IEEE Transactions on Information Theory

Exploring Quantum Average-Case Distances: Proofs, properties, and examples

Journal:	*IEEE Transactions on Information Theory*
Manuscript ID	IT-22-0500.R1
Manuscript Type:	Regular Manuscript
Date Submitted by the Author:	29-Dec-2022
Complete List of Authors:	Maciejewski, Filip; Polish Academy of Sciences, Center for Theoretical Physics Puchała, Zbigniew; Polish Academy of Sciences Institute of Theoretical and Applied Informatics; Jagiellonian University, Physics Oszmaniec, Michał; Polish Academy of Sciences, Center for Theoretical Physics
Subject Category:	Quantum
Keywords:	measures of distance in quantum information, statistical distinguishability protocols, random quantum circuits, NISQ devices, quantum advantage sampling

Exploring Quantum Average-Case Distances: Proofs, properties, and examples

Filip B. Maciejewski,[1] Zbigniew Puchała,[2,3] and Michał Oszmaniec[1]

[1]*Center for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warszawa, Poland**

[2]*Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, 44-100 Gliwice, Poland*

[3]*Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, 30-348 Kraków, Poland*

In this work, we present an in-depth study of average-case quantum distances introduced in [1]. The average-case distances approximate, up to the relative error, the average Total-Variation (TV) distance between measurement outputs of two quantum processes, in which quantum objects of interest (states, measurements, or channels) are intertwined with random quantum circuits. Contrary to conventional distances, such as trace distance or diamond norm, they quantify *average-case* statistical distinguishability via random quantum circuits.

We prove that once a family of random circuits forms an δ-approximate 4-design, with $\delta = o(d^{-8})$, then the average-case distances can be approximated by simple explicit functions that can be expressed via simple degree two polynomials in objects of interest. For systems of moderate dimension, they can be easily explicitly computed – no optimization is needed as opposed to diamond norm distance between channels or operational distance between measurements. We prove that those functions, which we call quantum average-case distances, have a plethora of desirable properties, such as subadditivity w.r.t. tensor products, joint convexity, and (restricted) data-processing inequalities. Notably, all distances utilize the Hilbert-Schmidt (HS) norm, which provides this norm with a new operational interpretation. We also provide upper bounds on the maximal ratio between worst-case and average-case distances, and for each of them, we provide an example that saturates the bound. Specifically, we show that for each dimension d this ratio is at most $d^{\frac{1}{2}}$, d, $d^{\frac{3}{2}}$ for states, measurements, and channels, respectively. To support the practical usefulness of our findings, we study multiple examples in which average-case quantum distances can be calculated analytically.

I. INTRODUCTION

Motivation

The question of how far away are two quantum objects (states, measurements, or channels) is of both fundamental and practical importance. That question is often phrased in terms of the statistical distinguishability of probability distributions corresponding to two objects in question (which is a problem of classical hypothesis testing [2]). Indeed, the most common distances, such as trace distance or diamond norm distance, are based on optimal protocols for such statistical discrimination. However, those protocols have limitations. In general, they might require a lot of resources (e.g., high-depth circuits) [3], thus they are not necessarily practical. Another perspective on limitations of common distance measures comes from study of noise on quantum devices. Specifically, a distance between a theoretical (ideal) model of an object in question, and a model for its experimental (noisy) implementation, can be used to study the potential effects of experimental imperfections on the protocol one wishes to implement. When that is the case, the distances based on optimal state/measurement/channel discrimination in fact inform about the worst-case performance of a protocol. However, in practice, one does not necessarily expect the worst-case to be representative of a typical device's performance (an in-depth study of the effects of noise on protocols involving random circuits that exploits average-case distances is presented in an accompanying manuscript [1]).

With this motivation in mind, we propose distance measures of distance based on *average statistical distinguishability* using random quantum circuits. Operationally, if the average-case distance between a pair of quantum objects is significant, this implies that they can be (statistically) distinguished almost perfectly using just a few implementations of random circuits. This provides a natural interpretation analogous to conventional distances, but we consider averages over random circuits instead of optimal scenarios. [removed two paragraphs] Such quantifiers can be more suitable for studying the performance of NISQ devices' performance than the above-mentioned conventional distances quantifying worst-case performance. In particular, one of the most promising near-term applications of quantum computing are hybrid quantum-classical variational algorithms [4], such as Quantum Approximate Optimization Algorithm (QAOA) [5–7] and Variational Quantum Eigensolver (VQE) [8–10]. Since NISQ devices are expected to suffer from a significant amount of noise, it is instrumental to understand how it can affect such algorithms (see, e.g., [11–14]). Our distance measures might prove particularly useful in this context because, as explained later, the random circuits we consider form unitary designs. Recently it was realized that circuits appearing in variational algorithms are expected to have, on average,

* This paper was presented in part at Algorithmiq Ltd (Finland, 2022), and QIP (US, 2022), and APS Physics March Meeting (US, 2022), and Workshop on Quantum t-Designs and Applications in Quantum Computing (Scotland, 2022), and Entanglement in Action workshop (Spain, 2022), and QPL (UK, 2022).

design-like properties. Thus we expect average-case quantum distances to be a good metric to quantify the average performance of such algorithms [15].

The manuscript is accompanied by a shorter paper that summarizes all the main results and contains discussion of practical applications of average-case distances, as well as exhaustive numerical studies (involving random circuits with QAOA-like structure) [1].

Summary of results

In this work, we study the average Total-Variation (TV) distance between measurement outputs (statistics) of two quantum processes, in which quantum objects of interest are intertwined with random quantum circuits. TV distance is well known to quantify the statistical distinguishability of two probability distributions. In general, as TV distance is not a polynomial function of underlying probability distributions, the relevant averages are hard to calculate. However, we derive lower and upper bounds for average TV distance and show that both bounds differ only by dimension-independent *constants*. The derivation of upper bounds requires the calculation of 2nd moments of quantities of interest, and lower bounds are derived using 2nd and 4th moments. Formally, this means that to get both upper and lower bounds, the random circuits must form an approximate 4-design. Importantly, our results are valid also for any (approximate) k-design with $k \geq 4$. The particular choice of 4-designs is of purely technical origin – as remarked above, our proof techniques require 4th degree polynomials to get lower bounds on average TV distance, while for upper bounds already 2-designs suffice. The above implies that for a broad family of random quantum circuits, the average TV distance is *approximated*, up to the known relative error, by a simple explicit function of the objects that we wish to compare (states, measurements, or channels). These functions, which we call average-case quantum distances, define bona fide distance measures with multiple desired properties, such as subadditivity w.r.p. to tensor products, joint convexity, or (restricted) data processing inequalities. Importantly, all of the proposed distances (between states, measurements, and channels) can be expressed via simple degree two polynomials in objects in question and can be easily explicitly computed for systems of moderate dimension. No optimization is needed as opposed to diamond norm distance between channels [16] or operational distance between measurements [17]. Notably, all of the distances utilize the Hilbert-Schmidt (HS) norm in some way. This gives the HS norm an operational interpretation that it did not possess before (especially for quantum states for which average-case distance is proportional to HS distance).

Finally, so-defined average-case quantum distances have sound operational interpretation. Namely, if a TV distance is bounded from below by a constant c (here proportional to average-case quantum distance), then there exists a strategy that uses random circuits which distinguishes between two objects with probability at least $\frac{1}{2}(1+c)$ in *single-shot* scenario. Thus from Hoeffding bound, it follows that having access to multiple copies (samples) allows to exponentially quickly approach success probability of discrimination equal to 1 using simple majority vote.

Related works

Let us now comment on some of the commonly used distances. The study of similarity measures between quantum objects has a long history [2, 18], and thus there are a lot of different metrics currently used in the field. Some of the most popular distances are based on the *optimal* statistical distinguishability of quantum objects – this includes trace distance between states [19], the operational distance between measurements [17], as well as diamond norm distance between channels [19]. While in those distances the optimization is done over all possible operations, there has been an interest also in distinguishability under restricted sets of operations – such as local POVMs for discrimination of quantum states [20, 21]. Recently, a quantum Wasserstein distance of order 1 was proposed as a measure of distance between quantum states. It generalizes a classical Wasserstein distance based on the Hamming weight and captures the notion of similarity of quantum states based on differences between their marginals [22].

For quantum states, the other very common similarity measure is quantum fidelity, which induces distance between states known as Bures distance [23, 24]. When one wants to compare unitary channel (quantum gate) with a general channel (noisy implementation of a gate), the relevant notions are worst-case [2] and average-case gate fidelity [25–29]. In both cases, the relevant optimization/averaging is over all quantum states. [removed unfinished sentence] For distance measures between measurements, one of the natural choices is to treat measurement as a quantum-classical channel and compute diamond norm distance [17, 30]. In the context of detector tomography sometimes fidelities between theoretical and experimental POVM's elements were considered [31–33]. When the target measurement is a computational basis, it is customary to use single-qubit error probabilities as a simplified quantifier of measurement's quality [34]. See [18] for an extensive overview of distinguishability measures between quantum objects.

The distance measures introduced by us rely on random quantum circuits which have many applications in the context of practical quantum computing. A notable example is shadow tomography, where random circuits are exploited to estimate multiple properties of quantum states with relatively low sample complexity [35–39]. Another example are generalizations of the classical randomized-benchmarking scheme [40–43] that use random circuits to estimate averaged quality metrics of quantum gates [44–46].

In Ref. [47] the authors prove that two states distant in Hilbert-Schmidt norm can be distinguished by a POVM constructed from approximate 4-design. Our proofs concerning average TV distances for quantum states and measurements were inspired by proofs

therein. In Ref. [48] the authors derived lower bounds (also containing HS distance) for TV distance in the same scenario for Haar-random POVMs, investigating applications for hidden subgroup problems. In Ref. [49] the "total operational distance" between states was introduced. It is based on the differences in obtained statistics when one performs mutually complementary projective measurements (see Ref. [49] for the notion of complementarity that is used). Importantly, the authors show that such distance is equivalent to HS distance between states of interest.

Structure of the paper

Let us now outline the structure of the paper. We start by introducing necessary theoretical concepts in Section II. This includes a discussion of common distance measures based on optimal statistical distinguishability, exact and approximate unitary k-designs, as well as stating several auxiliary Lemmas. From those, Lemma 5 is one of the important technical results of the work. In Section III we define the average Total-Variation distance between two states, measurements, and channels. We also outline the general methodology of the proofs presented in the main section of our work – Section IV. In that section, we prove the main results of our work. Namely, that the average Total-Variation distances between quantum objects can be approximated by explicit functions of the objects in question – quantum states in Theorem 1, quantum measurements in Theorem 2, and quantum channels in Theorem 3. Those functions are what we call average-case quantum distances. The main section is followed by Section V where we prove that average-case quantum distances posses variety of desired properties, such as subadditivity, joint convexity, and restricted data-processing inequalities – summarized in Table I for states, Table II for measurements, and Table III for channels. In this section we also prove asymptotic separations between average-case and worst-case distances, together with examples that saturate derived bounds. In Section VI we study exemplary scenarios where average-case quantum distances can be calculated analytically. We also show that average-case distances can be used to study average convergence of noisy distribution to uniform (trivial) distribution. We conclude the paper with Section VII where we discuss possible future research directions.

II. THEORETICAL BACKGROUND

In this section we give theoretical background for our main results. We start by introducing basic concepts and notation. Then we discuss in detail common distance measures based on optimal statistical distinguishability, and we recall notions of exact and approximate unitary k-designs. Finally, we state a number of auxiliary lemmas that will prove useful in later parts of the work.

A. Notation and basic concepts

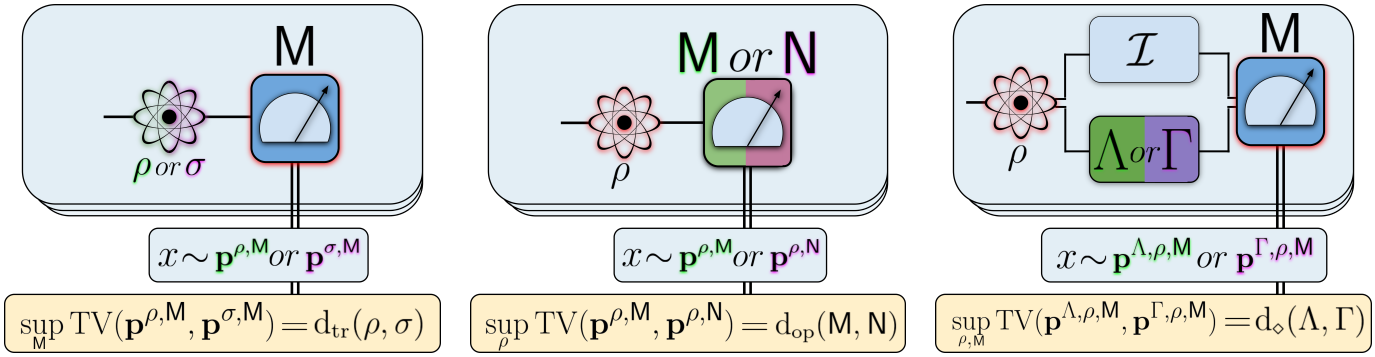
We start by recalling basic quantum-mechanical concepts used throughout the paper. We will be interested in d-dimensional Hilbert space $\mathcal{H}_d \approx \mathbb{C}^d$. We will omit subscript "d" if the dimension is not of importance. By $\mathrm{Herm}(\mathcal{H}_d)$ we denote a space of Hermitian operators on $\mathcal{H}_d$. A quantum state ρ is a positive-semi-definite operator with a trace equal to 1. We denote set of all quantum states on $\mathcal{H}_d$ as $\mathrm{D}(\mathcal{H}_d)$, and subset of pure states as $\mathcal{S}(\mathcal{H}_d)$. An n-outcome POVM [50] (Positive Operator-Valued Measure, or simply a quantum measurement) M is a tuple of n operators (called *effects*) $\mathsf{M} = (M_1, \ldots, M_n)$ that fulfill $M_i \geq 0$ and $\sum_{i=1}^n M_i = \mathbb{I}_d$, where $\mathbb{I}_d$ is an identity on $\mathcal{H}_d$. The set of all n-outcome POVMs on $\mathcal{H}_d$ will be denoted as $\mathrm{P}(\mathcal{H}_d, n)$. We will omit the symbol "n" if the number of outcomes is not of importance. An important example of a measurement that will be useful later is a computational-basis measurement defined as $\mathsf{M}^{\mathrm{comp}} = (|1\rangle\langle 1|, \ldots, |d\rangle\langle d|)$. A quantum channel Λ is a linear CTPT (Completely-Positive Trace-Preserving) map [2]. Trace-preserving condition means that for any quantum state $\rho, \Lambda(\rho) \in \mathrm{D}(\mathcal{H}_d)$, we have $\mathrm{tr}(\Lambda(\rho)) = \mathrm{tr}(\rho)$. Complete-positivity means that $(\Lambda \otimes \mathcal{I}_{d'})\, \tilde{\rho} \geq 0$ for any d' and any $\tilde{\rho} \in \mathrm{D}(\mathcal{H}_{dd'})$, where $\mathcal{I}_{d'}$ denotes identity channel on $\mathcal{H}_{d'}$. [removed repetition here] Quantum channel Λ is described via corresponding Choi-Jamiołkowski state defined as $\mathcal{J}_\Lambda := (\mathcal{I}_d \otimes \Lambda)(|\Phi^+\rangle\langle\Phi^+|)$, where we extend Hilbert space by its copy and act with channel Λ on a half of the maximally entangled state $|\Phi^+\rangle := \frac{1}{\sqrt{d}} \sum_{i=1}^d |ii\rangle$. We denote the set of all quantum channels from $\mathcal{H}$ to itself as $\mathrm{CPTP}(\mathcal{H}_d)$. An *unital* quantum channels is a channel $\Phi \in \mathrm{CPTP}(\mathcal{H}_d)$ such that $\Phi(\tau_d) = \tau_d$, where τ_d is the maximally mixed state in $\mathcal{H}_d$. When quantum state $\rho \in \mathrm{D}(\mathcal{H}_d)$ undergoes process $\Lambda \in \mathrm{CPTP}(\mathcal{H}_d)$ followed by measurement described by POVM $\mathsf{M} \in \mathrm{P}(\mathcal{H}_d, n)$, the probability of outcome labeled as "i" is given by Born's rule $p_i(i|\rho, \Lambda, \mathsf{M}) = \mathrm{tr}(\Lambda(\rho) M_i)$.

