Postselected quantum hypothesis testing

SCHOLARONE™
Manuscripts

Postselected quantum hypothesis testing

Bartosz Regula,[1, *] Ludovico Lami,[2, †] and Mark M. Wilde[3, ‡]

[1]*Department of Physics, Graduate School of Science,*
The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan
[2]*Institut für Theoretische Physik und IQST, Universität Ulm,*
Albert-Einstein-Allee 11, D-89069 Ulm, Germany
[3]*School of Electrical and Computer Engineering,*
Cornell University, Ithaca, New York 14850, USA

We study a variant of quantum hypothesis testing wherein an additional 'inconclusive' measurement outcome is added, allowing one to abstain from attempting to discriminate the hypotheses. The error probabilities are then conditioned on a successful attempt, with inconclusive trials disregarded. We completely characterise this task in both the single-shot and asymptotic regimes, providing exact formulas for the optimal error probabilities. In particular, we prove that the asymptotic error exponent of discriminating any two quantum states ρ and σ is given by the Hilbert projective metric $D_{\max}(\rho\|\sigma) + D_{\max}(\sigma\|\rho)$ in asymmetric hypothesis testing, and by the Thompson metric $\max\{D_{\max}(\rho\|\sigma), D_{\max}(\sigma\|\rho)\}$ in symmetric hypothesis testing. This endows these two quantities with fundamental operational interpretations in quantum state discrimination. Our findings extend to composite hypothesis testing, where we show that the asymmetric error exponent with respect to any convex set of density matrices is given by a regularisation of the Hilbert projective metric. We apply our results also to quantum channels, showing that no advantage is gained by employing adaptive or even more general discrimination schemes over parallel ones, in both the asymmetric and symmetric settings. Our state discrimination results make use of no properties specific to quantum mechanics and are also valid in general probabilistic theories.

I. INTRODUCTION

A. Background

Quantum hypothesis testing — a generalisation of the crucial statistical primitive of hypothesis testing to the non-commutative setting — is a fundamental task that directly underlies many protocols in quantum information science [1–3]. It is concerned with the problem of distinguishing different quantum states or channels, and in particular asks about the ultimate limits of such distinguishability.

In quantum state discrimination, after receiving one of two possible states, ρ (null hypothesis) or σ (alternative hypothesis), one aims to determine which of the two was actually obtained. This is typically done by performing a quantum measurement, or positive operator-valued measure (POVM), given by the tuple $M := \{M_1, M_2\}$ where $M_2 = \mathbb{1} - M_1$. If measurement outcome 1 is obtained (corresponding to the operator M_1), one guesses that the state is ρ; if outcome 2 is obtained, one guesses σ. There are then two types of errors that can occur: the *type I error*

$$\alpha(M) := \operatorname{Tr} M_2 \rho, \tag{1}$$

which corresponds to the probability of incorrectly guessing σ when ρ was true (false positive), and the *type II error*

$$\beta(M) := \operatorname{Tr} M_1 \sigma, \tag{2}$$

* bartosz.regula@gmail.com
† ludovico.lami@gmail.com
‡ wilde@cornell.edu

which is the probability of guessing ρ when σ was true (false negative). Two settings naturally emerge here: one is *asymmetric hypothesis testing*, which asks about how small one of the errors can be subject to constraints on the other error; another setting is *symmetric hypothesis testing*, where the two errors are treated equally and one aims to minimise their average. Specifically, the asymmetric case is concerned with the study of the optimised type II error

$$
\begin{aligned}
\beta_\varepsilon(\rho, \sigma) &:= \min_{M \in \mathcal{M}_2} \left\{ \beta(M) \mid \alpha(M) \le \varepsilon \right\} \\
&= \min_{M_1, M_2 \ge 0} \left\{ \operatorname{Tr} M_1 \sigma \mid M_1 + M_2 = \mathbb{1}, \ \operatorname{Tr} M_2 \rho \le \varepsilon \right\},
\end{aligned}
\tag{3}
$$

where $\mathcal{M}_2$ denotes the set of all two-outcome measurements. The symmetric error depends also on the prior probabilities associated with the two hypotheses, that is, the probability that the state to be discriminated is ρ or that it is σ. Denoting the respective priors by p and $q := 1 - p$, we define the symmetric error as

$$
\begin{aligned}
p_{\mathrm{err}}(\rho, \sigma \mid p, q) &:= \min_{M \in \mathcal{M}_2} \left[p\alpha(M) + q\beta(M) \right] \\
&= \min_{M_1, M_2 \ge 0} \left\{ p \operatorname{Tr} M_2 \rho + q \operatorname{Tr} M_1 \sigma \mid M_1 + M_2 = \mathbb{1} \right\}.
\end{aligned}
\tag{4}
$$

The two optimised errors defined above are not too difficult to characterise: they are both efficiently computable semidefinite programs, and the symmetric error even admits an exact solution given by the celebrated Helstrom–Holevo theorem [4, 5], which showed that

$$
p_{\mathrm{err}}(\rho, \sigma \mid p, q) = \frac{1}{2} \left(1 - \| p\rho - q\sigma \|_1 \right),
\tag{5}
$$

where $\|\cdot\|_1$ denotes the trace norm (Schatten 1-norm).

However, a limitation of the above formulations is that they assume access to only a single copy of the state to be discriminated. This is not reflective of situations encountered in practice, where one typically deals with sources that emit many copies of quantum states, which we can use to our advantage in attempting to discriminate them. That is, we can assume that we receive either $\rho^{\otimes n}$ or $\sigma^{\otimes n}$ and then perform a joint measurement over all n systems. The crucial question then is: how exactly does the performance in discriminating two states improve when we have access to more copies of them? The ultimate limit of such protocols is described by the asymptotic setting with $n \to \infty$, and the figures of merit then are the decay rates of $\beta_\varepsilon(\rho^{\otimes n}, \sigma^{\otimes n})$ and $p_{\mathrm{err}}(\rho^{\otimes n}, \sigma^{\otimes n} \mid p, q)$, known as the error exponents. These two limits give rise to two of the most fundamental quantities in quantum information. The first one is the *quantum relative entropy* $D(\rho\|\sigma) := \operatorname{Tr} \rho(\log \rho - \log \sigma)$ [6], which is known to be exactly equal to the error exponent in asymmetric hypothesis testing:

$$
\lim_{n \to \infty} -\frac{1}{n} \log \beta_\varepsilon(\rho^{\otimes n}, \sigma^{\otimes n}) = D(\rho\|\sigma) \qquad \forall \varepsilon \in (0, 1).
\tag{6}
$$

This was shown in the works of Hiai and Petz [7] (achievability and weak converse[1]) and Ogawa and Nagaoka [8] (strong converse). The above result, known as the quantum Stein's lemma, generalises a seminal result due to Stein and Chernoff [9, 10]. It provides an important operational interpretation of the quantum relative entropy, and in a large part motivates its extensive use in

---

[1] In the setting of asymptotic asymmetric hypothesis testing, a 'weak converse' for the error exponent is understood as one that holds in the limit $\varepsilon \to 0$, while a 'strong converse' holds for every $\varepsilon \in (0, 1)$. The motivation for considering the latter is that, for every type II error exponent larger than a strong converse bound, the type I error must be prohibitively large — it must asymptotically converge to 1, making discrimination effectively impossible.

quantum information as a quantifier of the distinguishability between two quantum states and a meaningful generalisation of the Kullback–Leibler divergence [11].

An analogous result in symmetric hypothesis testing, generalising a finding of Chernoff [12], was obtained in the works of Nussbaum and Szkoła [13] (converse) and Audenaert *et al.* [14] (achievability), yielding

$$\lim_{n\to\infty} -\frac{1}{n} \log p_{\text{err}}(\rho^{\otimes n}, \sigma^{\otimes n} \mid p, q) = \xi(\rho\|\sigma) := -\log \inf_{0\leq s\leq 1} \operatorname{Tr} \rho^{1-s}\sigma^s \qquad \forall p, q = 1 - p \in (0, 1). \quad (7)$$

This is commonly known as the quantum Chernoff bound, with the Chernoff divergence $\xi(\rho\|\sigma)$ constituting another important measure of distinguishability between quantum states.

Due to the intricate mathematical methods needed to establish the quantum Stein's lemma and the quantum Chernoff bound, the results have found few generalisations. There are, for instance, important cases where the alternative hypothesis is not given simply by i.i.d. quantum states $(\sigma^{\otimes n})_n$, but rather by a sequence $(\sigma_n)_n$ of states taken from some subset of density matrices — this is known as composite hypothesis testing. A generalised quantum Stein's lemma in this setting was proposed [15], but so far no valid proof of such a relation has been found [16] beyond some special cases [17, 18] or related but incomparable composite settings [16, 18–20].

Perhaps even more challenging is the case of quantum channel discrimination. This task is inherently much more complex than state discrimination: any manipulation of n copies of a quantum state ρ is fully equivalent to having access to the tensor product $\rho^{\otimes n}$, while being able to invoke n uses of a channel $\mathcal{M}$ is not the same as manipulating n parallel copies $\mathcal{M}^{\otimes n}$, since schemes for channel discrimination much more general than naïve parallel protocols can be employed. A common example are the so-called adaptive protocols, which work by processing the channel n times, such that information obtained in previous rounds can be used to enhance the successive rounds of processing [21]. It is known that adaptive discrimination of channels can yield strict improvements over parallel protocols, although a very recent line of work [22–24] established that the advantage disappears asymptotically in asymmetric hypothesis testing, where the error exponent is given by the regularised channel relative entropy

$$\lim_{n\to\infty} -\frac{1}{n} \log \beta_\varepsilon(\mathcal{M}^{\otimes n}, \mathcal{N}^{\otimes n}) = D^\infty(\mathcal{M}\|\mathcal{N}) := \lim_{n\to\infty} \frac{1}{n} \sup_\rho D(\operatorname{id} \otimes \mathcal{M}^{\otimes n}(\rho)\|\operatorname{id} \otimes \mathcal{N}^{\otimes n}(\rho))$$
$$= \sup_{\rho,\omega} \left[D(\operatorname{id} \otimes \mathcal{M}(\rho)\|\operatorname{id} \otimes \mathcal{N}(\omega)) - D(\rho\|\omega) \right]. \quad (8)$$

In symmetric hypothesis testing, it is known that adaptive strategies *can* improve over parallel ones both non-asymptotically and asymptotically [25, 26], although in general no exact form of the asymptotic error exponent — and thus no extension of the Chernoff bound — is known in the setting of quantum channels [27]. The landscape of channel discrimination is further complicated by the fact that there exist even more general ways of manipulating quantum channels than adaptive protocols, for instance ones without a definite causal order structure [28, 29], and they can yield strict improvements in some settings [28, 30–32]; the asymptotic performance of such strategies, however, has not been characterised yet.

B. Setting

In this work, we propose a variant of quantum hypothesis testing that adds an additional, 'inconclusive' measurement outcome, allowing us to abstain from discriminating the states if we are not sufficiently confident in the result. This can be motivated, for instance, by an experimental situation in which a 'no detection' event occurs, that is, a detector responsible for distinguishing

4

two states fails to recognise either of them. More generally, a motivation to consider such a setting could be any situation where avoiding a wrong result is of utmost importance, even if this means having to repeat the trial many times, such as would be the case e.g. with medical diagnoses.

To be specific, let us consider the case of quantum state discrimination. Given either of the two states ρ or σ, we perform a *three*-outcome measurement $M = \{M_1, M_2, M_?\}$. The first two outcomes are as before — outcome 1 corresponds to the guess 'ρ' and outcome 2 corresponds to the guess 'σ' — while the third outcome '?' is designated as an inconclusive result, that is, a case in which we are unable to decisively distinguish the hypotheses. There is therefore some probability that our attempt to perform hypothesis testing fails: this probability is $\text{Tr}\, M_? \rho$ when measuring ρ, and $\text{Tr}\, M_? \sigma$ when measuring σ.

Such an approach dates back to the early works of Ivanovic [33], Dieks [34], and Peres [35], where it was shown that all linearly independent pure states can be discriminated with no error, but only through inconclusive protocols whose probability of success may be strictly less than one. Extensions of this idea were studied in a number of settings [36–44], but mostly restricted to non-asymptotic, symmetric state discrimination.

Here we establish a complete and rigorous formalisation of such inconclusive state discrimination schemes in quantum hypothesis testing. Our approach is based on characterising the corresponding errors conditioned on a conclusive result; that is, assuming that either of the outcomes '1' or '2' was obtained, how likely is it that we made an incorrect guess? This can be understood as postselecting on a conclusive outcome of the measurement, and ignoring all measurement results where the outcome '?' is obtained. Depending on whether discussing the asymmetric or symmetric setting, we are concerned with the error probability when measuring a particular state ρ or σ, or the average probability when measuring either of them.

To clarify how our setting differs from the conventional one of conclusive hypothesis testing, let us first consider the case of asymmetric error. Observe that the type I error, for instance, is defined as $\alpha(M) = P(\text{error} \mid \rho, M)$, i.e. the probability of guessing incorrectly given that the measured state is ρ. In our approach here, we further condition the incorrect guess event on the attainment of a conclusive outcome. We thus define the following counterparts of the conventional hypothesis testing errors: the *conditional type I error* and *conditional type II error* probabilities are given by

$$\overline{\alpha}(M) := P(\text{error} \mid \rho, M, \text{conclusive}) = \frac{\alpha(M)}{1 - \text{Tr}\, M_? \rho} = \frac{\text{Tr}\, M_2 \rho}{\text{Tr}\,[(M_1 + M_2)\rho]},$$

$$\overline{\beta}(M) := P(\text{error} \mid \sigma, M, \text{conclusive}) = \frac{\beta(M)}{1 - \text{Tr}\, M_? \sigma} = \frac{\text{Tr}\, M_1 \sigma}{\text{Tr}\,[(M_1 + M_2)\sigma]}, \tag{9}$$

where 'conclusive' denotes the event of a conclusive measurement outcome. The minimal type II error probability is then

$$\overline{\beta}_\varepsilon(\rho, \sigma) := \inf_{M \in \mathcal{M}_3} \left\{ \overline{\beta}(M) \,\middle|\, \overline{\alpha}(M) \leq \varepsilon \right\}$$

$$= \inf_{M_1, M_2 \geq 0} \left\{ \frac{\text{Tr}\, M_1 \sigma}{\text{Tr}\,[(M_1 + M_2)\sigma]} \,\middle|\, M_1 + M_2 \leq \mathbb{1}, \; \frac{\text{Tr}\, M_2 \rho}{\text{Tr}\,[(M_1 + M_2)\rho]} \leq \varepsilon \right\}, \tag{10}$$

where $\mathcal{M}_3$ denotes the set of all three-outcome measurements of the form $\{M_1, M_2, M_?\}$, and to avoid pathological cases the optimisation is implicitly restricted to be over operators such that $\text{Tr}\,[(M_1 + M_2)\sigma]$, $\text{Tr}\,[(M_1 + M_2)\rho] > 0$. By construction, this error can never be larger than the conventional error $\beta_\varepsilon(\rho, \sigma)$, since choosing $M_? = 0$ reduces to the usual definitions of $\alpha(M)$ and $\beta(M)$.

