ln^2\frac{p_i}{q_i} - \left(\sum_i w_i\ln\frac{p_i}{q_i}\right)^2. \tag{B8b}$$

So now we can uniformly bound both in terms of $|\ln p_i/q_i| \leq \max\{-\ln\min_i p_i, -\ln\min_i q_i\}$; specifically,

$$|((\alpha-1)D_\alpha(p\|q))'| \leq \max\left\{-\ln\min_i p_i, -\ln\min_i q_i\right\}, \tag{B9a}$$

$$|((\alpha-1)D_\alpha(p\|q))''| \leq \max\left\{-\ln\min_i p_i, -\ln\min_i q_i\right\}^2. \tag{B9b}$$

Now, we want to return to f_n, wherein $p = \mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n})$ and $q = \sigma^{\otimes n}$. Before that, we need to deal with the pinching. Specifically, if we use the the pinching inequality

$$\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n}) \geq \frac{\rho^{\otimes n}}{|\mathrm{spec}(\sigma^{\otimes n})|} \geq \frac{\rho^{\otimes n}}{n^d}, \tag{B10}$$

we can see that

$$\begin{aligned} -\frac{1}{n}\ln\lambda_{\min}\left(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n})\right) &\leq -\frac{1}{n}\ln\frac{\lambda_{\min}^n(\rho)}{n^d} \\ &= \frac{d\ln n}{n} - \ln\lambda_{\min}(\rho) \\ &\leq d - \ln\lambda_{\min}(\rho) \end{aligned} \tag{B11}$$

and thus

$$\max\left\{-\frac{1}{n}\ln\lambda_{\min}\left(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n})\right), -\ln\lambda_{\min}(\sigma)\right\} \leq M, \tag{B12}$$

where $M := \max\{d - \ln\lambda_{\min}(\rho), -\ln\lambda_{\min}(\sigma)\}$.

Now, we return to f_n. We take the unnormalized version of this,

$$(\alpha-1)f_n(\alpha) = \frac{1}{n}(\alpha-1)D\big(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n})\big\|\sigma^{\otimes n}\big), \tag{B13}$$

so by the above arguments the derivatives can be bounded:

$$\begin{aligned} &|((\alpha-1)f_n(\alpha))'| \\ &\quad\leq \max\left\{-\frac{1}{n}\ln\lambda_{\min}\left(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n})\right), -\ln\lambda_{\min}(\sigma)\right\} \\ &\quad\leq M, \end{aligned} \tag{B14a}$$

$$\begin{aligned} &|((\alpha-1)f_n(\alpha))''| \\ &\quad\leq \max\left\{-\frac{1}{n}\ln\lambda_{\min}\left(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n})\right), -\ln\lambda_{\min}(\sigma)\right\}^2 \\ &\quad\leq M^2. \end{aligned} \tag{B14b}$$

We thus have that the derivatives of the $(\alpha-1)f_n(\alpha)$ are bounded, so all that is left is to show is that this necessarily extends to $f_n(\alpha)$ itself. By straightforward algebraic manipulation, we can write the derivatives of the latter quantity in terms of those of the former; specifically,

$$f_n{}'(\alpha) = \frac{((\alpha-1)f_n(\alpha))' - f_n(\alpha)}{\alpha-1}, \tag{B15a}$$

$$f_n{}''(\alpha) = \frac{((\alpha-1)f_n(\alpha))'' - 2f_n{}'(\alpha)}{\alpha-1}. \tag{B15b}$$

Given that $\alpha \leq 0$ and is therefore gapped away from $\alpha = 1$, this causes no issues. Specifically, if we take $C_1 := M + C_0$ and $C_2 := M^2 + 2C_1$, then the desired uniform bounds hold as required. ■

Before attacking the existence and properties of $\overleftarrow{D}_\alpha$, we need one final theorem that allows us to leverage these uniform bounds to extend properties of $\{f_n\}_n$ through to $\overleftarrow{D}_\alpha$. This theorem is a corollary of the Arzelà-Ascoli theorem.

Lemma 11 (Arzelà-Ascoli theorem [128, Corollary 11.6.11]). Let $\{f_n\}_n$ be a sequence of differentiable functions on a compact domain that are uniformly bounded, and the derivative of which is also uniformly bounded. Then, there exists a uniformly convergent subsequence $\{f_{m_n}\}_n$.

With this in hand, we turn to proving the properties of the pinched relative entropies.

Theorem 11 (Properties of the pinched relative entropy). Define the left-pinched relative entropy as

$$\overleftarrow{D}_\alpha(\rho\|\sigma) := \lim_{n\to\infty}\frac{1}{n}D_\alpha\big(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n})\big\|\sigma^{\otimes n}\big). \tag{B16}$$

For full-rank states and $\alpha \in \overline{\mathbb{R}}$, the pinched relative entropy has the following properties:

(i) Existence: $\overleftarrow{D}_\alpha(\rho\|\sigma)$ exists.
(ii) (Non)positivity: $\overleftarrow{D}_\alpha(\rho\|\sigma)$ is non-negative for $\alpha \geq 0$, and nonpositive for $\alpha \leq 0$.
(iii) Subminimality: $\overleftarrow{D}_\alpha(\rho\|\sigma) \leq \check{D}_\alpha(\rho\|\sigma)$ for $\alpha \geq 0$ and $\overleftarrow{D}_\alpha(\rho\|\sigma) \geq \check{D}_\alpha(\rho\|\sigma)$ for $\alpha \leq 0$.
(iv) Differentiability: $\alpha \mapsto \overleftarrow{D}_\alpha(\rho\|\sigma)$ is differentiable.

Moreover, all of these properties also extend to the right-pinched relative entropy:

$$\overrightarrow{D}_\alpha(\rho\|\sigma) := \lim_{n\to\infty}\frac{1}{n}D_\alpha\big(\rho^{\otimes n}\big\|\mathcal{P}_{\rho^{\otimes n}}(\sigma^{\otimes n})\big). \tag{B17}$$

Proof. The sandwiched relative entropy has all of the above properties [87] and coincides with the pinched relative entropy with $\alpha \geq 0$, so we need only show that these properties hold for $\alpha \leq 0$ as well.

If we consider the composition of pinching a composite system, we have

$$(\mathcal{P}_X \otimes \mathcal{P}_Y)\,(\mathcal{P}_{X\otimes Y}(A)) = (\mathcal{P}_X \otimes \mathcal{P}_Y)\,(A)\,. \tag{B18}$$

This, together with the data-processing inequality, gives

$$D_\alpha\big(\mathcal{P}_{\sigma^{\otimes(n+m)}}\left(\rho^{\otimes(n+m)}\right)\big\|\sigma^{\otimes(n+m)}\big) \le D_\alpha\big((\mathcal{P}_{\sigma^{\otimes n}} \otimes \mathcal{P}_{\sigma^{\otimes m}})\left(\rho^{\otimes(n+m)}\right)\big\|\sigma^{\otimes(n+m)}\big), \tag{B19a}$$

$$= D_\alpha\big(\mathcal{P}_{\sigma^{\otimes n}}\left(\rho^{\otimes n}\right)\big\|\sigma^{\otimes n}\big) + D_\alpha\big(\mathcal{P}_{\sigma^{\otimes m}}\left(\rho^{\otimes m}\right)\big\|\sigma^{\otimes m}\big) \tag{B19b}$$

or, in other words, $(n+m)f_{n+m}(\alpha) \ge nf_n(\alpha) + mf_m(\alpha)$. Applying Fekete's superadditive lemma [129], this superaddivity implies that $f_n(\alpha)$ is convergent in n for each α.

Next, we want to apply the Arzelà-Ascoli theorem (Lemma 12). Lemma 11 gives us the uniform boundedness required to apply Lemma 12 to $\{f_n\}_n$, which gives that there exists a uniformly convergent subsequence $\{f_{a_n}\}_n$. But as we already have established that $\{f_n\}_n$ is also convergent, this implies that this convergence is uniform. Next, using the uniform bound on $\{f_n''\}_n$ from Lemma 11, we can also apply Lemma 12 to $\{f_n'\}_n$, which gives a uniformly convergent subsequence $\{f_{b_n}'\}_n$. As $\{f_{b_n}\}_n$ and $\{f_{b_n}'\}_n$ are both uniformly convergent, we can commute through the limit and the derivative. Using this together with the convergence of $\{f_n\}_n$, we can see that $\{f_n'\}_n$ must also be (uniformly) convergent and thus that $\overleftarrow{D}_\alpha$ is differentiable:

$$\lim_n f_{b_n}'(\alpha) = \left(\lim_n f_{b_n}(\alpha)\right)' = \left(\lim_n f_n(\alpha)\right)' = \left(\overleftarrow{D}_\alpha\right)'. \tag{B20}$$

To extend all of these properties to the right-pinched relative entropy, we can simply use the identity

$$\overrightarrow{D}_\alpha(\rho\|\sigma) = \frac{\alpha}{1-\alpha}\overleftarrow{D}_{1-\alpha}(\sigma\|\rho). \tag{B21}$$

We might suspect that the $1-\alpha$ denominator causes issues around $\alpha = 1$ but as $\overrightarrow{D}_\alpha$ reduces to the reverse sandwiched relative entropy for $\alpha > 1/2$, it therefore inherits the above properties within that range. ■

Next, we prove a nice relationship between the two pinched relative entropies and the minimal relative entropy.

Lemma 12. The maximum (in magnitude) of the pinched Rényi relative entropies corresponds to the minimal relative entropy:

$$\max\left\{|\overleftarrow{D}_\alpha(\rho\|\sigma)|, |\overrightarrow{D}_\alpha(\rho\|\sigma)|\right\} = |\breve{D}_\alpha(\rho\|\sigma)|. \tag{B22}$$

Proof. From the subminimality property of the pinched relative entropy (see Theorem 11), we have

$$|\breve{D}_\alpha(\rho\|\sigma)| \ge \max\left\{|\overleftarrow{D}_\alpha(\rho\|\sigma)|, |\overrightarrow{D}_\alpha(\rho\|\sigma)|\right\}. \tag{B23}$$

Next, Ref. [87, Proposition 4.12] gives that $\overleftarrow{D}_\alpha$ corresponds to the sandwiched entropy for $\alpha \ge 0$, which in turn corresponds to the minimal entropy for $\alpha > 1/2$ [87, Sec. 4.3]. By duality, this means that $\overrightarrow{D}_\alpha$ corresponds to the reverse sandwiched relative entropy for $\alpha \le 1$ and also to the minimal for $\alpha \le 1/2$. Thus, this inequality is satisfied for all α, as required. ■

APPENDIX C: TWO-SIDED ERROR

In Theorems 2–7, we have only considered transformations involving an error on the first state in the dichotomy. One reason why this has been done is because such transformations are the relevant transformations for the resource-theoretic applications of concern (see Secs. II B and II C). Another is that, as we will see, the more general problem in which we allow errors on both states is no more rich. In this appendix, we will give a summary of the asymptotic rate scalings for two nonzero errors. In lieu of giving rigorous proofs of these rates, we will instead mention how things change from the proofs of Theorems 2–7.

One of the reasons why two nonzero errors do not give a much richer problem is that there exist errors for which the rate becomes infinite. To be clear, we do not mean that the rate diverges as $n \to \infty$ (*a la* Theorem 7) but, instead, a situation in which the rate is unbounded for a finite n. This occurs because if the errors as sufficiently large, the Blackwell order breaks down in its entirety.

Lemma 13 (Breakdown of Blackwell ordering). If $\beta_{\epsilon_\rho}(\rho\|\sigma) \le \epsilon_\sigma$, then (ρ, σ), Blackwell dominates *all* dichotomies, i.e.,

$$\beta_{\epsilon_\rho}(\rho_1\|\sigma_1) \le \epsilon_\sigma \quad \Longleftrightarrow \quad (\rho,\sigma) \succeq_{(\epsilon_\rho,\epsilon_\sigma)} (\rho',\sigma')\ \forall \rho',\sigma'. \tag{C1}$$

Proof. Using the definition of β_x, we have that there exists a test Q such that

$$\mathrm{Tr}((I-Q)\rho_1) \le \epsilon_\rho \quad \text{and} \quad \mathrm{Tr}(Q\sigma_1) \le \epsilon_\sigma. \tag{C2}$$

Consider a measure-and-prepare channel based on that very test, specifically

$$\mathcal{E}(\tau) := \rho_2\mathrm{Tr}(Q\tau) + \sigma_2\mathrm{Tr}((I-Q)\tau). \tag{C3}$$

Applying this channel, we can easily see that it has the desired error properties for any output dichotomy:

$$T(\mathcal{E}(\rho_1), \rho_2) = T(\rho_2, \sigma_2) \times \mathrm{Tr}((I-Q)\rho_1) \le \epsilon_\rho, \tag{C4a}$$

$$T(\mathcal{E}(\sigma_1), \sigma_2) = T(\rho_2, \sigma_2) \times \mathrm{Tr}(Q\sigma_1) \le \epsilon_\sigma. \quad \text{(C4b)}$$

As for the reverse direction, this simply follows from using the data-processing inequality for β_x and the output states $\rho' = |0\rangle\langle 0|$ and $\sigma' = |1\rangle\langle 1|$, as $\beta_x(|0\rangle\langle 0|\|\,|1\rangle\langle 1|) \equiv 0$. ■

So now let us move on to transformation rates. Similar to the one-sided error case, let $R_n^*(\epsilon_n^{(\rho)}, \epsilon_n^{(\sigma)})$ denote the largest R_n such that

$$\left(\rho_1^{\otimes n}, \sigma_1^{\otimes n}\right) \succeq_{(\epsilon_n^{(\rho)}, \epsilon_n^{(\sigma)})} \left(\rho_2^{\otimes R_n n}, \sigma_2^{\otimes R_n n}\right), \quad \text{(C5)}$$

where we note that $R_n^*(\epsilon_n) := R_n^*(\epsilon_n, 0)$.