In the next subsection, we define distances induced by the following norms. Denote by $L(\mathcal{H}_d)$ a space of linear operators on $\mathcal{H}_d$. Then for $A \in L(\mathcal{H}_d)$, the trace norm is defined as

$$||A||_1 = \mathrm{tr}\left(\sqrt{AA^\dagger}\right). \tag{1}$$

For a channel $\Lambda \in \mathrm{CPTP}(\mathcal{H}_d)$, the diamond norm is defined through optimization of a trace norm as

$$||\Lambda||_\diamond = \max_{A \in L(\mathcal{H}_{d^2}),\ ||A||_1 \leq 1} ||(\Lambda \otimes \mathcal{I}_d)\, A||_1 . \tag{2}$$



(a) Trace distance between quantum states. (b) Operational distance between POVMs. (c) Diamond distance between quantum channels.

FIG. 1. Depiction of measures of distance between quantum objects based on *optimal* statistical distinguishability – which can be also interpreted as "worst-case" distance. For quantum states (1a), we optimize over all POVMs, while for measurements (1b) we optimize over all states. For quantum channels (1c) we optimize over both states and measurements on the extended Hilbert space.

B. Worst-case distance measures

Total-Variation Distance between two probability distributions $\mathbf{p} = \{p_i\}_{i=1}^n$ and $\mathbf{q} = \{q_i\}_{i=1}^n$ is defined by

$$\mathrm{TV}\left(\mathbf{p},\mathbf{q}\right) = \frac{1}{2}\sum_{i=1}^{n} |p_i - q_i| \,. \tag{3}$$

The TV distance quantifies the maximal statistical distinguishability of $\mathbf{p}$ and $\mathbf{q}$. Specifically, in a task when we are asked to decide whether the provided samples come from $\mathbf{p}$ or $\mathbf{q}$ (where both are promised to be given with probability $\frac{1}{2}$), the optimal success probability (i.e., probability of correctly guessing using the best possible strategy) is $\frac{1}{2}\left(1 + \mathrm{TV}\left(\mathbf{p},\mathbf{q}\right)\right)$ [2]. In quantum mechanics, the analogous task is to distinguish between two quantum objects, which can be either states, measurements, or channels (and, again, both are promised to be given with probability $\frac{1}{2}$), provided samples from the probability distributions that the objects of interest generate (via Born's rule). In all cases, the optimal success probability of performing this task is related to the optimal (maximized) TV distance between relevant probability distributions. This success probability is given by similar formula $\frac{1}{2}\left(1 + \mathrm{d}\left(\alpha_1,\alpha_2\right)\right)$, where α_1 and α_2 denote two objects to be distinguished, and the distance $\mathrm{d}\left(\,.\,,\,.\right)$ depends on the scenario. In Fig. 1, we pictorially present the most important distances based on optimal statistical distinguishability.

In the task where we want to distinguish between quantum states ρ and σ, we optimize over measurements (POVMs) performed on them, and the relevant distance is *trace distance* defined as [2]

$$\mathrm{d}_{\mathrm{tr}}\left(\rho,\sigma\right) = \sup_{\mathsf{M}\in\mathrm{P}(\mathcal{H})} \mathrm{TV}\big(\mathbf{p}^{\rho,\mathsf{M}}, \mathbf{p}^{\sigma,\mathsf{M}}\big) = \frac{1}{2}||\rho-\sigma||_1 \,, \tag{4}$$

where by $\mathbf{p}^{\rho,\mathsf{M}}$ we denote probability distribution obtained via Born's rule when measurement M is performed on state ρ. In this case, the optimal measurement, known as Helstrom's measurement, is projective with 2 outcomes [51].

In the case of quantum measurements, we want to decide whether the measurement performed is a POVM M or N, and we are optimizing over input states. The relevant distance is so called *operational distance* defined as [17, 30, 52]

$$\mathrm{d}_{\mathrm{op}}\left(\mathsf{M},\mathsf{N}\right) = \sup_{\rho\in\mathrm{D}(\mathcal{H})} \mathrm{TV}\big(\mathbf{p}^{\rho,\mathsf{M}}, \mathbf{p}^{\sigma,\mathsf{N}}\big) \,. \tag{5}$$

Finally, for distinguishing between two quantum channels Λ and Γ, we are optimizing over both input states (with ancillas) and measurements. In this case, the relevant distance is known as *diamond distance* defined as [2]

$$\mathrm{d}_{\diamond}(\Lambda,\Gamma) = \sup_{\rho\in\mathrm{D}(\mathcal{H}^{\otimes 2}),\ \mathsf{M}\in\mathrm{P}(\mathcal{H}^{\otimes 2})} \mathrm{TV}\big(\mathbf{p}^{\rho,\Lambda,\mathsf{M}}, \mathbf{p}^{\sigma,\Gamma,\mathsf{M}}\big) \,, \tag{6}$$

where we extended channel $\Lambda \otimes \mathcal{I}_d$ via identity channel $\mathcal{I}_d$ acting on ancillary system. While for the above distance we do not have a simple expression as a function of underlying objects, its calculation can be formulated as an SDP program that can be efficiently computed for moderate system sizes [16].

C. Exact and approximate unitary k-designs

In our work, we will be interested in expected values (integrals) $\mathop{\mathbb{E}}_{\beta\sim\nu} f(\beta) = \int_{\mathrm{U}(\mathcal{H})} d\nu(\beta) f(\beta)$ of a random variable f with respect to measure ν defined on unitary group $\mathrm{U}(\mathcal{H})$. The measure ν on unitary group induces measure on the set of pure quantum states in the following way – choose arbitrary fixed state ψ_0 and apply to it unitary $U \sim \nu$ drawn from measure ν, obtaining random state $\psi = U\psi_0 U^\dagger$. In short, we denote $\psi \sim \nu_{\mathcal{S}}$. The unique left-and right-invariant probability measure on $\mathrm{U}(\mathcal{H})$ is known as Haar measure and it will be denoted as μ. Random states obtained from the induced measure on states are called Haar-random states and the corresponding measure will be denoted by $\mu_{\mathcal{S}}$. Instrumental in our considerations, will be the notion of (approximate) unitary 4-designs. Unitary k-designs are, by definition, measures on $\mathrm{U}(\mathcal{H})$ that reproduce averages of Haar measure μ on balanced polynomials of degree k in entries of U [47]. For approximate k-designs these averages agree only approximately, and the quantitative notion of approximation can be defined differently (see, e.g.[53]). Here we adapt the notion of approximation based on the diamond norm. We say that a measure ν on $\mathrm{U}(\mathcal{H})$ is δ-approximate k-design if

$$\|\mathcal{T}_{k,\nu} - \mathcal{T}_{k,\mu}\|_\diamond \leq \delta \,, \tag{7}$$

where $\mathcal{T}_{k,\nu}$ is the quantum channel acting on $\mathcal{H}^{\otimes k}$ defined as $\mathcal{T}_{k,\nu}(\rho) = \int_{\mathrm{U}(\mathcal{H})} d\nu(U) U^{\otimes k} \rho (U^\dagger)^{\otimes k}$. An important example of approximate k-designs are the 1D architecture random quantum circuits formed from *arbitrary* universal gates that randomly couple neighboring qubits. These easy-to-implement circuits approximate k-designs efficiently with the number of qubits N [54–56]. Specifically, δ-approximate 4-designs are generated by local random quantum circuits of depth $O\left(N\left(N+\log\frac{1}{\delta}\right)\right)$ and by the random brickwork architecture in depth $O(N+\log(1/\delta))$, with moderate numerical constants (see Table 1 of [56] for the exact scaling).

D. Auxiliary lemmas

We will be interested in bounding from below and from above the expected values of some random variables. In bounding from above, we will use the following

Lemma 1. *(Jensen's inequality [57]) Let f be a concave function, and X a random variable. Then we have*

$$f\left(\mathbb{E}X\right) \geq \mathbb{E} f\left(X\right) \,. \tag{8}$$

On the other hand, in bounding from below, we will use the following

Lemma 2. *(Berger's inequality [58]) Let X be a random variable with well-defined second and fourth moments. Then we have*

$$\frac{(\mathbb{E}[X^2])^{\frac{3}{2}}}{(\mathbb{E}[X^4])^{\frac{1}{2}}} \leq \mathbb{E}|X| \,. \tag{9}$$

We will also make use of the following auxiliary lemmas.

Lemma 3 (Auxiliary integral involving k-th moment [59, Prop. 6])**.** *Let $X \in \mathrm{Herm}((\mathcal{H}_d)^{\otimes k})$ [removed redundant explanation here] and μ be a Haar measure. Then we have*

$$\mathop{\mathbb{E}}_{U\sim\mu} \mathrm{tr}\left(U^{\otimes k}|i\rangle\langle i|^{\otimes k}(U^\dagger)^{\otimes k} X\right) = \frac{1}{\binom{d+k-1}{k}} \mathrm{tr}\left(\mathbb{P}^{(k)}_{\mathrm{sym}} X\right) \,, \tag{10}$$

where $\mathbb{P}^{(k)}_{\mathrm{sym}}$ is the projector onto k-fold symmetric subspace $\mathcal{H}^{(k)}_{\mathrm{sym}} \subset \mathcal{H}_d^{\otimes k}$.

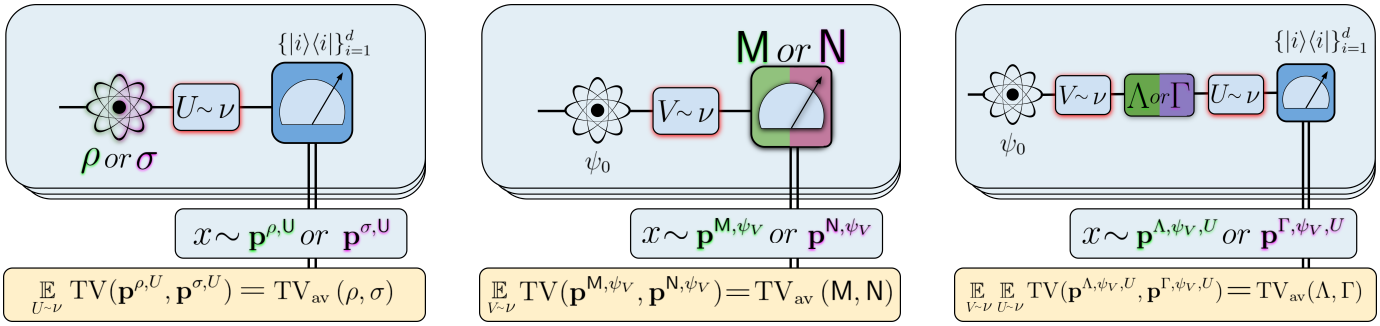
Corollary 1 (Auxiliary integral for 2nd moment)**.** *Let $X \in \mathrm{Herm}(\mathcal{H}_d)$. [removed redundant explanation here] Then we have*

$$\mathop{\mathbb{E}}_{U\sim\mu} \mathrm{tr}(|i\rangle\langle i| U X U^\dagger)^2 = \frac{1}{d(d+1)}\left(\mathrm{tr}(X^2) + \mathrm{tr}(X)^2\right) \,. \tag{11}$$

Proof. The above identity follows from Lemma 3. We use the identities $\mathbb{P}^{(2)}_{\mathrm{sym}} = \frac{1}{2}(\mathbb{I}\otimes\mathbb{I} + \mathbb{S})$ and $\mathrm{tr}(\mathbb{S}\rho\otimes\rho) = \mathrm{tr}(\rho^2)$, where $\mathbb{S}$ denotes the swap operator acting on $\mathcal{H}^{\otimes 2}$. □

Lemma 4 (Lemma 2 from [60])**.** *Let $X \in \mathrm{Herm}(\mathcal{H})$. [removed redundant explanation here] Let $\mathbb{P}^{(k)}_{\mathrm{sym}}$ denotes orthogonal projector onto k-fold symmetrization of $\mathcal{H}^{(k)}_{\mathrm{sym}} \subset \mathcal{H}^{\otimes k}$. We then have the following inequality*

$$\mathrm{tr}\left(X^{\otimes 4}\,\mathbb{P}^{(4)}_{\mathrm{sym}}\right) \leq C\,\mathrm{tr}\left(X^{\otimes 2}\,\mathbb{P}^{(2)}_{\mathrm{sym}}\right)^2 \,, \text{ where } C = \frac{10.1}{6} \,. \tag{12}$$



(a) Quantum states. (b) Quantum measurements. (c) Quantum channels.

FIG. 2. Measures of distance between quantum objects based on *average* statistical distinguishability. For quantum states (2a), we take the average over random unitaries applied to the state, followed by measurement in a standard basis. For quantum measurements (2a), we take the average over random pure states measured on the detector. Finally, for quantum channels (1c) we take the average over random input states, and random unitaries applied *after* the action of the channel. Note the difference with Fig. 1, where for common distance measures the *optimal* protocol is chosen, while here we consider random protocols.

Finally, the following Lemma 5, proved in Appendix A of Supplementary Material (SM), generalizes Lemma 4 and can be of independent interest. This result will be instrumental in proofs regarding average-case distances between quantum channels.

Lemma 5 (Inequality involving two operators and projections onto 2-fold symmetric subspaces)**.** *Let $X, Y \in \mathrm{Herm}(\mathcal{H})$. [removed redundant explanation here] Let $\mathbb{P}_{\mathrm{sym}}^{(k)}$ denotes the orthogonal projector onto k-fold symmetrization of $\mathcal{H}_{\mathrm{sym}}^{(k)} \subset \mathcal{H}^{\otimes k}$. We then have the following inequality*

$$\mathrm{tr}\left(X^{\otimes 2}\otimes Y^{\otimes 2}\,\mathbb{P}_{\mathrm{sym}}^{(4)}\right) \leq C\,\mathrm{tr}\left(X^{\otimes 2}\,\mathbb{P}_{\mathrm{sym}}^{(2)}\right)\mathrm{tr}\left(Y^{\otimes 2}\,\mathbb{P}_{\mathrm{sym}}^{(2)}\right)\ ,\ \textit{where } C=\frac{13}{6}\ . \tag{13}$$

Remark 1. *Note that the constant appearing on the right-hand side of (13) is slightly worse than the one from (12).*

III. METHODOLOGY

The goal of this section is to provide an overview of the main results of this work. We will describe the notion of average Total-Variation, and the general methodology for proofs given in Section IV.

A. Average Total Variation distances

We will be interested in establishing bounds for *average Total-Variation distance* between probability distributions generated by two quantum objects (states, measurements, or general channels). The average will be taken over an ensemble of *random* circuits. These notions are represented pictorially in Fig. 2, and we will now formally define them.