The analogous symmetric error is defined as follows. First, given two states, ρ with prior probability p and σ with prior q, a measurement $M = \{M_1, M_2, M_?\}$ has the following probability

of obtaining an incorrect outcome:

$$
\begin{aligned}
p_{\mathrm{err}}(\rho, \sigma \,|\, p, q \,|\, M) &:= P(\mathrm{error}\,|\,M)\\
&= p \operatorname{Tr} M_2 \rho + q \operatorname{Tr} M_1 \sigma,
\end{aligned}
\tag{11}
$$

while the probability of obtaining a conclusive outcome is

$$
\begin{aligned}
p_{\mathrm{conc}}(\rho, \sigma \,|\, p, q \,|\, M) &:= P(\mathrm{conclusive}\,|\,M)\\
&= 1 - \left(p \operatorname{Tr} M_? \rho + q \operatorname{Tr} M_? \sigma \right)\\
&= p \operatorname{Tr}[(M_1 + M_2)\rho] + q \operatorname{Tr}[(M_1 + M_2)\sigma]\\
&= \operatorname{Tr}[(M_1 + M_2)(p\rho + q\sigma)].
\end{aligned}
\tag{12}
$$

We then define the postselected symmetric probability of error $\overline{p}_{\mathrm{err}}$ as the optimal conditional probability of error given a conclusive outcome — in other words, the smallest ratio of incorrect outcomes to all conclusive outcomes on average:

$$
\begin{aligned}
\overline{p}_{\mathrm{err}}(\rho, \sigma \,|\, p, q) &:= \inf_{M \in \mathcal{M}_3} \frac{p_{\mathrm{err}}(\rho, \sigma \,|\, p, q \,|\, M)}{p_{\mathrm{conc}}(\rho, \sigma \,|\, p, q \,|\, M)}\\
&= \inf_{\substack{M_1, M_2 \geq 0,\\ M_1 + M_2 \leq \mathbb{1}}} \frac{p \operatorname{Tr} M_2 \rho + q \operatorname{Tr} M_1 \sigma}{\operatorname{Tr}[(M_1 + M_2)(p\rho + q\sigma)]}.
\end{aligned}
\tag{13}
$$

Before stating our main results, we need to introduce several quantities which we will find to characterise the hypothesis testing errors defined above.

Underlying all of them is the *max-relative entropy* [45], defined as

$$
\begin{aligned}
D_{\max}(\rho\|\sigma) &:= \log \inf \left\{ \lambda \in \mathbb{R}_+ \;\middle|\; \rho \leq \lambda\sigma \right\}\\
&= \begin{cases} \log \left\| \sigma^{-1/2} \rho \sigma^{-1/2} \right\|_\infty & \operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)\\ \infty & \text{otherwise.} \end{cases}
\end{aligned}
\tag{14}
$$

Here, the inequality between operators is understood in terms of the positive semidefinite cone, i.e. $\rho \leq \lambda\sigma$ if and only if $\lambda\sigma - \rho$ is a positive semidefinite operator, and $\|\cdot\|_\infty$ denotes the operator norm (largest singular value). This quantity is typically defined for normalised quantum states ρ and σ, although we will extend the same definition also to unnormalised positive semidefinite operators.

The max-relative entropy gives rise to two important quantities: the **Hilbert projective metric** [46]

$$
D_\Omega(\rho\|\sigma) := D_{\max}(\rho\|\sigma) + D_{\max}(\sigma\|\rho)
\tag{15}
$$

and the **Thompson metric** [47]

$$
D_\Xi(\rho\|\sigma) := \max \left\{ D_{\max}(\rho\|\sigma), D_{\max}(\sigma\|\rho) \right\}.
\tag{16}
$$

The two quantities have enjoyed a number of applications in various branches of geometry and analysis [46, 48]. Within quantum information theory, the Hilbert projective metric found operational use in characterising the convertibility of quantum states in different scenarios [49–51], often constituting the unique function that determines the existence of probabilistic transformations between quantum resources [51, 52]; the Thompson metric, on the other hand, was recently identified with the rate of a task known as symmetric distinguishability dilution [53], concerned with transformations of pairs of quantum states. To date, they have not been connected with rates of quantum hypothesis testing tasks.

Throughout this work, we assume that the considered quantum systems are finite dimensional.

C. Results

Our main results are the exact evaluation of the asymptotic error exponents in the setting of postselected hypothesis testing, both for asymmetric and symmetric discrimination, and both for quantum states and channels. The framework is shown to enjoy remarkably simplified properties compared to conventional hypothesis testing, allowing us to completely characterise the errors at the one-shot level as well as their asymptotic exponents. Notably, we show that the asymptotic exponents in asymmetric and symmetric postselected hypothesis testing are given exactly by the Hilbert and Thompson metrics, respectively, endowing the two quantities with operational meanings mirroring that of the quantum relative entropy and the Chernoff divergence. Our methods straightforwardly extend to the setting of composite hypothesis testing and channel discrimination, establishing results that are strictly stronger and more general than corresponding known results in conventional quantum hypothesis testing.

To our knowledge, this is the first time that the asymptotic setting of postselected hypothesis testing has been studied, even in the classical case; our results are therefore new also when the discrimination of quantum states is replaced with the discrimination of probability distributions (equivalently, commuting states).

We provide an overview of our main findings below.

Asymmetric hypothesis testing. In Section II we show that the asymptotic exponent of the conditional type II error in discriminating any two quantum states is given by the Hilbert projective metric D_Ω between them; specifically,

$$\lim_{n\to\infty} -\frac{1}{n}\log\overline{\beta}_\varepsilon(\rho^{\otimes n}, \sigma^{\otimes n}) = D_\Omega(\rho\|\sigma) \qquad \forall\varepsilon\in(0,1). \tag{17}$$

This establishes an equivalent of quantum Stein's lemma in the setting of postselected hypothesis testing. We in fact establish a closed formula for the asymmetric hypothesis testing error already at the one-shot level, leading directly to the above asymptotic result. Specifically, our main result in Theorem 1 shows that

$$\overline{\beta}_\varepsilon(\rho, \sigma) = \left(\frac{\varepsilon}{1-\varepsilon}2^{D_\Omega(\rho\|\sigma)} + 1\right)^{-1} \qquad \forall\varepsilon\in(0,1). \tag{18}$$

In Section II D, we extend the above results to composite hypotheses, showing that when the alternative hypothesis $\sigma^{\otimes n}$ is replaced by a family of convex sets of density matrices $(\mathcal{F}_n)_n$ that is closed under tensor product, then the asymptotic exponent is given by

$$\lim_{n\to\infty} -\frac{1}{n}\log\overline{\beta}_{\varepsilon,\mathcal{F}_n}(\rho^{\otimes n}) = D^\infty_{\Omega,\mathcal{F}}(\rho) := \lim_{n\to\infty}\frac{1}{n}\min_{\sigma_n\in\mathcal{F}_n} D_\Omega(\rho^{\otimes n}\|\sigma_n) \qquad \forall\varepsilon\in(0,1), \tag{19}$$

where $\overline{\beta}_{\varepsilon,\mathcal{F}_n}$ denotes the least conditional type II error in discriminating a state against *all* states in the set $\mathcal{F}_n$. The above is the analogue of the conjectured generalised quantum Stein's lemma of [15], which is not known to hold in the setting of conventional quantum hypothesis testing, but which we show to be valid in the postselected setting.

Analogous findings are also valid for the discrimination of quantum channels. We show in Section IV that

$$\lim_{n\to\infty} -\frac{1}{n}\log\overline{\beta}_\varepsilon(\mathcal{M}^{\otimes n}, \mathcal{N}^{\otimes n}) = \lim_{n\to\infty} -\frac{1}{n}\log\overline{\beta}_\varepsilon^{\,\mathrm{any}}(\mathcal{M}, \mathcal{N}\,|\,n) = D_\Omega(\mathcal{M}\|\mathcal{N}) \qquad \forall\varepsilon\in(0,1), \tag{20}$$

where $\overline{\beta}_\varepsilon^{\,\mathrm{any}}(\mathcal{M}, \mathcal{N}\,|\,n)$ denotes the error exponent obtained using the most general channel discrimination schemes that use n channel evaluations, and where $D_\Omega(\mathcal{M}\|\mathcal{N})$ is given by the Hilbert

projective metric between the Choi operators of the two channels, which does not require any regularisation and can be readily evaluated. Importantly, our results show that the performance of channel hypothesis testing is the same regardless of whether the channels are used in parallel or discriminated using broader schemes such as adaptive protocols. This equivalence holds both asymptotically and in the finite-copy regime.

Symmetric hypothesis testing. In Section III we establish that the asymptotic exponent of the postselected symmetric error in distinguishing any two states equals their Thompson metric:

$$\lim_{n\to\infty} -\frac{1}{n} \log \overline{p}_{\mathrm{err}}(\rho^{\otimes n}, \sigma^{\otimes n} \mid p, q) = D_\Xi(\rho\|\sigma) \qquad \forall p, q = 1 - p \in (0, 1). \tag{21}$$

This is the equivalent of the Chernoff bound in the setting of postselected quantum hypothesis testing and gives a direct operational interpretation of the Thompson metric in state discrimination. As in the asymmetric case, the result is based on an exact quantification of the one-shot symmetric error, which we show in Theorem 6 to be

$$\overline{p}_{\mathrm{err}}(\rho, \sigma \mid p, q) = \left(2^{D_\Xi(p\rho\|q\sigma)} + 1\right)^{-1}. \tag{22}$$

The above is generalised to quantum channels in Section IV, where we show that

$$\lim_{n\to\infty} -\frac{1}{n} \log \overline{p}_{\mathrm{err}}(\mathcal{M}^{\otimes n}, \mathcal{N}^{\otimes n} \mid p, q) = \lim_{n\to\infty} -\frac{1}{n} \log \overline{p}_{\mathrm{err}}^{\mathrm{any}}(\mathcal{M}, \mathcal{N} \mid p, q \mid n) = D_\Xi(\mathcal{M}\|\mathcal{N}), \tag{23}$$

again demonstrating that no advantage is gained by employing protocols more general than parallel ones.

Section VI discusses how the above results in asymmetric and symmetric state discrimination are actually valid in physical theories broader than quantum mechanics, namely, settings under the umbrella of general probabilistic theories. This is, to the best of our knowledge, the first time that asymptotic error exponents in hypothesis testing have been computed in such general theories — previous such results made explicit use of the properties of quantum mechanics or classical probability theory.

II. ASYMMETRIC CASE

A. One-shot discrimination error

Let us recall the definition of the minimal postselected type II error,

$$\overline{\beta}_\varepsilon(\rho, \sigma) = \inf_{M_1, M_2 \geq 0} \left\{ \frac{\mathrm{Tr}\, M_1 \sigma}{\mathrm{Tr}\left[(M_1 + M_2)\sigma\right]} \;\middle|\; M_1 + M_2 \leq \mathbb{1},\; \frac{\mathrm{Tr}\, M_2 \rho}{\mathrm{Tr}\left[(M_1 + M_2)\rho\right]} \leq \varepsilon \right\}. \tag{24}$$

Our first result is to connect it with the Hilbert projective metric. We will frequently encounter the non-logarithmic variant of the latter, for which we introduce the notation

$$\Omega(\rho\|\sigma) := 2^{D_\Omega(\rho\|\sigma)}. \tag{25}$$

Theorem 1. *For all $\varepsilon \in (0, 1)$, we have*

$$\overline{\beta}_\varepsilon(\rho, \sigma) = \left(\frac{\varepsilon}{1 - \varepsilon} \Omega(\rho\|\sigma) + 1\right)^{-1}. \tag{26}$$

Proof. We begin by noticing that the constraint $M_1 + M_2 \leq \mathbb{1}$ in (24) is superfluous: for all $M_1, M_2 \geq 0$ that satisfy $\frac{\operatorname{Tr} M_2 \rho}{\operatorname{Tr}[(M_1 + M_2)\rho]} \leq \varepsilon$, one can always rescale them as $M_i' := M_i / \|M_1 + M_2\|_\infty$ and they are then feasible for the original program with the same value of the objective function. Thus,[2]

$$\overline{\beta}_\varepsilon(\rho, \sigma)^{-1} = \sup_{M_1, M_2 \geq 0} \left\{ \frac{\operatorname{Tr}[(M_1 + M_2)\sigma]}{\operatorname{Tr} M_1 \sigma} \,\middle|\, \frac{\operatorname{Tr} M_2 \rho}{\operatorname{Tr}[(M_1 + M_2)\rho]} \leq \varepsilon \right\}. \tag{27}$$

In order to relate this optimisation with the Hilbert projective metric, we will use the dual form of the optimisation problem $\Omega(\rho \| \sigma)$. Writing

$$\Omega(\rho \| \sigma) = \Omega(\sigma \| \rho) = \inf \left\{ \gamma \in \mathbb{R}_+ \,\middle|\, \sigma \leq \lambda \rho \leq \gamma \sigma, \, \lambda \in \mathbb{R}_+ \right\}, \tag{28}$$

the dual optimisation problem can be obtained as [52]

$$\Omega(\rho \| \sigma) = \sup \left\{ \frac{\operatorname{Tr} A \sigma}{\operatorname{Tr} B \sigma} \,\middle|\, A, B \geq 0, \, \frac{\operatorname{Tr} A \rho}{\operatorname{Tr} B \rho} \leq 1 \right\}. \tag{29}$$

The similarity between the two problems can be noticed by rewriting the objective function of (27) as

$$\frac{\operatorname{Tr}[(M_1 + M_2)\sigma]}{\operatorname{Tr} M_1 \sigma} = 1 + \frac{\operatorname{Tr} M_2 \sigma}{\operatorname{Tr} M_1 \sigma}. \tag{30}$$

To make the relation explicit, let $M_1, M_2 \geq 0$ be arbitrary operators feasible for (27), and define

$$A := (1 - \varepsilon)M_2,$$
$$B := \varepsilon M_1. \tag{31}$$

Notice now that

$$\frac{\operatorname{Tr} A \rho}{\operatorname{Tr} B \rho} = \frac{(1 - \varepsilon) \operatorname{Tr} M_2 \rho}{\varepsilon \operatorname{Tr} M_1 \rho} \leq 1, \tag{32}$$

using the fact that

$$\frac{\operatorname{Tr} M_2 \rho}{\operatorname{Tr}[(M_1 + M_2)\rho]} \leq \varepsilon \iff \frac{\operatorname{Tr} M_1 \rho}{\operatorname{Tr} M_2 \rho} \geq \frac{1 - \varepsilon}{\varepsilon}. \tag{33}$$

A and B are therefore feasible operators for the dual optimisation problem for $\Omega(\rho \| \sigma)$ in (29), giving

$$\Omega(\rho \| \sigma) \geq \frac{1 - \varepsilon}{\varepsilon} \frac{\operatorname{Tr} M_2 \sigma}{\operatorname{Tr} M_1 \sigma} = \frac{1 - \varepsilon}{\varepsilon} \left(\frac{\operatorname{Tr}[(M_1 + M_2)\sigma]}{\operatorname{Tr} M_1 \sigma} - 1 \right). \tag{34}$$

Working backwards through the above reasoning, we can analogously show that any feasible solution $\{A, B\}$ for $\Omega(\rho \| \sigma)$ in (29) gives a feasible solution $\{M_1, M_2\}$ for $\overline{\beta}_\varepsilon(\rho, \sigma)^{-1}$ in (27). The two problems are therefore equivalent, in the sense that

$$\Omega(\rho \| \sigma) = \frac{1 - \varepsilon}{\varepsilon} \left(\overline{\beta}_\varepsilon(\rho, \sigma)^{-1} - 1 \right), \tag{35}$$

which is precisely the statement of the theorem. ∎

---

[2] A careful reader might wonder whether there is no danger of dividing by zero in this rewriting. That is not the case, as we can always constrain the optimisation to be over M_1 such that $\operatorname{Tr} M_1 \sigma > 0$. This can be seen by noting that, for every M_1 such that $\operatorname{Tr} M_1 \sigma = 0$, $M_1 + \delta \mathbb{1}$ is also feasible for all $\delta > 0$, and the error value of the original M_1 is recovered in the limit $\delta \to 0$.