In Theorems 2–7, we have dealt with the cases in which one error was exactly zero, splitting the results up by the scaling of the other error into seven different regimes (small, moderate low or high, large low or high, and extreme low or high). Naively, one might think that we then need to consider 49 different regimes for the general two-sided error problem (see Table I). However, Lemma 14 will allow us to instantly rule out 25 of these regimes in which neither error is exponentially small.

Lemma 14 (Rate breakdown). Suppose that neither $\epsilon_n^{(\rho)}$ nor $\epsilon_n^{(\sigma)}$ is exponentially bounded, i.e.,

$$\lim_{n\to\infty} \frac{1}{n} \log \epsilon_n^{(\rho)} = \lim_{n\to\infty} \frac{1}{n} \log \epsilon_n^{(\rho)} = 0. \quad \text{(C6)}$$

Then, the rate is eventually infinite: $R_n^*(\epsilon_n^{(\rho)}, \epsilon_n^{(\sigma)}) \overset{\text{ev.}}{=} +\infty$.

Proof. The idea here is to show that if neither error is exponentially shrinking, then eventually we see a breakdown of the Blackwell ordering in the sense of Lemma 14. As neither error is exponententially decaying, then we can take any arbitrarily small constant $\delta > 0$ and have that

$$\epsilon_n^{(\rho)} \overset{\text{ev.}}{>} \exp(-\delta n) \quad \text{and} \quad \epsilon_n^{(\sigma)} \overset{\text{ev.}}{>} \exp(-\delta n). \quad \text{(C7)}$$

From Lemma 5, we have

$$\gamma_{-\delta}\left(\rho_1^{\otimes n} \middle\| \sigma_1^{\otimes n}\right) = \Gamma_{-\delta}(\rho_1 \| \sigma_1). \quad \text{(C8)}$$

Given that $\Gamma_0(\rho\|\sigma) = -D(\rho\|\sigma) < 0$ and Γ_λ is continuous in λ, then for sufficiently small δ, we will also have $\Gamma_{-\delta}(\rho\|\sigma) < -\delta$. In terms of the type-II error probability,

$$\beta_{\exp(-\delta n)}\left(\rho_1^{\otimes n} \middle\| \sigma_1^{\otimes n}\right) \overset{\text{ev.}}{<} \exp(-\delta n). \quad \text{(C9)}$$

Lastly, we can use the monotonicity of β_x, which gives

$$\beta_{\epsilon_n^{(\rho)}}\left(\rho_1^{\otimes n} \middle\| \sigma_1^{\otimes n}\right) \overset{\text{ev.}}{\le} \beta_{L^{-1}[-\delta n]}\left(\rho_1^{\otimes n} \middle\| \sigma_1^{\otimes n}\right) \overset{\text{ev.}}{<} \exp(-\delta n) \overset{\text{ev.}}{<} \epsilon_n^{(\sigma)}, \quad \text{(C10)}$$

and so for sufficiently large n, Lemma 14 applies and thus the rate becomes infinite: $R_n^*(\epsilon_n^{(\rho)}, \epsilon_n^{(\sigma)}) \overset{\text{ev.}}{=} +\infty$. ■

As such, the only regimes left are the cases in which one error is exponentially small and the other is nonzero. For simplicity, we will assume that the second error is the exponentially small error for the rest of this appendix,

$$\epsilon_n^{(\sigma)} := \exp(-n\lambda_\sigma), \quad \text{(C11)}$$

and will discuss how this nonzero $\epsilon_n^{(\sigma)}$ modifies the results of Theorems 2–7 for different regimes of $\epsilon_n^{(\rho)}$. As we shall see below, in the large-deviation low-error regime, we obtain a nontrivial change in the asymptotic rate (Lemma 17) but in all other regimes we obtain that the one-sided error results hold unchanged up to a critical value of $\lambda_n^{(\rho)}$, beyond which we see a breakdown similar to Lemma 15, resulting in an eventually infinite rate (Lemma 16). A classification of the 49 two-sided error regimes is given in Table I.

TABLE I. A summary of the two-sided error results. The green results, corresponding to Theorems 2–7, are the one-sided error results presented in Sec. III. The red region, corresponding to Lemma 15, denotes where the Blackwell order breaks down, resulting in eventually infinite transformation rates. The yellow region, corresponding to Lemma 16, denotes the regimes in which the one-sided rates hold until a critical error exponent is reached, beyond which the Blackwell order once again breaks down. Finally the blue region, corresponding to Lemma 17, denotes the sole regime in which there is a nontrivial change in the transformation rate from the one-sided error case.

	Zero error	**Large$_<$**	**Moderate$_<$**	**Small**	**Moderate$_>$**	**Large$_>$**	**Extreme**
Zero error	Theorem 6	Theorem 4	Theorem 3	Theorem 3	Theorem 3	Theorem 5	Theorem 7
Large$_<$	Theorem 4	Lemma 16	Lemma 15				
Moderate$_<$	Theorem 3	Lemma 15	Lemma 14				
Small	Theorem 2						
Moderate$_>$	Theorem 3						
Large$_>$	Theorem 5						
Extreme	Theorem 7						

1. High errors

We will start with the small- and moderate-deviation results. In these cases, we will see that as long as the exponent of the second error, λ_σ, is above a certain critical exponent, then these regimes are left unchanged. But if it crosses, we also obtain a complete breakdown.

Lemma 15 (Unchanged two-sided rates). If $\epsilon_n^{(\rho)}$ is in the small- or moderate-deviation regime, $e^{\omega(n)} \leq \epsilon_n^{(\rho)} \leq 1 - e^{O(n)}$, and $\lambda_\sigma > D(\rho_1\|\sigma_1)$, then the small- and moderate-deviation results of Theorems 2 and 3 remain unchanged, and if $\lambda_\sigma < D(\rho_1\|\sigma_1)$, then $R_n^*(\epsilon_n^{(\rho)}, \epsilon_n^{(\sigma)}) \overset{\text{ev.}}{=} +\infty$.

If $\epsilon_n^{(\rho)}$ is in the high-error large-deviation regime, $\epsilon_n^{(\rho)} := 1 - \exp(-n\lambda_n^{(\rho)})$, and $\lambda_\sigma > -\Gamma_{\lambda_\rho}(\rho_1\|\sigma_1)$, then the high-error large-deviation results of Theorem 5 remain unchanged, and if $\lambda_\sigma < -\Gamma_{\lambda_\rho}(\rho_1\|\sigma_1)$, then $R_n^*(\epsilon_n^{(\rho)}, \epsilon_n^{(\sigma)}) \overset{\text{ev.}}{=} +\infty$.

If $\epsilon_n^{(\rho)}$ is in the extreme-deviation regime, $\epsilon_n^{(\rho)} = 1 - \exp(\omega(n))$, then Theorem 7 remains unchanged for any $\lambda_n^{(\sigma)}$.

Proof sketch. We start with the small and moderate cases. When $\lambda_\sigma < D(\rho_1\|\sigma_1)$, we can once again use Lemma 14. Specifically, the first-order contributions of Lemmas 4 and 6 give that

$$\lim_{n\to\infty} -\frac{1}{n}\log\beta_{\epsilon_n^{(\rho)}}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) = D(\rho_1\|\sigma_1), \tag{C12}$$

for any $\epsilon_n^{(\rho)}$ that is not exponentially approaching either 0 or 1. So, if $\lambda_\sigma < D(\rho_1\|\sigma_1)$, then $\epsilon_n^{(\sigma)}$ is decaying with a smaller exponent and must dominate this expression; specifically,

$$\beta_{\epsilon_n^{(\rho)}}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{\text{ev.}}{<} \epsilon_n^{(\sigma)}. \tag{C13}$$

Thus, by Lemma 14, the Blackwell order breaks down and $R_n^*(\epsilon_n^{(\rho)}, \epsilon_n^{(\sigma)}) \overset{\text{ev.}}{=} +\infty$.

Next, we want to argue that for $\lambda_\sigma > D(\rho_1\|\sigma_1)$, the results of Theorems 2 and 3 remain unchanged. Clearly, allowing errors on the second state can only *increase* the optimal transformation rate and so to demonstrate that this rate remains unchanged, we need only show that the upper bound (optimality) remains unchanged. The optimality bound of Theorems 2 and 3 comes from applying Lemma 2, which bounds the rate R_n by

$$\forall x \in \left(\epsilon_n^{(\rho)}, 1\right): \quad \beta_x\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \leq \beta_{x-\epsilon_n^{(\rho)}}\left(\rho_2^{\otimes R_n n}\middle\|\sigma_2^{\otimes R_n n}\right). \tag{C14}$$

In the presence of two-sided errors, this changes to

$$\forall x \in \left(\epsilon_n^{(\rho)}, 1\right): \\ \beta_x\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) - \epsilon_n^{(\sigma)} \leq \beta_{x-\epsilon_n^{(\rho)}}\left(\rho_2^{\otimes R_n n}\middle\|\sigma_2^{\otimes R_n n}\right). \tag{C15}$$

Now, consider the two terms on the left-hand side. By Eq. (C12), we know that the first β_x term is exponentially decaying with n, with an exponent of $D(\rho_1\|\sigma_1)$, and $\epsilon_n^{(\sigma)}$ is decaying with an exponent of λ_σ. As $\lambda_\sigma > D(\rho_1\|\sigma_1)$, we have that this error term is asymptotically dominated; specifically,

$$\lim_{n\to\infty} \frac{\beta_x\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) - \epsilon_n^{(\sigma)}}{\beta_x\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right)} = 1. \tag{C16}$$

As such, the $\epsilon_n^{(\sigma)}$ term in Eq. (C15) is asymptotically irrelevant, reducing this optimality bound to that given in Theorems 2 and 3.

We now turn to the high-error large-deviation regime. From Lemma 5, we have

$$\lim_{n\to\infty} -\frac{1}{n}\log\beta_{\epsilon_n^{(\rho)}}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) = \Gamma_{\lambda_\rho}(\rho_1\|\sigma_1). \tag{C17}$$

So, if $\lambda_\rho < -\Gamma_{\lambda_\rho}(\rho_1\|\sigma_1)$, then

$$\beta_{\epsilon_n^{(\rho)}}\left(\rho_1^{\otimes n}\middle\|\sigma_1^{\otimes n}\right) \overset{\text{ev.}}{<} \epsilon_n^{(\sigma)} \tag{C18}$$

and so by Lemma 14, we can conclude that $R_n^*(\epsilon_n^{(\rho)}, \epsilon_n^{(\sigma)}) \overset{\text{ev.}}{=} +\infty$. If, however, $\lambda_\rho > -\Gamma_{\lambda_\rho}(\rho_1\|\sigma_1)$, then the error term will be exponentially dominated by all of the relevant hypothesis-testing quantities in the optimality proof and therefore Theorem 5 will remain unchanged.

Lastly, Theorem 7 trivially remains unchanged, as the rate in that regime is unbounded, and introducing error on the second state can only increase the rate further. ■

2. Low errors

Finally, we are left with large deviation and low error. This is the one regime where a nontrivial change in the asymptotic rate occurs. For $\epsilon_n^{(\sigma)} = 0$, we have seen that the rate is given by $\overline{r}/\check{r}$ optimized over a range of type-I log odds determined by the error on the first state. Similarly, we will see that the optimal rate is once again an optimization of $\overline{r}/\check{r}$, this time optimized over a range of type-I log odds determined by the first state error *and* type-II log odds determined by the second state error. Before we can give the modified result, we first need to define $\check{\Gamma}_\lambda(\rho\|\sigma) := \min\left\{\overleftarrow{\Gamma}_\lambda(\rho\|\sigma), \overrightarrow{\Gamma}_\lambda(\rho\|\sigma)\right\}$, which we can evaluate using Lemmas 5 and 13 to be given by

$$\check{\Gamma}_\lambda(\rho\|\sigma) \\ = \begin{cases} \sup_{t<0} \check{D}_t(\rho\|\sigma) + \dfrac{t}{1-t}\lambda, & \lambda < -D(\sigma\|\rho), \\ \inf_{0<t<1} -\check{D}_t(\rho\|\sigma) - \dfrac{t}{1-t}\lambda, & -D(\sigma\|\rho) < \lambda < 0, \\ \sup_{t>1} -\check{D}_t(\rho\|\sigma) + \dfrac{t}{1-t}\lambda, & \lambda > 0, \end{cases}. \tag{C19}$$

Using this, we can now give the full two-sided low-error large-deviation result.