Consider a general quantum protocol that consists of a state preparation, an evolution of the system, and a quantum measurement. Now we consider *average Total-Variation distance* between two quantum objects:

1. (*States*) Two quantum states $\rho, \sigma \in \mathrm{D}(\mathcal{H})$ are fixed, rotated by a random unitary, and measured in the computational basis. Let us denote by $\mathbf{p}^{\rho,U}$ probability distribution obtained in this process, i.e., $p_i^{\rho,U} = \mathrm{tr}\left(|i\rangle\langle i|\,U\rho U^\dagger\right)$. The average TV distance between ρ and σ is

$$\mathrm{TV}_{\mathrm{av}}(\rho,\sigma) \coloneqq \mathop{\mathbb{E}}_{U\sim\nu}\mathrm{TV}\left(\mathbf{p}^{\rho,U},\mathbf{p}^{\sigma,U}\right)\ . \tag{14}$$

2. (*Measurements*) Two n-outcome quantum measurements $\mathsf{M},\mathsf{N} \in \mathrm{P}(\mathcal{H},n)$ are fixed, while states are taken to be random. Let us denote by $\mathbf{p}^{\mathsf{M},\psi_V}$ probability distribution obtained in this process, i.e., $p_i^{\mathsf{M},\psi_V} = \mathrm{tr}\left(M_i V\psi_0 V^\dagger\right)$, where ψ_0 is a fixed pure state. The average TV distance between M and N is

$$\mathrm{TV}_{\mathrm{av}}(\mathsf{M},\mathsf{N}) \coloneqq \mathop{\mathbb{E}}_{V\sim\nu}\mathrm{TV}\left(\mathbf{p}^{\mathsf{M},\psi_V},\mathbf{p}^{\mathsf{N},\psi_V}\right)\ . \tag{15}$$

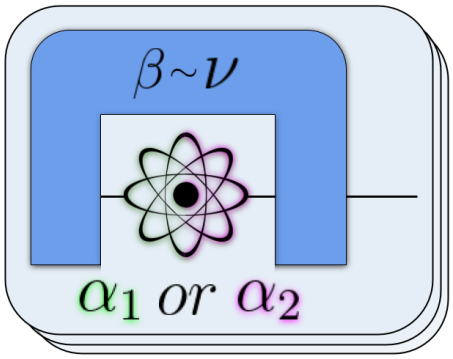


FIG. 3. Illustration of the general setup we consider in this work. Two quantum objects α_1, α_2 that can be either quantum states, measurements, or channels, are surrounded by random circuits β drawn from probability measure ν.

3. (*Channels*) Two quantum channels $\Lambda, \Gamma \in \mathrm{CPTP}(\mathcal{H})$ are fixed. Input state is taken to be a random pure state $V\psi_0 V^\dagger$ for fixed ψ_0. Output state is rotated by independent random unitary U (hence we have random unitary rotations before and after application of a channel), followed by measurement in a standard basis. Let us denote by $\mathbf{p}^{\Lambda,\psi_V,U}$ probability distribution obtained in this process, i.e., $p_i^{\Lambda,\psi_V,U} = \mathrm{tr}\left(|i\rangle\langle i| U\Lambda\left(V\psi_0 V^\dagger\right)U^\dagger\right)$ The average TV distance between Λ and Γ is

$$\mathrm{TV}_{\mathrm{av}}\left(\Lambda,\Gamma\right) \coloneqq \mathop{\mathbb{E}}_{U\sim\nu}\mathop{\mathbb{E}}_{V\sim\nu} \mathrm{TV}\left(\mathbf{p}^{\Lambda,\psi_V,U}, \mathbf{p}^{\Gamma,\psi_V,U}\right) . \tag{16}$$

Remark 2. *If the average TV distance is bounded from below by a constant c, then there exists a strategy that uses random circuits which distinguishes between two objects with probability at least* $\frac{1}{2}(1+c)$ *in single-shot scenario. Thus, by the virtue of Hoeffding's inequality, having access to* s *copies (samples) gives an error probability of the majority vote strategy dropping as* $2\exp(-\frac{c^2}{2}s)$*. We note that while the lower bound implies existence of such strategy, it does not tell what is the exact protocol for realizing this success probability.*

Remark 3. *The value of the average* TV *distance for quantum states can be reinterpreted as* TV*-distance of output statistics resulting from a measurement of a* single *POVM with effects* $M_{i,U_j} = \nu_j U_j^\dagger|i\rangle\langle i|U_j$*, where* ν_j *is the probability of occurrence of circuit* U_j *in the ensemble* ν *(for simplicity of presentation we assume that ensemble* ν *is discrete). This POVM can be interpreted as a convex combination [61] of projective measurements* M^{U_j} *with effects* $M_i^{U_j} = U_j^\dagger|i\rangle\langle i|U_j$*. Analogous interpretation holds also for the average* TV *distances for quantum measurements and channels – they can be interpreted as* TV *distances between output statistics of the corresponding randomized protocols [47]. Recall from Remark 2 that a lower bound on average TV distance implies that such randomized protocol distinguishes between quantum states with high probability. We note that it immediately follows that there also exists a deterministic (not randomized)* optimal *distinguishability protocol that achieves the same success probability.*

B. General methodology of proofs

Consider a general quantum protocol that results in probability distribution $\mathbf{p}^{\alpha,\beta}$ where α denotes a fixed quantum object (state, measurement, or channel), and β is a random variable (usually specifying quantum circuit) distributed according to probability distribution ν (typically Haar measure, approximate k-design, or random instances of variational circuits). See Fig. 3 for illustration. We will be interested in bounding quantities of the type

$$\mathrm{TV}_{\mathrm{av}}(\alpha_1,\alpha_2) \coloneqq \mathop{\mathbb{E}}_{\beta\sim\nu} \mathrm{TV}(\mathbf{p}^{\alpha_1,\beta}, \mathbf{p}^{\alpha_2,\beta}) . \tag{17}$$

For example, in the case of the distance between quantum states, α would correspond to two fixed quantum states that we want to calculate the distance between, and β would correspond to random quantum measurements (as in Eq. (14)).

To estimate $\mathrm{TV}_{\mathrm{av}}(\alpha_1,\alpha_2)$ from above we first expand

$$\mathop{\mathbb{E}}_{\beta\sim\nu} \mathrm{TV}(\mathbf{p}^{\alpha_1,\beta}, \mathbf{p}^{\alpha_2,\beta}) = \frac{1}{2}\sum_{i=1}^{n} \mathop{\mathbb{E}}_{\beta\sim\nu} |p_i^{\alpha_1,\beta} - p_i^{\alpha_2,\beta}|, \tag{18}$$

and use Jensen's inequality (see Lemma 1) for the concave function $f(x) = \sqrt{x}$ to upper bound the average of each of the summands

$$\mathrm{TV}_{\mathrm{av}}(\alpha_1, \alpha_2) \leq \frac{1}{2} \sum_{i=1}^{n} \sqrt{\mathop{\mathbb{E}}_{\beta\sim\nu} (p_i^{\alpha_1,\beta} - p_i^{\alpha_2,\beta})^2} \,. \tag{19}$$

To establish a lower bound for $\mathrm{TV}_{\mathrm{av}}(\alpha_1, \alpha_2)$ we will apply Berger's inequality (see Lemma 2) to random variables $x_i = p_i^{\alpha_1,\beta} - p_i^{\alpha_2,\beta}$ and insert the obtained result to (18). Importantly, in Section IV it will turn out that for probabilities and measures involved in our considerations we will have

$$\mathop{\mathbb{E}}_{\beta\sim\nu} (p_i^{\alpha_1,\beta} - p_i^{\alpha_2,\beta})^4 \leq C \left[\mathop{\mathbb{E}}_{\beta\sim\nu} (p_i^{\alpha_1,\beta} - p_i^{\alpha_2,\beta})^2 \right]^2 \,, \tag{20}$$

where $C > 0$ is a constant independent of the dimension of the Hilbert space or the number of measurement outcomes. This fact, together with Berger's inequality (Eq. (9), yields the bound

$$\frac{1}{C^{1/2}} \frac{1}{2} \sum_{i=1}^{n} \sqrt{\mathop{\mathbb{E}}_{\beta\sim\nu} (p_i^{\alpha_1,\beta} - p_i^{\alpha_2,\beta})^2} \leq \mathrm{TV}_{\mathrm{av}}(\alpha_1, \alpha_2) \,. \tag{21}$$

Therefore, we have

$$\frac{1}{C^{1/2}} \frac{1}{2} \sum_{i=1}^{n} \sqrt{\mathop{\mathbb{E}}_{\beta\sim\nu} (p_i^{\alpha_1,\beta} - p_i^{\alpha_2,\beta})^2} \leq \mathrm{TV}_{\mathrm{av}}(\alpha_1, \alpha_2) \leq \frac{1}{2} \sum_{i=1}^{n} \sqrt{\mathop{\mathbb{E}}_{\beta\sim\nu} (p_i^{\alpha_1,\beta} - p_i^{\alpha_2,\beta})^2} \,, \tag{22}$$

which makes it clear that to calculate both lower and upper bounds for average TV distance we will need to calculate $\mathop{\mathbb{E}}_{\beta\sim\nu} (p_i^{\alpha_1,\beta} - p_i^{\alpha_2,\beta})^2$. Importantly, since both bounds will differ only by a constant (independent of dimension), it will motivate the introduction of average-case quantum distances defined as

$$\mathrm{d}_{\mathrm{av}}(\alpha_1, \alpha_2) \coloneqq \frac{1}{2} \sum_{i=1}^{n} \sqrt{\mathop{\mathbb{E}}_{\beta\sim\nu} (p_i^{\alpha_1,\beta} - p_i^{\alpha_2,\beta})^2} \,. \tag{23}$$

Fortunately, as will be shown in Section IV, such terms can be expressed via simple, explicit functions of the quantum objects that we want to calculate the distance between, provided that ν forms an approximate 4-design.

Remark 4. *We note that depending on the perspective one adopts, either the upper bound or lower bound on average TV distance might be of particular interest. Namely, if one wishes to compare the ideal implementation of some protocol with its noisy version, then the upper bound might be satisfactory. In such a scenario, the ensemble of random circuits suffices to be approximate 2-design, since only 2nd moments are needed for its calculation. On the other hand, for statistical distinguishability, the lower bound is important (see Remark 2) and thus 4-design property is necessary.*

IV. AVERAGE-CASE QUANTUM DISTANCES

In this section, we present our main technical results. We prove that if random circuits form approximate unitary 4-designs, the average TV distances (see Section III) can be approximated, up to the relative error, by simple functions that can be expressed by degree-2 polynomials in quantum objects in question. We provide explicit expressions for those functions (which we call average-case quantum distances), as well as numerical constants for the relative errors. The proofs given in this section concern *exact* unitary 4-designs, while derivations for approximate designs are relegated to Appendix B of SM.

A. Quantum states

Let $\mathbf{p}^{\rho,U}$ denote the probability distribution obtained when the state ρ (σ) undergoes a unitary transformation according to U and is subsequently measured in the computational basis of $\mathcal{H}_d$. In other words $p_i^{\rho,U} = \mathrm{tr}\left(|i\rangle\langle i|U\rho U^\dagger\right)$, where $\{|i\rangle\}_{i=1}^d$ is a computational basis of $\mathcal{H}$.

Theorem 1 (Average-case distinguishability of quantum sates)**.** *Let $\rho, \sigma \in \mathrm{D}(\mathcal{H}_d)$ be states on $\mathcal{H}_d$ and let U be a random unitary in $\mathcal{H}_d$ drawn from measure ν that forms a δ-approximate 4-design, with $\delta \coloneqq \frac{\delta'}{2d^4}$, $\delta' \in \left[0, \frac{1}{3}\right]$. We then have the following inequalities*

$$\ell(\delta')\, a\, \mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\rho, \sigma) \leq \mathop{\mathbb{E}}_{\mathrm{U}\sim\nu} \mathrm{TV}(\mathbf{p}^{\rho,U}, \mathbf{p}^{\sigma,U}) \leq u(\delta')\, A\, \mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\rho, \sigma) \,, \tag{24}$$

where we define the average-case quantum distance between states

$$\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho,\sigma) = \frac{1}{2}\sqrt{\mathrm{tr}([\rho-\sigma]^2)} = \frac{1}{2}\|\rho-\sigma\|_{\mathrm{HS}}\,, \tag{25}$$

and $a = 0.31$, $A = \sqrt{\frac{d}{d+1}} \leq 1$, $\ell(\delta') = \sqrt{\frac{(1-\frac{\delta'}{d^2})^3}{1+\delta'}}$, $u(\delta') = \left(1+\frac{\delta'}{d^2}\right)^{\frac{1}{2}}$.

Proof. In what follows we prove a version of the theorem for exact 4-designs (i.e., setting $\delta = 0$). We start by proving the upper bound in (24). To this aim, we utilize the upper bound in (19) (from Jensen's inequality) to obtain

$$\mathrm{TV}_{\mathrm{av}}(\rho,\sigma) \leq \frac{1}{2}\sum_{i=1}^{d}\sqrt{\mathop{\mathbb{E}}_{U\sim\nu}\mathrm{tr}(|i\rangle\langle i|U\Delta U^\dagger)^2}\,, \tag{26}$$

where we set $\Delta = \rho - \sigma$. Using the assumed 2-design property of ν and the standard techniques of Haar measure integration (cf. Corrolary 1) we get

$$\mathop{\mathbb{E}}_{U\sim\nu}\mathrm{tr}(|i\rangle\langle i|U\Delta U^\dagger)^2 = \frac{1}{d(d+1)}\,\mathrm{tr}(\Delta^2)\,, \tag{27}$$

which follows directly from Eq. (11) and the fact that Δ is traceless. Inserting the above into (26) we obtain the upper bound from (24).

In order to prove the lower bound we use Berger's inequality (cf. Eq. (9)) for variable $X = \mathrm{tr}(U|i\rangle\langle i|U^\dagger\Delta)$:

$$\mathop{\mathbb{E}}_{U\sim\nu}|\,\mathrm{tr}(U|i\rangle\langle i|U^\dagger\Delta)| \geq \frac{\left(\mathop{\mathbb{E}}_{U\sim\nu}[\mathrm{tr}(U|i\rangle\langle i|U^\dagger\Delta)]^2\right)^{3/2}}{\left(\mathop{\mathbb{E}}_{U\sim\nu}[\mathrm{tr}(U|i\rangle\langle i|U^\dagger\Delta)]^4\right)^{1/2}}\,. \tag{28}$$

The numerator of the above fraction contains power of the second moment already calculated in Eq. (27), hence we get that it is equal to $K = [\frac{1}{d(d+1)}\,\mathrm{tr}(\Delta^2)]^{3/2}$. To get the upper bound for the denominator, we first note that from Lemma 3 it follows directly that the denominator is equal to $L = [\binom{d+3}{4}^{-1}\,\mathrm{tr}(\mathbb{P}^{(4)}_{\mathrm{sym}}\Delta^{\otimes 4})]^{1/2}$, where $\mathbb{P}^{(4)}_{\mathrm{sym}}$ is a projector onto 4-fold symmetric subspace of $\mathcal{H}_d^{\otimes 4}$. Now we get

$$\mathrm{tr}(\mathbb{P}^{(4)}_{\mathrm{sym}}\Delta^{\otimes 4}) \leq C\,\left(\mathrm{tr}\left(\mathbb{P}^{(2)}_{\mathrm{sym}}\Delta^{\otimes 2}\right)\right)^2 = \frac{C}{4}\left(\mathrm{tr}(\Delta^2)\right)^2\,, \tag{29}$$

with $C = \frac{10.1}{6}$. The inequality above is direct application of Lemma 4, while the equality follows from the fact that Δ is traceless and explicit form of $\mathbb{P}^{(2)}_{\mathrm{sym}}$. Inserting everything into Eq. (28) we obtain

$$\mathop{\mathbb{E}}_{U\sim\nu}|\,\mathrm{tr}(U|i\rangle\langle i|U^\dagger\Delta)| \geq \frac{K}{L} \geq \frac{w}{d}\sqrt{\mathrm{tr}\,\Delta^2}\,, \tag{30}$$

for $w = \sqrt{\frac{4}{C}\,\frac{\binom{d+3}{4}}{d^2(d+1)^2}} \geq 0.31 = a$. Finally, summing over $i = 1,\ldots,d$, we obtain lower bound on average TV distance

$$\mathop{\mathbb{E}}_{U\sim\nu}\frac{1}{2}\sum_{i=1}^{d}|\,\mathrm{tr}(U|i\rangle\langle i|U^\dagger\Delta)| \geq a\,\frac{1}{2}||\Delta||_{\mathrm{HS}}\,, \tag{31}$$

which concludes the proof. □

Remark 5. *The proof of Theorem 1 is inspired by the proof of Theorem 4 from [47] where Berger inequality was used to prove that two states far apart in Hilbert-Schmidt norm can be information-theoretically distinguished by a POVM constructed from approximate 4-design.*

Remark 6. *We note that in existing literature, the trace distance was usually preferred to Hilbert-Schmidt distance, one of the reasons being lack of the operational interpretation for the latter. The above considerations provide such interpretation for H-S distance in terms of average statistical distinguishability between quantum states, thus providing a sound physical motivation for its use.*

Remark 7. *We note that random quantum circuits in the 1D architecture formed from* arbitrary *universal gates that randomly couple neighbouring qubits, generate approximate k-designs efficiently with the number of qubits N [54–56]. Specifically, δ- approximate 4-designs are generated by the 1D random brickwork architecture in depth $O(N + \log(1/\delta))$, with moderate numerical constants [56]. This implies that ensembles appearing in Theorem 1 can be easily realized. Furthermore, it is expected that some of the classes of variational quantum circuits are expected to have, on average, unitary design-like properties [15]. This suggests that average-case quantum distances might be used to quantify average-case performance of hybrid quantum-classical algorithms. Naturally, the same remarks hold for Theorem 2 for measurements and Theorem 3 for channels.*

B. Quantum measurements

Let $\mathbf{p}^{\mathsf{M},\psi_V}$ denote the probability distribution of a quantum process in which a fixed pure quantum state ψ_0 on $\mathcal{H}_d$ is evolved according to unitary V and is subsequently measured via a n-outcome POVM $\mathsf{M} = (M_1, M_2, \ldots, M_n)$. In other words $p_i^{\mathsf{M},\psi_V} = \operatorname{tr}(V\psi_0 V^\dagger M_i)$.