Achievability. As long as $\overline{\beta}_\varepsilon(\rho, \sigma) \neq 0$, i.e. $\Omega(\rho\|\sigma) < \infty$, a measurement that achieves the optimal error $\overline{\beta}_\varepsilon(\rho, \sigma)$ can be readily constructed using the fact that $D_{\max}(\sigma\|\rho) = \log \left\|\rho^{-1/2}\sigma\rho^{-1/2}\right\|_\infty$. Let $|\psi_{\max}\rangle$ and $|\psi_{\min}\rangle$ be eigenvectors corresponding, respectively, to the largest and the smallest non-zero eigenvalue of the operator $\rho^{-1/2}\sigma\rho^{-1/2}$. Then, define

$$\widetilde{M}_1 := \frac{\rho^{-1/2}|\psi_{\min}\rangle\langle\psi_{\min}|\rho^{-1/2}}{\varepsilon}, \qquad \widetilde{M}_2 := \frac{\rho^{-1/2}|\psi_{\max}\rangle\langle\psi_{\max}|\rho^{-1/2}}{1 - \varepsilon}. \tag{36}$$

Now we only need to rescale these operators into a valid POVM, that is, define the three-outcome measurement

$$M := \left\{ \frac{\widetilde{M}_1}{\left\|\widetilde{M}_1 + \widetilde{M}_2\right\|_\infty}, \frac{\widetilde{M}_2}{\left\|\widetilde{M}_1 + \widetilde{M}_2\right\|_\infty}, \mathbb{1} - \frac{\widetilde{M}_1 + \widetilde{M}_2}{\left\|\widetilde{M}_1 + \widetilde{M}_2\right\|_\infty} \right\}. \tag{37}$$

It can be verified that $\overline{\alpha}(M) = \varepsilon$ and

$$\overline{\beta}(M) = \left(1 + \frac{\varepsilon}{1 - \varepsilon}\left\|\rho^{-1/2}\sigma\rho^{-1/2}\right\|_\infty \left\|\sigma^{-1/2}\rho\sigma^{-1/2}\right\|_\infty\right)^{-1} = \overline{\beta}_\varepsilon(\rho, \sigma), \tag{38}$$

meaning that the measurement is indeed optimal.

An interesting point for the discrimination of many-copy states, i.e. $\rho^{\otimes n}$ and $\sigma^{\otimes n}$, is as follows. An optimal measurement strategy can readily be constructed following the above recipe; remarkably, the operational procedure corresponding to the measurement one obtains in this way is a product strategy, i.e. it involves separate measurements on each individual copy of the state, which is much simpler to implement from a practical and technological perspective. To see this, let us define a single-copy measurement similarly to (37), except that we replace $\widetilde{M}_1$ and $\widetilde{M}_2$ with

$$\widetilde{M}_1' := \frac{\rho^{-1/2}|\psi_{\min}\rangle\langle\psi_{\min}|\rho^{-1/2}}{\varepsilon^{1/n}}, \qquad \widetilde{M}_2' := \frac{\rho^{-1/2}|\psi_{\max}\rangle\langle\psi_{\max}|\rho^{-1/2}}{(1 - \varepsilon)^{1/n}}. \tag{39}$$

We then measure each individual copy of ρ or σ. If the outcome '1' is obtained n times, we guess $\rho^{\otimes n}$; if the outcome '2' is obtained n times, we guess $\sigma^{\otimes n}$; any other combination of outcomes is deemed inconclusive. Conditioned on a conclusive result, the probability of success of such a strategy is clearly the same as if we used the measurement (37) directly.

B.　Asymptotic error exponent

The extension of Theorem 1 to the asymptotic setting is almost immediate. The crucial property of the Hilbert projective metric, and indeed also the Thompson metric, is that they are additive quantities: for all states ρ and σ,

$$\begin{aligned} D_\Omega(\rho^{\otimes n}\|\sigma^{\otimes n}) &= n\, D_\Omega(\rho\|\sigma), \\ D_\Xi(\rho^{\otimes n}\|\sigma^{\otimes n}) &= n\, D_\Xi(\rho\|\sigma). \end{aligned} \tag{40}$$

This follows from the additivity of $D_{\max}$, which in turn follows directly from the definition and the fact that

$$\left\|\left(\sigma^{\otimes n}\right)^{-1/2}\rho^{\otimes n}\left(\sigma^{\otimes n}\right)^{-1/2}\right\|_\infty = \left\|\left(\sigma^{-1/2}\rho\sigma^{-1/2}\right)^{\otimes n}\right\|_\infty = \left\|\sigma^{-1/2}\rho\sigma^{-1/2}\right\|_\infty^n. \tag{41}$$

Using this additivity and the two-sided inequality

$$\frac{\varepsilon}{1-\varepsilon}\,\Omega(\rho\|\sigma) \le \overline{\beta}_\varepsilon(\rho,\sigma)^{-1} \le \frac{1}{1-\varepsilon}\,\Omega(\rho\|\sigma), \tag{42}$$

itself a consequence of the fact that $\Omega(\rho\|\sigma) \ge 1$ for all ρ, σ, we obtain

$$n D_\Omega(\rho\|\sigma) + \log\frac{\varepsilon}{1-\varepsilon} \le -\log\overline{\beta}_\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n}) \le n D_\Omega(\rho\|\sigma) + \log\frac{1}{1-\varepsilon}. \tag{43}$$

This immediately yields an expression for the asymptotic error exponent, which is the equivalent of quantum Stein's lemma in the setting of this work.

> **Corollary 2** (Asymptotic error exponent of asymmetric postselected hypothesis testing).
> *For all $\varepsilon \in (0,1)$, we have that*
>
> $$\lim_{n\to\infty} -\frac{1}{n}\log\overline{\beta}_\varepsilon(\rho^{\otimes n}, \sigma^{\otimes n}) = D_\Omega(\rho\|\sigma). \tag{44}$$

What is noteworthy about the expression in (43) is that there are no lower-order terms in n: the convergence to the asymptotic limit of Corollary 2 does not depend on any terms other than ones linear in n, which strongly contrasts with the case encountered in conventional quantum hypothesis testing [54, 55].

C. Properties of the asymmetric error

As the asymptotic error exponent is given exactly by D_Ω, while the one-shot error $\overline{\beta}_\varepsilon$ is an inversely monotonic function of it, it suffices to characterise the properties of D_Ω to fully understand the behaviour of the errors in both the one-shot and exponential settings.

Proposition 3. *For all states ρ and σ, and for all $\varepsilon \in (0,1)$:*

(i) *$D_\Omega(\rho\|\sigma) \ge 0$, with equality if and only if $\rho = \sigma$. Consequently, $\overline{\beta}_\varepsilon(\rho,\sigma) \le 1 - \varepsilon$, with equality if and only if $\rho = \sigma$.*

(ii) *Both $D_\Omega(\rho\|\sigma)$ and $\overline{\beta}_\varepsilon(\rho,\sigma)$ are symmetric under the exchange of ρ and σ, i.e. $D_\Omega(\rho\|\sigma) = D_\Omega(\sigma\|\rho)$ and $\overline{\beta}_\varepsilon(\rho,\sigma) = \overline{\beta}_\varepsilon(\sigma,\rho)$.*

(iii) *For all positive numbers λ, μ, we have $D_\Omega(\lambda\rho\|\mu\sigma) = D_\Omega(\rho\|\sigma)$, and analogously $\overline{\beta}_\varepsilon(\lambda\rho, \mu\sigma) = \overline{\beta}_\varepsilon(\rho,\sigma)$.*

(iv) *$D_\Omega(\rho\|\sigma) = \infty$ if and only if $\operatorname{supp}\rho \ne \operatorname{supp}\sigma$. Hence, $\overline{\beta}_\varepsilon(\rho,\sigma) = 0$ if and only if ρ and σ have non-identical supports.*

(v) *D_Ω satisfies the data-processing inequality under all positive maps; specifically, for every positive linear map $\mathcal{E}$,*

$$D_\Omega\!\left(\frac{\mathcal{E}(\rho)}{\operatorname{Tr}\mathcal{E}(\rho)}\,\middle\|\,\frac{\mathcal{E}(\sigma)}{\operatorname{Tr}\mathcal{E}(\sigma)}\right) = D_\Omega(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) \le D_\Omega(\rho\|\sigma). \tag{45}$$

Hence, $\overline{\beta}_\varepsilon\!\left(\frac{\mathcal{E}(\rho)}{\operatorname{Tr}\mathcal{E}(\rho)}, \frac{\mathcal{E}(\sigma)}{\operatorname{Tr}\mathcal{E}(\sigma)}\right) \ge \overline{\beta}_\varepsilon(\rho,\sigma).$

Proof. (i) holds because $\rho \leq \sigma$ and $\sigma \leq \rho$ can only be true if $\rho = \sigma$. (ii) is an immediate consequence of the manifestly symmetric definition (15), together with Theorem 1. (iii) follows from the fact that $D_{\max}(\lambda\rho\|\mu\sigma) = \log\frac{\lambda}{\mu} + D_{\max}(\rho\|\sigma)$, which is immediate from the definition. (iv) holds because $D_{\max}(\rho\|\sigma) < \infty$ if and only if $\operatorname{supp}\rho \subseteq \operatorname{supp}\sigma$. (v) is a consequence of (iii) coupled with the fact that for all λ such that $\rho \leq \lambda\sigma$, the positivity of $\mathcal{E}$ ensures that $\lambda\mathcal{E}(\sigma) - \mathcal{E}(\rho) = \mathcal{E}(\lambda\sigma - \rho) \geq 0$, so $D_{\max}(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) \leq D_{\max}(\rho\|\sigma)$. ∎

Remark. The symmetry shown in Proposition 3(ii) is *a priori* highly non-obvious, and indeed one of the remarkable consequences of Theorem 1. Indeed, we started by looking at the error in *asymmetric* postselected hypothesis testing, but the expression we obtained is *symmetric* under the exchange of the two states. This can be understood as a consequence of the fact that, for all states ρ, σ with $\operatorname{supp}\rho = \operatorname{supp}\sigma$, there exists a completely positive map $\mathcal{E}$ such that $\mathcal{E}(\rho) = \sigma$ and $\mathcal{E}(\sigma) = k\rho$ for some $k \in (0, \infty)$ [49, Theorem 21]. Applying the data-processing inequality of Proposition 3(v) together with the scaling invariance (iii), we see that we must have $\beta_\varepsilon(\rho\|\sigma) = \beta_\varepsilon(\sigma\|\rho)$.

In all of the above, we have assumed that $\varepsilon \in (0, 1)$. The case $\varepsilon = 0$ requires a separate treatment.

Proposition 4. *It holds that*

$$\overline{\beta}_0(\rho, \sigma) = \begin{cases} 1 & \operatorname{supp}\sigma \subseteq \operatorname{supp}\rho, \\ 0 & \text{otherwise.} \end{cases} \tag{46}$$

Proof. If $\operatorname{supp}\sigma \subseteq \operatorname{supp}\rho$, then $\operatorname{Tr}M_2\rho = 0$ implies that $\operatorname{Tr}M_2\sigma = 0$. Any feasible pair $\{M_1, M_2\}$ such that $\operatorname{Tr}M_1\sigma \neq 0$ then satisfies

$$\frac{\operatorname{Tr}M_1\sigma}{\operatorname{Tr}[(M_1 + M_2)\sigma]} = \frac{\operatorname{Tr}M_1\sigma}{\operatorname{Tr}M_1\sigma} = 1. \tag{47}$$

For the case $\operatorname{supp}\sigma \not\subseteq \operatorname{supp}\rho$, by hypothesis we know that there exists a pure state ψ such that $\operatorname{Tr}\psi\sigma > 0$ but $\operatorname{Tr}\psi\rho = 0$. Pick $M_1 = \delta\mathbb{1}, M_2 = (1 - \delta)\psi$ for some $\delta \in (0, 1)$ to conclude that any value of

$$\frac{\operatorname{Tr}M_1\sigma}{\operatorname{Tr}[(M_1 + M_2)\sigma]} = \frac{\delta}{\delta + (1 - \delta)\operatorname{Tr}\psi\sigma} \tag{48}$$

is achievable. Taking the limit $\delta \to 0$ gives the stated result. ∎

It is a curious fact that $\overline{\beta}_\varepsilon(\rho, \sigma)$ is symmetric in its arguments for $\varepsilon > 0$ (Proposition 3(ii)), but no longer so for $\varepsilon = 0$. This asymmetry is due to the situation where $\operatorname{supp}\sigma$ is strictly contained in $\operatorname{supp}\rho$. In this case, ρ can be distinguished from σ, but not the other way around: consider the POVM $\{\mathbb{1}-\Pi_\sigma, 0, \Pi_\sigma\}$ which allows us to conclude, whenever we obtain a conclusive outcome, that the state in our possession is ρ; however, the probability of a conclusive outcome when measuring σ is zero. This problem is avoided for $\varepsilon > 0$.

D.   Composite postselected hypothesis testing

In the setting of composite hypothesis testing, we wish to distinguish between a given state ρ and a whole set of quantum states, denoted $\mathcal{F}$. Let us first consider the case of conventional (non-postselected) hypothesis testing. Here, for a given measurement M, the performance is quantified with the worst-case type II error, i.e.

$$\beta_{\mathcal{F}}(M) := \sup_{\sigma\in\mathcal{F}} \beta(M) = \sup_{\sigma\in\mathcal{F}} \operatorname{Tr}M_1\sigma, \tag{49}$$

and the optimised error is defined as $\beta_{\varepsilon,\mathcal{F}}(\rho) := \inf\left\{\beta_{\mathcal{F}}(M) \mid \alpha(M) \le \varepsilon\right\}$. Brandão and Plenio [15] conjectured a remarkable extension of the quantum Stein's lemma to composite hypothesis testing: namely, that for every family of sets $(\mathcal{F}_n)_n$ satisfying a small number of regularity properties (including convexity and closedness under tensor product, partial trace, and permutations of systems), it holds that

$$\lim_{n\to\infty} -\frac{1}{n}\log\beta_{\varepsilon,\mathcal{F}_n}(\rho^{\otimes n}) \overset{?}{=} D_{\mathcal{F}}^\infty(\rho), \tag{50}$$

where $D_{\mathcal{F}}^\infty$ is the regularised relative entropy, $D_{\mathcal{F}}^\infty(\rho) := \lim_{n\to\infty}\frac{1}{n}\min_{\sigma_n\in\mathcal{F}_n}D(\rho^{\otimes n}\|\sigma_n)$. Although a proof of this result was claimed in [15], issues with the derivation were recently uncovered [16], and (50) is therefore not known to hold in general.

We show that an analogue of the above result can be shown to hold true in the setting of postselected hypothesis testing. Define

$$\overline{\beta}_{\varepsilon,\mathcal{F}}(\rho) := \inf_{M\in\mathcal{M}_3}\left\{\overline{\beta}_{\mathcal{F}}(M) \mid \overline{\alpha}(M) \le \varepsilon\right\}$$

$$= \inf_{M_1,M_2\ge 0}\left\{\sup_{\sigma\in\mathcal{F}}\frac{\operatorname{Tr}M_1\sigma}{\operatorname{Tr}\left[(M_1+M_2)\sigma\right]} \;\middle|\; M_1+M_2\le\mathbb{1},\ \frac{\operatorname{Tr}M_2\rho}{\operatorname{Tr}\left[(M_1+M_2)\rho\right]}\le\varepsilon\right\}. \tag{51}$$

We then show that the asymptotic error exponent of composite postselected hypothesis testing is given exactly by the regularisation of the Hilbert projective metric $D_\Omega(\rho\|\sigma)$ optimised over the sets $\mathcal{F}_n$.

> **Theorem 5.** *For every convex and closed set of quantum states $\mathcal{F}$,*
>
> $$\overline{\beta}_{\varepsilon,\mathcal{F}}(\rho) = \left(\frac{\varepsilon}{1-\varepsilon}\min_{\sigma\in\mathcal{F}}\Omega(\rho\|\sigma)+1\right)^{-1}. \tag{52}$$
>
> *As a consequence, for every family of sets $(\mathcal{F}_n)_n$ that are convex and closed, we have that*
>
> $$\limsup_{n\to\infty} -\frac{1}{n}\log\overline{\beta}_{\varepsilon,\mathcal{F}_n}(\rho^{\otimes n}) = D_{\Omega,\mathcal{F}}^\infty(\rho), \tag{53}$$
>
> *where*
>
> $$D_{\Omega,\mathcal{F}}^\infty(\rho) = \limsup_{n\to\infty}\frac{1}{n}\min_{\sigma_n\in\mathcal{F}_n}D_\Omega(\rho^{\otimes n}\|\sigma_n). \tag{54}$$
>
> *If the sets $(\mathcal{F}_n)_n$ are closed under tensor product, i.e. if $\sigma_n\otimes\sigma_m\in\mathcal{F}_{n+m}$ for all $\sigma_n\in\mathcal{F}_n$ and $\sigma_m\in\mathcal{F}_m$, then $\limsup$ in (53)–(54) can be replaced with $\lim$, because the limits are achieved.*

We remark here that the quantity $\min_{\sigma\in\mathcal{F}}\Omega(\rho\|\sigma)$ can be computed as the optimal value of a conic linear optimisation problem [52]; when membership in the set $\mathcal{F}$ can be expressed using linear matrix inequalities, this reduces to a semidefinite program.