Lemma 16 (Two-sided large deviation, low error). For any error of the form $\epsilon_n^{(\rho)} = \exp(-\lambda_\rho n)$ with constant $\lambda_\rho > 0$, if $[\rho_2, \sigma_2] = 0$, then the optimal rate is lower bounded:

$$\liminf_{n\to\infty} R_n^*(\epsilon_n^{(\rho)}, \epsilon_n^{(\sigma)}) \geq \inf_{\substack{-\lambda_\rho < \mu < \lambda_\rho \\ -\lambda_\sigma < \check{\Gamma}_\mu(\rho_1\|\sigma_1) < \lambda_\sigma}} \check{r}(\mu). \tag{C20}$$

Furthermore, if we consider general output dichotomies, $[\rho_2, \sigma_2] \neq 0$, then the optimal rate is upper bounded by

$$\limsup_{n\to\infty} R_n^*(\epsilon_n^{(\rho)}, \epsilon_n^{(\sigma)}) \leq \inf_{\substack{-\lambda_\rho < \mu < \lambda_\rho \\ -\lambda_\sigma < \Gamma_\mu(\rho_1\|\sigma_1) < \lambda_\sigma}} \bar{r}(\mu). \tag{C21}$$

In the above, $\bar{r}$ and $\check{r}$ are defined in Sec. V C 2. Moreover, these expressions hold even if these domains are empty; i.e., if $\Gamma_{-\lambda_\rho}(\rho_1\|\sigma_1) < -\lambda_\sigma$, then $R_n^*(\epsilon_n^{(\rho)}, \epsilon_n^{(\sigma)}) \overset{\text{ev.}}{=} +\infty$.

Proof sketch. Here, we are just going to provide a sketch of the proof: for a more rigorous treatment of this argument, see the proof of Theorem 4. We start with optimality. Lemma 2 gives that, for any achievable rate R,

$$\begin{aligned} &\forall x \in (\epsilon_n^{(\rho)}, 1) : \\ &\beta_x\big(\rho_1^{\otimes n}\big\|\sigma_1^{\otimes n}\big) - \epsilon_n^{(\sigma)} \leq \beta_{x-\epsilon_n^{(\rho)}}\big(\rho_2^{\otimes Rn}\big\|\sigma_2^{\otimes Rn}\big). \end{aligned} \tag{C22}$$

First, we reparametrize $x \to x + \epsilon_n^{(\rho)}$, which gives

$$\begin{aligned} &\forall x \in (\epsilon_n^{(\rho)}/2, 1 - \epsilon_n^{(\rho)}/2) : \\ &\beta_{x+\epsilon_n^{(\rho)}/2} - \epsilon_n^{(\sigma)} \leq \beta_{x-\epsilon_n^{(\rho)}/2}\big(\rho_2^{\otimes Rn}\big\|\sigma_2^{\otimes Rn}\big). \end{aligned} \tag{C23}$$

Next, we want to reparametrize again by the log odds per copy instead of a probability. Specifically, we will switch from x to μ, where $x = L^{-1}[\mu n]$. Doing so gives

$$\begin{aligned} &\forall \mu \in (-\lambda_\rho, +\lambda_\rho) : \\ &\beta_{L^{-1}[\mu n]+\epsilon_n^{(\rho)}/2}\big(\rho_1^{\otimes n}\big\|\sigma_1^{\otimes n}\big) - \epsilon_n^{(\sigma)} \\ &\qquad \leq \beta_{L^{-1}[\mu n]-\epsilon_n^{(\rho)}/2}\big(\rho_2^{\otimes Rn}\big\|\sigma_2^{\otimes Rn}\big). \end{aligned} \tag{C24}$$

As $|\mu| < \lambda_\rho$, we have that the $L^{-1}[\mu n]$ terms must dominate over the $\epsilon_n^{(\rho)}$ terms; specifically,

$$\lim_{n\to\infty} \frac{1}{n} L\left[L^{-1}[\mu n] \pm \epsilon_n^{(\rho)}/2\right] = \mu \tag{C25}$$

and so this essentially reduces to

$$\begin{aligned} &\forall \mu \in (-\lambda_\rho, +\lambda_\rho) : \\ &\beta_{L^{-1}[\mu n]}\big(\rho_1^{\otimes n}\big\|\sigma_1^{\otimes n}\big) - \epsilon_n^{(\sigma)} \leq \beta_{L^{-1}[\mu n]}\big(\rho_2^{\otimes Rn}\big\|\sigma_2^{\otimes Rn}\big). \end{aligned} \tag{C26}$$

Put in terms of log odds per copy, this is

$$\begin{aligned} &\forall \mu \in (-\lambda_\rho, +\lambda_\rho) : \\ &L^{-1}\left[\gamma_{\mu n}\big(\rho_1^{\otimes n}\big\|\sigma_1^{\otimes n}\big)\right] - \epsilon_n^{(\sigma)} \leq L^{-1}\left[\gamma_{\mu n}\big(\rho_2^{\otimes Rn}\big\|\sigma_2^{\otimes Rn}\big)\right]. \end{aligned} \tag{C27}$$

Now, we can use Lemma 5, which gives

$$\lim_{n\to\infty} \frac{1}{n}\gamma_{\mu n}\big(\rho_1^{\otimes n}\big\|\sigma_1^{\otimes n}\big) = \Gamma_\mu(\rho_1\|\sigma_1), \tag{C28a}$$

$$\lim_{n\to\infty} \frac{1}{n}\gamma_{\mu n}\big(\rho_2^{\otimes Rn}\big\|\sigma_2^{\otimes Rn}\big) = R\Gamma_{\mu/R}(\rho_2\|\sigma_2). \tag{C28b}$$

So in Eq. (C27) we have that the log-odds terms on both sides are exponentially scaling. In the absence of $\epsilon_n^{(\sigma)}$, we can directly compare this, giving the optimality presented in Theorem 4.

We can break the analysis of Eq. (C27) into three cases based on how $\Gamma_\mu(\rho_1\|\sigma_1)$ compares to $\pm\lambda_\sigma$. If $\Gamma_\mu(\rho_1\|\sigma_1) < -\lambda_\sigma$, then the left-hand side of Eq. (C27) is eventually negative and thus trivially satisfied. If $-\lambda_\sigma \leq \Gamma_\mu(\rho_1\|\sigma_1) \leq +\lambda_\sigma$, then Eq. (C27) reduces to

$$\Gamma_\mu(\rho\|\sigma) \leq R\Gamma_{\mu/R}(\rho_2\|\sigma_2), \tag{C29}$$

as we saw in the absence of ϵ_n^σ, which in turn gives the bound $R \leq \bar{r}(\mu)$. Lastly, we have $\Gamma_\mu(\rho_1\|\sigma_1) > +\lambda_\sigma$, in which case the left-hand side of Eq. (C27) scales as $1 - \epsilon_n^\sigma$, so this reduces to

$$\lambda_\sigma \leq \Gamma_{\mu/R}(\rho_2\|\sigma_2), \tag{C30}$$

which is strictly weaker than the constraint $R \leq \bar{r}(\mu)$ for the μ for which $\Gamma_\mu(\rho_1\|\sigma_1) = +\lambda_\sigma$. The upshot is that we are left with an expression similar to Theorem 4, with the rate being an optimization of $\bar{r}$, this time with a constraint both on μ (coming from $\epsilon_n^{(\rho)}$) *and* on $\Gamma_\mu(\rho_1\|\sigma_1)$ (coming from $\epsilon_n^{(\sigma)}$). Specifically,

$$R \leq \inf_{\substack{-\lambda_\rho < \mu < \lambda_\rho \\ -\lambda_\sigma < \Gamma_\mu(\rho_1\|\sigma_1) < \lambda_\sigma}} \bar{r}(\mu). \tag{C31}$$

The same sort of argumentation works for the achievability, where we find that any rate r such that

$$r \leq \inf_{\substack{-\lambda_\rho < \mu < \lambda_\rho \\ -\lambda_\sigma < \check{\Gamma}_\mu(\rho_1\|\sigma_1) < \lambda_\sigma}} \check{r}(\mu) \tag{C32}$$

is achievable.

Finally, we note that the above arguments also hold if the domains of the infima are empty. Specifically, if

$$\Gamma_{-\lambda_\rho}(\rho_1\|\sigma_1) < -\lambda_\sigma, \tag{C33}$$

then Lemma 14 gives us that the rate breaks down and so for sufficiently large n, the optimal rate becomes infinite. ■

APPENDIX D: PROOF OF THEOREM 8

In order to prove Theorem 8, we start with the following lemma.

Lemma 17. Consider the initial and target states, ρ_1 and ρ_2, together with the corresponding thermal states, γ_1 and γ_2, such that $[\rho_2, \gamma_2] = 0$. Then, the condition

$$\forall x \in (\epsilon, 1): \quad \overleftarrow{\beta}_x(\rho_1\|\gamma_1) \leq \beta_{x-\epsilon}(\rho_2\|\gamma_2) \tag{D1}$$

implies that there exists a thermal operation mapping ρ_1 into a state ϵ-close to ρ_2 in trace distance:

$$\rho_1 \xrightarrow[\text{TO}]{\epsilon} \rho_2. \tag{D2}$$

Note that in general, however, the right-pinched variant

$$\forall x \in (\epsilon, 1): \quad \overrightarrow{\beta}_x(\rho_1\|\gamma_1) \leq \beta_{x-\epsilon}(\rho_2\|\gamma_2) \tag{D3}$$

does not necessarily similarly yield a TO-achievable Blackwell order.

Proof. First, note that a pinching map with respect to the eigenspaces of the thermal state γ_1 is a thermal operation and so $\mathcal{P}_{\gamma_1}(\rho_1)$ can be obtained from ρ_1. Since $[\mathcal{P}_{\gamma_1}(\rho_1), \gamma_1] = 0$ and $[\rho_2, \gamma_2] = 0$ by assumption, we are dealing with initial and target states commuting with the respective thermal states. For such states, however, it is known from Ref. [70] that the condition

$$\forall x \in (\epsilon, 1): \quad \beta_x(\mathcal{P}_{\gamma_1}(\rho_1)\|\gamma_1) \leq \beta_{x-\epsilon}(\rho_2\|\gamma_2) \tag{D4}$$

is equivalent to the existence of a thermal operation $\mathcal{E}$ mapping $\mathcal{P}_{\gamma_1}(\rho_1)$ into a state ϵ-close to ρ_2. However, given the definition of $\overleftarrow{\beta}_x$ from Eq. (62), the above is equivalent to Eq. (D1). Thus, assuming that Eq. (D1) holds, such $\mathcal{E}$ exists and a composition of thermal operations $\mathcal{E} \circ \mathcal{P}_{\gamma_1}$, which is itself a thermal operation, maps ρ_1 into a state ϵ-close to ρ_2. While the right-pinched condition similarly yields a Blackwell order, the right-pinching operation $\mathcal{P}_{\rho_1}(\cdot)$ is not a thermal operation (unless ρ_1 and γ_1 commute). ■

We now need to recall the general strategy used to prove Theorems 2, 3, 5, and 7 in Sec. V C. In these cases, the achievability has exclusively used the left-pinched sufficient condition of Lemma 2, showing that this condition gives a rate with the same asymptotic expansion as the optimality bound given by the necessary condition of Lemma 2. Thus, using Lemma 18, we conclude that the optimal rates from Theorems 2, 3, 5, and 7 in Sec. V C can be achieved by thermal operations.

In the achievability proofs of Theorems 4 and 6, we have needed to leverage both left and right pinching and thus these results are not necessarily TO achievable. In both proofs, however, we have started by proving separate achievability results using left and right pinching separately and we have constructed the final bound by combining the two. If we eschew the right-pinch-based bound and stick to the left-pinch-based bound, then these proofs do yield weaker, but TO-achievable, rates. For the low-error large-deviation case of Theorem 4, the TO-achievable rate is

$$\liminf_{n\to\infty} R_n^*\big(\exp(-\lambda n)\big) \geq \min_{-\lambda \leq \mu \leq \lambda} \overleftarrow{r}(\mu), \tag{D5}$$

where $\overleftarrow{r}(\mu)$ is defined in Sec. V C 2. Similarly, for the zero-error case of Theorem 6 the TO-achievable rate is

$$\liminf_{n\to\infty} R_n^*(0) \geq \inf_{\alpha\in\mathbb{R}} \frac{\overleftarrow{D}_\alpha(\rho_1\|\sigma_1)}{D_\alpha(\rho_2\|\sigma_2)}. \tag{D6}$$

Furthermore, when dealing with energy-incoherent states, there is no need for pinching and so one can stick only to thermal operations. Moreover, for commuting input *and* output states, the lower and upper bounds for the optimal rates in Theorems 4 and 6 coincide (as they only differ by choice of Rényi divergence) and thus these theorems yield optimal transformation rates in their respective error regimes.