Theorem 2 (Average-case quantum distance between quantum measurements)**.** *Let M, N be n-outcome POVMs on $\mathcal{H}_d$ and V be a random unitary on $\mathcal{H}_d$ drawn from measure ν that forms a δ-approximate 4-design, with $\delta \coloneqq \frac{\delta'}{2d^4}$, $\delta' \in \left[0, \frac{1}{3}\right]$. We then have the following inequalities*

$$\ell(\delta')\, a\, \mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{M},\mathsf{N}) \leq \mathop{\mathbb{E}}_{V\sim\nu} \mathrm{TV}(\mathbf{p}^{\mathsf{M},\psi_V}, \mathbf{p}^{\mathsf{N},\psi_V}) \leq u(\delta')\, A\, \mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{M},\mathsf{N})\,, \tag{32}$$

where we define average-case quantum distance between measurements

$$\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{M},\mathsf{N}) = \frac{1}{2d}\sum_{i=1}^{n} \sqrt{\|M_i - N_i\|^2_{\mathrm{HS}} + \operatorname{tr}(M_i - N_i)^2}\,. \tag{33}$$

and $a = 0.31$, $A = \sqrt{\frac{d}{d+1}} \leq 1$, $\ell(\delta') = \sqrt{\frac{(1-\frac{\delta'}{d^2})^3}{1+\delta'}}$, $u(\delta') = \left(1 + \frac{\delta'}{d^2}\right)^{\frac{1}{2}}$.

Proof. In what follows we prove a version of the theorem for exact 4-designs. The proof is in fact almost exactly the same as of Theorem 1. We can define $\Delta_i = M_i - N_i$, now each Δ_i having role of previous Δ, namely in each summand appearing in the TV distance is of the form $|\operatorname{tr}(V\psi_0 V^\dagger \Delta_i)|$. We now note that arbitrary fixed pure state $\psi_0 = U_0 |0\rangle\langle 0| U_0^\dagger$ is unitarily equivalent to computational basis state via some unitary U_0, and that Haar measure is invariant under transformation $U \to UU_0$. From those facts it follows that we can apply exactly the same steps as for proof of Theorem 1. For second moment we obtain

$$\mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}} \operatorname{tr}(\psi\Delta_i)^2 = \frac{1}{d(d+1)}\left(\operatorname{tr}(\Delta_i^2) + \operatorname{tr}(\Delta_i)^2\right)\,, \tag{34}$$

which differs from Eq. (27) by additional summand, because now Δ_i is not necessarily traceless. Rest of the steps is exactly analogous to the proof of Theorem 1.

□

C. Quantum channels

Let $\mathbf{p}^{\Lambda,\psi_V,U}$ be the probability distribution associated to a quantum process in which a fixed pure state $\psi_0 \in \mathcal{S}(\mathcal{H})$ is transformed by unitary transformation V, channel Λ, unitary U, and is subsequently measured in the computational basis of $\mathcal{H}_d$. In other words we have $p_i^{\Lambda,\psi_V,U} = \operatorname{tr}(|i\rangle\langle i| U \Lambda(V\psi_0 V^\dagger) U^\dagger)$.

Theorem 3 (Average-case distinguishability of quantum channels)**.** *Let Λ, Γ be quantum channels acting on $\mathcal{H}_d$. let ν be a distribution on on $\mathrm{U}(\mathcal{H}_d)$ forming δ-approximate 4-design for $\delta = \frac{\delta'}{(2d)^8}$. Then we have the following inequalities*

$$\ell^{\mathrm{ch}}(\delta')\, a^{\mathrm{ch}}\, \mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma) \leq \mathop{\mathbb{E}}_{V\sim\nu}\mathop{\mathbb{E}}_{U\sim\nu} \mathrm{TV}(\mathbf{p}^{\Lambda,\psi_V,U}, \mathbf{p}^{\Gamma,\psi_V,U}) \leq u^{\mathrm{ch}}(\delta')\, A^{\mathrm{ch}}\, \mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma)\,, \tag{35}$$

where we defined average-case quantum distance between channels

$$\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma) \coloneqq \frac{1}{2}\sqrt{\|\mathcal{J}_\Lambda - \mathcal{J}_\Gamma\|^2_{\mathrm{HS}} + \operatorname{tr}\left((\Lambda-\Gamma)[\tau_d]^2\right)}, \tag{36}$$

and $a^{\mathrm{ch}} = 0.087$, $A^{\mathrm{ch}} = \frac{d}{d+1} \leq 1$, $\ell^{\mathrm{ch}}(\delta') = \frac{\left(1-\frac{\delta'}{d^2}\right)^3}{1+\delta'}$, $u^{\mathrm{ch}}(\delta') = 1 + \frac{\delta'}{d^2}$.

Proof. In what follows we prove version of a theorem for exact 4-designs (i.e., setting $\delta = 0$). In order to simplify the notation we will use the notation $\Delta \coloneqq \Lambda - \Gamma$ (note that Δ is a superoperator and has a different meaning then Δ used in the proof of Theorem 1). We will make use of the Theorem 1 which implies that for fixed $\psi_V \in \mathcal{S}(\mathcal{H})$ the following inequalities hold

$$\frac{a}{2}\|\Delta[\psi_V]\|_{\mathrm{HS}} \leq \mathop{\mathbb{E}}_{U\sim\nu} \mathrm{TV}(\mathbf{p}^{\Lambda,\psi_V,U}, \mathbf{p}^{\Gamma,\psi_V,U})\,, \tag{37}$$

$$\frac{A}{2}\|\Delta[\psi_V]\|_{\mathrm{HS}} \geq \mathop{\mathbb{E}}_{U\sim\nu} \mathrm{TV}(\mathbf{p}^{\Lambda,\psi_V,U}, \mathbf{p}^{\Gamma,\psi_V,U})\,. \tag{38}$$

In what follows we prove bounds on $\mathop{\mathbb{E}}_{V\sim\nu}\|\Delta[\psi_V]\|_{\mathrm{HS}} = \mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}}\|\Delta[\psi]\|_{\mathrm{HS}}$. We first establish the upper bound by employing Jensen's inequality

$$\mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}}\|\Delta[\psi]\|_{\mathrm{HS}} \leq \sqrt{\mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}}\mathrm{tr}(\Delta[\psi]^2)}\,. \tag{39}$$

The average of $\mathrm{tr}(\Delta[\psi]^2)$ can be computed explicitly using the 2-design property and Lemma 3. We first rewrite using the same trick as in the proof of Corollary 1

$$\mathrm{tr}\left(\Delta[\psi]^2\right) = \mathrm{tr}\left(\Delta[\psi]^{\otimes 2}\mathbb{S}\right) \tag{40}$$

where $\mathbb{S} = \sum_{i,j=1}^{d}|i\rangle\langle j|\otimes|j\rangle\langle i|$ is the swap operator acting on $\mathcal{H}_d^{\otimes 2}$. Inserting the above into Lemma 3 yields

$$\frac{2}{d(d+1)}\,\mathrm{tr}\left(\mathbb{S}\,\Delta^{\otimes 2}[\mathbb{P}^{(2)}_{\mathrm{sym}}]\right) = \frac{2}{d(d+1)}\frac{1}{2}\left(\mathrm{tr}\left(\mathbb{S}\,\Delta^{\otimes 2}[\mathbb{I}]\right) + \mathrm{tr}\left(\mathbb{S}\,\Delta^{\otimes 2}[\mathbb{S}]\right)\right)\,, \tag{41}$$

where we used the identity $\mathbb{P}^{(2)}_{\mathrm{sym}} = \frac{1}{2}(\mathbb{I}\otimes\mathbb{I}+\mathbb{S})$. We now rewrite the first term as $\mathrm{tr}\left(\mathbb{S}\Delta^{\otimes 2}[\mathbb{I}]\right) = \mathrm{tr}\left(\Delta[\mathbb{I}]^2\right)$. Inserting the above into the integral (with multiplication and division by d^2) gives

$$\mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}}\mathrm{tr}(\Delta[\psi]^2) = \frac{d^2}{d(d+1)}\left(\mathrm{tr}\left(\Delta\left[\frac{\mathbb{I}}{d}\right]^2\right) + \mathrm{tr}\left(\mathbb{S}\Delta^{\otimes 2}\left[\frac{\mathbb{S}}{d^2}\right]\right)\right)\,. \tag{42}$$

Recall that $\mathcal{J}_\Delta = (\mathbb{I}\otimes\Delta)(\Phi_+)$, where $|\Phi_+\rangle = \frac{1}{\sqrt{d}}\sum_{i=1}^{d}|i\rangle\,|i\rangle$ is the maximally entangled state in $\mathcal{H}_d\otimes\mathcal{H}_d$. Explicit computation gives $\|\mathcal{J}_\Delta\|^2_{\mathrm{HS}} = \mathrm{tr}\left(\mathbb{S}\Delta^{\otimes 2}\left[\frac{\mathbb{S}}{d^2}\right]\right)$.

$$\mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}}\mathrm{tr}(\Delta[\psi]^2) = \frac{d^2}{d(d+1)}\left(\mathrm{tr}\left(\Delta\left[\frac{\mathbb{I}}{d}\right]^2\right) + \|\mathcal{J}_\Delta\|^2_{\mathrm{HS}}\right)\,. \tag{43}$$

Inserting this expression into (39) and using (38) finally gives upper bound in Eq. (35).

To prove the lower bound integrate both sides of (37) and apply Berger's inequality for $X_\psi = \|\Delta[\psi]\|_{\mathrm{HS}} = \sqrt{\mathrm{tr}(\Delta[\psi]^2)}$

$$\mathop{\mathbb{E}}_{\psi\sim\mu_{\mathcal{S}}}\|\Delta[\psi]\|_{\mathrm{HS}} \geq \frac{\left(\mathop{\mathbb{E}}_{\psi\sim\mu_{\mathcal{S}}}\mathrm{tr}(\Delta[\psi]^2)\right)^{3/2}}{\left(\mathop{\mathbb{E}}_{\psi\sim\mu_{\mathcal{S}}}\mathrm{tr}(\Delta[\psi]^2)^2\right)^{1/2}}\,. \tag{44}$$

We proceed by rewriting the integral in the denominator of the above expression

$$\mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}}\mathrm{tr}(\Delta[\psi]^2)^2 = 4\mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}}\mathrm{tr}(\Delta[\psi]^{\otimes 4}\mathbb{P}^{(2)}_{\mathrm{sym}}\otimes\mathbb{P}^{(2)}_{\mathrm{sym}})\,. \tag{45}$$

where we used the identity $2\,\mathrm{tr}\,\mathbb{P}^{(2)}_{\mathrm{sym}}X\otimes Y = \mathrm{tr}\,X\,\mathrm{tr}(Y) + \mathrm{tr}(XY)$ and $\mathrm{tr}(\Delta[\psi]) = 0$. By expanding $\Delta[\psi]^{\otimes 4} = \Delta^{\otimes 4}\left[\psi^{\otimes 4}\right]$ and integrating over ψ (cf. Lemma 3) we obtain

$$\mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}}\mathrm{tr}(\Delta[\psi]^2)^2 = \frac{4}{\binom{d+3}{4}}\,\mathrm{tr}\left(\Delta^{\otimes 4}\left[\mathbb{P}^{(4)}_{\mathrm{sym}}\right]\mathbb{P}^{(2)}_{\mathrm{sym}}\otimes\mathbb{P}^{(2)}_{\mathrm{sym}}\right)\,. \tag{46}$$

Substituting $\mathbb{P}^{(2)}_{\mathrm{sym}} = \binom{d+1}{2}\mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}}\psi^{\otimes 2}$ we get

$$\mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}}\mathrm{tr}(\Delta[\psi]^2)^2 = \frac{4\binom{d+1}{2}^2}{\binom{d+3}{4}}\mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}}\mathop{\mathbb{E}}_{\varphi\sim\nu_{\mathcal{S}}}\mathrm{tr}\left(\mathbb{P}^{(4)}_{\mathrm{sym}}\Delta^\dagger[\psi]^{\otimes 2}\otimes\Delta^\dagger[\varphi]^{\otimes 2}\right)\,. \tag{47}$$

We now utilize Lemma 5 to upper bound the function inside the integral

$$\mathrm{tr}\left(\mathbb{P}^{(4)}_{\mathrm{sym}}\Delta^\dagger[\psi]^{\otimes 2}\otimes\Delta^\dagger[\varphi]^{\otimes 2}\right) \leq C\,\mathrm{tr}(\mathbb{P}^{(2)}_{\mathrm{sym}}\Delta^\dagger[\psi]^{\otimes 2})\,\mathrm{tr}(\mathbb{P}^{(2)}_{\mathrm{sym}}\Delta^\dagger[\varphi]^{\otimes 2})\,, \tag{48}$$

where $C = \frac{13}{6}$. Inserting this into (47) and carrying over the integrals over ψ and ϕ (with the help of Corollary 1) we get

$$\mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}} \operatorname{tr}(\Delta[\psi]^2)^2 \leq \frac{4C}{\binom{d+3}{4}} \operatorname{tr}\left(\mathbb{P}^{(2)}_{\mathrm{sym}} \Delta^{\otimes 2} \left[\mathbb{P}^{(2)}_{\mathrm{sym}}\right]\right)^2 . \tag{49}$$

We now calculate

$$\begin{aligned}\operatorname{tr}\left(\mathbb{P}^{(2)}_{\mathrm{sym}} \Delta^{\otimes 2} \left[\mathbb{P}^{(2)}_{\mathrm{sym}}\right]\right) &= \binom{d+1}{2} \mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}} \operatorname{tr}\left(\mathbb{P}^{(2)}_{\mathrm{sym}} \Delta[\psi]^{\otimes 2}\right) = \\ &= \frac{1}{2}\binom{d+1}{2} \mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}} \operatorname{tr}\left(\Delta[\psi]^{\otimes 2}\mathbb{S}\right) = \\ &= \frac{1}{2}\binom{d+1}{2} \mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}} \operatorname{tr}\left(\Delta[\psi]^2\right) . \end{aligned} \tag{50}$$

In the first equality we used the fact that since $\nu_{\mathcal{S}}$ forms a 2-design we can substitute the projector onto 2-fold symmetric subspace with corresponding (renormalized) average. In the second equality we exploited the fact that states of the type $\psi^{\otimes 2}$ are in invariant subspace of $\mathbb{P}^{(2)}_{\mathrm{sym}}$. Third equality follows directly from Corollary 1 and fact that $\operatorname{tr}(\Delta[\psi]) = 0$. Combining (49) and (50) we obtain

$$\mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}} \operatorname{tr}(\Delta[\psi]^2)^2 \leq \frac{C\binom{d+1}{2}^2}{\binom{d+3}{4}} \left(\mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}} \operatorname{tr}\left(\Delta[\psi]^2\right)\right)^2 . \tag{51}$$