Proof. The basic idea is exactly the same as in the proof of Theorem 1. Let us first consider the

one-shot case. We have that

$$
\begin{aligned}
\overline{\beta}_{\varepsilon,\mathcal{F}}(\rho) &\overset{(i)}{=} \inf_{M_1,M_2\geq 0} \left\{ \sup_{\sigma\in\mathcal{F}} \frac{\operatorname{Tr} M_1\sigma}{\operatorname{Tr}\left[(M_1+M_2)\sigma\right]} \ \middle| \ \frac{\operatorname{Tr} M_2\rho}{\operatorname{Tr}\left[(M_1+M_2)\rho\right]} \leq \varepsilon \right\} \\
&= \inf_{\substack{M_1,M_2\geq 0 \\ t>0}} \left\{ t \ \middle| \ \frac{\operatorname{Tr} M_1\sigma}{\operatorname{Tr}\left[(M_1+M_2)\sigma\right]} \leq t \ \forall\sigma\in\mathcal{F}, \ \frac{\operatorname{Tr} M_2\rho}{\operatorname{Tr}\left[(M_1+M_2)\rho\right]} \leq \varepsilon \right\} \\
&\overset{(ii)}{=} \inf_{\substack{M_1,M_2\geq 0 \\ t>0}} \left\{ t \ \middle| \ \frac{\operatorname{Tr} M_2\sigma}{\operatorname{Tr} M_1\sigma} \geq \frac{1}{t}-1 \ \forall\sigma\in\mathcal{F}, \ \frac{\operatorname{Tr} M_1\rho}{\operatorname{Tr} M_2\rho} \geq \frac{1}{\varepsilon}-1 \right\} \\
&\overset{(iii)}{=} \inf_{\substack{M_1,M_2'\geq 0 \\ t>0}} \left\{ t \ \middle| \ \frac{\operatorname{Tr} M_2'\sigma}{\operatorname{Tr} M_1\sigma} \geq 1 \ \forall\sigma\in\mathcal{F}, \ \frac{\operatorname{Tr} M_1\rho}{\operatorname{Tr} M_2'\rho} \geq \frac{1-\varepsilon}{\varepsilon}\left(\frac{1}{t}-1\right) \right\} \\
&\overset{(iv)}{=} \inf_{\substack{M_1,M_2'\geq 0 \\ t'>0}} \left\{ \frac{1}{t'} \ \middle| \ \frac{\operatorname{Tr} M_1\sigma}{\operatorname{Tr} M_2'\sigma} \leq 1 \ \forall\sigma\in\mathcal{F}, \ \frac{\operatorname{Tr} M_1\rho}{\operatorname{Tr} M_2'\rho} \geq \frac{1-\varepsilon}{\varepsilon}(t'-1) \right\},
\end{aligned}
\tag{55}
$$

where: (i) we removed the superfluous constraint $M_1 + M_2 \leq \mathbb{1}$ as before; (ii) we rewrote $\frac{\operatorname{Tr} M_1\sigma}{\operatorname{Tr}[(M_1+M_2)\sigma]} = \left(1 + \frac{\operatorname{Tr} M_2\rho}{\operatorname{Tr} M_1\sigma}\right)^{-1}$ and analogously for ρ; (iii) we made the change of variables $M_2' := \frac{t}{1-t}M_2$; (iv) we made the change of variables $t' := \frac{1}{t}$. Thus,

$$
\begin{aligned}
\overline{\beta}_{\varepsilon,\mathcal{F}}(\rho)^{-1} &= \sup_{\substack{M_1,M_2'\geq 0 \\ t'>0}} \left\{ t' \ \middle| \ \frac{\operatorname{Tr} M_1\sigma}{\operatorname{Tr} M_2'\sigma} \leq 1 \ \forall\sigma\in\mathcal{F}, \ \frac{\operatorname{Tr} M_1\rho}{\operatorname{Tr} M_2'\rho} \geq \frac{1-\varepsilon}{\varepsilon}(t'-1) \right\} \\
&= \sup_{M_1,M_2'\geq 0} \left\{ \frac{\varepsilon}{1-\varepsilon}\frac{\operatorname{Tr} M_1\rho}{\operatorname{Tr} M_2'\rho} + 1 \ \middle| \ \frac{\operatorname{Tr} M_1\sigma}{\operatorname{Tr} M_2'\sigma} \leq 1 \ \forall\sigma\in\mathcal{F} \right\}.
\end{aligned}
\tag{56}
$$

Now, for every convex and closed set $\mathcal{F}$, we have [52, Theorem 1]

$$
\min_{\sigma\in\mathcal{F}} \Omega(\rho\|\sigma) = \sup_{A,B\geq 0} \left\{ \frac{\operatorname{Tr} A\rho}{\operatorname{Tr} B\rho} \ \middle| \ \frac{\operatorname{Tr} A\sigma}{\operatorname{Tr} B\sigma} \leq 1 \ \forall\sigma\in\mathcal{F} \right\},
\tag{57}
$$

from which we immediately get that

$$
\overline{\beta}_{\varepsilon,\mathcal{F}}(\rho)^{-1} = \frac{\varepsilon}{1-\varepsilon}\min_{\sigma\in\mathcal{F}}\Omega(\rho\|\sigma) + 1.
\tag{58}
$$

For the n-copy case, this means that

$$
\overline{\beta}_{\varepsilon,\mathcal{F}_n}(\rho^{\otimes n})^{-1} = \frac{\varepsilon}{1-\varepsilon}\min_{\sigma_n\in\mathcal{F}_n}\Omega(\rho^{\otimes n}\|\sigma_n) + 1,
\tag{59}
$$

implying, as in (42), that

$$
\frac{\varepsilon}{1-\varepsilon}\min_{\sigma_n\in\mathcal{F}_n}\Omega(\rho^{\otimes n}\|\sigma_n) \leq \overline{\beta}_{\varepsilon,\mathcal{F}_n}(\rho^{\otimes n})^{-1} \leq \frac{1}{1-\varepsilon}\min_{\sigma_n\in\mathcal{F}_n}\Omega(\rho^{\otimes n}\|\sigma_n).
\tag{60}
$$

The result follows by taking the logarithm of the above, dividing by n, and taking the lim sup.

When the sets $\mathcal{F}_n$ are closed under tensor product, then $D_{\Omega,\mathcal{F}_{m+n}}(\rho^{\otimes m+n}) \leq D_{\Omega,\mathcal{F}_m}(\rho^{\otimes m}) + D_{\Omega,\mathcal{F}_n}(\rho^{\otimes n})$ [52, Theorem 1]. By Fekete's lemma [56], this means that the limit $\lim_{n\to\infty}\frac{1}{n}D_{\Omega,\mathcal{F}_n}(\rho^{\otimes n})$ exists. ∎

In addition to the conjectured generalisation of quantum Stein's lemma, the works of Brandão and Plenio [15, 57] established an interesting connection between quantum hypothesis testing and operational aspects of quantum resource transformations. It was shown, in particular, that the asymptotic error exponent in (conventional) asymmetric hypothesis testing (Eq. (50)) gives exactly the rate at which quantum entanglement can be distilled, i.e. converted to pure maximally entangled states, under the class of non-entangling quantum channels. This relation is mirrored exactly in postselected hypothesis testing: the quantity $D_{\Omega,\mathcal{F}}^{\infty}(\rho)$ in Theorem 5 was recently shown to give the rate of distillation under probabilistic non-entangling operations [51], and here we have shown that it is also precisely the asymptotic error exponent.

III. SYMMETRIC CASE

A. One-shot symmetric hypothesis testing error

Consider the discrimination of two states, ρ and σ, with their respective priors being p and q. Let us recall the definition of the postselected probability of error $\overline{p}_{\mathrm{err}}$:

$$\overline{p}_{\mathrm{err}}(\rho, \sigma \mid p, q) = \inf_{M_1, M_2 \geq 0, \, M_1 + M_2 \leq \mathbb{1}} \frac{p \operatorname{Tr} M_2 \rho + q \operatorname{Tr} M_1 \sigma}{p \operatorname{Tr}[(M_1 + M_2)\rho] + q \operatorname{Tr}[(M_1 + M_2)\sigma]}. \tag{61}$$

We note that a closely related figure of merit appeared previously in [37], motivated by a framework generalising the setting of unambiguous state discrimination (cf. also [41]). There, the authors computed the postselected probability of *success*, which is just $1 - \overline{p}_{\mathrm{err}}(\rho, \sigma \mid p, q)$ in our notation. However, the asymptotic properties of this quantity were not studied, nor was its relation with the Thompson metric observed.

Our first result establishes that the one-shot postselected discrimination error is given by a simple function of the Thompson metric, or rather its non-logarithmic variant

$$\Xi(\rho \| \sigma) := 2^{D_{\Xi}(\rho \| \sigma)}. \tag{62}$$

Note that below we will apply the Thompson metric also to rescaled quantum states as $\Xi(p\rho \| q\sigma)$. In such a case, we use the exact same definition of the max-relative entropy as for normalised states, which gives

$$\begin{aligned} D_{\max}(p\rho \| q\sigma) &= \log \inf \left\{ \lambda \in \mathbb{R}_+ \mid p\rho \leq \lambda q\sigma \right\} \\ &= \log \frac{p}{q} + D_{\max}(\rho \| \sigma). \end{aligned} \tag{63}$$

Theorem 6. *For all states ρ, σ and for all priors p, $q = 1 - p \in (0, 1)$, we have*

$$\overline{p}_{\mathrm{err}}(\rho, \sigma \mid p, q) = \left(\Xi(p\rho \| q\sigma) + 1 \right)^{-1}. \tag{64}$$

Proof. We first derive the dual formulation of the Thompson metric. Writing

$$\begin{aligned} \Xi(p\rho \| q\sigma) &= \inf \left\{ \max\{\lambda_1, \lambda_2\} \, \Big| \, \frac{1}{\lambda_1} q\sigma \leq p\rho \leq \lambda_2 \, q\sigma \right\} \\ &= \inf \left\{ \lambda \mid q\sigma \leq \lambda \, p\rho, \, p\rho \leq \lambda q\sigma \right\}, \end{aligned} \tag{65}$$

we obtain the dual program as

$$\Xi(p\rho\|q\sigma) = \sup\left\{ q\operatorname{Tr}A\sigma + p\operatorname{Tr}B\rho \mid A, B \geq 0, \ q\operatorname{Tr}B\sigma + p\operatorname{Tr}A\rho = 1 \right\}$$
$$= \sup_{A,B \geq 0} \frac{q\operatorname{Tr}A\sigma + p\operatorname{Tr}B\rho}{q\operatorname{Tr}B\sigma + p\operatorname{Tr}A\rho}. \tag{66}$$

The fact that we were justified in claiming that the optimal value of (65) equals the optimal value of (66) follows from the strong duality of the optimisation problem, easily seen by noticing that $A = B = \mathbb{1}$ are strongly feasible solutions and invoking Slater's theorem [58, Thm. 28.2].

As before, we see that the constraint $M_1 + M_2 \leq \mathbb{1}$ in the definition of $\overline{p}_{\mathrm{err}}$ is superfluous and we can ignore it without loss of generality. Notice then that

$$\overline{p}_{\mathrm{err}}(\rho, \sigma \mid p, q)^{-1} = \sup_{M_1,M_2 \geq 0} \frac{q\operatorname{Tr}[(M_1 + M_2)\sigma] + p\operatorname{Tr}[(M_1 + M_2)\rho]}{q\operatorname{Tr}M_1\sigma + p\operatorname{Tr}M_2\rho}$$
$$= \sup_{M_1,M_2 \geq 0}\left(1 + \frac{q\operatorname{Tr}M_2\sigma + p\operatorname{Tr}M_1\rho}{q\operatorname{Tr}M_1\sigma + p\operatorname{Tr}M_2\rho} \right) \tag{67}$$
$$= \left(1 + \Xi(p\rho\|q\sigma) \right),$$

where we used the dual program for the Thompson metric (66). This establishes the desired equivalence between the two problems. ∎

Achievability. A measurement that achieves the error $\overline{p}_{\mathrm{err}}(\rho, \sigma \mid p, q)$ can be constructed again using the fact that $D_{\max}(p\rho\|q\sigma) = \log\frac{p}{q}\left\|\sigma^{-1/2}\rho\sigma^{-1/2}\right\|_\infty$. Below, we will assume that $D_\Xi(p\rho\|q\sigma) = D_{\max}(p\rho\|q\sigma)$ without loss of generality. Let $|\psi_{\max}\rangle$ be an eigenvector corresponding to the largest eigenvalue of the operator $\sigma^{-1/2}\rho\sigma^{-1/2}$. Then, define

$$W = \sigma^{-1/2}|\psi_{\max}\rangle\langle\psi_{\max}|\sigma^{-1/2}. \tag{68}$$

This operator satisfies $\operatorname{Tr}W\sigma = 1$ and $\operatorname{Tr}W\rho = \left\|\sigma^{-1/2}\rho\sigma^{-1/2}\right\|_\infty$. Then, define the three-outcome POVM

$$\left\{ \frac{W}{\langle\psi_{\max}|\sigma^{-1}|\psi_{\max}\rangle}, \ 0, \ \mathbb{1} - \frac{W}{\langle\psi_{\max}|\sigma^{-1}|\psi_{\max}\rangle} \right\}, \tag{69}$$

where we used that the largest (and the only non-zero) eigenvalue of W is $\langle\psi_{\max}|\sigma^{-1}|\psi_{\max}\rangle$. Although the probability of conclusively discriminating ρ is zero, the average error $\overline{p}_{\mathrm{err}}$ indeed matches the optimal value of $(1 + \frac{p}{q}\left\|\sigma^{-1/2}\rho\sigma^{-1/2}\right\|_\infty)^{-1}$. Note that when $D_\Xi(p\rho\|q\sigma) = D_{\max}(q\sigma\|p\rho)$, the role of ρ and σ is interchanged in the above discussion.

B. Symmetric hypothesis testing error exponent

Using the additivity of $D_{\max}$ and the property that $D_{\max}(p\rho\|q\sigma) = D_{\max}(\rho\|\sigma) + \log\frac{p}{q}$, we have that

$$D_\Xi(p\rho^{\otimes n}\|q\sigma^{\otimes n}) = \max\left\{ \log\frac{p}{q} + nD_{\max}(\rho\|\sigma), \ \log\frac{q}{p} + nD_{\max}(\sigma\|\rho) \right\}. \tag{70}$$

For large enough n, the maximum on the right-hand side will be determined only by the maximum between $D_{\max}(\rho\|\sigma)$ and $D_{\max}(\sigma\|\rho)$, which implies that

$$\lim_{n\to\infty} \frac{1}{n}D_\Xi(p\rho^{\otimes n}\|q\sigma^{\otimes n}) = D_\Xi(\rho\|\sigma), \tag{71}$$

as long as the priors p, q are non-zero. Theorem 6 then immediately gives the asymptotic regularisation of the symmetric distinguishability error.

> **Corollary 7** (Asymptotic error exponent of symmetric postselected hypothesis testing). *For all states ρ, σ and for all priors p, $q = 1 - p \in (0, 1)$,*
>
> $$\lim_{n \to \infty} -\frac{1}{n} \log \overline{p}_{\mathrm{err}}(\rho^{\otimes n}, \sigma^{\otimes n} \mid p, q) = D_\Xi(\rho \| \sigma). \tag{72}$$

As in the asymmetric case, the expression for $-\log \overline{p}_{\mathrm{err}}(\rho^{\otimes n}, \sigma^{\otimes n} \mid p, q)$ contains no n-dependent terms of order lower than 1. Furthermore, and again in analogy with what happens in the asymmetric case, there is a product strategy achieving the asymptotically optimal conditional error probability in symmetric discrimination. It consists in making the measurement $\{M_1, M_2, M_?\}$ detailed in (69) on each copy of the state, guessing ρ if the outcome '1' is obtained n times, and declaring an inconclusive result in any other case.