APPENDIX E: PROOF OF THEOREM 9

In this appendix, we present how one can modify the reasoning used to prove Theorem 2 to prove Theorem 9. Our aim is thus to find R_n^*, which is the largest rate R_n for which the following transformation can be performed by thermal operations:

$$\rho_1^{\otimes n} \otimes |0\rangle\langle 0|_W \xrightarrow[\text{TO}]{\epsilon} \rho_2^{\otimes nR_n} \otimes |1\rangle\langle 1|_W. \tag{E1}$$

Let us recall that here W denotes the ancillary battery system with energy levels $|0\rangle_W$ and $|1\rangle_W$ separated by energy gap w, so that the thermal state of the battery is given by

$$\gamma_W = \lambda|0\rangle\langle 0|_W + (1-\lambda)|1\rangle\langle 1|_W \quad \text{with } \lambda = \frac{1}{1+e^{-\beta w}}. \tag{E2}$$

From Lemma 2, we know that the necessary condition for that is given by

$$\begin{aligned} &\forall x \in (\epsilon, 1): \\ &\beta_x(\rho_1^{\otimes n} \otimes |0\rangle\langle 0|_W \| \gamma_1^{\otimes n} \otimes \gamma_W) \\ &\quad \leq \beta_{x-\epsilon}(\rho_2^{\otimes nR_n} \otimes |1\rangle\langle 1|_W \| \gamma_2^{\otimes nR_n} \otimes \gamma_W), \end{aligned} \tag{E3}$$

whereas from Lemmas 2 and 18, we know that the sufficient condition is given by

$$\forall x \in (\epsilon, 1):$$
$$\overleftarrow{\beta}_x(\rho_1^{\otimes n} \otimes |0\rangle\langle 0|_W \| \gamma_1^{\otimes n} \otimes \gamma_W) \leq \beta_{x-\epsilon}(\rho_2^{\otimes nR_n} \otimes |1\rangle\langle 1|_W \| \gamma_2^{\otimes nR_n} \otimes \gamma_W). \quad \text{(E4)}$$

We will simplify these conditions by using the fact that

$$\overleftarrow{\beta}_x(\rho_1^{\otimes n} \otimes |0\rangle\langle 0|_W \| \gamma_1^{\otimes n} \otimes \gamma_W) = \beta_x(\mathcal{P}_{\gamma_1^{\otimes n}}(\rho_1^{\otimes n}) \otimes |0\rangle\langle 0|_W \| \gamma_1^{\otimes n} \otimes \gamma_W) \quad \text{(E5)}$$

and employing the following lemma.

Lemma 18. Consider three quantum states, ρ, σ, and γ, where $\gamma = \sum_i \gamma_i |i\rangle\langle i|$. Then, for all $x \in [0, 1]$ we have

$$\beta_x(\rho \otimes |i\rangle\langle i| \, \| \sigma \otimes \gamma) = \gamma_i \beta_x(\rho \| \sigma). \quad \text{(E6)}$$

Proof. Expanding out the left-hand side of Eq. (E6) using the definition from Eqs. (3a)–(3c) and decomposing the test as $Q := \sum_{i,j} Q_{ij} \otimes |i\rangle\langle j|$ yields

$$\beta_x(\rho \otimes |i\rangle\langle i| \, \| \sigma \otimes \gamma) = \min_Q \{\mathrm{Tr}[(\sigma \otimes \gamma)Q] \,|\, \mathrm{Tr}[(\rho \otimes |i\rangle\langle i|)Q] \geq 1 - x \quad \text{and} \quad 0 \leq Q \leq 1\} \quad \text{(E7a)}$$
$$= \min_{Q_{kk}} \left\{ \sum_k \gamma_k \mathrm{Tr}[\sigma Q_{kk}] \,\middle|\, \mathrm{Tr}[\rho Q_{ii}] \geq 1 - x \quad \text{and} \quad 0 \leq Q_{kk} \leq 1 \quad \text{for all } k \right\} \quad \text{(E7b)}$$
$$= \gamma_i \min_{Q_{ii}} \{\mathrm{Tr}[\sigma Q_{ii}] \,|\, \mathrm{Tr}[\rho Q_{ii}] \geq 1 - x \quad \text{and} \quad 0 \leq Q_{ii} \leq 1\} \quad \text{(E7c)}$$
$$= \gamma_i \beta_x(\rho \| \sigma), \quad \text{(E7d)}$$

which proves the claim. ■

We can then rewrite Eqs. (E3) and (E4) as

$$\forall x \in (\epsilon, 1): \quad \lambda \beta_x(\rho_1^{\otimes n} \| \gamma_1^{\otimes n}) \leq (1 - \lambda) \beta_{x-\epsilon}(\rho_2^{\otimes nR_n^*} \| \gamma_2^{\otimes nR_n^*}), \quad \text{(E8a)}$$
$$\forall x \in (\epsilon, 1): \quad \lambda \overleftarrow{\beta}_x(\rho_1^{\otimes n} \| \gamma_1^{\otimes n}) \leq (1 - \lambda) \beta_{x-\epsilon}(\rho_2^{\otimes nR_n^*} \| \gamma_2^{\otimes nR_n^*}). \quad \text{(E8b)}$$

Taking the minus log of both sides and dividing by n, we thus obtain that the necessary condition and the sufficient condition for the transformation in Eq. (E1) are given by

$$\forall x \in (\epsilon, 1): \quad -\frac{1}{n} \log\left(\beta_x(\rho_1^{\otimes n} \| \gamma_1^{\otimes n})\right) - \frac{\beta w}{n} \geq R_n \left(-\frac{1}{nR_n} \log\left(\beta_{x-\epsilon}(\rho_2^{\otimes nR_n} \| \gamma_2^{\otimes nR_n})\right) \right), \quad \text{(E9a)}$$
$$\forall x \in (\epsilon, 1): \quad -\frac{1}{n} \log\left(\overleftarrow{\beta}_x(\rho_1^{\otimes n} \| \gamma_1^{\otimes n})\right) - \frac{\beta w}{n} \geq R_n \left(-\frac{1}{nR_n} \log\left(\beta_{x-\epsilon}(\rho_2^{\otimes nR_n} \| \gamma_2^{\otimes nR_n})\right) \right). \quad \text{(E9b)}$$

Crucially now, as Lemma 4 tells us that the second-order asymptotic expansions of $-(1/n)\log \beta_x$ and $-(1/n)\log \overleftarrow{\beta}_x$ are the same, in the small-deviation regime, the above necessary and sufficient conditions coincide and are given by

$$\forall x \in (\epsilon, 1): \quad D(\rho_1 \| \gamma_1) + \sqrt{\frac{V(\rho_1 \| \gamma_1)}{n}} \Phi^{-1}(x) - \frac{\beta w}{n} \gtrsim R_n D(\rho_2 \| \gamma_2) + \sqrt{\frac{R_n V(\rho_2 \| \gamma_2)}{n}} \Phi^{-1}(x - \epsilon), \quad \text{(E10)}$$

where $\gtrsim$ denotes inequality up to terms $o(1/\sqrt{n})$. Introducing

$$\xi' := \frac{V(\rho_1 \| \gamma_1)}{R_n V(\rho_2 \| \gamma_2)}, \quad \text{(E11)}$$

using the definition and properties of the sesquinormal distribution, one can rearrange Eq. (E10) to arrive at the following equivalent condition:

$$\frac{\beta w}{n} \lesssim D(\rho_1\|\gamma_1) - R_n D(\rho_2\|\gamma_2) + \sqrt{\frac{V(\rho_1\|\gamma_1)}{n}} S^{-1}_{1/\xi'}(\epsilon). \tag{E12}$$

Clearly, if $\rho_2 = \gamma_2$, then the above is satisfied for any rate R_n as long as

$$\frac{\beta w}{n} \lesssim D(\rho_1\|\gamma_1) + \sqrt{\frac{V(\rho_1\|\gamma_1)}{n}} \Phi^{-1}(\epsilon), \tag{E13}$$

which proves the second part of Theorem 9. If $\rho_2 \neq \gamma_2$, then we can expand w and rearrange Eq. (E10) to obtain

$$R_n \lesssim \frac{D(\rho_1\|\gamma_1) - \beta w_1}{D(\rho_2\|\gamma_2)} + \frac{\sqrt{V(\rho_1\|\gamma_1)} S^{-1}_{1/\xi'}(\epsilon) - \beta w_2}{\sqrt{n} D(\rho_2\|\gamma_2)}. \tag{E14}$$

Now, we note that in the expression for ξ', we only need to account for the constant term of R_n, as any higher-order terms will result in corrections of the order $o(1/\sqrt{n})$. Thus, a transformation from Eq. (E1) exists for every rate R_n satisfying the above inequality with

$$\xi' = \frac{V(\rho_1\|\gamma_1)}{D(\rho_1\|\gamma_1) - \beta w_1} \Bigg/ \frac{V(\rho_2\|\gamma_2)}{D(\rho_2\|\gamma_2)}, \tag{E15}$$

which proves the first part of Theorem 9.

APPENDIX F: PROOF SKETCH OF THEOREM 10

The proof of Theorem 10 largely follows the proofs covered in Theorems 2–7 and so instead of reproducing all the gory details, we will instead point out some key differences and then give the resulting rate expressions. Consider a transformation $|\psi_1\rangle^{\otimes n} \xrightarrow[\text{LOCC}]{\epsilon} |\psi_2\rangle^{\otimes Rn}$ for bipartite states $|\psi_1\rangle$ and $|\psi_2\rangle$ with local dimensions d_1 and d_2, and with Schmidt spectra $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$. Recalling Eq. (8), such a transformation is possible if and only if

$$d_2^{Rn} \beta_x\big(\boldsymbol{p_2}^{\otimes Rn}\big\|\boldsymbol{f_2}^{\otimes Rn}\big) \leq d_1^n \beta_{x-\epsilon}\big(\boldsymbol{p_1}^{\otimes n}\big\|\boldsymbol{f_1}^{\otimes n}\big) \quad \forall x \in (\epsilon, 1), \tag{F1}$$

where $\boldsymbol{f}_i$ denotes a uniform distributions of dimension d_i and $\beta_x(\boldsymbol{p}\|\boldsymbol{q})$ should be understood as $\beta_x(\rho\|\sigma)$, with ρ and σ being diagonal states and with the diagonals given by $\boldsymbol{p}$ and $\boldsymbol{q}$, respectively. Applying the techniques of Sec. V C to convert hypothesis-testing asymptotics into transformation-rate asymptotics, we can extract from this second-order expressions for transformation rates in the entanglement setting.

Importantly, this condition has three major differences that will influence the resulting rate expressions. First, the order of the expression is backward compared to that seen in the thermodynamic setting, so the resulting rates will be reciprocated. Second, all hypothesis testing is relative to uniform states, meaning that all of our rates will involve information-theoretic quantities relative to the uniform states. All of these can be expressed in terms of their nonrelative analogues, e.g.,

$$D_\alpha(\boldsymbol{p}_i\|\boldsymbol{f}_i) = \frac{\alpha}{\alpha-1}\log d_i - H_\alpha(\boldsymbol{p}_i). \tag{F2}$$

And then, third, we have the lingering dimensional factors, which happen to all cancel out in such a way as to yield rate expressions broadly similar to those seen in Theorems 2–7.

Taking these modifications into account, if one was to follow our techniques from Sec. V C *mutatis mutandis*, the entanglement transformation rates, for $\lambda > 0$ and $a, \epsilon \in (0,1)$, scale as follows:

Zero error: $$R_n^*(0) = \min_{0\leq\alpha\leq\infty} \frac{H_\alpha(p)}{H_\alpha(q)} + o(1), \tag{F3a}$$

Large deviation (lo): $$R_n^*(\exp(-\lambda n)) = \min_{-\lambda\leq\mu\leq\lambda} r(\mu), \tag{F3b}$$

Moderate deviation (lo): $$R_n^*(\exp(-\lambda n^a)) = \frac{H(p) - |1-\xi^{-1/2}|\sqrt{2V(p)n^{a-1}} \times S^{-1}_{1/\xi}(\epsilon)}{H(q)} + o\left(\sqrt{n^{a-1}}\right), \tag{F3c}$$

Small deviation: $$R_n^*(\epsilon) = \frac{H(p) + \sqrt{V(p)/n} \times S^{-1}_{1/\xi}(\epsilon)}{H(q)} + o(1/\sqrt{n}), \tag{F3d}$$

Moderate deviation (hi): $$R_n^*(1-\exp(-\lambda n^a)) = \frac{H(p) + \left[1+\xi^{-1/2}\right]\sqrt{2V(p)n^{a-1}} \times S^{-1}_{1/\xi}(\epsilon)}{H(q)} + o\left(\sqrt{n^{a-1}}\right), \tag{F3e}$$

Large deviation (hi): $$R_n^*(1-\exp(-\lambda n)) = \inf_{\substack{t_1>1\\0<t_2<1}} \frac{H_{t_1}(p) - \left(\frac{t_1}{1-t_1} + \frac{t_2}{1-t_2}\right)\lambda}{H_{t_2}(q)} + o(1), \tag{F3f}$$

where

$$\xi = \frac{V(p)}{H(p)} \Big/ \frac{V(q)}{H(q)} \quad \text{and}$$

$$r(\mu) = \begin{cases} 1, & \mu \leq -D(f\,\|p), \\ \sup_{0<t_2<1} \inf_{0<t_1<1} \dfrac{H_{t_1}(p) + \left(\frac{t_2}{1-t_2} - \frac{t_1}{1-t_1}\right)\mu}{H_{t_2}(q)}, & -D(f\,\|p) \leq \mu \leq 0, \\ \inf_{t_2>1} \sup_{t_1>1} \dfrac{H_{t_1}(p) + \left(\frac{t_1}{1-t_1} - \frac{t_2}{1-t_2}\right)\mu}{H_{t_2}(q)}, & \mu \geq 0. \end{cases} \tag{F4}$$

We note that the small- and moderate-deviation rates are consistent in form with the existing infidelity results of Refs. [89,117], respectively, and the zero-error rate is a restatement of Ref. [102].