Inserting the above bound into (44) gives

$$\mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}} \|\Delta[\psi]\|_{\mathrm{HS}} \geq b_d \sqrt{\mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}} \operatorname{tr}(\Delta[\psi]^2)} , \tag{52}$$

with $b_d = \frac{1}{\sqrt{13}}\sqrt{\frac{(d+2)(d+3)}{d(d+1)}}$. Integrating both sides of (37) over $\psi \sim \nu_{\mathcal{S}}$ and using the the above inequality we finally obtain

$$a^{\mathrm{ch}}\, \mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda, \Gamma) \leq \mathop{\mathbb{E}}_{V\sim\nu} \mathop{\mathbb{E}}_{U\sim\nu} \mathrm{TV}(\mathbf{p}^{\Lambda,\psi_V,U}, \mathbf{p}^{\Gamma,\psi_V,U}) , \tag{53}$$

with $a^{\mathrm{ch}} = a \cdot b_d \approx 0.087$. □

Remark 8. *We note that one can view quantum states and measurements as special types of quantum channels. While for state preparation channels the operational procedure of discrimination is equivalent and one gets the same expression (see Example 11), for measurements it is not the case (this is because for measurement channels one can average only over input states). Moreover, in the case of δ-approximate designs, applying the above Theorem 3 for the average-case distance between state preparation channels gives worse than Theorem 1 constants and functional dependence on δ. This approach thus leads to less tight bounds for states than treating them separately.*

V. PROPERTIES OF DISTANCE MEASURES

While expressions for average-case distances introduced in Section IV might seem abstract (especially in the case of measurements and channels), it turns out that they share multiple desired properties with common distances used in quantum information [2, 62]. In particular, our distances indeed fulfill metric axioms, they are subadditive with respect to tensor products, and have a joint convexity property. They are also non-increasing under *unital* quantum channels. Finally, average-case quantum distance between *unital* channels possesses two additional physically well-motivated properties – stability (it does not change when both channels are extended by identity channel) and chaining (distance between compositions of multiple channels is at most the sum of distances between constituting channels) [62]. In this section we state and prove those properties for states (Section V A), measurements (Section V B), and channels (Section V C). To make navigation easier, each subsection starts with a table of properties, comparison with relevant worst-case distance, and the text reference in which the properties are are proved – Table I for states, Table II for measurements, and Table III for channels.

Property	**Worst-case distance** $\mathrm{d}_{\mathrm{tr}}(\rho,\sigma)$	**Average-case distance** $\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho,\sigma)$	**Text reference for average-case distance**
Function	$\frac{1}{2}\|\|\rho-\sigma\|\|_1$	$\frac{1}{2}\|\rho-\sigma\|_{\mathrm{HS}}$	Theorem 1, Lemma 6
Subadditivity	$\mathrm{d}_{\mathrm{tr}}(\rho_1\otimes\rho_2,\sigma_1\otimes\sigma_2)\leq \mathrm{d}_{\mathrm{tr}}(\rho_1,\sigma_1)+\mathrm{d}_{\mathrm{tr}}(\rho_2,\sigma_2)$	same as for $\mathrm{d}_{\mathrm{tr}}(\rho,\sigma)$	Lemma 7
Joint convexity	$\mathrm{d}_{\mathrm{tr}}\left(\sum_\alpha p_\alpha\rho_\alpha,\sum_\alpha p_\alpha\sigma_\alpha\right)\leq\sum_\alpha p_\alpha \mathrm{d}_{\mathrm{tr}}\left(\rho_\alpha,\sigma_\alpha\right)$	same as for $\mathrm{d}_{\mathrm{tr}}(\rho,\sigma)$	Lemma 8
Data processing inequality	$\mathrm{d}_{\mathrm{tr}}(\Lambda(\rho),\Lambda(\sigma))\leq\mathrm{d}_{\mathrm{tr}}(\rho,\sigma)$ for CPTP Λ	$\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\Phi(\rho),\Phi(\sigma))\leq\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho,\sigma)$ for unital Φ	Lemma 9

TABLE I.

A. Quantum states

The following Table I summarizes properties of average-case quantum distance between states, and compares it to the worst-case trace distance.

Lemma 6 ($\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}$ fulfils axioms of a metric)**.** *Let $\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}$, denote average distances between states (Eq. (25)). Then $\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}$ satisfies axioms of a metric in space of quantum states. Specifically, it satisfies triangle inequality, symmetry and identity of indiscernibles:*

$$\mathrm{d}^{s}_{\mathrm{av}}(\rho,\sigma)\leq\mathrm{d}^{s}_{\mathrm{av}}(\rho,\tau)+\mathrm{d}^{s}_{\mathrm{av}}(\tau,\sigma)\quad \textit{for all}\quad \rho,\sigma,\tau\in\mathrm{D}(\mathcal{H}_{dim}) \tag{54}$$

$$\mathrm{d}^{s}_{\mathrm{av}}(\rho,\sigma)=\mathrm{d}^{s}_{\mathrm{av}}(\sigma,\rho)\quad \textit{for all}\quad \rho,\sigma\in\mathrm{D}(\mathcal{H}_{dim}) \tag{55}$$

$$\mathrm{d}^{s}_{\mathrm{av}}(\rho,\sigma)=0\iff\rho=\sigma\quad \textit{for all}\quad \rho,\sigma\in\mathrm{D}(\mathcal{H}_{dim})\ . \tag{56}$$

Proof. The result follows directly from the fact, that $\mathrm{d}^{s}_{\mathrm{av}}$ is a Hilbert Schmidt distance. □

Lemma 7 ($\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}$ is subadditive)**.** *For arbitrary quantum states $\rho_1,\sigma_1\in\mathrm{D}(\mathcal{H}),\rho_2,\sigma_2\in\mathrm{D}(\mathcal{H})$, we have*

$$\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho_1\otimes\rho_2,\sigma_1\otimes\sigma_2)\leq\mathrm{d}_{\mathrm{av}}(\rho_1,\sigma_1)+\mathrm{d}_{\mathrm{av}}(\rho_2,\sigma_2)\ . \tag{57}$$

Proof. The proof follows from triangle inequality and multiplicativity with respect to tensor product, i.e.

$$\begin{aligned}\|\rho_1\otimes\rho_2-\sigma_1\otimes\sigma_2\|_{\mathrm{HS}}&=\|\rho_1\otimes(\rho_2-\sigma_2)-(\sigma_1-\rho_1)\otimes\sigma_2\|_{\mathrm{HS}}\\&\leq\|\rho_1\|_{\mathrm{HS}}\|\rho_2-\sigma_2\|_{\mathrm{HS}}+\|\sigma_1\|_{\mathrm{HS}}\|\sigma_1-\rho_1\|_{\mathrm{HS}}\\&\leq\|\rho_2-\sigma_2\|_{\mathrm{HS}}+\|\sigma_1-\rho_1\|_{\mathrm{HS}}.\end{aligned} \tag{58}$$

□

Lemma 8 ($\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}$ has joint-convextiy property)**.** *For arbitrary sets of quantum states $\{\rho_\alpha\}_\alpha$, $\{\sigma_\alpha\}_\alpha$ and probability distributions $\{p_\alpha\}$, we have*

$$\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}\left(\sum_\alpha p_\alpha\rho_\alpha,\sum_\alpha p_\alpha\sigma\alpha\right)\leq\sum_\alpha p_\alpha\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}\left(\rho_\alpha,\sigma_\alpha\right)\ . \tag{59}$$

Proof. The proof follows directly from triangle inequality,

$$\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}\left(\sum_\alpha p_\alpha\rho_\alpha,\sum_\alpha p_\alpha\sigma\alpha\right)=\left\|\sum_\alpha p_\alpha(\rho_\alpha-\sigma_\alpha)\right\|_{\mathrm{HS}}\leq\sum_\alpha p_\alpha\left\|\rho_\alpha-\sigma_\alpha\right\|_{\mathrm{HS}}=\sum_\alpha p_\alpha\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}\left(\rho_\alpha,\sigma_\alpha\right)\ . \tag{60}$$

□

Lemma 9 (Data-processing inequalities for average-case distance between states)**.** *Average-case distance between states is monotonic with respect to unital maps, i.e. for a unital Φ, we have*

$$\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho,\sigma)\geq\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\Phi(\rho),\Phi(\sigma))\ . \tag{61}$$

Proof. We begin the proof by reminding celebrated Uhlmann theorem [63], that for unital channel Φ and a Hermitian operator H, we have

$$\Phi(H) \prec H \ , \tag{62}$$

where the majorization relation above can be seen as a majorization between real vectors of eigenvalues. We also note, that using the fact, that Hilbert-Schmidt norm is a Schur-convex function of eigenvalues, we get

$$\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho,\sigma) \geq \mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\Phi(\rho),\Phi(\sigma)) \ . \tag{63}$$

□

Lemma 10 (Separation between $\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}$ and d_{tr})**.** *Let $\rho,\sigma \in \mathcal{H}_d$ be quantum states. Then from standard inequalities between 1 and 2 norms it follows that*

$$\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho,\sigma) \ \leq \ \mathrm{d}_{\mathrm{tr}}(\rho,\sigma) \ \leq \ \sqrt{d}\, \mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho,\sigma) \ . \tag{64}$$

We now consider an example that attains the bound in Lemma 10.

Example 1 (Two orthogonal maximally mixed states of rank $\frac{d}{2}$)**.** *Consider two states $\rho,\sigma \in \mathcal{H}_d$, such that $\rho = \frac{\mathbb{I}_{d'}}{d'}, \sigma = \frac{\mathbb{I}_{d'}}{d'}$, where $d' = \frac{d}{2}$ and $\mathrm{tr}(\rho\sigma) = 0$. Direct calculation yields*

$$\begin{aligned} \mathrm{d}^{\mathrm{s}}_{\mathrm{av}}\left(\rho,\sigma\right) &= \frac{1}{\sqrt{d}}, \\ \mathrm{d}_{\mathrm{tr}}\left(\rho,\sigma\right) &= 1 \ . \end{aligned} \tag{65}$$

Clearly, the above shows that in the asymptotic limit the average-case distance between states goes to 0. From perspective of statistical distinguishability, it means that the states can be distinguished perfectly with optimal strategy, while randomized strategy fails dramatically.

Example 2 (Counterexample for data-procesiing inequality for quantum states)**.** *Consider two mixed states $\rho,\sigma \in \mathrm{D}(\mathcal{H})$ from previous Example 1. Now consider a* non-unital *quantum channel Λ s.t. $\Lambda(\rho) = |0\rangle\langle 0|$ and $\Lambda(\sigma) = |1\rangle\langle 1|$. Explicit computation combined with results of previous example yields*

$$\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}\left(\Lambda(\rho),\Lambda(\sigma)\right) = \frac{1}{\sqrt{2}} \ > \ \frac{1}{\sqrt{d}} = \mathrm{d}^{\mathrm{s}}_{\mathrm{av}}\left(\rho,\sigma\right) \ , \tag{66}$$

for $d > 2$.

B. Quantum measurements

The following Table II summarizes properties of average-case quantum distance between measurements, and compares it to the worst-case operational distance. For POVMs M and N, symbol $\mathsf{M} \otimes \mathsf{N}$ denotes a POVM with effects $\{M_i \otimes N_j\}_{i,j}$. Pre-processing channel Γ acts on the state just before measurement M. This is equivalent to performing new POVM with effects transformed via dual channel $M_i \to \Gamma^*(M_i)$ on the original state [64]. The fact that channel Γ is trace-preserving implies that dual channel Γ^* is unital, which ensures that $\{\Gamma^*(M_i)\}_i$ a proper POVM. The post-processing stochastic map described by matrix Λ transforms POVM's effects as $M_i \to \sum_j \Lambda_{ij} M_j$ (this can be interpreted as classical post-processing of classical outputs of the measurement).

Lemma 11 ($\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}$ fulfils axioms of a metric)**.** *Let $\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}$, denote average distances between quantum measurements (Eq. (33)). Then $\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}$ satisfies axioms of a metric in space of POVMs. Specifically, it satisfies triangle inequality, symmetry and identity of indiscernibles:*

$$\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{M},\mathsf{N}) \leq \mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{M},\mathsf{L}) + \mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{L},\mathsf{N}) \quad \textit{for all} \ \ \mathsf{M},\mathsf{N},\mathsf{L} \in \mathrm{P}(\mathcal{H}) \tag{67}$$

$$\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{M},\mathsf{N}) = \mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{N},\mathsf{M}) \quad \textit{for all} \ \ \mathsf{M},\mathsf{N} \in \mathrm{P}(\mathcal{H}) \tag{68}$$

$$\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{M},\mathsf{N}) = 0 \iff \mathsf{M} = \mathsf{N} \quad \textit{for all} \ \ \mathsf{M},\mathsf{N} \in \mathrm{P}(\mathcal{H}) \ . \tag{69}$$

Note, that $\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{M},\mathsf{N})$ is absolute homogeneous, i.e. if we extend the definition of $\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}$ to arbitrary collections of operators, we see, that $\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(s\mathsf{M}, s\mathsf{N}) = |s|\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{M},\mathsf{N})$.

Property	**Worst-case distance** $\mathrm{d}_{\mathrm{op}}(\mathsf{M},\mathsf{N})$	**Average-case distance** $\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{M},\mathsf{N})$	**Text reference for average-case distance**
Function	$\frac{1}{2}\sup_{\rho\in\mathrm{D}(\mathcal{H})}\sum_{i=1}^{n}\lvert\operatorname{tr}(M_i\rho)-\operatorname{tr}(N_i\rho)\rvert$	$\frac{1}{2d}\sum_{i=1}^{n}\sqrt{\lVert M_i-N_i\rVert_{\mathrm{HS}}^2+\operatorname{tr}(M_i-N_i)^2}$	Theorem 2, Lemma 11
Subadditivity	$\mathrm{d}_{\mathrm{op}}(\mathsf{M}_1\otimes\mathsf{M}_2,\mathsf{N}_1\otimes\mathsf{N}_2)\leq\mathrm{d}_{\mathrm{op}}(\mathsf{M}_1,\mathsf{N}_1)+\mathrm{d}_{\mathrm{op}}(\mathsf{M}_2,\mathsf{N}_2)$	same as for $\mathrm{d}_{\mathrm{op}}(\mathsf{M},\mathsf{N})$	Lemma 12
Joint convexity	$\mathrm{d}_{\mathrm{op}}\left(\sum_\alpha p_\alpha\mathsf{M}_\alpha,\sum_\alpha p_\alpha\mathsf{N}_\alpha\right)\leq\sum_\alpha p_\alpha\mathrm{d}_{\mathrm{op}}\left(\mathsf{M}_\alpha,\mathsf{N}_\alpha\right)$	same as for $\mathrm{d}_{\mathrm{op}}(\mathsf{M},\mathsf{N})$	Lemma 13
Data processing inequality	$\mathrm{d}_{\mathrm{op}}(\Lambda\circ\mathsf{M}\circ\Gamma,\Lambda\circ\mathsf{N}\circ\Gamma)\leq\mathrm{d}_{\mathrm{op}}(\mathsf{M},\mathsf{N})$ for CPTP Γ, stochastic Λ	$\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\Lambda\circ\mathsf{M}\circ\Phi,\Lambda\circ\mathsf{N}\circ\Phi)\leq\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{M},\mathsf{N})$ for unital Φ, stochastic Λ	Lemma 14

TABLE II.