Finally, we remark that the exact expression for many-copy discrimination error simplifies slightly for the case $p = q = \frac{1}{2}$, where the fact that $D_\Xi(\rho^{\otimes n} \| \sigma^{\otimes n}) = n\, D_\Xi(\rho \| \sigma)$ yields

$$-\log \overline{p}_{\mathrm{err}}\left(\rho^{\otimes n}, \sigma^{\otimes n} \,\middle|\, \frac{1}{2}, \frac{1}{2}\right) = \log\left(\Xi(\rho \| \sigma)^n + 1\right) \tag{73}$$

for every n.

C. Properties of the symmetric error

Proposition 8. *For all states ρ and σ, and all priors p, $q = 1 - p \in (0, 1)$:*

(i) *$D_\Xi(p\rho \| q\sigma) \geq 0$, with equality if and only if $\rho = \sigma$ and $p = q$. Hence, $\overline{p}_{\mathrm{err}}(\rho, \sigma \mid p, q) \leq \frac{1}{2}$, with equality if and only if $\rho = \sigma$ and $p = q$.*

(ii) *$D_\Xi(p\rho \| q\sigma) = \infty$ if and only if $\operatorname{supp} \rho \neq \operatorname{supp} \sigma$. Hence, $\overline{p}_{\mathrm{err}}(\rho, \sigma \mid p, q) = 0$ if and only if ρ and σ have different supports.*

(iii) *For every positive linear map $\mathcal{E}$, the inequality $D_\Xi(p\mathcal{E}(\rho) \| q\mathcal{E}(\sigma)) \leq D_\Xi(p\rho \| q\sigma)$ holds true.*

We note that our conclusion in point (ii) was previously shown in [38].

Proof. (i) follows from the fact that $p\rho \leq q\sigma$ and $p\rho \geq q\sigma$ can both be true iff $p\rho = q\sigma$, which is possible if and only if $p = q$ and $\rho = \sigma$ due to ρ and σ being normalised states. (ii) is a consequence of the fact that $D_{\max}(\rho \| \sigma) < \infty \iff \operatorname{supp} \rho \subseteq \operatorname{supp} \sigma$. (iii) is immediate because $\mathcal{E}(\lambda q\sigma - p\rho) = \lambda q\mathcal{E}(\sigma) - p\mathcal{E}(\rho) \geq 0$ for every λ such that $p\rho \leq \lambda q\sigma$ due to the positivity of $\mathcal{E}$. ∎

Note that the monotonicity $D_\Xi(p\rho \| q\sigma) \geq D_\Xi(p\mathcal{E}(\rho) \| q\mathcal{E}(\sigma))$ does not mean that we can renormalise the state at the output, as we did in Proposition 3 for the asymmetric error. Indeed, in such a case we only get the inequality

$$\Xi\left(p\frac{\mathcal{E}(\rho)}{\operatorname{Tr}\mathcal{E}(\rho)} \,\middle\|\, q\frac{\mathcal{E}(\sigma)}{\operatorname{Tr}\mathcal{E}(\sigma)}\right) \leq \max\left\{\frac{\operatorname{Tr}\mathcal{E}(\sigma)}{\operatorname{Tr}\mathcal{E}(\rho)}2^{D_{\max}(p\rho \| q\sigma)}, \frac{\operatorname{Tr}\mathcal{E}(\rho)}{\operatorname{Tr}\mathcal{E}(\sigma)}2^{D_{\max}(q\sigma \| p\rho)}\right\}$$

$$\leq \max\left\{\frac{\operatorname{Tr}\mathcal{E}(\sigma)}{\operatorname{Tr}\mathcal{E}(\rho)}, \frac{\operatorname{Tr}\mathcal{E}(\rho)}{\operatorname{Tr}\mathcal{E}(\sigma)}\right\}\Xi(p\rho \| q\sigma). \tag{74}$$

To see that $\Xi\left(\frac{\mathcal{E}(\rho)}{\operatorname{Tr}\mathcal{E}(\rho)} \,\middle\|\, \frac{\mathcal{E}(\sigma)}{\operatorname{Tr}\mathcal{E}(\sigma)}\right) \not\leq \Xi(\rho\|\sigma)$ in general, consider e.g. $\rho = \begin{pmatrix} 2/3 & 0 \\ 0 & 1/3 \end{pmatrix}$ and $\sigma = \mathbb{1}/2$, in which case it is easy to verify that $\Xi(\rho\|\sigma) = 3/2$. Setting $\mathcal{E}(\cdot) := D(\cdot)D$ with $D := \begin{pmatrix} \sqrt{\gamma} & 0 \\ 0 & 1/\sqrt{\gamma} \end{pmatrix}$ and $\gamma > 1$, we find $\Xi\left(\frac{\mathcal{E}(\rho)}{\operatorname{Tr}\mathcal{E}(\rho)} \,\middle\|\, \frac{\mathcal{E}(\sigma)}{\operatorname{Tr}\mathcal{E}(\sigma)}\right) = \frac{2\gamma^2+1}{\gamma^2+1} > 3/2 = \Xi(\rho\|\sigma)$.

IV.　QUANTUM CHANNEL DISCRIMINATION

A.　One-shot channel hypothesis testing

We will now consider quantum channels, that is, completely positive and trace-preserving linear maps acting between the spaces of operators on the Hilbert spaces of two quantum systems, A and B. We denote such channels as $\mathcal{M} : A \to B$ for simplicity. The set of quantum states of systems A and B will be denoted $\mathcal{D}_A$ and $\mathcal{D}_B$, respectively.

If we only have access to a single copy of two channels $\mathcal{M}, \mathcal{N} : A \to B$, there is only one thing we can do to discriminate them: feed in some state ρ into the channel, and then perform a distinguishing measurement at the output. However, a crucial insight here is that using a larger, entangled state and measurement can help discriminate the channels better than simply measuring $\mathcal{N}(\rho)$ for some $\rho \in \mathcal{D}_A$ [59]. Because of this, one performs channel hypothesis testing by measuring $\mathrm{id} \otimes \mathcal{M}(\rho)$ for some state $\rho \in \mathcal{D}_{RA}$, where id denotes the identity channel on the ancillary space R, with the latter space arbitrary. We therefore define the postselected asymmetric and symmetric hypothesis testing errors as

$$\overline{\beta}_\varepsilon(\mathcal{M}, \mathcal{N}) := \inf_{\rho\in\mathcal{D}_{RA}} \overline{\beta}_\varepsilon(\mathrm{id} \otimes \mathcal{M}(\rho), \mathrm{id} \otimes \mathcal{N}(\rho)),$$

$$\overline{p}_{\mathrm{err}}(\mathcal{M}, \mathcal{N} \,|\, p, q) := \inf_{\rho\in\mathcal{D}_{RA}} \overline{p}_{\mathrm{err}}(\mathrm{id} \otimes \mathcal{M}(\rho), \mathrm{id} \otimes \mathcal{N}(\rho) \,|\, p, q). \tag{75}$$

From Theorems 1 and 6, we immediately have that

$$\overline{\beta}_\varepsilon(\mathcal{M}, \mathcal{N}) = \inf_{\rho\in\mathcal{D}_{RA}} \left[\frac{\varepsilon}{1-\varepsilon}\, \Omega(\mathrm{id} \otimes \mathcal{M}(\rho)\|\mathrm{id} \otimes \mathcal{N}(\rho)) + 1\right]^{-1},$$

$$\overline{p}_{\mathrm{err}}(\mathcal{M}, \mathcal{N} \,|\, p, q) = \inf_{\rho\in\mathcal{D}_{RA}} \left[\Xi\big(p\,\mathrm{id} \otimes \mathcal{M}(\rho)\|q\,\mathrm{id} \otimes \mathcal{N}(\rho)\big) + 1\right]^{-1}. \tag{76}$$

It may not be immediately obvious if these optimisation problems can be evaluated. However, by exploiting a property of the max-relative entropy noticed in [22], we can show that it suffices to take as input the maximally entangled state $\Phi_+ = |\Phi_+\rangle\langle\Phi_+|$ with $|\Phi_+\rangle = \frac{1}{\sqrt{d_A}} \sum_{i=1}^{d_A} |ii\rangle$, which reduces the problem to evaluating $D_{\max}$ between two fixed states. We state this in the form of the following lemma.

> **Lemma 9.** *For all channels* $\mathcal{M}, \mathcal{N} : A \to B$, *define*
>
> $$\Omega(\mathcal{M}\|\mathcal{N}) := \sup_{\rho\in\mathcal{D}_{RA}} \Omega(\mathrm{id} \otimes \mathcal{M}(\rho)\|\mathrm{id} \otimes \mathcal{N}(\rho))$$
>
> $$\Xi(\mathcal{M}\|\mathcal{N}) := \sup_{\rho\in\mathcal{D}_{RA}} \Xi\big(\mathrm{id} \otimes \mathcal{M}(\rho)\|\mathrm{id} \otimes \mathcal{N}(\rho)\big). \tag{77}$$
>
> *Then,*
>
> $$\Omega(\mathcal{M}\|\mathcal{N}) = \Omega(J_\mathcal{M}\|J_\mathcal{N})$$
>
> $$\Xi(\mathcal{M}\|\mathcal{N}) = \Xi(J_\mathcal{M}\|J_\mathcal{N}), \tag{78}$$

where $J_{\mathcal{M}}$ and $J_{\mathcal{N}}$ denote the Choi states of the channels, defined as $J_{\mathcal{M}} := \mathrm{id} \otimes \mathcal{M}(\Phi_+)$.

Proof. The proof mostly follows [22, Lemma 12]. We will consider the Hilbert projective metric Ω, with the case of the Thompson metric Ξ being analogous.

First, notice that $\Omega(\mathcal{M}\|\mathcal{N}) \geq \Omega(J_{\mathcal{M}}\|J_{\mathcal{N}})$ by definition. For the other direction, we first notice that it suffices to restrict the optimisation to pure states $\psi = |\psi\rangle\langle\psi| \in \mathcal{D}_{RA}$ due to the fact that $\Omega(\cdot\|\cdot)$ is jointly quasiconvex [52, Theorem 1], that is,

$$\Omega\left(\sum_i p_i \mathrm{id} \otimes \mathcal{M}(\psi_i) \middle\| \sum_i p_i \mathrm{id} \otimes \mathcal{N}(\psi_i)\right) \leq \max_i \Omega(\mathrm{id} \otimes \mathcal{M}(\psi_i)\|\mathrm{id} \otimes \mathcal{N}(\psi_i)). \tag{79}$$

Consider then the optimal values of λ, μ such that $\lambda\mu = \Omega(J_{\mathcal{M}}\|J_{\mathcal{N}})$ and

$$J_{\mathcal{M}} \leq \lambda J_{\mathcal{N}} \leq \lambda\mu J_{\mathcal{M}}. \tag{80}$$

Let $\psi \in \mathcal{D}_{RA}$ be any pure state. Using the fact that every such state $|\psi\rangle$ can be written as $|\psi\rangle = (X_R \otimes \mathbb{1}_A)|\Phi_+\rangle$ for some operator X_R, we have that

$$\mathrm{id} \otimes \mathcal{M}(|\Phi_+\rangle\langle\Phi_+|) \leq \lambda\, \mathrm{id} \otimes \mathcal{N}(|\Phi_+\rangle\langle\Phi_+|)$$
$$\Rightarrow (X_R \otimes \mathbb{1}_A)\,\mathrm{id} \otimes \mathcal{M}(|\Phi_+\rangle\langle\Phi_+|)(X_R^\dagger \otimes \mathbb{1}_A) \leq \lambda\,(X_R^\dagger \otimes \mathbb{1}_A)\,\mathrm{id} \otimes \mathcal{N}(|\Phi_+\rangle\langle\Phi_+|)(X_R^\dagger \otimes \mathbb{1}_A)$$
$$\Rightarrow \mathrm{id} \otimes \mathcal{M}(X_R \otimes \mathbb{1}_A\,|\Phi_+\rangle\langle\Phi_+|\,X_R^\dagger \otimes \mathbb{1}_A) \leq \lambda\,\mathrm{id} \otimes \mathcal{N}(X_R \otimes \mathbb{1}_A\,|\Phi_+\rangle\langle\Phi_+|\,X_R^\dagger \otimes \mathbb{1}_A)$$
$$\Rightarrow \mathrm{id} \otimes \mathcal{M}(|\psi\rangle\langle\psi|) \leq \lambda\,\mathrm{id} \otimes \mathcal{N}(|\psi\rangle\langle\psi|), \tag{81}$$

and analogously for the inequality $\mathrm{id} \otimes \mathcal{N}(|\Phi_+\rangle\langle\Phi_+|) \leq \mu\,\mathrm{id} \otimes \mathcal{M}(|\Phi_+\rangle\langle\Phi_+|)$. Here, the first implication follows because conjugation by an operator preserves positive semidefiniteness, and the second implication follows because the channels $\mathcal{M}, \mathcal{N}$ act only on the system A. This entails that

$$\Omega(\mathrm{id} \otimes \mathcal{M}(\psi)\|\mathrm{id} \otimes \mathcal{N}(\psi)) \leq \lambda\mu = \Omega(J_{\mathcal{M}}\|J_{\mathcal{N}}), \tag{82}$$

and since this holds for every feasible ψ, it holds also for the supremum. ∎

Lemma 9 then immediately gives the following result, which establishes closed-form expressions for the asymmetric and symmetric error in postselected channel hypothesis testing.

Corollary 10. *For all $\varepsilon \in (0,1)$ and $p, q = 1 - p \in (0,1)$, we have that*

$$\bar{\beta}_\varepsilon(\mathcal{M}, \mathcal{N}) = \left(\frac{\varepsilon}{1-\varepsilon}\,\Omega(J_{\mathcal{M}}\|J_{\mathcal{N}}) + 1\right)^{-1},$$
$$\bar{p}_{\mathrm{err}}(\mathcal{M}, \mathcal{N}\,|\,p, q) = \left(\Xi(pJ_{\mathcal{M}}\|qJ_{\mathcal{N}}) + 1\right)^{-1}. \tag{83}$$

B. Asymptotic channel hypothesis testing

When given n uses of a channel, the basic strategy is the so-called *parallel channel discrimination*, which simply puts the channel to be discriminated in a tensor product $\mathcal{M}^{\otimes n}$ and optimises over

all input states and measurements. Let us then define the n-copy parallel distinguishability errors as

$$
\begin{aligned}
\overline{\beta}_\varepsilon^{\text{parallel}}(\mathcal{M}, \mathcal{N} \mid n) &:= \overline{\beta}_\varepsilon(\mathcal{M}^{\otimes n}, \mathcal{N}^{\otimes n}) \\
&= \inf_{\rho \in \mathcal{D}_{RA^{\otimes n}}} \overline{\beta}_\varepsilon(\text{id} \otimes \mathcal{M}^{\otimes n}(\rho), \text{id} \otimes \mathcal{N}^{\otimes n}(\rho)), \\
\overline{p}_{\text{err}}^{\text{parallel}}(\mathcal{M}, \mathcal{N} \mid p, q \mid n) &:= \overline{p}_{\text{err}}(\mathcal{M}^{\otimes n}, \mathcal{N}^{\otimes n} \mid p, q) \\
&= \inf_{\rho \in \mathcal{D}_{RA^{\otimes n}}} \overline{p}_{\text{err}}(\text{id} \otimes \mathcal{M}^{\otimes n}(\rho), \text{id} \otimes \mathcal{N}^{\otimes n}(\rho) \mid p, q).
\end{aligned}
\tag{84}
$$

where the ancillary space R is *a priori* arbitrary. We stress that, although the channels are in a tensor product as $\mathcal{M}^{\otimes n}$ and $\mathcal{N}^{\otimes n}$, the input states ρ may *a priori* be entangled, and the measurements at the output may be joint over all systems. Exact expressions for these quantities follow immediately from Corollary 10, showing also that it suffices to take the n-copy maximally entangled input state $\Phi_+^{\otimes n}$ in the above. In particular, no entanglement between different copies is required in the preparation of the optimal input states. Furthermore, due to the discussion in Sections II A and III B, we know that also joint measurements are not necessary, and that an asymptotically optimal performance can be achieved by product strategies alone. These two facts, together, would considerably simplify any practical realisation of the above channel discrimination scheme.