APPENDIX G: NUMERICAL EXAMPLES OF STRONG AND WEAK RESONANCE

In this appendix, we will give a numerical example of a dichotomy transformation that exhibits both weak and strong resonance in the sense discussed in Sec. IV C 3. Following Ref. [101], we can construct examples with resonance by considering two different input states and varying the relative ratio of their numbers. That is, instead of just considering the rates R_n and errors ϵ_n such that

$$(\rho_1^{\otimes n}, \sigma^{\otimes n}) \succeq_{(\epsilon_n,0)} (\rho_2^{\otimes R_n n}, \sigma^{\otimes R_n n}), \tag{G1}$$

we can instead consider

$$(\rho_1^{\otimes \lambda n} \otimes \rho_1'^{\otimes(1-\lambda)n}, \sigma^{\otimes n}) \succeq_{(\epsilon_n,0)} (\rho_2^{\otimes R_n n}, \sigma^{\otimes R_n n}), \tag{G2}$$

for some $\lambda \in (0, 1)$. Consider the states

$$\rho_1 = \mathrm{Diag}(0.4309, 0.4300, 0.1391), \tag{G3a}$$
$$\rho_1' = \mathrm{Diag}(0.5499, 0.2300, 0.2201), \tag{G3b}$$
$$\rho_2 = \mathrm{Diag}(0.5121, 0.3300, 0.1579), \tag{G3c}$$
$$\sigma = \mathrm{Diag}(0.3333, 0.3333, 0.3333). \tag{G3d}$$

These states exhibit *weak* resonance, as shown in the left panel of Fig. 6. Alternatively, if we consider the reverse process of attempting to make a mixture of two possible output states,

$$(\rho_2^{\otimes n}, \sigma^{\otimes n}) \succeq_{(\epsilon_n,0)} (\rho_1^{\otimes \lambda R_n n} \otimes \rho_1'^{\otimes(1-\lambda)R_n n}, \sigma^{\otimes R_n n}), \tag{G4}$$

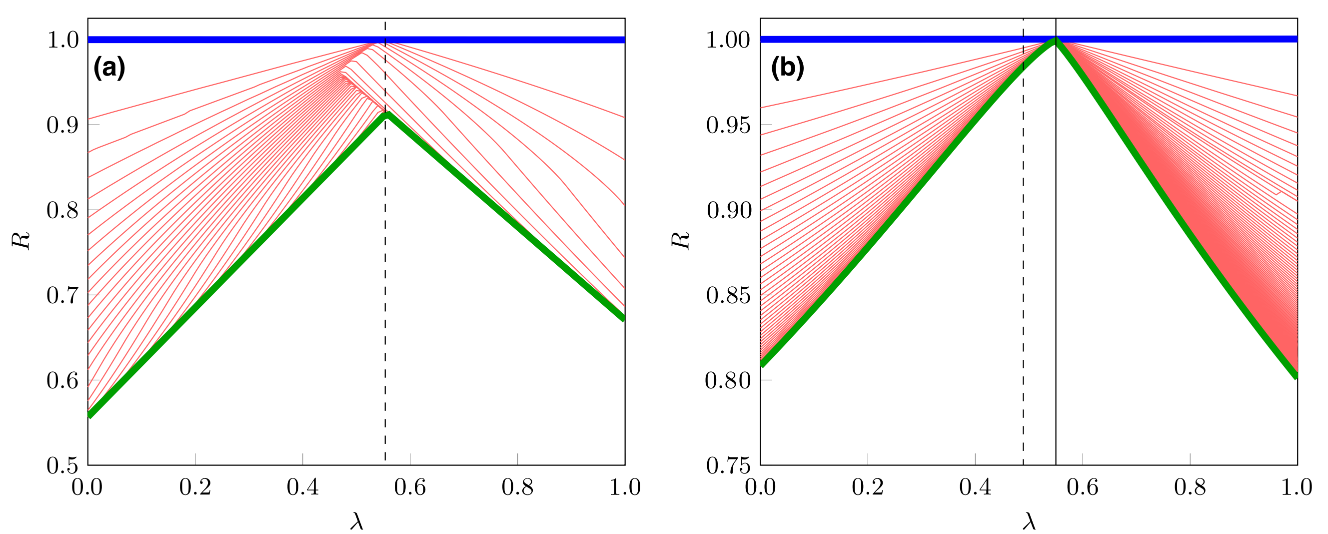


FIG. 6. Examples of (a) strong and (b) weak resonance. The upper blue lines correspond to the first-order rate, *à la* Theorem 1. The lower green lines correspond to the zero-error transformation rates, *à la* Theorem 6. The internal red lines correspond to the optimal rates at an error level of $\exp(-\mu n)$, with each line corresponding to a different value of μ. In the weak case $\mu \in \{0, 0.05, \ldots, 2\}$ and in the strong case $\mu \in \{0, 0.01, \ldots, 1\}$. Lastly, the vertical dashed black lines correspond to the mixture at which the weak-resonance condition is satisfied and the vertical solid black line to where the strong-resonance condition is met.

then this in fact exhibits *strong* resonance, as shown in the right panel of Fig. 6. As we can see, the weak-resonance condition determines the behavior of rates for high errors. But when the strong resonance is present, it dominates over this and in fact determines the behavior of rates at all error levels.

APPENDIX H: ASYMPTOTIC CONSISTENCY

In this appendix, we will show rather satisfying "asymptotic consistencies" among our results, namely, the hypothesis-testing results of Lemmas 4–6 and Lemma 8, the transformation-rate results of Theorems 2–7, and the resonance phenomena considered in Sec. IV C 3.

In Sec. V we have considered deriving the asymptotic behavior of both hypothesis testing and transformation rates for dichotomies in several different error regimes (Figs. 1 and 4). Formally, these results must be separately proven in each of these distinct regimes. Eschewing rigor for the moment [130], we might ask what happens if we take results from each error regime and naively take the limit approaching the neighboring regime. By *asymptotic consistency*, we mean that this blasphemous and heretical procedure manages to reproduce the results of the rigorous treatments given in Sec. V.

1. Small and moderate deviation

The small-deviation error regime refers to errors $\epsilon \in (0,1)$ that are constant in n and the moderate-deviation regime concerns errors ϵ_n that are subexponentially approaching either 0 (low error) or 1 (high error). Here, we will consider starting with the small-deviation results (Lamma 4 and Theorem 2) and then applying expansions of this result around $\epsilon = 0, 1$, showing that this gives an entirely nonrigorous reproduction of the moderate-deviation results (Lemma 6 and Theorem 3).

As noted in Ref. [117], the inverse cdf of the standard Gaussian can be expanded for small positive ϵ as

$$\Phi^{-1}(\epsilon) \approx -\sqrt{\ln 1/\epsilon^2} \quad \text{and} \quad \Phi^{-1}(1-\epsilon) \approx +\sqrt{\ln 1/\epsilon^2}. \tag{H1}$$

Next, consider the small-deviation expansion of the type-II hypothesis-testing error given in Lemma 4:

$$-\frac{1}{n}\beta_\epsilon\left(\rho^{\otimes}\middle\|\sigma^{\otimes n}\right) \approx D(\rho\|\sigma) + \sqrt{\frac{V(\rho\|\sigma)}{n}} \times \Phi^{-1}(\epsilon). \tag{H2}$$

If we simply substitute into these moderate error sequences $\epsilon_n := \exp(-\lambda n^a)$ or $1-\epsilon_n$ for $\lambda > 0, a \in (0,1)$ and use the above expansions, then we recover the moderate-deviation expansion given in Lemma 6:

$$-\frac{1}{n}\beta_{\epsilon_n}\left(\rho^{\otimes}\middle\|\sigma^{\otimes n}\right) \approx D(\rho\|\sigma) - \sqrt{2V(\rho\|\sigma)\lambda n^{a-1}}, \tag{H3a}$$

$$-\frac{1}{n}\beta_{1-\epsilon_n}\left(\rho^{\otimes}\middle\|\sigma^{\otimes n}\right) \approx D(\rho\|\sigma) + \sqrt{2V(\rho\|\sigma)\lambda n^{a-1}}. \tag{H3b}$$

As for the dichotomy-transformation rates, we need to consider expansions not just of the standard Gaussian but also of the sesquinormal distribution considered in Sec. III A. In Lemma 1, we have seen that the sesquinormal distribution can be expressed in terms of the standard Gaussian distribution. Using this, we can expand the sesquinormal inverse cdf for small positive ϵ as

$$S_{1/\xi}^{-1}(\epsilon) \approx -|1-\xi^{-1/2}|\sqrt{\ln 1/\epsilon^2} \quad \text{and}$$
$$S_{1/\xi}^{-1}(1-\epsilon) \approx +\left[1+\xi^{-1/2}\right]\sqrt{\ln 1/\epsilon^2}. \tag{H4}$$

Similar to the case of hypothesis testing, if we take the small-deviation dichotomy-transformation rate (Theorem 2),

$$R_n^*(\epsilon) \approx \frac{D(\rho_1\|\sigma_1) + \sqrt{V(\rho_1\|\sigma_1)/n} \times S_{1/\xi}^{-1}(\epsilon)}{D(\rho_2\|\sigma_2)}, \tag{H5}$$

and substitute moderate error rates, we reproduce the moderate-deviation results (Theorem 3):

$$R_n^*\left(\exp(-\lambda n^a)\right) \approx \frac{D(\rho_1\|\sigma_1) - |1-\xi^{-1/2}| \times \sqrt{2\lambda V(\rho_1\|\sigma_1) n^{a-1}}}{D(\rho_2\|\sigma_2)}, \tag{H6a}$$

$$R_n^*\left(1-\exp(-\lambda n^a)\right) \approx \frac{D(\rho_1\|\sigma_1) + \left[1+\xi^{-1/2}\right] \times \sqrt{2\lambda V(\rho_1\|\sigma_1) n^{a-1}}}{D(\rho_2\|\sigma_2)}. \tag{H6b}$$

2. Large and moderate deviation

The moderate-deviation regime serves as a barrier between the small- and large-deviation regimes. As such, an alternative way of recovering the moderate-deviation results is to consider the limit of large deviations—specifically, errors that are exponentially approaching 0 or 1—but then consider the limit where we treat that exponent as arbitrarily small.

In the case of hypothesis testing, the large-deviation results (Lemma 5) are

$$\frac{1}{n}\gamma_{\lambda n}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big) \to \begin{cases} \sup_{t<0} \breve{D}_t(\rho\|\sigma) + \dfrac{t}{1-t}\lambda, & \lambda \leq -D(\sigma\|\rho), \\ \inf_{0<t<1} -\overline{D}_t(\rho\|\sigma) - \dfrac{t}{1-t}\lambda, & -D(\sigma\|\rho) \leq \lambda \leq 0, \\ \sup_{t>1} -\breve{D}_t(\rho\|\sigma) + \dfrac{t}{1-t}\lambda, & \lambda \geq 0. \end{cases} \tag{H7}$$

Substituting moderate errors into the large-deviation result gives the expressions

$$\frac{1}{n}\gamma_{-\lambda n^a}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big) \approx \inf_{0<t<1} -\overline{D}_t(\rho\|\sigma) + \frac{t}{1-t}\lambda n^{a-1}, \tag{H8a}$$

$$\frac{1}{n}\gamma_{+\lambda n^a}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big) \approx \sup_{t>1} -\breve{D}_t(\rho\|\sigma) + \frac{t}{1-t}\lambda n^{a-1}, \tag{H8b}$$

where $\lambda > 0$ and $a \in (0,1)$. In both cases, the optimizations approach $t \approx 1$ in this moderate regime, so we can expand the Rényi entropies using

$$\breve{D}_t(\rho\|\sigma) \approx \overline{D}_t(\rho\|\sigma) \approx D(\rho\|\sigma) + \frac{t-1}{2}V(\rho\|\sigma), \tag{H9}$$

which gives

$$\frac{1}{n}\gamma_{-\lambda n^a}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big) \approx \inf_{t<1} -D(\rho\|\sigma) + \frac{1-t}{2}V(\rho\|\sigma) + \frac{t}{1-t}\lambda n^{a-1}, \tag{H10a}$$

$$\frac{1}{n}\gamma_{+\lambda n^a}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big) \approx \sup_{t>1} -D(\rho\|\sigma) + \frac{1-t}{2}V(\rho\|\sigma) + \frac{t}{1-t}\lambda n^{a-1}. \tag{H10b}$$

These optimizations can now be explicitly evaluated. To leading order, they give

$$\frac{1}{n}\gamma_{\pm\lambda n^a}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big) \approx -D(\rho\|\sigma) \mp \sqrt{2V(\rho\|\sigma)\lambda n^{a-1}}, \tag{H11}$$

which is Lemma 6.