Proof. We note first that according to Eq. (33), $\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{M},\mathsf{N})$ is proportional to the sum of non-negative terms of the form

$$\sqrt{\lVert M_i-N_i\rVert_{\mathrm{HS}}^2+\operatorname{tr}(M_i-N_i)^2}\,. \tag{70}$$

First we note, that both terms, treated as a functions $(M,N)\mapsto\lVert M-N\rVert_{\mathrm{HS}}$ and $(M,N)\mapsto\lvert\operatorname{tr}(M-N)\rvert$ satisfies triangle inequality, moreover the function $(a,b)\mapsto\sqrt{\lvert a\rvert^2+\lvert b\rvert^2}$ is subadditive and increasing in each argument. Therefore $\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{M},\mathsf{N})$ obeys triangle inequality. Symmetry, absolute homogeneity and identity of indiscernibles follows from direct inspection. □

Lemma 12 ($\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}$ is subadditive)**.** *For arbitrary quantum measurements* $\mathsf{M}_1,\mathsf{N}_1\in\mathrm{P}(\mathcal{H}_d,n),\mathsf{M}_2,\mathsf{N}_2\in\mathrm{P}(\mathcal{H}_{d'},n')$*, we have*

$$\mathrm{d}^{m}_{\mathrm{av}}(\mathsf{M}_1\otimes\mathsf{M}_2,\mathsf{N}_1\otimes\mathsf{N}_2)\leq\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{M}_1,\mathsf{N}_1)+\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{M}_2,\mathsf{N}_2)\,. \tag{71}$$

Proof. By triangle inequality we have

$$\mathrm{d}^{m}_{\mathrm{av}}(\mathsf{M}_1\otimes\mathsf{M}_2,\mathsf{N}_1\otimes\mathsf{N}_2)\leq\mathrm{d}^{m}_{\mathrm{av}}(\mathsf{M}_1\otimes\mathsf{M}_2,\mathsf{N}_1\otimes\mathsf{M}_2)+\mathrm{d}^{m}_{\mathrm{av}}(\mathsf{N}_1\otimes\mathsf{N}_2,\mathsf{N}_1\otimes\mathsf{M}_2). \tag{72}$$

Now we consider one of the terms form the right hand side of the inequality above and bound it by

$$\mathrm{d}^{m}_{\mathrm{av}}(\mathsf{M}_1\otimes\mathsf{M}_2,\mathsf{N}_1\otimes\mathsf{M}_2)\leq\mathrm{d}^{m}_{\mathrm{av}}(\mathsf{M}_1,\mathsf{N}_1). \tag{73}$$

The above inequality follows from direct calculations, since $\mathrm{d}^{m}_{\mathrm{av}}$ is a sum of square-roots of the formulas for single effect, for which we can write

$$\begin{aligned}&\lVert(M_1)_i\otimes(M_2)_j-(N_1)_i\otimes(M_2)_j\rVert_{\mathrm{HS}}^2+\operatorname{tr}((M_1)_i\otimes(M_2)_j-(N_1)_i\otimes(M_2)_j)^2\\&=\lVert(M_1)_i-(N_1)_i\rVert_{\mathrm{HS}}^2\lVert(M_2)_j\rVert_{\mathrm{HS}}^2+\operatorname{tr}((M_1)_i-(N_1)_i)^2\operatorname{tr}((M_2)_j)^2\\&\leq\operatorname{tr}((M_2)_j)^2\left(\lVert(M_1)_i-(N_1)_i\rVert_{\mathrm{HS}}^2+\operatorname{tr}((M_1)_i-(N_1)_i)^2\right)\,.\end{aligned} \tag{74}$$

Combining the terms above together with the fact, that $\sum_j\operatorname{tr}(M_2)_j=d'$, we obtain Eq. (73). Similarly we can bound $\mathrm{d}^{m}_{\mathrm{av}}(\mathsf{N}_1\otimes\mathsf{N}_2,\mathsf{N}_1\otimes\mathsf{M}_2)\leq\mathrm{d}^{m}_{\mathrm{av}}(\mathsf{N}_2,\mathsf{M}_2)$ which, together with Eq. (72) gives us the result. □

Lemma 13 ($\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}$ has joint-convexity property)**.** *For arbitrary sets of quantum measurements* $\{\mathsf{M}_\alpha\}_\alpha$, $\{\mathsf{N}_\alpha\}_\alpha$ *and probability distributions* $\{p_\alpha\}$*, we have*

$$\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}\left(\sum_\alpha p_\alpha\mathsf{M}_\alpha,\sum_\alpha p_\alpha\mathsf{N}_\alpha\right)\leq\sum_\alpha p_\alpha\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}\left(\mathsf{M}_\alpha,\mathsf{N}_\alpha\right)\,. \tag{75}$$

Proof. The proof is analogous to the one for states, and follows from triangle inequality, and absolute homogeneity:

$$\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}\left(\sum_\alpha p_\alpha\mathsf{M}_\alpha,\sum_\alpha p_\alpha\mathsf{N}_\alpha\right)\leq\sum_\alpha\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}\left(p_\alpha\mathsf{M}_\alpha,p_\alpha\mathsf{N}_\alpha\right)=\sum_\alpha p_\alpha\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}\left(\mathsf{M}_\alpha,\mathsf{N}_\alpha\right)\,. \tag{76}$$

□

Lemma 14 (Data-processing inequalities for average-case distance between measurements)**.** *Average-case distance between quantum measurements is monotonic with respect to unital pre- and general post-processing, i.e. for a stochastic matrix* Λ *and a general unital CPTP map* Φ*, we have*

$$\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\Lambda\circ\mathsf{M}\circ\Phi,\Lambda\circ\mathsf{N}\circ\Phi)\leq\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{M},\mathsf{N})\,. \tag{77}$$

Proof. We will show, that average-case distance between measurements is monotonic with respect to post-processing. Since the outcome of measurement is classical, we will consider only classical post-processing, given by a stochastic matrix Λ. We denote by $\Delta_j = M_j - N_j$. We will use the fact, that each term in the sum which defines $\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathsf{M},\mathsf{N})$, is absolute homogeneous and obeys triangle inequality (see Eq. (33) and discussion under Eq. (70))

$$\begin{aligned}
&\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\Lambda\circ\mathsf{M},\Lambda\circ\mathsf{N}) = \\
&= \frac{1}{2d}\sum_i^n \sqrt{\mathrm{tr}\,(\sum_j \Lambda_{ij}\Delta_j)^2 + (\mathrm{tr}\sum_j \Lambda_{ij}\Delta_j)^2} \leq \\
&\leq \frac{1}{2d}\sum_{i,j}^n \Lambda_{ij}\sqrt{\mathrm{tr}\,(\Delta_j)^2 + (\mathrm{tr}\,\Delta_j)^2} = \\
&= \frac{1}{2d}\sum_j^n \sqrt{\mathrm{tr}\,(\Delta_j)^2 + (\mathrm{tr}\,\Delta_j)^2} = \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathsf{M},\mathsf{N})\,.
\end{aligned} \tag{78}$$

In order to show that average-case distance between quantum measurements is monotonic with respect to unital pre-processing, we consider a general unital CPTP map Φ. Note, that the adjoint map Φ^* is also unital and CPTP. Recall that we can look at adjoint action of the channel on effects of M as

$$\mathrm{tr}\,M_i\Phi(\rho) = \mathrm{tr}\,\Phi^*(M_i)\rho\,. \tag{79}$$

The fact, that Φ^* is unital assures us that M' with effects $\{\Phi^*(M_i)\}_i$ forms a POVM . Now we consider the basic terms, which define $\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}$, first we see (again using Uhlmann's theorem [63] and Schur convexity of HS-norm)

$$\|\Phi^*(\Delta_i)\|_{\mathrm{HS}}^2 \leq \|\Delta_i\|_{\mathrm{HS}}^2\,. \tag{80}$$

Next, since Φ^* is trace preserving we have

$$\mathrm{tr}(\Phi^*(\Delta_i))^2 = \mathrm{tr}(\Delta_i)^2\,, \tag{81}$$

which finishes the proof of monotonicity with respect to unital pre-processing. □

Lemma 15 (Separation between $\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}$ and d_{op})**.** *For any quantum measurements* $\mathsf{M},\mathsf{N}\in\mathrm{P}(\mathcal{H}_d)$*, we have*

$$a\;\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathsf{M},\mathsf{N}) \;\leq\; \mathrm{d}_{\mathrm{op}}(\mathsf{M},\mathsf{N}) \leq d\;\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathsf{M},\mathsf{N})\,, \tag{82}$$

where $a = 0.31$.

Proof. The lower bound follows from Theorem 2. For upper bound, we directly calculate

$$\begin{aligned}
\mathrm{d}_{\mathrm{op}}(\mathsf{M},\mathsf{N}) &= \frac{1}{2}\max_\rho \sum_i |\,\mathrm{tr}(M_i - N_i)\rho| \leq \frac{1}{2}\sum_i \sqrt{\|(M_i - N_i)\|_{\mathrm{HS}}^2} \leq \\
&\leq \frac{1}{2}\sum_i \sqrt{\|(M_i - N_i)\|_{\mathrm{HS}}^2 + \mathrm{tr}(M_i - N_i)^2} = d\;\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(M,N)\,.
\end{aligned}$$

□

The following example attains bound in Lemma 15 up to a constant.

Example 3 (Swapping two outcomes of standard measurement)**.** *Consider computational-basis measurement* P *in* $\mathcal{H}_d$ *with effects* $P_i = |i\rangle\langle i|$*), and second measurement* M *that is obtained from* P *by exchanging first two effects, leaving others intact, i.e.,* $M_1 = |2\rangle\langle 2|\,, M_2 = |1\rangle\langle 1|$*, and* $M_i = |i\rangle\langle i|$ *for* $i = 3,\ldots,d$*. In this scenario, direct calculation yields*

$$\begin{aligned}
\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathsf{P},\mathsf{M}) &= \sqrt{2}\,\frac{1}{d}\,, \\
\mathrm{d}_{\mathrm{op}}(\mathsf{P},\mathsf{M}) &= 1\,.
\end{aligned}$$

The above implies that in asymptotic limit, similarly to Example 1 for states, considered measurements can be distinguished perfectly with optimal strategy, while randomized one will not work. On the other hand, if we interpret second measurement M *as noisy version of target* P*, then this peculiar type of noise (that swaps two measurements outcomes) will not highly affect results of generic experiments.*

We note that the above example, together with asymptotic separation, can be easily generalized to a scenario where second measurement, instead of swapping only 2 outcomes of P*, swaps some* constant *number of them.*

Example 4 (Counterexample for data-procesing inequality for quantum measurements)**.** *Consider POVMs* P *and* M *from previous Example 3. Consider now a* non-unital *channel* Λ *that regardless of the input state prepares a state* $|1\rangle\langle 1|$ *(which is a possible choice for optimal discriminator of POVMs* M *and* P*). Dual action of this channel on POVM's effects is* $\Lambda^\dagger(M_i) = \operatorname{tr}(M_i\,|1\rangle\langle 1|)\,\mathbb{I}$*. Direct calculation, together with results from previous example, yields*

$$\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{P}\circ\Lambda, \mathsf{M}\circ\Lambda) = \sqrt{1+\frac{1}{d}} \;>\; \sqrt{2}\,\frac{1}{d} = \mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{P},\mathsf{M})\,. \tag{83}$$

C. Quantum channels

The following Table III summarizes properties of average-case quantum distance between channels, and compares it to the worst-case diamond-norm distance. Compared to previous Tables, here we also consider two additional properties relevant for quantum channels, namely stability and chaining [62] which for average-case distance hold for *unital* quantum channels. Stability means that a given distance measure does not change if channel is extended by an identity channel. In other words trivial extensions of maps by ancillary system do not affect their distance measure. Chaining means that distance between multiple compositions of channel is at most sum of distances between constituting channels. If one sequence is a composition of target gates, and the other is their noisy version, this property implies that the total error is at most additive in a given distance measure.

Property	**Worst-case distance** $\mathrm{d}_\diamond(\Lambda,\Gamma)$	**Average-case distance** $\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma)$	**Text reference for average-case distance**
Function	$\|\|\Lambda-\Gamma\|\|_\diamond$	$\frac{1}{2}\sqrt{\|\mathcal{J}_\Lambda-\mathcal{J}_\Gamma\|^2_{\mathrm{HS}}+\operatorname{tr}\left((\Lambda-\Gamma)[\tau_d]^2\right)}$	Theorem 3, Lemma 16
Subadditivity	$\mathrm{d}_\diamond(\Lambda_1\otimes\Lambda_2,\Gamma_1\otimes\Gamma_2)\leq \mathrm{d}_\diamond(\Lambda_1,\Gamma_1)+\mathrm{d}_\diamond(\Lambda_2,\Gamma_2)$	same as for $\mathrm{d}_\diamond(\Lambda,\Gamma)$	Lemma 17
Joint convexity	$\mathrm{d}_\diamond\left(\sum_\alpha p_\alpha\Lambda_\alpha,\sum_\alpha p_\alpha\Gamma_\alpha\right)\leq\sum_\alpha p_\alpha\mathrm{d}_\diamond\left(\Lambda_\alpha,\Gamma_\alpha\right)$	same as for $\mathrm{d}_\diamond(\Lambda,\Gamma)$	Lemma 18
Data processing inequality	$\mathrm{d}_\diamond(\Phi_o\circ\Lambda\circ\Phi_i,\Phi_o\circ\Gamma\circ\Phi_i)\leq\mathrm{d}_\diamond(\Lambda,\Gamma)$ for CPTP $\Phi_i,\ \Phi_o$	$\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Phi_o\circ\Lambda\circ\Phi_i,\Phi_o\circ\Gamma\circ\Phi_i)\leq\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma)$ for unital $\Phi_i,\ \Phi_o$	Lemma 19
Stability	$\mathrm{d}_\diamond(\Lambda\otimes\mathcal{I},\Gamma\otimes\mathcal{I})=\mathrm{d}_\diamond(\Lambda,\Gamma)$ for CPTP Λ,Γ	$\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda\otimes\mathcal{I},\Gamma\otimes\mathcal{I})=\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma)$ for unital Λ,Γ	Lemma 20
Chaining	$\mathrm{d}_\diamond(\Lambda_1\circ\Lambda_2,\Gamma_1\circ\Gamma_2)\leq\mathrm{d}_\diamond(\Lambda_1,\Gamma_1)+\mathrm{d}_\diamond(\Lambda_2,\Gamma_2)$ for CPTP $\Lambda_1,\Lambda_2,\Gamma_1,\Gamma_2$	$\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda_1\circ\Lambda_2,\Gamma_1\circ\Gamma_2)\leq\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda_1,\Gamma_1)+\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda_2,\Gamma_2)$ for unital $\Lambda_1,\Lambda_2,\Gamma_1,\Gamma_2$	Lemma 21

TABLE III.

Lemma 16 ($\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}$ fulfils axioms of a metric)**.** *Let* $\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}$*, denote average distances between channels (Eq. (36)). Then* $\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}$ *satisfies axioms of a metric in space of quantum channels. Specifically, it satisfies triangle inequality, symmetry and identity of indiscernibles:*

$$\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma)\leq\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Phi)+\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Phi,\Gamma)\quad \textit{for all}\quad \Lambda,\Gamma,\Phi\in\mathrm{CPTP}(\mathcal{H}_d) \tag{84}$$

$$\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma)=\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Gamma,\Lambda)\quad \textit{for all}\quad \Lambda,\Gamma\in\mathrm{CPTP}(\mathcal{H}_d) \tag{85}$$

$$\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma)=0\iff\Lambda=\Gamma\quad \textit{for all}\quad \Lambda,\Gamma\in\mathrm{CPTP}(\mathcal{H}_d)\,. \tag{86}$$

Note, that $\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma)$ *is absolute homogeneous, i.e.* $\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(s\Lambda,s\Gamma)=|s|\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma)$.