To consider more general channel discrimination schemes, most works employ the formalism of *adaptive* discrimination schemes [21, 22, 25, 60–63]. However, as we mentioned previously, it is known that broader types of channel manipulation protocols — in particular, ones that use more exotic transformation structures such as superposition of causal orders — can provide advantages over adaptive protocols [28, 30–32]. We will therefore aim to characterise the most general protocols allowed by the laws of quantum physics, without assuming *anything whatsoever* about their structure.

To discuss a general n-channel manipulation protocol, let us consider an n-linear map Υ that takes n channels as input and outputs a single quantum state. Specifically, using $(A \to B)$ to denote the space of all linear maps from A to B, the map Υ can be considered as a map from the space $(A \to B)^{\times n}$ to $(\mathbb{C} \to R)$, where the latter is just the space of linear operators on some quantum system R of arbitrary dimension. In order for Υ to be a valid channel transformation, we assume that it satisfies

$$
\mathcal{M}_1, \ldots, \mathcal{M}_n \in \text{CP}(A \to B) \implies \Upsilon(\mathcal{M}_1, \ldots, \mathcal{M}_n) \geq 0,
\tag{85}
$$

with $\text{CP}(A \to B)$ denoting the set of all completely positive maps between the two spaces. That is, the sole assumption we make is that the map Υ is positive: it maps any n-tuple of completely positive maps into a positive operator. We do not enforce any form of *complete* positivity of this transformation, nor do we assume normalisation (trace preservation). By construction, any adaptive protocol is included under such a definition. We will use $\mathbb{P}_n$ to denote the set of all n-linear maps $\Upsilon : (A \to B)^{\times n} \to (\mathbb{C} \to R)$ satisfying (85).

We then define

$$
\begin{aligned}
\overline{\beta}_\varepsilon^{\text{any}}(\mathcal{M}, \mathcal{N} \mid n) &:= \inf_{\Upsilon \in \mathbb{P}_n} \overline{\beta}_\varepsilon\big(\Upsilon(\mathcal{M}^{\times n}), \Upsilon(\mathcal{N}^{\times n})\big) \\
\overline{p}_{\text{err}}^{\text{any}}(\mathcal{M}, \mathcal{N} \mid p, q \mid n) &:= \inf_{\Upsilon \in \mathbb{P}_n} \overline{p}_{\text{err}}\big(\Upsilon(\mathcal{M}^{\times n}), \Upsilon(\mathcal{N}^{\times n}) \mid p, q\big),
\end{aligned}
\tag{86}
$$

where $\mathcal{M}^{\times n}$ denotes the n-tuple $(\mathcal{M}, \ldots, \mathcal{M})$. With these definitions in place, we are ready to state our main result on channel discrimination.

Theorem 11. *For all $n \in \mathbb{N}$, we have that*

$$\overline{\beta}_\varepsilon^{\text{any}}(\mathcal{M}, \mathcal{N} \mid n) = \overline{\beta}_\varepsilon^{\text{parallel}}(\mathcal{M}, \mathcal{N} \mid n),$$
$$\overline{p}_{\text{err}}^{\text{any}}(\mathcal{M}, \mathcal{N} \mid p, q \mid n) = \overline{p}_{\text{err}}^{\text{parallel}}(\mathcal{M}, \mathcal{N} \mid p, q \mid n). \tag{87}$$

In particular, the asymptotic exponent of asymmetric postselected channel discrimination is given by the Hilbert projective metric between the Choi operators of the channels, regardless of whether only parallel or any general discrimination schemes are employed: for all $\varepsilon \in (0, 1)$,

$$\lim_{n \to \infty} -\frac{1}{n} \log \overline{\beta}_\varepsilon^{\text{any}}(\mathcal{M}, \mathcal{N} \mid n) = \lim_{n \to \infty} -\frac{1}{n} \log \overline{\beta}_\varepsilon^{\text{parallel}}(\mathcal{M}, \mathcal{N} \mid n)$$
$$= D_\Omega(\mathcal{M} \| \mathcal{N}) \tag{88}$$
$$= D_\Omega(J_\mathcal{M} \| J_\mathcal{N}).$$

Similarly, for all discrimination protocols, the asymptotic exponent of symmetric postselected channel discrimination is given by the Thompson metric between the Choi operators of the channels. That is, for all priors $p, q = 1 - p \in (0, 1)$,

$$\lim_{n \to \infty} -\frac{1}{n} \log \overline{p}_{\text{err}}^{\text{any}}(\mathcal{M}, \mathcal{N} \mid p, q \mid n) = \lim_{n \to \infty} -\frac{1}{n} \log \overline{p}_{\text{err}}^{\text{parallel}}(\mathcal{M}, \mathcal{N} \mid p, q \mid n)$$
$$= D_\Xi(\mathcal{M} \| \mathcal{N}) \tag{89}$$
$$= D_\Xi(J_\mathcal{M} \| J_\mathcal{N}).$$

The result establishes a complete equivalence of parallel discrimination strategies and more general protocols, not only at the asymptotic level but also for any finite number of channel uses. The computation of the asymmetric (Stein) exponent is an analogue of a corresponding result in conventional discrimination of quantum channels [22–24], but strictly stronger: Theorem 11 applies even beyond adaptive discrimination strategies, and furthermore proves the strong converse property of postselected channel discrimination, which is not known to hold in the conventional setting. The evaluation of the symmetric (Chernoff) exponent has no known analogue in conventional hypothesis testing of quantum channels.

Proof. The inclusion of parallel protocols in general channel discrimination protocols implies that

$$\overline{\beta}_\varepsilon^{\text{any}}(\mathcal{M}, \mathcal{N} \mid n) \leq \overline{\beta}_\varepsilon^{\text{parallel}}(\mathcal{M}, \mathcal{N} \mid n) = \left(\frac{\varepsilon}{1 - \varepsilon} \Omega(J_\mathcal{M}^{\otimes n} \| J_\mathcal{N}^{\otimes n}) + 1 \right)^{-1},$$
$$\overline{p}_{\text{err}}^{\text{any}}(\mathcal{M}, \mathcal{N} \mid p, q \mid n) \leq \overline{p}_{\text{err}}^{\text{parallel}}(\mathcal{M}, \mathcal{N} \mid p, q \mid n) = \left(\Xi(p J_\mathcal{M}^{\otimes n} \| q J_\mathcal{N}^{\otimes n}) + 1 \right)^{-1}. \tag{90}$$

The equalities in the above follow by Corollary 10.

For the other direction, we employ the result of [64, Theorem 13], which states that

$$D_{\max} \left(\Upsilon(\mathcal{M}^{\times n}) \| \Upsilon(\mathcal{N}^{\times n}) \right) \leq n \, D_{\max}(J_\mathcal{M} \| J_\mathcal{N}) \tag{91}$$

for every $\Upsilon \in \mathbb{P}_n$. This gives

$$
\begin{aligned}
\overline{\beta}_{\varepsilon}^{\mathrm{any}}(\mathcal{M}, \mathcal{N} \mid n) &= \inf_{\Upsilon \in \mathbb{P}_n} \left(\frac{\varepsilon}{1-\varepsilon} \Omega\left(\Upsilon(\mathcal{M}^{\times n}) \| \Upsilon(\mathcal{N}^{\times n})\right) + 1 \right)^{-1} \\
&\geq \left(\frac{\varepsilon}{1-\varepsilon} \Omega(J_{\mathcal{M}} \| J_{\mathcal{N}})^n + 1 \right)^{-1} \\
&= \left(\frac{\varepsilon}{1-\varepsilon} \Omega(J_{\mathcal{M}}^{\otimes n} \| J_{\mathcal{N}}^{\otimes n}) + 1 \right)^{-1} \\
&= \overline{\beta}_{\varepsilon}^{\mathrm{parallel}}(\mathcal{M}, \mathcal{N} \mid n)
\end{aligned}
\tag{92}
$$

where we used the additivity property

$$
D_{\max}\left(J_{\mathcal{M}^{\otimes n}} \| J_{\mathcal{N}^{\otimes n}}\right) = D_{\max}(J_{\mathcal{M}}^{\otimes n} \| J_{\mathcal{N}}^{\otimes n}) = n\, D_{\max}(J_{\mathcal{M}} \| J_{\mathcal{N}}),
\tag{93}
$$

and analogously

$$
\begin{aligned}
\overline{p}_{\mathrm{err}}^{\mathrm{any}}(\mathcal{M}, \mathcal{N} \mid p, q \mid n) &= \inf_{\Upsilon \in \mathbb{P}_n} \left(\Xi(p\Upsilon(\mathcal{M}^{\times n}) \| q\Upsilon(\mathcal{N}^{\times n})) + 1 \right)^{-1} \\
&\geq \left(\max\left\{ \frac{p}{q} 2^{n D_{\max}(J_{\mathcal{M}} \| J_{\mathcal{N}})}, \frac{q}{p} 2^{n D_{\max}(J_{\mathcal{M}} \| J_{\mathcal{N}})} \right\} + 1 \right)^{-1} \\
&= \left(\Xi(p J_{\mathcal{M}}^{\otimes n} \| q J_{\mathcal{N}}^{\otimes n}) + 1 \right)^{-1} \\
&= \overline{p}_{\mathrm{err}}^{\mathrm{parallel}}(\mathcal{M}, \mathcal{N} \mid p, q \mid n).
\end{aligned}
\tag{94}
$$

The proof is concluded by taking the limits as in the state cases (Sections II and III). ∎

C. Composite channel hypothesis testing

The case of composite channel discrimination can be considered very similarly. Given some subset of channels $\mathcal{F} \subseteq \mathrm{CPTP}(A \to B)$, we define

$$
\overline{\beta}_{\varepsilon, \mathcal{F}}(\mathcal{M}) := \inf_{\rho \in D_{RA}} \inf_{\substack{M_1, M_2 \geq 0 \\ M_1 + M_2 \leq \mathbb{1}}} \left\{ \sup_{\mathcal{N} \in \mathcal{F}} \frac{\operatorname{Tr} M_1(\mathrm{id} \otimes \mathcal{N}(\rho))}{\operatorname{Tr}\left[(M_1 + M_2)(\mathrm{id} \otimes \mathcal{N}(\rho))\right]} \;\middle|\; \frac{\operatorname{Tr} M_2(\mathrm{id} \otimes \mathcal{M}(\rho))}{\operatorname{Tr}\left[(M_1 + M_2)(\mathrm{id} \otimes \mathcal{M}(\rho))\right]} \leq \varepsilon \right\}.
\tag{95}
$$

Using the result for the state case (Theorem 5), we immediately get that

$$
\overline{\beta}_{\varepsilon, \mathcal{F}}(\mathcal{M}) = \left(\frac{\varepsilon}{1-\varepsilon} \Omega_{\mathcal{F}}(\mathcal{M}) + 1 \right)^{-1},
\tag{96}
$$

where

$$
\Omega_{\mathcal{F}}(\mathcal{M}) := \sup_{\rho \in D_{RA}} \inf_{\mathcal{N} \in \mathcal{F}} \Omega(\mathrm{id} \otimes \mathcal{M}(\rho) \| \mathrm{id} \otimes \mathcal{N}(\rho)).
\tag{97}
$$

The following Lemma can then be established in full analogy with Lemma 9.

Lemma 12. *Let* $\mathcal{M} : A \to B$ *be a channel and* $\mathcal{F}$ *a convex and closed set of quantum channels.*

Then

$$\Omega_{\mathcal{F}}(\mathcal{M}) = \min_{\mathcal{N} \in \mathcal{F}} \Omega(J_{\mathcal{M}} \| J_{\mathcal{N}}). \tag{98}$$

As a consequence, we obtain the following generalisation of quantum Stein's lemma for parallel quantum channel discrimination.

Corollary 13. *For all $\varepsilon \in (0,1)$, every channel $\mathcal{M} : A \to B$, and every family $(\mathcal{F}_n)_n$ of convex and closed sets $\mathcal{F}_n \subseteq \mathrm{CPTP}(A^{\otimes n} \to B^{\otimes n})$, we have*

$$\limsup_{n \to \infty} -\frac{1}{n} \log \overline{\beta}_{\varepsilon, \mathcal{F}}(\mathcal{M}^{\otimes n}) = D_{\Omega, \mathcal{F}}^{\infty}(\mathcal{M}) \tag{99}$$

where

$$D_{\Omega, \mathcal{F}}^{\infty}(\mathcal{M}) = \limsup_{n \to \infty} \frac{1}{n} \min_{\mathcal{N} \in \mathcal{F}} D_{\Omega}\left(J_{\mathcal{M}}^{\otimes n} \| J_{\mathcal{N}}^{\otimes n} \right). \tag{100}$$

If the sets $\mathcal{F}_n$ are closed under tensor product, i.e. $\mathcal{N}_n \in \mathcal{F}_n, \mathcal{N}_m \in \mathcal{F}_m \Rightarrow \mathcal{N}_n \otimes \mathcal{N}_m \in \mathcal{F}_{n+m}$, then $\limsup$ *can be replaced with* $\lim$ *in the above.*

V. RESOURCE THEORY OF ASYMMETRIC DISTINGUISHABILITY

The resource theory of asymmetric distinguishability was introduced in [65, 66] as a way to treat the distinguishability of quantum states operationally — just as general resource theories [67] are concerned with transforming quantum states under some set of restricted, 'free' operations, the theory of distinguishability studies the transformations of pairs of quantum states ('boxes') under channels that act as $(\rho, \sigma) \to (\mathcal{M}(\rho), \mathcal{M}(\sigma))$, and therefore do not increase the distinguishability between the two states. The standard resourceful box, representing a unit of 'perfect distinguishability', is represented by the pair $(|0\rangle\langle 0|, \frac{1}{2}\mathbb{1}_2)$. In this context, the task of distinguishability distillation was defined as transforming a given box (ρ, σ) into tensor powers of the standard box, $(|0\rangle\langle 0|^{\otimes m}, \frac{1}{2^m}\mathbb{1}_2^{\otimes m})$. An equivalent variant of this task can be defined by considering the transformation $(\rho, \sigma) \to (|0\rangle\langle 0|, \pi_{2^m})$ instead, where $\pi_k := \frac{1}{k}|0\rangle\langle 0| + \frac{k-1}{k}|1\rangle\langle 1|$, with the advantage that the latter approach allows the parameter m to vary continuously.

A result of [66] was to identify the one-shot distillable distinguishability with the hypothesis testing relative entropy D_H^{ε} [68, 69], which is equal to the optimised type II error in hypothesis testing: $D_H^{\varepsilon}(\rho\|\sigma) := -\log \beta_{\varepsilon}(\rho\|\sigma)$. A natural extension of this result to the asymptotic case where $(\rho^{\otimes n}, \sigma^{\otimes n}) \to (|0\rangle\langle 0|^{\otimes rn}, \frac{1}{2^{rn}}\mathbb{1}_2^{\otimes rn})$, coupled with the quantum Stein's lemma, shows that the value of asymptotically distillable distinguishability can be identified with the relative entropy $D(\rho\|\sigma)$, recovering and extending a finding of [65] (cf. also [70]).

We can obtain an analogous result in our setting, giving an operational interpretation in the resource theory of asymmetric distinguishability to the *postselected hypothesis testing relative entropy*

$$\overline{D}_H^{\varepsilon}(\rho\|\sigma) := -\log \overline{\beta}_{\varepsilon}(\rho\|\sigma). \tag{101}$$

To do so, the transformations under quantum channels need to be replaced by transformations under probabilistic maps.