Next, we turn to the dichotomy-transformation rates. We start with the high-error large-deviation result given in Theorem 5,

$$R_n^*\big(1-\exp(-\lambda n)\big) \approx \inf_{0<t_2<1}\inf_{t_1>1} \frac{\overline{D}_{t_1}(\rho_1\|\sigma_1) + \left(\frac{t_1}{t_1-1} + \frac{t_2}{1-t_2}\right)\lambda}{D_{t_2}(\rho_2\|\sigma_2)} \tag{H12}$$

for $\lambda > 0$. Substituting moderate errors, this becomes

$$R_n^*\big(1-\exp(-\lambda n^a)\big) \approx \inf_{0<t_2<1}\inf_{t_1>1} \frac{\overline{D}_{t_1}(\rho_1\|\sigma_1) + \left(\frac{t_1}{t_1-1} + \frac{t_2}{1-t_2}\right)\lambda n^{a-1}}{D_{t_2}(\rho_2\|\sigma_2)}. \tag{H13}$$

As with hypothesis testing, the optimizations will both approach $t_1, t_2 \approx 1$, so we can expand the Rényi entropies around $t_1, t_2 = 1$:

$$R_n^*\big(1-\exp(-\lambda n^a)\big) \approx \inf_{t_2<1}\inf_{t_1>1} \frac{D(\rho_1\|\sigma_1) + \frac{t_1-1}{2}V(\rho_1\|\sigma_1) + \left(\frac{t_1}{t_1-1} + \frac{t_2}{1-t_2}\right)\lambda n^{a-1}}{D(\rho_2\|\sigma_2) + \frac{t_2-1}{2}V(\rho_2\|\sigma_2)} \tag{H14}$$

$$\approx \inf_{t_2<1}\inf_{t_1>1} \frac{D(\rho_1\|\sigma_1) + \left[\frac{t_1-1}{2}V(\rho_1\|\sigma_1) + \frac{\lambda n^{a-1}}{t_1-1}\right] + \left[-\frac{t_2-1}{2}\frac{D(\rho_1\|\sigma_1)}{D(\rho_2\|\sigma_2)}V(\rho_2\|\sigma_2) + \frac{\lambda n^{a-1}}{1-t_2}\right]}{D(\rho_2\|\sigma_2)} \tag{H15}$$

$$\approx \frac{D(\rho_1\|\sigma_1) + \inf_{t_1>1}\left[\frac{t_1-1}{2}V(\rho_1\|\sigma_1) + \frac{\lambda n^{a-1}}{t_1-1}\right] + \inf_{t_2<1}\left[\frac{1-t_2}{2}\frac{D(\rho_1\|\sigma_1)}{D(\rho_2\|\sigma_2)}V(\rho_2\|\sigma_2) + \frac{\lambda n^{a-1}}{1-t_2}\right]}{D(\rho_2\|\sigma_2)} \tag{H16}$$

$$\approx \frac{D(\rho_1\|\sigma_1) + \sqrt{2V(\rho_1\|\sigma_1)\times\lambda n^{a-1}} + \sqrt{2V(\rho_2\|\sigma_2)\frac{D(\rho_1\|\sigma_1)}{D(\rho_2\|\sigma_2)}\times\lambda n^{a-1}}}{D(\rho_2\|\sigma_2)} \tag{H17}$$

$$\approx \frac{D(\rho_1\|\sigma_1) + [1+\xi^{-1/2}]\sqrt{2V(\rho_1\|\sigma_1)\times\lambda n^{a-1}}}{D(\rho_2\|\sigma_2)}, \tag{H18}$$

which is Theorem 3. For the low-error case, we can use the same arguments for $\overline{r}_2/\breve{r}_2$ and r_3. Specifically, for small negative μ, we have

$$\overline{r}_2(-\mu) \approx \breve{r}_2(-\mu) \approx \frac{D(\rho_1\|\sigma_1) - [1-\xi^{-1/2}]\sqrt{-2V(\rho_1\|\sigma_1)\times\mu}}{D(\rho_2\|\sigma_2)}, \tag{H19a}$$

$$r_3(\mu) \approx \frac{D(\rho_1\|\sigma_1) + [1-\xi^{-1/2}]\sqrt{2V(\rho_1\|\sigma_1)\times\mu}}{D(\rho_2\|\sigma_2)} \tag{H19b}$$

and thus

$$R_n^*\big(\exp(-\lambda n^a)\big) \approx \min_{-\lambda n^{a-1}\le\mu\le\lambda n^{a-1}} \begin{cases} r_2(\mu), & \mu<0,\\ r_3(\mu), & \mu>0,\end{cases} \tag{H20a}$$

$$\approx \frac{D(\rho_1\|\sigma_1) + \min\left\{\xi^{-1/2}-1, 1-\xi^{-1/2}\right\}\sqrt{2V(\rho_1\|\sigma_1)\times\lambda n^{a-1}}}{D(\rho_2\|\sigma_2)} \tag{H20b}$$

$$\approx \frac{D(\rho_1\|\sigma_1) - |1-\xi^{-1/2}|\sqrt{2V(\rho_1\|\sigma_1)\times\lambda n^{a-1}}}{D(\rho_2\|\sigma_2)}, \tag{H20c}$$

once again rederiving Theorem 3.

3. Large and extreme deviation

The other regime neighboring large deviations is extreme deviations. Here, instead of taking the limit of an arbitrarily small error exponent, we will instead take the limit of an arbitrarily large error exponent, as a crude model of superexponential error.

We start with hypothesis testing. For $\lambda > 0$, Lemma 5 gives that

$$\frac{1}{n}\gamma_{\lambda n}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big) \approx \sup_{t>1} -\breve{D}_t(\rho\|\sigma) + \frac{t}{1-t}\lambda. \tag{H21}$$

As $\breve{D}_t(\rho\|\sigma)$ is monotonically increasing in t and bounded, as we take $\lambda \to +\infty$, the optimizing t must also keep increasing. If we take $t\to\infty$, then this gives

$$\frac{1}{n}\gamma_{\lambda n}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big) \approx -\breve{D}_{+\infty}(\rho\|\sigma) - \lambda. \tag{H22}$$

Application of the same argument for $-\lambda$ gives

$$\frac{1}{n}\gamma_{-\lambda n}\big(\rho^{\otimes n}\big\|\sigma^{\otimes n}\big) \approx -\breve{D}_{-\infty}(\rho\|\sigma) + \lambda. \tag{H23}$$

Both of these are precisely the extreme-deviation results given in Lemma 8.

Next, we turn to the zero-error transformation rate of dichotomies. Recall that Theorem 4 gives that

$$R_n^*\big(\exp(-\lambda n)\big) \ge \min_{-\lambda\le\mu\le\lambda} \begin{cases} r_1(\mu), & \mu < -D(\sigma_1\|\rho_1),\\ \breve{r}_2(\mu), & -D(\sigma_1\|\rho_1) < \mu < 0,\\ r_3(\mu), & \mu > 0,\end{cases} \tag{H24}$$

where

$$r_1(\mu) := \sup_{t_2<0}\inf_{t_1<0} \frac{-\overleftarrow{D}_{t_1}(\rho_1\|\sigma_1) + \left(\frac{t_1}{t_1-1} - \frac{t_2}{t_2-1}\right)\mu}{-D_{t_2}(\rho_2\|\sigma_2)}, \tag{H25a}$$

$$\breve{r}_2(\mu) := \inf_{0<t_2<1}\sup_{0<t_1<1} \frac{\overleftarrow{D}_{t_1}(\rho_1\|\sigma_1) + \left(\frac{t_1}{1-t_1} - \frac{t_2}{1-t_2}\right)\mu}{D_{t_2}(\rho_2\|\sigma_2)}, \tag{H25b}$$

$$r_3(\mu) := \sup_{t_2>1}\inf_{t_1>1}\frac{\overleftarrow{D}_{t_1}(\rho_1\|\sigma_1)+\left(\frac{t_1}{t_1-1}-\frac{t_2}{t_2-1}\right)\mu}{D_{t_2}(\rho_2\|\sigma_2)}. \tag{H25c}$$

If we take the limit of $\lambda\to\infty$, then the optimization of μ becomes unconstrained and μ can be seen as a Lagrange multiplier in the above optimizations. Ignoring issues of order of limits, this means that optimizations of μ can be converted into constrained optimizations, with the constraint being that $t_1 = t_2$; specifically,

$$\inf_{\mu<-D(\sigma_1\|\rho_1)} r_1(\mu) = \inf_{t<0}\frac{\overleftarrow{D}_t(\rho_1\|\sigma_1)}{D_t(\rho_2\|\sigma_2)}, \tag{H26a}$$

$$\inf_{-D(\sigma_1\|\rho_1)<\mu<0} \breve{r}_2(\mu) = \inf_{0<t<1}\frac{\overleftarrow{D}_t(\rho_1\|\sigma_1)}{D_t(\rho_2\|\sigma_2)}, \tag{H26b}$$

$$\inf_{\mu>0} r_3(\mu) = \inf_{t>1}\frac{\overleftarrow{D}_t(\rho_1\|\sigma_1)}{D_t(\rho_2\|\sigma_2)}. \tag{H26c}$$

This means that

$$R_n^*(0) \gtrsim \inf_{t\in\mathbb{R}}\frac{\overleftarrow{D}(\rho_1\|\sigma_1)}{D(\rho_2\|\sigma_2)}. \tag{H27}$$

Following the discussion in Theorem 6 about pinching, this would extend to

$$R_n^*(0) \gtrsim \max\left\{\inf_{t\in\mathbb{R}}\frac{\overleftarrow{D}(\rho_1\|\sigma_1)}{D(\rho_2\|\sigma_2)}, \inf_{t\in\mathbb{R}}\frac{\overrightarrow{D}(\rho_1\|\sigma_1)}{D(\rho_2\|\sigma_2)}\right\}. \tag{H28}$$

4. Strong and weak resonance

In Sec. IV C 3, we have discussed a strong-resonance phenomenon that arises in the large- and extreme-deviation regimes, and complements the (weak) resonance discussed in Ref. [101]. We will now explain how the weak-resonance condition can be seen as an edge case of the strong condition. The strong-resonance condition is

$$\arg\min_{\alpha\in\overline{\mathbb{R}}}\frac{\breve{D}_\alpha(\rho_1\|\sigma_1)}{D_\alpha(\rho_2\|\sigma_2)} = 1 \tag{H29}$$

or, in other words,

$$\min_{\alpha\in\overline{\mathbb{R}}}\frac{\breve{D}_\alpha(\rho_1\|\sigma_1)}{D_\alpha(\rho_2\|\sigma_2)} = \frac{D(\rho_1\|\sigma_1)}{D(\rho_2\|\sigma_2)}. \tag{H30}$$

Weak resonance is a phenomenon that appears in the small- and moderate-deviation regimes. As we have shown before, these regimes can be seen as corresponding to values of α close to 1. So, if we consider only such α values and expand around $\alpha = 1$, then this condition becomes

$$\frac{D(\rho_1\|\sigma_1)}{D(\rho_2\|\sigma_2)} = \min_{\alpha\in\overline{\mathbb{R}}}\frac{\breve{D}_\alpha(\rho_1\|\sigma_1)}{D_\alpha(\rho_2\|\sigma_2)} \tag{H31a}$$

$$\approx \min_{\alpha}\frac{D(\rho_1\|\sigma_1)+\frac{\alpha-1}{2}V(\rho_1\|\sigma_1)+O((\alpha-1)^2)}{D(\rho_2\|\sigma_2)+\frac{\alpha-1}{2}V(\rho_2\|\sigma_2)+O((\alpha-1)^2)} \tag{H31b}$$

$$\approx \frac{D(\rho_1\|\sigma_1)}{D(\rho_2\|\sigma_2)}\left[1+\frac{\alpha-1}{2}\left(\frac{V(\rho_1\|\sigma_1)}{D(\rho_1\|\sigma_1)}-\frac{V(\rho_2\|\sigma_2)}{D(\rho_2\|\sigma_2)}\right)+O((\alpha-1)^2)\right], \tag{H31c}$$

which clearly then reduces to the weak-resonance condition

$$\frac{V(\rho_1\|\sigma_1)}{D(\rho_1\|\sigma_1)} = \frac{V(\rho_2\|\sigma_2)}{D(\rho_2\|\sigma_2)}. \tag{H32}$$

APPENDIX I: UNIFORM HYPOTHESIS-TESTING CONVERGENCE

An important feature of Lemma 2 is that it requires an ordering of the type-II errors simultaneously *for all* values of x. If one were to naively apply the hypothesis-testing results in Sec. V A, however, these would only provide pointwise convergence, instead of the uniform convergence that such a statement would require. In this appendix, we show that the hypothesis-testing results of Sec. V A can all be extended to uniform results as required essentially for free. This comes from the fact that the quantities being considered are monotonic, in such a way that rules out the pathologies necessary for nonuniform convergence. Specifically, we will use the following lemma.

Lemma 19 (Proposition 2.1 of [131]). Convergence of monotone functions on a compact set to a continuous function is uniform. In other words, if a sequence of functions $\{f_n\}_n$ from $[a,b]$ to $\mathbb{R}$ are all monotone and pointwise converge to a continuous function f, then that convergence is in fact uniform.

Lemma 20 (Uniform small-deviation analysis of hypothesis testing). For any $\delta > 0$, there exists a finite $N(\rho, \sigma, \delta)$ such that both inequalities

$$\left| -\log \beta_\epsilon\left(\rho^{\otimes n} \middle\| \sigma^{\otimes n}\right) - nD(\rho\|\sigma) - \sqrt{nV(\rho\|\sigma)}\Phi^{-1}(\epsilon)\right| \leq \delta\sqrt{n}, \tag{I1a}$$

$$\left| -\log \overleftarrow{\beta}_\epsilon\left(\rho^{\otimes n} \middle\| \sigma^{\otimes n}\right) - nD(\rho\|\sigma) - \sqrt{nV(\rho\|\sigma)}\Phi^{-1}(\epsilon)\right| \leq \delta\sqrt{n} \tag{I1b}$$

hold for all $n \geq N$ and $\epsilon \in [\delta, 1-\delta]$.