Proof. Note that $\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}$ is a function of a distance measure ($\|\mathcal{J}_\Lambda-\mathcal{J}_\Gamma\|_{HS}$) and a term ($\sqrt{\operatorname{tr}\left((\Lambda-\Gamma)[\tau_d]^2\right)}$), which treated as a function, obeys triangle inequality. Since the function $(a,b)\mapsto\sqrt{|a|^2+|b|^2}$ is subadditive and increasing in each argument, thus $\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}$ obeys triangle inequality. Symmetry and identity of indiscernibles follows from direct inspection. □

Lemma 17 ($\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}$ is subadditive)**.** *For arbitrary quantum channels* $\Lambda_1,\Gamma_1\in\mathrm{CPTP}(\mathcal{H}),\Lambda_2,\Gamma_2\in\mathrm{CPTP}(\mathbb{C}^{d'})$*, we have*

$$\mathrm{d}^{ch}_{\mathrm{av}}(\Lambda_1\otimes\Lambda_2,\Gamma_1\otimes\Gamma_2)\leq\mathrm{d}_{\mathrm{av}}(\Lambda_1,\Gamma_1)+\mathrm{d}_{\mathrm{av}}(\Lambda_2,\Gamma_2)\,. \tag{87}$$

Proof. We begin with triangle inequality

$$\mathrm{d}_{\mathrm{av}}^{ch}(\Lambda_1 \otimes \Lambda_2, \Gamma_1 \otimes \Gamma_2) \leq \mathrm{d}_{\mathrm{av}}^{ch}(\Lambda_1 \otimes \Lambda_2, \Gamma_1 \otimes \Lambda_2) + \mathrm{d}_{\mathrm{av}}^{ch}(\Gamma_1 \otimes \Gamma_2, \Gamma_1 \otimes \Lambda_2)\,. \tag{88}$$

Now we consider

$$\mathrm{d}_{\mathrm{av}}^{ch}(\Lambda_1 \otimes \Lambda_2, \Gamma_1 \otimes \Lambda_2) = \sqrt{\|\mathcal{J}_{\Lambda_1\otimes\Lambda_2} - \mathcal{J}_{\Gamma_1\otimes\Lambda_2}\|_{\mathrm{HS}}^2 + \mathrm{tr}\left(((\Lambda_1 - \Gamma_1)\otimes\Lambda_2)\left(\frac{\mathbb{I}_{dd'}}{dd'}\right)^2\right)}\,. \tag{89}$$

First we note, that $\mathcal{J}_{\Lambda_1\otimes\Lambda_2}$ is permutationally similar, to $\mathcal{J}_{\Lambda_1}\otimes\mathcal{J}_{\Lambda_2}$, in fact we have $\mathcal{J}_{\Lambda_1\otimes\Lambda_2} = \mathbb{S}_{23}\left(\mathcal{J}_{\Lambda_1}\otimes\mathcal{J}_{\Lambda_2}\right)\mathbb{S}_{23}$, where $\mathbb{S}_{23}$ is the swap of the second and third subsystems, i.e. Λ_1-input system and Λ_2-output system. Therefore

$$\begin{aligned}\|\mathcal{J}_{\Lambda_1\otimes\Lambda_2} - \mathcal{J}_{\Gamma_1\otimes\Lambda_2}\|_{\mathrm{HS}} &= \|\mathcal{J}_{\Lambda_1}\otimes\mathcal{J}_{\Lambda_2} - \mathcal{J}_{\Gamma_1}\otimes\mathcal{J}_{\Lambda_2}\|_{\mathrm{HS}}\\ &= \|\mathcal{J}_{\Lambda_1} - \mathcal{J}_{\Gamma_1}\|_{\mathrm{HS}}\|\mathcal{J}_{\Lambda_2}\|_{\mathrm{HS}}\\ &\leq \|\mathcal{J}_{\Lambda_1} - \mathcal{J}_{\Gamma_1}\|_{\mathrm{HS}}\,.\end{aligned} \tag{90}$$

Next we note

$$\begin{aligned}\mathrm{tr}\left(((\Lambda_1 - \Gamma_1)\otimes\Lambda_2)\left(\frac{\mathbb{I}_{dd'}}{dd'}\right)^2\right) &= \mathrm{tr}\left((\Lambda_1 - \Gamma_1)\left(\frac{\mathbb{I}_d}{d}\right)^2\right)\mathrm{tr}\left(\Lambda_2\left(\frac{\mathbb{I}_{d'}}{d'}\right)^2\right)\\ &\leq \mathrm{tr}\left((\Lambda_1 - \Gamma_1)\left(\frac{\mathbb{I}_d}{d}\right)^2\right).\end{aligned} \tag{91}$$

Combining above we get

$$\mathrm{d}_{\mathrm{av}}^{ch}(\Lambda_1 \otimes \Lambda_2, \Gamma_1 \otimes \Lambda_2) \leq \sqrt{\|\mathcal{J}_{\Lambda_1} - \mathcal{J}_{\Gamma_1}\|_{\mathrm{HS}}^2 + \mathrm{tr}\left((\Lambda_1 - \Gamma_1)\left(\frac{\mathbb{I}}{d}\right)^2\right)} = \mathrm{d}_{\mathrm{av}}^{ch}(\Lambda_1, \Gamma_1)\,. \tag{92}$$

We can analogously bound $\mathrm{d}_{\mathrm{av}}^{ch}(\Gamma_1\otimes\Gamma_2, \Gamma_1\otimes\Lambda_2) \leq \mathrm{d}_{\mathrm{av}}^{ch}(\Gamma_2, \Lambda_2)$ and using Eq. (88) we obtain the result. □

Lemma 18 ($\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}$ has joint-convexity property)**.** *For arbitrary sets of quantum channels* $\{\Lambda_\alpha\}_\alpha$, $\{\Gamma_\alpha\}_\alpha$ *and probability distributions* $\{p_\alpha\}$*, we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}\left(\sum_\alpha p_\alpha\Lambda_\alpha, \sum_\alpha p_\alpha\Gamma_\alpha\right) \leq \sum_\alpha p_\alpha \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}\left(\Lambda_\alpha, \Gamma_\alpha\right)\,. \tag{93}$$

Proof. The proof is analogous to the one for states and measurements, and follows from triangle inequality, and absolute homogeneity of $\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}$. □

Lemma 19 (Data-processing inequalities for average-case distance between channels)**.** *Average-case distance between quantum channels is monotonic with respect to unital pre- and postprocessing, i.e. for a unital maps* Φ_o, Φ_i*, we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Phi_o\circ\Lambda\circ\Phi_i, \Phi_o\circ\Gamma\circ\Phi_i) \leq \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda, \Gamma). \tag{94}$$

Proof. The inequality related to the postprocessing follows directly analogous results for for states, in order to show monotonicity with respect to preprocessing inequality we write, for unital Φ

$$\begin{aligned}&\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda\circ\Phi, \Gamma\circ\Phi)\\ &= \frac{1}{2}\sqrt{\|\mathcal{J}_{\Lambda\circ\Phi} - \mathcal{J}_{\Gamma\circ\Phi}\|_{\mathrm{HS}}^2 + \mathrm{tr}\left((\Lambda - \Gamma)[\frac{\mathbb{I}}{d}]\right)^2}.\end{aligned} \tag{95}$$

We can consider only the term $\|\mathcal{J}_{\Lambda\circ\Phi} - \mathcal{J}_{\Gamma\circ\Phi}\|_{\mathrm{HS}}$, since the second one does not change under preprocessing by a unital map. First we write a norm in terms of superoperators, i.e.

$$\|\mathcal{J}_{\Lambda\circ\Phi} - \mathcal{J}_{\Gamma\circ\Phi}\|_{\mathrm{HS}} = \left\|(\hat{\Lambda} - \hat{\Gamma})\hat{\Phi}\right\|_{\mathrm{HS}}, \tag{96}$$

where $\hat{\Lambda}$ denotes the superoperator matrix ([19]) of channel Λ. Now we use inequality

$$\|AB\|_{\mathrm{HS}} \leq \|A\|_{\mathrm{HS}} \|B\|_{\infty} \tag{97}$$

and write

$$\|\mathcal{J}_{\Lambda\circ\Phi} - \mathcal{J}_{\Gamma\circ\Phi}\|_{\mathrm{HS}} \leq \left\|\hat{\Lambda} - \hat{\Gamma}\right\|_{\mathrm{HS}} \left\|\hat{\Phi}\right\|_{\infty} . \tag{98}$$

Now since for any unital map we have $\left\|\hat{\Gamma}\right\|_{\infty} = 1$ (see [65, Theorem 1]), we obtain the result. □

Lemma 20 (Stability property of average-case distance between unital channels)**.** *Average-case distance between unital quantum channels fulfills stability property [62], i.e. for unital maps Λ, Γ, and identity channel $\mathcal{I}$ acting on arbitrary dimension, we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda \otimes \mathcal{I}, \Gamma \otimes \mathcal{I}) = \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda, \Gamma) . \tag{99}$$

Proof. For unital channels we have $\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda, \Gamma) = \frac{1}{2}||\mathcal{J}_{\Lambda} - \mathcal{J}_{\Gamma}||_{HS}$. We now recall that for any channels Λ, Γ, Choi matrix $\mathcal{J}_{\Lambda\otimes\Gamma}$ is permutationally similar to $\mathcal{J}_{\Lambda} \otimes \mathcal{J}_{\Gamma}$. This allows to rewrite HS norm as

$$||\mathcal{J}_{\Lambda\otimes\mathcal{I}} - \mathcal{J}_{\Gamma\otimes\mathcal{I}}||_{\mathrm{HS}} = ||(\mathcal{J}_{\Lambda} - \mathcal{J}_{\Gamma}) \otimes \mathcal{J}_{\mathcal{I}}||_{\mathrm{HS}} = ||\mathcal{J}_{\Lambda} - \mathcal{J}_{\Gamma}||_{\mathrm{HS}} \underbrace{||\mathcal{J}_{\mathcal{I}}||_{\mathrm{HS}}}_{=1} = ||\mathcal{J}_{\Lambda} - \mathcal{J}_{\Gamma}||_{\mathrm{HS}} ,$$

which concludes the proof. □

Remark 9. *For generic, non-unital channels, the expression for average-case distance has additional term* $\mathrm{tr}\left((\Gamma - \Lambda)(\frac{\mathbb{I}}{d})\right)^2$ *. If we extend our channels by identity $\mathcal{I}_{d'}$ on dimension d', this 'non-unitality' term changes to* $\mathrm{tr}\left(((\Gamma - \Lambda) \otimes \mathcal{I}_{d'})(\frac{\mathbb{I}}{dd'})\right)^2 = \frac{1}{d'}\,\mathrm{tr}\left((\Gamma - \Lambda)(\frac{\mathbb{I}}{d})\right)^2$*. Therefore, the contribution to the average-case distance of the 'non-unitality' decreases as d' increases. Note that this scenario corresponds to channel discrimination (via random circuits) with the use of an ancillary system.*

Lemma 21 (Chaining property of average-case distance between unital channels)**.** *Average-case distance between unital quantum channels fulfills chaining property [62], i.e. for unital maps $\Lambda_1, \Lambda_2, \Gamma_1, \Gamma_2$, we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_1 \circ \Lambda_2, \Gamma_1 \circ \Gamma_2) \leq \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_1, \Gamma_1) + \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_2, \Gamma_2). \tag{100}$$

Proof. To prove the theorem, we apply triangle inequality followed by data-processing inequality for unital channels (Lemma 19)

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_1 \circ \Lambda_2, \Gamma_1 \circ \Gamma_2) \leq \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_1 \circ \Gamma_2, \Gamma_1 \circ \Gamma_2) + \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_1 \circ \Gamma_2, \Lambda_1 \circ \Lambda_2) \leq \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_1, \Gamma_1) + \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_2, \Gamma_2) . \tag{101}$$

□

Remark 10. *We note that for generic, non-unital channels, the chaining property of average-case distance does not hold. To see that, we note that if we choose channels $\Lambda_1 = \Gamma_1$ to be the same, the chaining property effectively reduces to data-processing inequality, which we know that does not hold for generic channels (see below for counterexample).*

Lemma 22 (Separation between $\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}$ and $\mathrm{d}_{\diamond}$)**.** *For any quantum measurements $\Lambda, \Gamma \in \mathrm{CPTP}(\mathcal{H}_d)$, we have*

$$a^{\mathrm{ch}}\, \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda, \Gamma) \;\leq\; \mathrm{d}_{\diamond}(\Lambda, \Gamma) \leq d^{\frac{3}{2}}\, \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda, \Gamma) , \tag{102}$$

where $a^{\mathrm{ch}} = 0.087$.

Proof. The lower bound is a consequence of Theorem 3, note that the constant here can be improved. To show the other inequality we begin with the upper bound for diamond norm (see [66, Thm. 2] and [67, Prop. 1]), which for Hermiticity preserving operation can be written in our notation as

$$\mathrm{d}_{\diamond}(\Lambda, \Gamma) \leq \frac{d}{2} \left\|\mathrm{tr}_2(|\mathcal{J}_{\Lambda} - \mathcal{J}_{\Gamma}|)\right\|_{\infty} , \tag{103}$$

Next express the operator norm via maximisation over pure states on the first subsystem

$$\left\|\mathrm{tr}_2(|\mathcal{J}_{\Lambda} - \mathcal{J}_{\Gamma}|)\right\|_{\infty} = \max_{\psi\in\mathcal{S}(\mathcal{H}_d)} |\,\mathrm{tr}\,(\psi \otimes \mathbb{I}_d |\mathcal{J}_{\Lambda} - \mathcal{J}_{\Gamma}|)\,| . \tag{104}$$

Applying to the above Cauchy-Schwarz inequality we obtain

$$\left\|\mathrm{tr}_2(|\mathcal{J}_{\Lambda} - \mathcal{J}_{\Gamma}|)\right\|_{\infty} \leq \max_{\psi\in\mathcal{S}(\mathcal{H}_d)} \|\psi \otimes \mathbb{I}_d\|_{\mathrm{HS}} \|\mathcal{J}_{\Lambda} - \mathcal{J}_{\Gamma}\|_{\mathrm{HS}} = \sqrt{d}\, \|\mathcal{J}_{\Lambda} - \mathcal{J}_{\Gamma}\|_{\mathrm{HS}} . \tag{105}$$

Combining above we obtain the desired result

$$\mathrm{d}_{\diamond}(\Lambda, \Gamma) \leq \frac{d}{2}\sqrt{d}\, \|\mathcal{J}_{\Lambda} - \mathcal{J}_{\Gamma}\|_{\mathrm{HS}} \leq d^{\frac{3}{2}}\, \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda, \Gamma) . \tag{106}$$

□

Example 5 (Separation example)**.** *Let us consider even dimensional Hilbert space $\mathcal{H}_d$ and a Hermitian matrix A , such that* $\operatorname{tr} A = 0$ *and $A^2 = \mathbb{I}_d$. Next we define a pair of channels Λ and Γ by their Jamiołkowski states as*

$$\begin{aligned} \mathcal{J}_\Lambda &= \frac{1}{d^2}\,\mathbb{I}_{d^2}\,, \\ \mathcal{J}_\Gamma &= \frac{1}{d^2}\,\mathbb{I}_{d^2} - \frac{1}{d^2}\,|\psi\rangle\langle\psi| \otimes A\,, \end{aligned} \tag{107}$$

where $\psi \in \mathcal{S}(\mathcal{H}_d)$ is an arbitrary pure state. The diamond norm between Λ and Γ can be calculated easily, using alternative formula for diamond norm for Hermiticity preserving operations (see e.g. [68, Eqn. (11)]), i.e.

$$\begin{aligned} \|\Lambda - \Gamma\|_\diamond &= d \max_{\rho\in \mathrm{D}(\mathcal{H}_d)} \|(\sqrt{\rho}\otimes\mathbb{I})\mathcal{J}_{\Lambda-\Gamma}(\sqrt{\rho}\otimes\mathbb{I})\|_1 = d \max_{\rho\in \mathrm{D}(\mathcal{H}_d)} \|(\sqrt{\rho}\otimes\mathbb{I})(\frac{1}{d^2}\,|\psi\rangle\langle\psi|\otimes A)(\sqrt{\rho}\otimes\mathbb{I})\|_1 \\ &= \frac{1}{d}\|A\|_1 = 1. \end{aligned} \tag{108}$$

The average distance, can be evaluated as

$$\begin{aligned} \mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma) &= \frac{1}{2}\sqrt{\|\mathcal{J}_{\Lambda-\Gamma}\|^2_{\mathrm{HS}} + \|\Lambda(\mathbb{I}/d) - \Gamma(\mathbb{I}/d)\|^2_{\mathrm{HS}}} \\ &= \frac{1}{2}\sqrt{\|\frac{1}{d^2}\,|\psi\rangle\langle\psi|\otimes A\|^2_{\mathrm{HS}} + \|\frac{1}{d^2}\operatorname{tr}_1(|\psi\rangle\langle\psi|\otimes A)\|^2_{\mathrm{HS}}} \\ &= \frac{1}{2}\sqrt{\frac{1}{d^4}\|A\|^2_{\mathrm{HS}} + \frac{1}{d^4}\|A\|^2_{\mathrm{HS}}} = \frac{\sqrt{2}}{2d^{\frac{3}{2}}}. \end{aligned} \tag{109}$$

Which gives us finally the separation of order $d^{\frac{3}{2}}$,

$$\frac{1}{2} = \mathrm{d}_\diamond(\Lambda,\Gamma) = d^{\frac{3}{2}}\frac{1}{\sqrt{2}}\,\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma). \tag{110}$$

Example 6 (Counterexample for general post-processing monotonicity for quantum channels)**.** *Consider two state-preparation channels acting on N-qubit system as $\Lambda(\rho) = \operatorname{tr}(\rho)\,|0\rangle\langle 0| \otimes \frac{\mathbb{I}}{2^{N-1}}$ and $\Gamma(\rho) = \operatorname{tr}(\rho)\,|1\rangle\langle 1| \otimes \frac{\mathbb{I}}{2^{N-1}}$, for any input state $\rho \in \mathrm{D}(\mathcal{H})$. Then we have*

$$\begin{aligned} \mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma) &= \frac{1}{2}\sqrt{\frac{1}{d}(1+\frac{1}{d})} \\ \mathrm{d}_\diamond(\Lambda,\Gamma) &= 1\,, \end{aligned} \tag{111}$$

where expression for average-case distance follows from direct calculation, and the value of diamond norm follows from the fact that channels always prepare states that are orthogonal on first qubits, and thus can be perfectly distinguished.