Corollary 14. *Define the probabilistically distillable distinguishability as*

$$D_{d,\text{prob}}^{\varepsilon}(\rho,\sigma) := \log \sup_{\mathcal{E}\in\text{CPTNI}} \left\{ k \;\middle|\; F\left(\frac{\mathcal{E}(\rho)}{\text{Tr}\,\mathcal{E}(\rho)}, |0\rangle\langle 0|\right) \geq 1 - \varepsilon, \; \frac{\mathcal{E}(\sigma)}{\text{Tr}\,\mathcal{E}(\sigma)} = \pi_k \right\}, \qquad (102)$$

where $F(\omega,\tau) = \left\|\sqrt{\omega}\sqrt{\tau}\right\|_1^2$ is the fidelity and CPTNI *denotes all completely positive and trace–non-increasing maps. Then*

$$D_{d,\text{prob}}^{\varepsilon}(\rho,\sigma) = \overline{D}_H^{\varepsilon}(\rho\|\sigma). \qquad (103)$$

We remark that the fidelity can be replaced with the trace distance $\frac{1}{2}\left\|\frac{\mathcal{E}(\rho)}{\text{Tr}\,\mathcal{E}(\rho)} - |0\rangle\langle 0|\right\|_1$ in the above without affecting the result.

Proof. Applying [52, Corollary 14] gives that

$$D_{d,\text{prob}}^{\varepsilon}(\rho,\sigma) = \log\left(\frac{\varepsilon}{1-\varepsilon}\Omega(\rho\|\sigma) + 1\right). \qquad (104)$$

But in Theorem 1 we have identified the right-hand side with $\overline{D}_H^{\varepsilon}(\rho\|\sigma)$, and so the stated result follows. ∎

An asymptotic extension of this result can be used to show that the rate of distinguishability distillation using probabilistic (CPTNI) operations is given by $D_{\Omega}(\rho\|\sigma)$. This was also recently shown in [51] through a more general approach applicable to broader resource theories.

We remark that the work [66] studied also the reverse task of distinguishability dilution, that is, the transformation from the maximal box to a given pair of quantum states. In this context, a connection with the max-relative entropy $D_{\max}$ was obtained. In an analogous way, in the conceptually related resource theory of *symmetric* distinguishability [53], the corresponding task of dilution was connected with the Thompson metric D_{Ξ}. In contrast, in our probabilistic setting, the task of dilution trivialises — any state can be obtained from a pure state probabilistically [52].

VI. GENERAL PROBABILISTIC THEORIES

The framework of general probabilistic theories (GPTs) [71–73] encompasses both quantum and classical probability theory, providing a natural way to study the properties of very general physical theories, and in particular to formalise the similarities and differences between them. Although concepts such as states and measurements can be defined in any such setting, studying the optimal errors in the asymptotic regime did not appear possible — the optimal error exponents encountered in classical and quantum theory involve quantities such as the quantum relative entropy $D(\cdot\|\cdot)$ and the Chernoff divergence $\xi(\cdot\|\cdot)$, both of which require the properties of classical or quantum theory even to be defined. Because of this, no results whatsoever have been known about asymptotic hypothesis testing in broader GPTs. Our framework, however, can be straightforwardly adapted also to such general theories, and we will see that our findings on state discrimination in Sections II and III hold in arbitrary GPTs verbatim.

A. Definitions

To define a GPT, one first identifies the compact and convex set $\mathcal{S}$ of *states* within a real vector space $\mathcal{V}$, here assumed to be finite dimensional. In analogy with the geometry of quantum theory,

the affine hull of the set $\mathcal{S}$ is required to have co-dimension 1 within $\mathcal{V}$, and to not contain the zero vector.[3] The cone generated by $\mathcal{S}$, namely $C := \{\lambda\rho \mid \lambda \in \mathbb{R}_+, \rho \in \mathcal{S}\}$, is used to induce a partial order on the space $\mathcal{V}$ as $x \leq_C y \iff y - x \in C$. The compactness of $\mathcal{S}$, which implies that C is pointed, or $C \cap (-C) = \{0\}$, indeed means that this is a partial order. Furthermore, the fact that the affine hull of $\mathcal{S}$ is full dimensional entails that C is generating, i.e. $\text{span}(C) = \mathcal{V}$.

The dual space $\mathcal{V}^*$, i.e. the space of continuous linear functionals $M : \mathcal{V} \to \mathbb{R}$, then has a partial order $X \leq_{C^*} Y \iff Y - X \in C^*$ induced by the dual cone $C^* := \{M \in \mathcal{V}^* \mid \langle M, \rho \rangle \geq 0 \; \forall \rho \in \mathcal{S}\}$, where we write $\langle M, x \rangle = M(x)$ for every $x \in \mathcal{V}$, $M \in \mathcal{V}^*$. We use the notation $X <_{C^*} Y$ to denote $Y - X \in \text{int}(C^*)$.

In order to define measurements in this theory, we also need to identify the set of *effects*, that is, elements of the dual space which yield physical measurement outcomes. First, a fixed *unit effect* $U >_{C^*} 0$ is defined as the unique element of $\mathcal{V}^*$ such that $\langle U, \rho \rangle = 1 \; \forall \rho \in \mathcal{S}$. The so-called no restriction hypothesis [71, 75, 76] then allows us to assume that every element $M \in \mathcal{V}^*$ such that $\langle M, \rho \rangle \in [0, 1] \; \forall \rho \in \mathcal{S}$ is a valid effect, and thus a valid measurement operator; this can be equivalently understood as $0 \leq_{C^*} M \leq_{C^*} U$. A measurement is then a collection of effects $\{M_i\}_i$ such that $\sum_{i=1}^n M_i = U$ — in other words, a collection such that the measurement outcomes $\langle M_i, \rho \rangle$ sum to unity.

We will use the tuple $(\mathcal{V}, C, U)$ to denote a GPT as above, as the theory is uniquely determined by the three choices. Given two GPTs, $(\mathcal{V}_A, C_A, U_A)$ and $(\mathcal{V}_B, C_B, U_B)$, we then wish to define a bipartite structure that combines the two. Under natural assumptions including the so-called local tomography principle, which states that bipartite states should be fully determined by the statistics of local measurements, it holds that $\mathcal{V}_{AB}$ is isomorphic to $\mathcal{V}_A \otimes \mathcal{V}_B$ [77, 78], which means that we can identify $\mathcal{V}_{AB} = \mathcal{V}_A \otimes \mathcal{V}_B$ and $U_{AB} = U_A \otimes U_B$ without loss of generality. Interestingly, this does *not* uniquely determine a bipartite cone C_{AB}; the only thing that can be concluded is that [79]

$$C_A \underset{\min}{\otimes} C_B \subseteq C_{AB} \subseteq C_A \underset{\max}{\otimes} C_B \tag{105}$$

where the *minimal tensor product* and the *maximal tensor product* are defined, respectively, as

$$
\begin{aligned}
C_A \underset{\min}{\otimes} C_B &:= \text{conv}(C_A \otimes C_B), \\
C_A \underset{\max}{\otimes} C_B &:= \left(C_A^* \underset{\min}{\otimes} C_B^*\right)^*.
\end{aligned}
\tag{106}
$$

For every pair of non-classical GPTs, it holds that $C_A \underset{\min}{\otimes} C_B \neq C_A \underset{\max}{\otimes} C_B$ [80], and the difference between these two cones is paramount to understanding properties of multipartite GPTs, in particular the existence of phenomena such as entanglement [79, 81].

Despite the fact that many aspects of GPTs crucially depend on the specific choice of a multipartite cone C_{AB}, our results will apply to *any* choice of C_{AB} as in (105), making them universally applicable. We remark in particular that $C_A \otimes C_B \subseteq C_{AB}$ and $C_A^* \otimes C_B^* \subseteq C_{AB}^*$ for all valid choices of C_{AB}.

The extension of the above to the tensor product of more than two GPTs is immediate.

---

[3] This seemingly ad hoc set of axioms is actually well justified from a foundational perspective. We refer the interested reader to [74, Chapters 1–2] for more details on these points as well as complete derivations.

B. Extension of results

The task of state discrimination, in both the conventional and postselected settings, can be formulated in exactly the same way as we have done it in Section I; specifically,

$$
\begin{aligned}
\overline{\beta}_{\varepsilon}(\rho, \sigma) &= \inf_{M_1, M_2 \geq_{C^*} 0} \left\{ \frac{\langle M_1, \sigma \rangle}{\langle M_1 + M_2, \sigma \rangle} \,\middle|\, M_1 + M_2 \leq_{C^*} U, \ \frac{\langle M_2, \rho \rangle}{\langle M_1 + M_2, \rho \rangle} \leq \varepsilon \right\} \\
\overline{p}_{\mathrm{err}}(\rho, \sigma \,|\, p, q) &= \inf_{\substack{M_1, M_2 \geq_{C^*} 0, \\ M_1 + M_2 \leq_{C^*} U}} \frac{p \langle M_2, \rho \rangle + q \langle M_1, \sigma \rangle}{p \langle M_1 + M_2, \rho \rangle + q \langle M_1 + M_2, \sigma \rangle},
\end{aligned}
\tag{107}
$$

where $\rho, \sigma \in \mathcal{S}$ and $p, q = 1 - p \in (0, 1)$ as before. The equivalent of the max-relative entropy can also be defined in the same way,

$$
D_{\max}(\rho\|\sigma) := \log \inf \left\{ \lambda \in \mathbb{R}_+ \,\middle|\, \rho \leq_C \lambda \sigma \right\},
\tag{108}
$$

with the definitions of D_Ω and D_Ξ naturally extending to the GPT formalism using the above quantity.

The results of our Theorem 1 (asymmetric error) and Theorem 6 (symmetric error) then apply to any GPT *exactly as stated* under the replacement $\mathrm{Tr}(M\cdot) \mapsto \langle M, \cdot \rangle$ — there is no need to adjust any of the proofs, as we have not made use of any properties of quantum mechanics, with our proofs resting on the linearity of the trace and convex (Lagrange) duality, both of which hold for arbitrary GPTs defined as in Section VI A.

This correspondence is not too surprising, given that many results on one-shot state discrimination can be mapped between quantum mechanics and GPTs [82, 83]. What is more remarkable about the setting of postselected hypothesis testing is that the asymptotic results — namely, Corollary 2 and Corollary 7 — are also valid in arbitrary GPTs, with no other assumptions needed. This is because of the easily verified additivity of $D_{\max}$, which we formalise below.

Lemma 15. *Given a local GPT $(\mathcal{V}, C, U)$, consider any n-copy GPT $(\mathcal{V}_n, C_n, U_n)$, where $\mathcal{V}_n = \bigotimes_{i=1}^n \mathcal{V}$, $U_n = \bigotimes_{i=1}^n U$, and $\bigotimes_{i=1}^n{}_{\min} C \subseteq C_n \subseteq \bigotimes_{i=1}^n{}_{\max} C$. Then, for all $\rho, \sigma \in \mathcal{S}$,*

$$
D_{\max}(\rho^{\otimes n} \| \sigma^{\otimes n}) = n D_{\max}(\rho\|\sigma).
\tag{109}
$$

Proof. Consider any feasible λ such that $\rho \leq_C \lambda \sigma$. Then, $\rho^{\otimes n} \leq_{C_n} (\lambda \sigma)^{\otimes n} = \lambda^n \sigma^{\otimes n}$; this can be seen by using the fact that $C \underset{\min}{\otimes} C \subseteq C_2$ to show that

$$
\lambda^2(\sigma \otimes \sigma) - \rho \otimes \rho = (\lambda \sigma - \rho) \otimes \lambda \sigma + \rho \otimes (\lambda \sigma - \rho) \geq_{C_2} 0
\tag{110}
$$

and extending to arbitrary n by induction. This implies that $D_{\max}(\rho^{\otimes n}) \leq n \log \lambda$, and hence $D_{\max}(\rho^{\otimes n} \| \sigma^{\otimes n}) \leq n D_{\max}(\rho\|\sigma)$.

For the other direction, we write $D_{\max}$ in its dual form as

$$
D_{\max}(\rho\|\sigma) = \log \sup_{W \geq_{C^*} 0} \frac{\langle W, \rho \rangle}{\langle W, \sigma \rangle}.
\tag{111}
$$

The fact that $\otimes_{i=1}^n C^* \subseteq C_n^*$ immediately implies that any feasible solution W to $D_{\max}(\rho\|\sigma)$ gives a feasible solution to $D_{\max}(\rho^{\otimes n} \| \sigma^{\otimes n})$ as $W^{\otimes n}$, concluding the proof. ∎

The above lemma tells us that $D_\Omega(\rho^{\otimes n}\|\sigma^{\otimes n}) = nD_\Omega(\rho\|\sigma)$ and $D_\Xi(\rho^{\otimes n}\|\sigma^{\otimes n}) = nD_\Xi(\rho\|\sigma)$ for all normalised $\rho, \sigma \in \mathcal{S}$. The asymptotic results of Corollaries 2 and 7 thus follow directly; namely, we have that

$$\lim_{n\to\infty} -\frac{1}{n}\log\overline{\beta}_\varepsilon(\rho^{\otimes n}, \sigma^{\otimes n}) = D_\Omega(\rho\|\sigma) \tag{112}$$

and

$$\lim_{n\to\infty} -\frac{1}{n}\log\overline{p}_{\mathrm{err}}(\rho^{\otimes n}, \sigma^{\otimes n} \,|\, p, q) = D_\Xi(\rho\|\sigma) \tag{113}$$

for all $\rho, \sigma \in \mathcal{S}$, all $\varepsilon \in (0, 1)$, and all $p, q = 1 - p \in (0, 1)$.

Theorem 5, which establishes a postselected generalisation of the quantum Stein's lemma in the composite setting, also immediately holds in every GPT. The proof is, once again, noticed not to use any assumptions save for linearity and convex duality.

We thus see that virtually all of our results on state discrimination immediately extend to GPTs, without requiring any restrictive assumptions on the latter. On the other hand, although we expect at least some of our channel discrimination results to also be generalisable to GPTs under some reasonable assumptions, we should note that this is far from immediate — even defining a 'channel' is no trivial task in a framework so general, and tools such as the Choi–Jamiołkowski isomorphism might no longer be available to us. We thus leave the question of investigating channel discrimination in this setting as an open problem.

VII. DISCUSSION

The approach of our work is inherently different from conventional quantum hypothesis testing, yet we hope that it can ultimately prove helpful in shedding light also on the latter. The distinguishing feature of our results, and perhaps an unexpected finding, is how significantly the whole framework simplifies in the postselected setting: the optimal probabilities can be given straightforward closed-form expressions; the case of composite hypothesis testing reduces to a simple extension of the i.i.d. case; all terms but those linear in n disappear from the asymptotic expansions; and, importantly, the task of channel discrimination not only admits efficiently computable error probabilities and exponents, but in fact reduces all discrimination strategies to the much simpler case of parallel protocols. This allowed us to essentially characterise completely the task of postselected quantum hypothesis testing with just a few concise proofs — something that, in the conventional setting, has been a multi-decade endeavour requiring significant effort and the development of highly specialised methods.

An underlying property of our framework is the need to condition on a conclusive outcome of the measurement. This is an easily realisable operation: simply repeat the measurement protocol, disregarding measurement data corresponding to inconclusive outcomes. It is, however, conceivable that the conclusive outcome may have a low probability of occurrence, meaning that many repetitions of the measurement protocol would be needed to successfully discriminate the states or channels. Such a difference indeed makes direct comparisons between n-copy distinguishability protocols in the postselected and conventional settings difficult to perform. However, as we argued in [51], allowing postselection is much more easily justified in the limit of infinitely many i.i.d. copies, where the definition of a rate may be adjusted depending on what 'cost function' we wish to employ: if we are interested in error rates per number of copies of states *generated*, then the conventional setting appears more appropriate; if, however, we instead focus on how many copies of states need to be *manipulated* at once, then any protocol may be repeated without incurring an

additional cost, and a postselected setting is thus perfectly natural. The latter assumption is not necessarily an extravagant one — generating many copies of states is far easier than coherently manipulating them with current technologies.

The idea of postselection has previously appeared in a range of contexts in quantum information, including — besides the aforementioned state discrimination [33–35] —- quantum computing [84], metrology [85–89], connections with spacetime geometry [90–92], and manipulation of quantum entanglement [93–95] as well as more general quantum resources [49, 51, 52]. In such settings, our results can be viewed as ultimate limits that cannot be exceeded even if postselection is permitted: the optimal discrimination error bounds shown in our work cannot be beat by any measurement strategy, postselected or not, even in permissive frameworks such as ones allowing the increased distinguishing power of postselected closed timelike curves [90, 92].