Proof. Start by defining

$$f_n(x) := \frac{-\log \beta_x\left(\rho^{\otimes n} \middle\| \sigma^{\otimes n}\right) - nD(\rho\|\sigma)}{\sqrt{n}} \quad \text{and}$$

$$\overleftarrow{f}_n(x) := \frac{-\log \overleftarrow{\beta}_x\left(\rho^{\otimes n} \middle\| \sigma^{\otimes n}\right) - nD(\rho\|\sigma)}{\sqrt{n}}, \tag{I2}$$

and $f(x) := \sqrt{V(\rho\|\sigma)}\Phi^{-1}(x)$. Lemma 4 is equivalent to the statement that $f_n \to f$ and $\overleftarrow{f}_n \to f$ pointwise on $(0,1)$. However, because $\beta_x(\cdot\|\cdot)$ and $\overleftarrow{\beta}_x(\cdot\|\cdot)$ are monotone decreasing functions of x, we have that each f_n and $\overleftarrow{f}|_n$ is monotone increasing. Thus, if we constrain x to some compact subset of $(0,1)$—say, $x \in [\delta, 1-\delta]$—then the uniformity of $f_n \to f$ and $\overleftarrow{f}_n \to f$ follows from Lemma 20. This in turn implies that there exists an ϵ-independent constant $N(\rho,\sigma,\delta)$ for which $|f_n(\epsilon) - f(\epsilon)| \leq \delta$ and $|\overleftarrow{f}_n(\epsilon) - f(\epsilon)| \leq \delta$ hold for all $\epsilon \in [\delta, 1-\delta]$ and $n \geq N$. Expanding this out gives the required inequalities. ■

Lemma 21 (Uniform large-deviation analysis of hypothesis testing). For any constant $\delta > 0$, there exists an $N(\rho,\sigma,\delta)$ such that the nonpinched and pinched log odds error per copy are bounded as

$$|\frac{1}{n}\gamma_{\lambda n}\left(\rho^{\otimes n} \middle\| \sigma^{\otimes n}\right) - \Gamma_\lambda(\rho\|\sigma)| \leq \delta, \tag{I3a}$$

$$|\frac{1}{n}\overleftarrow{\gamma}_{\lambda n}\left(\rho^{\otimes n} \middle\| \sigma^{\otimes n}\right) - \overleftarrow{\Gamma}_\lambda(\rho\|\sigma)| \leq \delta, \tag{I3b}$$

$$|\frac{1}{n}\overrightarrow{\gamma}_{\lambda n}\left(\rho^{\otimes n} \middle\| \sigma^{\otimes n}\right) - \overrightarrow{\Gamma}_\lambda(\rho\|\sigma)| \leq \delta, \tag{I3c}$$

for all $-1/\delta \leq \lambda \leq 1/\delta$ and $n \geq N$.

Proof. This proof follows similarly to Lemma 21. Here, we define

$$f_n(x) := \frac{\gamma_{nx}\left(\rho^{\otimes n} \middle\| \sigma^{\otimes n}\right)}{n}, \tag{I4a}$$

$$\overleftarrow{f}_n(x) := \frac{\overleftarrow{\gamma}_{nx}\left(\rho^{\otimes n} \middle\| \sigma^{\otimes n}\right)}{n}, \tag{I4b}$$

$$\overrightarrow{f}_n(x) := \frac{\overrightarrow{\gamma}_{nx}\left(\rho^{\otimes n} \middle\| \sigma^{\otimes n}\right)}{n}, \tag{I4c}$$

as well as $f(x) := \Gamma_x(\rho\|\sigma)$, $\overleftarrow{f}(x) := \overleftarrow{\Gamma}_x(\rho\|\sigma)$, and $\overrightarrow{f}(x) := \overrightarrow{\Gamma}_x(\rho\|\sigma)$. Lemma 5 gives that $f_n \to f$, $\overleftarrow{f}_n \to \overleftarrow{f}$, and $\overrightarrow{f}_n \to \overrightarrow{f}$ pointwise on $\mathbb{R}$, and Lemma 20 allows us to make this uniform on $[-1/\delta, 1/\delta]$. This uniform convergence in turn implies the existence of a finite $N(\rho,\sigma,\delta)$ such that $|f_n(x) - f(x)| \leq \delta$, $|\overleftarrow{f}|_n(x) - \overleftarrow{f}(x) \leq \delta$, and $|\overrightarrow{f}|_n(x) - \overrightarrow{f}(x) \leq \delta$ for any $n \geq N$ and $x \in [-1/\delta, 1/\delta]$. Expanding this gives the required inequalities. ■

Lemma 22 (Uniform moderate-deviation analysis of hypothesis testing). For any constant $\delta > 0$ and $a \in (0,1)$, there exists an $N(\rho,\sigma,\delta,a)$ such that the nonpinched and pinched log odds error per copy are bounded as

$$\left|\frac{1}{n}\gamma_{\lambda n^a}\left(\rho^{\otimes n} \middle\| \sigma^{\otimes n}\right) + D(\rho\|\sigma) + \operatorname{sgn}(\lambda) \times \sqrt{2V(\rho\|\sigma)\lambda n^{a-1}}\right| \leq \delta\sqrt{n^{a-1}}, \tag{I5a}$$

$$\left|\frac{1}{n}\overleftarrow{\gamma}_{\lambda n^a}\left(\rho^{\otimes n} \middle\| \sigma^{\otimes n}\right) + D(\rho\|\sigma) + \operatorname{sgn}(\lambda) \times \sqrt{2V(\rho\|\sigma)\lambda n^{a-1}}\right| \leq \delta\sqrt{n^{a-1}}, \tag{I5b}$$

for all $-1/\delta \leq \lambda \leq 1/\delta$ and $n \geq N$.

Proof. For this, we define

$$f_n(x) := \frac{\frac{1}{n}\gamma_{xn^a}\left(\rho^{\otimes n} \middle\| \sigma^{\otimes n}\right) + D(\rho\|\sigma)}{\sqrt{n^{a-1}}} \quad \text{and}$$

$$\overleftarrow{f}_n(x) := \frac{\frac{1}{n}\overleftarrow{\gamma}_{xn^a}\left(\rho^{\otimes n} \middle\| \sigma^{\otimes n}\right) + D(\rho\|\sigma)}{\sqrt{n^{a-1}}}, \tag{I6}$$

and $f(x) := -\operatorname{sgn}(x) \times \sqrt{2V(\rho\|\sigma)|x|}$. Lemma 6 is equivalent to the statement that $f_n, \overleftarrow{f}_n \to f$ pointwise on $\mathbb{R}$. As γ_x and $\overleftarrow{\gamma}_x$ are monotone-increasing functions, we can apply Lemma 20 to upgrade this convergence to uniform, which gives that there exists a $N(\rho,\sigma,\delta,a)$ such that $|f_n(x) - f(x)| \leq \delta$ and $|\overleftarrow{f}|_n(x) - f(x) \leq \delta$ for any $n \geq N$ and $x \in [-1/\delta, 1/\delta]$. Expanding these out gives the desired bounds. ■

[1] D. G. Altman, *Practical Statistics for Medical Research* (CRC Press, New York, 1990).
[2] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice-Hall, Inc., USA, 1993).
[3] E. Walter, L. Pronzato, and J. Norton, *Identification of Parametric Models from Experimental Data* (Springer, New York, 1997), Vol. 1.
[4] G. Cumming, *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis* (Routledge, New York, 2013).
[5] K. Weise and W. Woger, A Bayesian theory of measurement uncertainty, Meas. Sci. Technol. **4**, 1 (1993).
[6] J. Taylor, *Introduction to Error Analysis, the Study of Uncertainties in Physical Measurements* (University Science Books, Sausalito, California, 1997).
[7] S. G. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice* (Springer Science & Business Media, New York, 2006).
[8] J. M. Bernardo and A. F. Smith, *Bayesian Theory* (John Wiley & Sons, New York, 2009), Vol. 405.
[9] R. Fisher, Statistical methods and scientific induction, J. R. Stat. Soc. B Stat. Methodol. **17**, 69 (1955).
[10] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses* (Springer, Cham, 1986), Vol. 3.
[11] J. Neyman and E. S. Pearson, IX. On the problem of the most efficient tests of statistical hypotheses, Philos. Trans. R. Soc. A **231**, 289 (1933).
[12] A. Wald, Contributions to the theory of statistical estimation and testing hypotheses, Ann. Math. Stat. **10**, 299 (1939).
[13] J. Von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton University Press, Oxford, 1947).
[14] P. J. Schoemaker, The expected utility model: Its variants, purposes, evidence and limitations, J. Econ. Lit. **20**, 529 (1982).
[15] E. J. Johnson and J. W. Payne, Effort and accuracy in choice, Manage. Sci. **31**, 395 (1985).
[16] R. B. Myerson, *Game Theory: Analysis of Conflict* (Harvard University Press, Cambridge, 1997).
[17] T. M. Mitchell, *Machine Learning* (McGraw-Hill, New York, 1997).
[18] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning* (Springer, New York, 2006).
[19] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning* (Springer, Cham, 2013), Vol. 112.
[20] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, SCIKIT-LEARN: Machine learning in PYTHON, J. Mach. Learn. Res. **12**, 2825 (2011).
[21] D. Blackwell, Equivalent comparisons of experiments, Ann. Math. Stat. **24**, 265 (1953).
[22] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities* (Cambridge University Press, New York, 1952).
[23] J. Cohen, J. H. Kempermann, and G. Zbaganu, *Comparisons of Stochastic Matrices with Applications in Information Theory, Statistics, Economics and Population* (Springer Science & Business Media, Boston, 1998).
[24] C. E. Shannon, A note on a partial ordering for communication channels, Inf. Control **1**, 390 (1958).
[25] E. Jorswieck and H. Boche, Majorization and matrix-monotone functions in wireless communications, Found. Trends Commun. Inf. Theory **3**, 553 (2007).
[26] C. W. Helstrom, Quantum detection and estimation theory, J. Stat. Phys. **1**, 231 (1969).
[27] P. Alberti and A. Uhlmann, A problem relating to positive linear maps on matrix algebras, Rep. Math. Phys. **18**, 163 (1980).
[28] A. Holevo, Testing statistical hypotheses in quantum theory, Probab. Math. Stat. **3**, 113 (1982).
[29] K. Matsumoto, Reverse test and characterization of quantum relative entropy, arXiv:1010:1030.
[30] F. Buscemi, Comparison of quantum statistical models: Equivalent conditions for sufficiency, Commun. Math. Phys. **310**, 625 (2012).
[31] Y. Aharonov and J. Anandan, Phase change during a cyclic quantum evolution, Phys. Rev. Lett. **58**, 1593 (1987).
[32] M. G. Paris, Quantum estimation for quantum technology, Int. J. Quant. Inf. **7**, 125 (2009).
[33] P. Hyllus, W. Laskowski, R. Krischek, C. Schwemmer, W. Wieczorek, H. Weinfurter, L. Pezzé, and A. Smerzi, Fisher information and multiparticle entanglement, Phys. Rev. A **85**, 022321 (2012).
[34] V. Giovannetti, S. Lloyd, and L. Maccone, Advances in quantum metrology, Nat. Photon. **5**, 222 (2011).
[35] J. M. Boss, K. Cujia, J. Zopes, and C. L. Degen, Quantum sensing with arbitrary frequency resolution, Science **356**, 837 (2017).
[36] G. Tóth and I. Apellaniz, Quantum metrology from a quantum information science perspective, J. Phys. A **47**, 424006 (2014).
[37] L. Pezzè, A. Smerzi, M. K. Oberthaler, R. Schmied, and P. Treutlein, Quantum metrology with nonclassical states of atomic ensembles, Rev. Mod. Phys. **90**, 035005 (2018).
[38] S. Pirandola, B. R. Bardhan, T. Gehring, C. Weedbrook, and S. Lloyd, Advances in photonic quantum sensing, Nat. Photon. **12**, 724 (2018).
[39] R. Demkowicz-Dobrzański, J. Kołodyński, and M. Guţă, The elusive Heisenberg limit in quantum-enhanced metrology, Nat. Commun. **3**, 1063 (2012).
[40] F. Reif, *Fundamentals of Statistical and Thermal Physics* (Waveland Press, Long Grove, 2009). https://books.google.ch/books?id=ObsbAAAAQBAJ.
[41] T. L. Hill, *An Introduction to Statistical Thermodynamics* (Courier Corporation, New York, 1986).
[42] J. A. Hertz, in *Basic Notions of Condensed Matter Physics* (CRC Press, New York, 2018), p. 525.
[43] R. P. Feynman, in *Feynman and computation* (CRC Press, Boca Raton, 2018), p. 133.
[44] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, Quantum algorithms revisited, Proc. R. Soc. A: Math. **454**, 339 (1998).
[45] S. Aaronson, *Quantum Computing Since Democritus* (Cambridge University Press, Cambridge, 2013).
[46] E. Shmaya, Comparison of information structures and completely positive maps, J. Phys. A Math. Theor. **38**, 9717 (2005).