Ultimately, although it might be contentious to claim that it has directly led to practical advancements, postselection has unquestionably contributed to the understanding of the foundations of quantum mechanics [96–98], and even formed the conceptual basis of the first proposals for quantum supremacy experiments [99]. In a similar manner, we hope that our results can find use both in the formalisation of the foundations of quantum information, as well as in the study of the limits of practical state and channel discrimination protocols.

[1] M. Hayashi, *Quantum Information Theory: Mathematical Foundation* (Springer, 2016).

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

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

[4] C. W. Helstrom, *Quantum detection and estimation theory*, J. Stat. Phys. **1**, 231 (1969).

[5] A. S. Holevo, *An analogue of statistical decision theory and noncommutative probability theory*, Trudy Moskov. Mat. Obsc. **26**, 133 (1972).

[6] H. Umegaki, *Conditional expectation in an operator algebra, IV (Entropy and information)*, Kodai Math. Sem. Rep. **14**, 59 (1962).

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

[8] T. Ogawa and H. Nagaoka, *Strong converse and Stein's lemma in quantum hypothesis testing*, IEEE Trans. Inf. Theory **46**, 2428 (2000).

[9] C. Stein, *Information and Comparison of Experiments* (unpublished), Charles Stein papers (SC1224). Box 12, Folder 7, Department of Special Collections and University Archives, Stanford University Libraries.

[10] H. Chernoff, *Large-Sample Theory: Parametric Case*, Ann. Math. Stat. **27**, 1 (1956).

[11] S. Kullback and R. A. Leibler, *On Information and Sufficiency*, Ann. Math. Stat. **22**, 79 (1951).

[12] H. Chernoff, *A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the sum of Observations*, Ann. Math. Stat. **23**, 493 (1952).

[13] M. Nussbaum and A. Szkoła, *The Chernoff lower bound for symmetric quantum hypothesis testing*, Ann. Stat. **37**, 1040 (2009).

[14] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, L. Masanes, A. Acin, and F. Verstraete, *Discriminating States: The Quantum Chernoff Bound*, Phys. Rev. Lett. **98**, 160501 (2007).

[15] F. G. S. L. Brandão and M. B. Plenio, *A Generalization of Quantum Stein's Lemma*, Commun. Math. Phys. **295**, 791 (2010).

[16] M. Berta, F. G. S. L. Brandão, G. Gour, L. Lami, M. B. Plenio, B. Regula, and M. Tomamichel, *On a gap in the proof of the generalised quantum Stein's lemma and its consequences for the reversibility of quantum resources*, arXiv:2205.02813 (2022).

[17] M. Tomamichel and M. Hayashi, *Operational Interpretation of Rényi Information Measures via Composite Hypothesis Testing Against Product and Markov Distributions*, IEEE Trans. Inf. Theory **64**, 1064 (2018).

[18] M. Berta, F. G. S. L. Brandão, and C. Hirche, *On Composite Quantum Hypothesis Testing*, Commun. Math. Phys. **385**, 55 (2021).

[19] F. G. S. L. Brandão, A. W. Harrow, J. R. Lee, and Y. Peres, *Adversarial Hypothesis Testing and a Quantum Stein's Lemma for Restricted Measurements*, IEEE Trans. Inf. Theory **66**, 5037 (2020).

[20] M. Mosonyi, Z. Szilágyi, and M. Weiner, *On the error exponents of binary state discrimination with composite hypotheses*, arXiv:2011.04645 (2021).

[21] G. Chiribella, G. M. D'Ariano, and P. Perinotti, *Quantum Circuit Architecture*, Phys. Rev. Lett. **101**, 060401 (2008).

[22] M. M. Wilde, M. Berta, C. Hirche, and E. Kaur, *Amortized channel divergence for asymptotic quantum channel discrimination*, Lett. Math. Phys. **110**, 2277 (2020).

[23] X. Wang and M. M. Wilde, *Resource theory of asymmetric distinguishability for quantum channels*, Phys. Rev. Research **1**, 033169 (2019).

[24] K. Fang, O. Fawzi, R. Renner, and D. Sutter, *Chain Rule for the Quantum Relative Entropy*, Phys. Rev. Lett. **124**, 100501 (2020).

[25] A. W. Harrow, A. Hassidim, D. W. Leung, and J. Watrous, *Adaptive versus nonadaptive strategies for quantum channel discrimination*, Phys. Rev. A **81**, 032339 (2010).

[26] F. Salek, M. Hayashi, and A. Winter, *Usefulness of adaptive strategies in asymptotic quantum channel discrimination*, Phys. Rev. A **105**, 022419 (2022).

[27] N. Yu and L. Zhou, *When is the Chernoff Exponent for Quantum Operations Finite?* IEEE Trans. Inf. Theory **67**, 4517 (2021).

[28] G. Chiribella, G. M. D'Ariano, P. Perinotti, and B. Valiron, *Quantum computations without definite causal structure*, Phys. Rev. A **88**, 022318 (2013).

[29] O. Oreshkov, F. Costa, and Č. Brukner, *Quantum correlations with no causal order*, Nat. Commun. **3**, 1092 (2012).

[30] D. Ebler, S. Salek, and G. Chiribella, *Enhanced Communication with the Assistance of Indefinite Causal Order*, Phys. Rev. Lett. **120**, 120502 (2018).

[31] M. T. Quintino, Q. Dong, A. Shimbo, A. Soeda, and M. Murao, *Reversing Unknown Quantum Transformations: Universal Quantum Circuit for Inverting General Unitary Operations*, Phys. Rev. Lett. **123**, 210502 (2019).

[32] J. Bavaresco, M. Murao, and M. T. Quintino, *Strict Hierarchy between Parallel, Sequential, and Indefinite-Causal-Order Strategies for Channel Discrimination*, Phys. Rev. Lett. **127**, 200504 (2021).

[33] I. D. Ivanovic, *How to differentiate between non-orthogonal states*, Phys. Lett. A **123**, 257 (1987).

[34] D. Dieks, *Overlap and distinguishability of quantum states*, Phys. Lett. A **126**, 303 (1988).

[35] A. Peres, *How to differentiate between non-orthogonal states*, Phys. Lett. A **128**, 19 (1988).

[36] A. Chefles and S. M. Barnett, *Strategies for discriminating between non-orthogonal quantum states*, J. Mod. Opt. **45**, 1295 (1998).

[37] J. Fiurášek and M. Ježek, *Optimal discrimination of mixed quantum states involving inconclusive results*, Phys. Rev. A **67**, 012321 (2003).

[38] T. Rudolph, R. W. Spekkens, and P. S. Turner, *Unambiguous discrimination of mixed states*, Phys. Rev. A **68**, 010301 (2003).

[39] S. Croke, E. Andersson, S. M. Barnett, C. R. Gilson, and J. Jeffers, *Maximum Confidence Quantum Measurements*, Phys. Rev. Lett. **96**, 070401 (2006).

[40] U. Herzog, *Discrimination of two mixed quantum states with maximum confidence and minimum probability of inconclusive results*, Phys. Rev. A **79**, 032323 (2009).

[41] E. Bagan, R. Muñoz-Tapia, G. A. Olivares-Rentería, and J. A. Bergou, *Optimal discrimination of quantum states with a fixed rate of inconclusive outcomes*, Phys. Rev. A **86**, 040303 (2012).

[42] Q. Zhuang, *Ultimate limits of approximate unambiguous discrimination*, Phys. Rev. Research **2**, 043276 (2020).

[43] S. M. Barnett and S. Croke, *Quantum state discrimination*, Adv. Opt. Photon., AOP **1**, 238 (2009).

[44] J. Bae and L.-C. Kwek, *Quantum state discrimination and its applications*, J. Phys. A: Math. Theor. **48**, 083001 (2015).

[45] N. Datta, *Min- and Max-Relative Entropies and a New Entanglement Monotone*, IEEE Trans. Inf. Theory **55**, 2816 (2009).

[46] P. J. Bushell, *Hilbert's metric and positive contraction mappings in a Banach space*, Arch. Rat. Mech. Anal. **52**, 330 (1973).

[47] A. C. Thompson, *On certain contraction mappings in a partially ordered vector space.* Proc. Amer. Math. Soc. **14**, 438 (1963).

[48] R. D. Nussbaum and C. Walsh, *A metric inequality for the Thompson and Hilbert geometries.* J. Inequal. Pure Appl. Math. **5**, 54 (2004).

[49] D. Reeb, M. J. Kastoryano, and M. M. Wolf, *Hilbert's projective metric in quantum information theory*, J. Math. Phys. **52**, 082201 (2011).

[50] B. Regula, *Probabilistic Transformations of Quantum Resources*, Phys. Rev. Lett. **128**, 110505 (2022).

[51] B. Regula, L. Lami, and M. M. Wilde, *Overcoming entropic limitations on asymptotic state transformations through probabilistic protocols*, arXiv:2209.03362 (2022).

[52] B. Regula, *Tight constraints on probabilistic convertibility of quantum states*, Quantum **6**, 817 (2022).

[53] R. Salzmann, N. Datta, G. Gour, X. Wang, and M. M. Wilde, *Symmetric distinguishability as a quantum resource*, New J. Phys. **23**, 083016 (2021).

[54] K. Li, *Second-order asymptotics for quantum hypothesis testing*, Ann. Stat. **42**, 171 (2014).

[55] 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).

[56] M. Fekete, *Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten*, Math Z **17**, 228 (1923).

[57] F. G. S. L. Brandão and M. B. Plenio, *A Reversible Theory of Entanglement and its Relation to the Second Law*, Commun. Math. Phys. **295**, 829 (2010).

[58] R. T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, 1970).

[59] A. Y. Kitaev, *Quantum computations: Algorithms and error correction*, Russ. Math. Surv. **52**, 1191 (1997).

[60] M. Hayashi, *Discrimination of Two Channels by Adaptive Methods and Its Application to Quantum System*, IEEE Trans. Inf. Theory **55**, 3807 (2009).

[61] R. Duan, Y. Feng, and M. Ying, *Perfect Distinguishability of Quantum Operations*, Phys. Rev. Lett. **103**, 210501 (2009).

[62] T. Cooney, M. Mosonyi, and M. M. Wilde, *Strong Converse Exponents for a Quantum Channel Discrimination Problem and Quantum-Feedback-Assisted Communication*, Commun. Math. Phys. **344**, 797 (2016).

[63] S. Pirandola, R. Laurenza, C. Lupo, and J. L. Pereira, *Fundamental limits to quantum channel discrimination*, npj Quantum Information **5**, 1 (2019).

[64] B. Regula and R. Takagi, *Fundamental limitations on distillation of quantum channel resources*, Nat. Commun. **12**, 4411 (2021).

[65] K. Matsumoto, *Reverse Test and Characterization of Quantum Relative Entropy*, arXiv:1010.1030 (2010).

[66] X. Wang and M. M. Wilde, *Resource theory of asymmetric distinguishability*, Phys. Rev. Research **1**, 033170 (2019).

[67] E. Chitambar and G. Gour, *Quantum resource theories*, Rev. Mod. Phys. **91**, 025001 (2019).

[68] L. Wang and R. Renner, *One-Shot Classical-Quantum Capacity and Hypothesis Testing*, Phys. Rev. Lett. **108**, 200501 (2012).

[69] F. Buscemi and N. Datta, *The Quantum Capacity of Channels With Arbitrarily Correlated Noise*, IEEE Trans. Inf. Theory **56**, 1447 (2010).

[70] F. Buscemi, D. Sutter, and M. Tomamichel, *An information-theoretic treatment of quantum dichotomies*, Quantum **3**, 209 (2019).

[71] G. Ludwig, *An Axiomatic Basis for Quantum Mechanics: Volume 1 Derivation of Hilbert Space Structure* (Springer-Verlag, Berlin Heidelberg, 1985).

[72] A. Hartkämper and H. Neumann, eds., *Foundations of Quantum Mechanics and Ordered Linear Spaces* (Springer, 1974).

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

[74] L. Lami, *Non-Classical Correlations in Quantum Mechanics and Beyond*, Ph.D. thesis, Universitat Autònoma de Barcelona (2017), arXiv:1803.02902.

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

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

[77] M. Kläy, C. Randall, and D. Foulis, *Tensor products and probability weights*, Int. J. Theor. Phys. **26**, 199 (1987).

[78] A. Wilce, *Tensor products in generalized measure theory*, Int. J. Theor. Phys. **31**, 1915 (1992).

[79] I. Namioka and R. R. Phelps, *Tensor products of compact convex sets*. Pac. J. Math. **31**, 469 (1969).

[80] G. Aubrun, L. Lami, C. Palazuelos, and M. Plávala, *Entangleability of cones*, Geom. Funct. Anal. **31**, 181 (2021).

[81] G. Aubrun, L. Lami, C. Palazuelos, and M. Plávala, *Entanglement and Superposition Are Equivalent Concepts in Any Physical Theory*, Phys. Rev. Lett. **128**, 160402 (2022).

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

[83] J. Bae, D.-G. Kim, and L.-C. Kwek, *Structure of Optimal State Discrimination in Generalized Probabilistic Theories*, Entropy **18**, 39 (2016).

[84] S. Aaronson, *Quantum computing, postselection, and probabilistic polynomial-time*, Proc. R. Soc. A **461**, 3473 (2005).

[85] J. Fiurášek, *Optimal probabilistic estimation of quantum states*, New J. Phys. **8**, 192 (2006).

[86] B. Gendra, E. Ronco-Bonvehi, J. Calsamiglia, R. Muñoz-Tapia, and E. Bagan, *Beating noise with abstention in state estimation*, New J. Phys. **14**, 105015 (2012).

[87] J. Combes, C. Ferrie, Z. Jiang, and C. M. Caves, *Quantum limits on postselected, probabilistic quantum metrology*, Phys. Rev. A **89**, 052117 (2014).

[88] J. Combes and C. Ferrie, *Cost of postselection in decision theory*, Phys. Rev. A **92**, 022117 (2015).

[89] D. R. M. Arvidsson-Shukur, N. Yunger Halpern, H. V. Lepage, A. A. Lasek, C. H. W. Barnes, and S. Lloyd, *Quantum advantage in postselected metrology*, Nat. Commun. **11**, 3775 (2020).

[90] S. Lloyd, L. Maccone, R. Garcia-Patron, V. Giovannetti, Y. Shikano, S. Pirandola, L. A. Rozema, A. Darabi, Y. Soudagar, L. K. Shalm, and A. M. Steinberg, *Closed Timelike Curves via Postselection: Theory and Experimental Test of Consistency*, Phys. Rev. Lett. **106**, 040403 (2011).

[91] S. Lloyd, L. Maccone, R. Garcia-Patron, V. Giovannetti, and Y. Shikano, *Quantum mechanics of time travel through post-selected teleportation*, Phys. Rev. D **84**, 025007 (2011).

[92] T. A. Brun and M. M. Wilde, *Perfect State Distinguishability and Computational Speedups with Postselected Closed Timelike Curves*, Found. Phys. **42**, 341 (2012).

[93] N. Gisin, *Hidden quantum nonlocality revealed by local filters*, Phys. Lett. A **210**, 151 (1996).

[94] A. Kent, *Entangled Mixed States and Local Purification*, Phys. Rev. Lett. **81**, 2839 (1998).

[95] M. Horodecki, P. Horodecki, and R. Horodecki, *General teleportation channel, singlet fraction, and quasidistillation*, Phys. Rev. A **60**, 1888 (1999).

[96] Y. Aharonov, D. Z. Albert, and L. Vaidman, *How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100*, Phys. Rev. Lett. **60**, 1351 (1988).

[97] M. F. Pusey, *Anomalous Weak Values Are Proofs of Contextuality*, Phys. Rev. Lett. **113**, 200401 (2014).

[98] O. Oreshkov and N. J. Cerf, *Operational formulation of time reversal in quantum theory*, Nat. Phys. **11**, 853 (2015).

[99] A. W. Harrow and A. Montanaro, *Quantum computational supremacy*, Nature **549**, 203 (2017).