[47] A. Chefles, The quantum Blackwell theorem and minimum error state discrimination, arXiv:0907.0866.
[48] R. Kubo, The fluctuation-dissipation theorem, Rep. Prog. Phys. **29**, 255 (1966).
[49] H. B. Callen and T. A. Welton, Irreversibility and generalized noise, Phys. Rev. **83**, 34 (1951).
[50] H. Stanley, *Introduction to Phase Transitions and Critical Phenomena*, International Series of Monographs on Physics (Oxford University Press, Oxford, 1987). https://books.google.ch/books?id=C3BzcUxoaNkC.
[51] K. Binder, Theory of first-order phase transitions, Rep. Prog. Phys. **50**, 783 (1987).
[52] I. Prigogine and R. Defay, *Chemical Thermodynamics* (Wiley, Dordrecht, 1962). https://books.google.ch/books?id=8wsJAQAAIAAJ.
[53] J. W. Gibbs, *Elementary Principles in Statistical Mechanics* (C. Scribner's Sons, New York, 1902).
[54] M. Ledoux, *The Concentration of Measure Phenomenon* (American Mathematical Society, Providence, 2001). https://bookstore.ams.org/view?ProductCode=SURV/89.
[55] H. Touchette, Equivalence and nonequivalence of ensembles: Thermodynamic, macrostate, and measure levels, J. Stat. Phys. **159**, 987 (2015).
[56] C. T. Chubb, M. Tomamichel, and K. Korzekwa, Beyond the thermodynamic limit: Finite-size corrections to state interconversion rates, Quantum **2**, 108 (2018).
[57] D. Janzing, P. Wocjan, R. Zeier, R. Geiss, and T. Beth, Thermodynamic cost of reliability and low temperatures: Tightening Landauer's principle and the second law, Int. J. Theor. Phys. **39**, 2717 (2000).
[58] M. Horodeckiand J. Oppenheim, Fundamental limitations for quantum and nanoscale thermodynamics, Nat. Commun. **4**, 2059 (2013).
[59] F. G. Brandao, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, Resource theory of quantum states out of thermal equilibrium, Phys. Rev. Lett. **111**, 250404 (2013).
[60] F. Brandão, M. Horodecki, N. Ng, J. Oppenheim, and S. Wehner, The second laws of quantum thermodynamics, Proc. Natl. Acad. Sci. USA **112**, 3275 (2015).
[61] A. M. Alhambra, L. Masanes, J. Oppenheim, and C. Perry, Fluctuating work: From quantum thermodynamical identities to a second law equality, Phys. Rev. X **6**, 041017 (2016).
[62] M. Lostaglio, An introductory review of the resource theory approach to thermodynamics, Rep. Prog. Phys. **82**, 114001 (2019).
[63] S. Vinjanampathy and J. Anders, Quantum thermodynamics, Contemp. Phys. **57**, 545 (2016).
[64] N. Y. Halpern, A. J. Garner, O. C. Dahlsten, and V. Vedral, Introducing one-shot work into fluctuation relations, New J. Phys. **17**, 095003 (2015).
[65] M. P. Müller, Correlating thermal machines and the second law at the nanoscale, Phys. Rev. X **8**, 041051 (2018).
[66] P. Lipka-Bartosik and P. Skrzypczyk, All states are universal catalysts in quantum thermodynamics, Phys. Rev. X **11**, 011061 (2021).
[67] M. Lostaglio, K. Korzekwa, D. Jennings, and T. Rudolph, Quantum coherence, time-translation symmetry, and thermodynamics, Phys. Rev. X **5**, 021001 (2015).
[68] N. Y. Halpern, P. Faist, J. Oppenheim, and A. Winter, Microcanonical and resource-theoretic derivations of the thermal state of a quantum system with noncommuting charges, Nat. Commun. **7**, 12051 (2016).
[69] Note that the problem of comparing quantum dichotomies is sometimes formulated as a resource theory of asymmetric distinguishability [91]. Since the two formulations are essentially equivalent, all of our results can be interpreted in this resource-theoretic way.
[70] J. M. Renes, Relative submajorization and its use in quantum resource theories, J. Math. Phys. **57**, 122202 (2016).
[71] K. Matsumoto, An example of a quantum statistical model which cannot be mapped to a less informative one by any trace preserving positive map, arXiv:1409.5658.
[72] A. Jencová, Comparison of quantum binary experiments, Rep. Math. Phys. **70**, 237 (2012).
[73] D. Reeb, M. J. Kastoryano, and M. M. Wolf, Hilbert's projective metric in quantum information theory, J. Math. Phys. **52**, 082201 (2011).
[74] A. Jencová, in *2016 IEEE International Symposium on Information Theory (ISIT)* (IEEE, Barcelona, 2016), p. 2249.
[75] G. Gour, D. Jennings, F. Buscemi, R. Duan, and I. Marvian, Quantum majorization and a complete set of entropic conditions for quantum thermodynamics, Nat. Commun. **9**, 1 (2018).
[76] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, Cambridge, 2010).
[77] M. A. Nielsen, Conditions for a class of entanglement transformations, Phys. Rev. Lett. **83**, 436 (1999).
[78] H.-K. Lo and S. Popescu, Concentrating entanglement by local actions: Beyond mean values, Phys. Rev. A **63**, 022301 (2001).
[79] H. Umegaki, Conditional expectation in an operator algebra, Kodai Math. Sem. Rep. **14**, 59 (1962).
[80] M. Tomamichel and M. Hayashi, A hierarchy of information quantities for finite block length analysis of quantum tasks, IEEE Trans. Inf. Theory **59**, 7693 (2013).
[81] K. Li, Second-order asymptotics for quantum hypothesis testing, Ann. Stat. **42**, 171 (2014).
[82] T. Biswas, A. de Oliveira Junior, M. Horodecki, and K. Korzekwa, Fluctuation-dissipation relations for thermodynamic distillation processes, Phys. Rev. E **105**, 054127 (2022).
[83] A. Rényi, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics* (University of California Press, Berkley, 1961), Vol. 4, p. 547. https://projecteuclid.org/ebook/Download?urlid=bsmsp/1200512181isFullBook=false.
[84] D. Petz, Quasi-entropies for finite quantum systems, Rep. Math. Phys. **23**, 57 (1986).
[85] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, On quantum Rényi entropies: A new generalization and some properties, J. Math. Phys. **54**, 122203 (2013).
[86] K. M. R. Audenaert and N. Datta, α-z-Rényi relative entropies, J. Math. Phys. **56**, 022202 (2015).

[87] M. Tomamichel, *Quantum Information Processing with Finite Resources—Mathematical Foundations* (Springer International Publishing, Cham, 2016).
[88] M. M. Wilde, A. Winter, and D. Yang, Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy, Commun. Math. Phys. **331**, 593 (2014).
[89] W. Kumagai and M. Hayashi, Second-order asymptotics of conversions of distributions and entangled states based on Rayleigh-normal probability distributions, IEEE Trans. Inf. Theory **63**, 1829 (2016).
[90] C. T. Chubb, M. Tomamichel, and K. Korzekwa, Moderate deviation analysis of majorization-based resource interconversion, Phys. Rev. A **99**, 032332 (2019).
[91] X. Wang and M. M. Wilde, Resource theory of asymmetric distinguishability, Phys. Rev. Res. **1**, 033170 (2019).
[92] G. Gour, Role of quantum coherence in thermodynamics, PRX Quantum **3**, 040323 (2022).
[93] F. Buscemi, D. Sutter, and M. Tomamichel, An information-theoretic treatment of quantum dichotomies, Quantum **3**, 209 (2019).
[94] It should be noted that there is some dispute on the status of proofs of Theorem 1. Reference [59] claims to prove it, for noncommuting inputs *and* outputs. However, Ref. [92] has raised questions of possible gaps in this proof, especially surrounding the case of noncommuting outputs.
[95] K. Li and Y. Yao, Operational interpretation of the sandwiched Rényi divergences of order 1/2 to 1 as strong converse exponents, arXiv:2209.00554.
[96] X. Mu, L. Pomatto, P. Strack, and O. Tamuz, From Blackwell dominance in large samples to Rényi divergences and back again, Econometrica **89**, 475 (2021).
[97] M. U. Farooq, T. Fritz, E. Haapasalo, and M. Tomamichel, Asymptotic and catalytic matrix majorization, arXiv:2301.07353.
[98] P. Lipka-Bartosik, P. Mazurek, and M. Horodecki, Second law of thermodynamics for batteries with vacuum state, Quantum **5**, 408 (2021).
[99] M. Łobejko, Work and fluctuations: Coherent vs. incoherent ergotropy extraction, Quantum **6**, 762 (2022).
[100] K. Korzekwa, Z. Puchała, M. Tomamichel, and K. Życzkowski, Encoding classical information into quantum resources, IEEE Trans. Inf. Theory **68**, 4518 (2022).
[101] K. Korzekwa, C. T. Chubb, and M. Tomamichel, Avoiding irreversibility: Engineering resonant conversions of quantum resources, Phys. Rev. Lett. **122**, 110403 (2019).
[102] A. K. Jensen, Asymptotic majorization of finite probability distributions, IEEE Trans. Inf. Theory **65**, 8131 (2019).
[103] S. Du, Z. Bai, and Y. Guo, Conditions for coherence transformations under incoherent operations, Phys. Rev. A **91**, 052120 (2015).
[104] T. Baumgratz, M. Cramer, and M. Plenio, Quantifying coherence, Phys. Rev. Lett. **113**, 140401 (2014).
[105] F. Hiai and D. Petz, The proper formula for relative entropy and its asymptotics in quantum probability, Commun. Math. Phys. **143**, 99 (1991).
[106] T. Ogawa and H. Nagaoka, *Strong Converse and Stein's Lemma in Quantum Hypothesis Testing* (World Scientific, Hackensack, 2005), p. 28.
[107] M. Tomamichel and V. Y. F. Tan, Second-order asymptotics for the classical capacity of image-additive quantum channels, Commun. Math. Phys. **338**, 103 (2015).
[108] M. Tomamichel and M. Hayashi, A hierarchy of information quantities for finite block length analysis of quantum tasks, IEEE Trans. Inf. Theory **59**, 7693 (2013).
[109] M. Hayashi, Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding, Phys. Rev. A **76**, 62301 (2006).
[110] H. Nagaoka, The converse part of the theorem for quantum Hoeffding bound, arXiv:quant-ph/0611289.
[111] M. Mosonyi and T. Ogawa, Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies, Commun. Math. Phys. **334**, 1617 (2015).
[112] M. Mosonyi and T. Ogawa, Two approaches to obtain the strong converse exponent of quantum hypothesis testing for general sequences of quantum states, IEEE Trans. Inf. Theory **61**, 6975 (2015).
[113] It should be noted that the results in Refs. [109–112] are given with the roles of type-I and -II errors reversed, but by swapping the states $\rho \leftrightarrow \sigma$ and inverting these bounds they can be shown to be equivalent to the expressions for Γ_λ given above.
[114] Strictly speaking, these functions are not necessarily *inverses* but only *quasiinverses*, which are functions that act as inverses on each other's domain or codomain, i.e., functions f and g such that $f \circ g \circ f = f$ and $g \circ f \circ g = g$.
[115] F. Hiai, M. Mosonyi, and T. Ogawa, Error exponents in hypothesis testing for correlated states on a spin chain, J. Math. Phys. **49**, 032112 (2008).
[116] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Stochastic Modelling and Applied Probability (Springer, Heidelberg, 1998), 2nd ed.
[117] C. T. Chubb, V. Y. Tan, and M. Tomamichel, Moderate deviation analysis for classical communication over quantum channels, Commun. Math. Phys. **355**, 1283 (2017).
[118] H.-C. Cheng and M.-H. Hsieh, Moderate deviation analysis for classical-quantum channels and quantum hypothesis testing, IEEE Trans. Inf. Theory **64**, 1385 (2018).
[119] A. S. Holevo, Analog of a theory of statistical decisions in a noncommutative theory of probability, Tr. Mosk. Mat. Obs. **26**, 133 (1972).
[120] G. Guarnieri, N. Ng, K. Modi, J. Eisert, M. Paternostro, and J. Goold, Quantum work statistics and resource theories: Bridging the gap through Rényi divergences, Phys. Rev. E **99**, 050101 (2019).
[121] M. Scandi, H. J. Miller, J. Anders, and M. Perarnau-Llobet, Quantum work statistics close to equilibrium, Phys. Rev. Res. **2**, 023377 (2020).
[122] M. Jarzyna and J. Kołodyński, Geometric approach to quantum statistical inference, IEEE J. Sel. Area Inf. Theory **1**, 367 (2020).
[123] A. Misra, U. Singh, S. Bhattacharya, and A. K. Pati, Energy cost of creating quantum coherence, Phys. Rev. A **93**, 052335 (2016).

[124] P. Faist, J. Oppenheim, and R. Renner, Gibbs-preserving maps outperform thermal operations in the quantum regime, New J. Phys. **17**, 043003 (2015).
[125] I. Marvian, Ph.D. thesis, University of Waterloo, 2012.
[126] G. Gour and R. W. Spekkens, The resource theory of quantum reference frames: Manipulations and monotones, New J. Phys. **10**, 033023 (2008).
[127] K. Szymański, Ph.D. thesis, Jagiellonian University, 2022.
[128] J. Lebel, *Basic Analysis II: Introduction to Real Analysis* (CreateSpace Independent Publishing, Palo Alto, 2018), Vol. 2.
[129] M. Fekete, Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten, Math. Z. **17**, 228 (1923).
[130] May the Lord forgive us....
[131] S. Resnick, *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling* (Springer, New York, 2007).