Now consider additional non-unital *conditional state-preparation channel $\tilde{\Lambda}$ that acts as $\tilde{\Lambda}(|0\rangle\langle 0|\otimes\sigma) = \psi$ and $\tilde{\Lambda}(|1\rangle\langle 1|\otimes\sigma) = \psi^\perp$ for any σ, where $\psi, \psi^\perp$ are two orthogonal pure states. Note that composed action of the channels reduces to state-preparation channels $\tilde{\Lambda}\circ\Lambda(\rho) = \psi$ and $\tilde{\Lambda}\circ\Gamma(\rho) = \psi^\perp$ for any ρ. Direct computation together with Eq.* (111) *yields*

$$\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\tilde{\Lambda}\circ\Lambda, \tilde{\Lambda}\circ\Gamma) = \sqrt{\frac{1}{2}(1+\frac{1}{d})} \;>\; \frac{1}{2}\sqrt{\frac{1}{d}(1+\frac{1}{d})} = \mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma)\,. \tag{112}$$

Example 7 (Counterexample for general pre-processing monotonicity for quantum channels)**.** *Consider two perfectly distinguishable unitary channels of size $d > 2$, $\Lambda_U : \rho \mapsto U\rho U^\dagger$ and $\Lambda_V : \rho \mapsto V\rho V^\dagger$.*

The average distance can be calculated directly and is equal to (see also Example 12)

$$\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda_U,\Lambda_V) = \frac{1}{2}\sqrt{2 - \frac{2}{d^2}|\operatorname{tr} U^\dagger V|^2}\,. \tag{113}$$

Since channels Λ_U and Λ_V are perfectly distinguishable, let $|\psi\rangle$ be the optimal discriminator, i.e. the state for which $\langle\psi|\,U^\dagger V\,|\psi\rangle = 0$. Note, that in the case of unitary channels one does not need to attach an additional system in order to perform optimal discrimination. Now we consider a channel $\Gamma : \rho \mapsto \operatorname{tr}(\rho)\,|\psi\rangle\langle\psi|$, which prepares the optimal discriminator. We then have

$$\begin{aligned} \mathcal{J}_{\Lambda_U\circ\Gamma} &= \mathbb{I}/d \otimes U\,|\psi\rangle\langle\psi|\,U^\dagger\,, \\ \mathcal{J}_{\Lambda_V\circ\Gamma} &= \mathbb{I}/d \otimes V\,|\psi\rangle\langle\psi|\,V^\dagger\,. \end{aligned} \tag{114}$$

Direct computations yields the following result

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_U \circ \Gamma, \Lambda_V \circ \Gamma) = \frac{1}{2}\sqrt{\|\mathbb{I}/d \otimes (U|\psi\rangle\langle\psi|U^\dagger - V|\psi\rangle\langle\psi|V^\dagger)\|_{\mathrm{HS}}^2 + \mathrm{tr}(U|\psi\rangle\langle\psi|U^\dagger - V|\psi\rangle\langle\psi|V^\dagger)^2} = \frac{1}{2}\sqrt{\frac{2}{d}+2}. \quad (115)$$

Finally we obtain, that the data processing inequality for general pre-processing does not hold.

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_U \circ \Gamma, \Lambda_V \circ \Gamma) \;\; > \;\; \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_U, \Lambda_V). \quad (116)$$

If we choose $U = \mathbb{I}, V = diag(1,-1,1,\ldots,1)$ *we get* $\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_U,\Lambda_V) = \frac{1}{2}\sqrt{2-\frac{2}{d^2}(d-2)^2} = \frac{1}{d}\sqrt{2(d-1)}$.

In fact similar calculations can be performed on any distinguishable channels, with the pre-processing channel chosen to be the preparation of the optimal discriminator.

VI. EXAMPLES

In previous parts of the work we discussed some specific scenarios in which scaling of average-case quantum distances with system-size provided some insight into various areas of quantum information. In this part we investigate some further exemplary scenarios, and we provide discussion of some of the consequences of our findings.

1. Convergence to uniform distribution

One particularly interesting consequence of our main theorems is that average-case distances allow to easily study a convergence of average Total-Variation distance between the noisy distribution and the uniform distribution. To this aim, one needs to calculate an average-case distance between a noisy state, measurement, or channel, and the maximally mixed state, trivial POVM, or maximally depolarizing channel, respectively. We summarize those observations in the following Lemmas 23, 24, 25 – the proofs follow directly from Theorems 1, 2, and 3, respectively. In what follows we denote uniform distribution as $\mathbf{p}^{\mathrm{uniform}}$, meaning $p_i^{\mathrm{uniform}} = \frac{1}{d}$ for all $i = 1,\ldots,d$. All of the above Lemmas follow directly from Theorem 1 (states),Theorem 2 (measurements), and Theorem 3.

Lemma 23. *[Noisy states – convergence to uniform distribution] Let* ψ *be a pure state and* Λ *a quantum channel. Then we have*

$$\mathop{\mathbb{E}}_{U\sim\nu} \mathrm{TV}(\mathbf{p}^{\Lambda(\psi),U}, \mathbf{p}^{\mathit{uniform}}) \approx \mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\Lambda(\psi), \frac{\mathbb{I}}{d}) = \frac{1}{2}\sqrt{\mathrm{tr}\left((\Lambda(\psi))^2\right) - \frac{1}{d}}\,. \quad (117)$$

In the above, the $\approx$ *sign means the approximation in a sense of Eq. (24). The notation* $\mathbf{p}^{\Lambda(\psi),U}$ *is the same as for Theorem 1.*

From the above it follows that the convergence of noisy distribution to the uniform in random circuits setting is controlled by the purity of the output state. For quantum measurements and channels we have similar expressions.

Lemma 24. *[Noisy measurements – convergence to uniform distribution] Let* M *be a generic* d*-outcome quantum measurement on* d*-dimensional space, and* $\mathsf{M}^{\mathcal{I}}$ *a trivial POVM s.t.* $M_i^{\mathcal{I}} = \frac{\mathbb{I}}{d}$ *for each* $i = 1,\ldots,d$*. Then we have*

$$\mathop{\mathbb{E}}_{V\sim\nu} \mathrm{TV}(\mathbf{p}^{\mathsf{M},\psi_V}, \mathbf{p}^{\mathit{uniform}}) \approx \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathsf{M}, \mathsf{M}^{\mathcal{I}}) = \frac{1}{2d}\sum_{i=1}^{d}\sqrt{\mathrm{tr}\,(M_i^2) + (\mathrm{tr}\,M_i - 1)^2 - \frac{1}{d}}\,. \quad (118)$$

In the above, the $\approx$ *sign means the approximation in a sense of Eq. (32). The notation* $\mathbf{p}^{\mathsf{M},\psi_V}$ *is the same as for Theorem 2.*

Lemma 25. *[Noisy channels – convergence to uniform distribution] Let* Λ *be a generic quantum channel and* Λ_{dep} *be a maximally depolarizing channel, i.e.,* $\Lambda_{\mathrm{dep}}(\rho) = \frac{\mathbb{I}}{d}$ *for any state* ρ*. Then we have*

$$\mathop{\mathbb{E}}_{V\sim\nu}\mathop{\mathbb{E}}_{U\sim\nu} \mathrm{TV}(\mathbf{p}^{\Lambda,\psi_V,U}, \mathbf{p}^{\mathit{uniform}}) \approx \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda, \Lambda_{\mathrm{dep}}) = \frac{1}{2}\sqrt{\mathrm{tr}\,(\mathcal{J}_\Lambda^2) + \mathrm{tr}\left(\left(\Lambda(\frac{\mathbb{I}}{d})\right)^2\right) - \frac{1}{d}\left(1+\frac{1}{d}\right)} \quad (119)$$

In the above, the $\approx$ *sign means the approximation in a sense of Eq. (35). The notation* $\mathbf{p}^{\Lambda,\psi_V,U}$ *is the same as for Theorem 3.*

2. More examples

Example 8 (Two pure states)**.** *For two pure states ψ and ϕ we have*

$$\begin{aligned} \mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\psi,\phi) &= \frac{1}{\sqrt{2}}\sqrt{1-\mathrm{tr}(\psi\phi)}\,, \\ \mathrm{d}_{\mathrm{tr}}(\psi,\phi) &= \sqrt{1-\mathrm{tr}(\psi\phi)}\,. \end{aligned} \tag{120}$$

Therefore, in this case we see that $\mathrm{d}_{\mathrm{av}}(\psi,\phi) = \frac{1}{\sqrt{2}}\mathrm{d}_{\mathrm{tr}}(\psi,\phi)$, which gives only constant separation between average-case and worst-case scenarios.

The consequences of the above example are twofold, depending on the perspective we adopt. First, if we wish to perform a task of state discrimination between two pure states, then the above identity implies that there exists strategy which uses random quantum circuits that is worse then optimal strategy only by a constant. Second, if we treat ψ as our target state and ϕ as its noisy version affected by unwanted unitary rotation, then this type of noise will highly affect the quality of our results. Specifically, for generic quantum states it will behave similarly to the worst-case scenario.

Example 9. *[Pauli eigenstates and separable Pauli noise] Consider state $\psi^{pauli} = \otimes_{i=1}^{N}|\pm r_i\rangle\langle\pm r_i|$, where $r_i \in \{x,y,z\}$, i.e., $|\pm r_i\rangle$ is any Pauli eigenstate on qubit i (with eigenvalue $+1$ or -1.). Consider separable Pauli channel $\Lambda^{pauli} = \otimes_{i=1}^{N}\Lambda_i^{pauli}$, where single-qubit channel is $\Lambda_i^{pauli}(\rho) = \sum_{j=1} p_j^{(i)}\sigma_j\rho\sigma_j$ with $j \in \{1,x,y,z\}$, $\sigma_1 = \mathbb{I}$, and $p_j^{(i)} \geq 0$, $\sum_j p_j^{(i)} = 1$. Define $q^{(i)} = p_1^{(i)} + p_{r_i}^{(i)}$, i.e., a probability of applying on qubit i a gate that stabilizes the state of that qubit (namely, either identity or Pauli matrix of which $|\pm r_i\rangle$ is an eigenstate). Furthermore, assume that for each qubit i we have $q^{(i)} \geq \frac{1}{2}$. Then we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\Lambda^{pauli}(\psi^{pauli}),\frac{\mathbb{I}}{d}) = \frac{1}{2}\sqrt{\Pi_{i=1}^{N}\left(1-2q^{(i)}(1-q^{(i)})\right)-\frac{1}{d}}\,, \tag{121}$$

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\Lambda^{pauli}(\psi^{pauli}),\psi^{pauli}) = \frac{1}{2}\sqrt{1-2\Pi_{i=1}^{N}q^{(i)}+\Pi_{i=1}^{N}(1-2q^{(i)}(1-q^{(i)}))}\,, \tag{122}$$

Proof. We start by analyzing effects of Pauli noise on single-qubit Pauli eigenstate. We first write $|\pm r_i\rangle\langle\pm r_i| = \frac{1}{2}\left(\mathbb{I}\pm\sigma_{r_i}\right)$ and evaluate

$$\Lambda_i^{\mathrm{pauli}}(|\pm r_i\rangle\langle\pm r_i|) = \frac{1}{2}\left(\mathbb{I}\pm\left((p_1^{(i)}+p_r^{(i)}-p_{k\neq r_i}^{(i)}-p_{l\neq r_i}^{(i)}\right)\right) = \frac{1}{2}\left(\mathbb{I}\pm\left((2(p_1^{(i)}+p_r^{(i)})-1\right)\right) = \frac{1}{2}\left(\mathbb{I}\pm\left(2q^{(i)}-1\right)\right)\,, \tag{123}$$

where $p_{k\neq r_i}^{(i)}$ and $p_{l\neq r_i}^{(i)}$ are error probabilities corresponding to two Pauli matrices that are not σ_{r_i}. We now notice that the above state has two eigenvalues which are $\frac{1}{2}(1\pm|2q^{(i)}-1|)$, which for assumed regime $q^{(i)} \geq \frac{1}{2}$ gives eigenvalues $q^{(i)}$ and $(1-q^{(i)})$.

To get Eq. (121) we refer to Lemma 23 and use the fact that purity of separable states is a product of purities. For a single qubit, purity of noisy state is $\mathrm{tr}\left(\Lambda_i^{\mathrm{pauli}}\left(|\pm r_i\rangle\langle\pm r_i|\right)^2\right) = (q^{(i)})^2+(1-q^{(i)})^2 = 1-2q^{(i)}(1-q^{(i)})$, which for multiple qubits yields Eq (121).

To get Eq. (122), we first diagonalize all noisy Pauli states, getting global state represented as $\bigotimes_{i=1}^{N}\left(q^{(i)}|0\rangle\langle 0|+(1-q^{(i)})|1\rangle\langle 1|\right)$. In this basis, noiseless Pauli eigenstate is simply $|0\rangle\langle 0|^{\otimes N}$ (note that both states are simultaneously diagonalizable). Having this in mind, we want to decompose the distance between states $||\Lambda^{\mathrm{pauli}}(\psi^{\mathrm{pauli}})-\psi^{\mathrm{pauli}})||_{HS}^2$ into parts that are easy to handle. To this aim, we use the fact that for any states ρ and $\tilde{\rho}$, the HS distance can be written as $||\rho-\tilde{\rho}||_{HS}^2 = \mathrm{tr}\,\rho^2+\mathrm{tr}\,\tilde{\rho}^2-2\,\mathrm{tr}\,(\rho\,\tilde{\rho})$. In our case $\rho = \Lambda^{\mathrm{pauli}}(\psi^{\mathrm{pauli}})$ and $\tilde{\rho} = \psi^{\mathrm{pauli}}$. Since Pauli state is pure we get $\mathrm{tr}\,\tilde{\rho}^2 = 1$, while the purity of ρ was already calculated above. The cross-term can by evaluated by recalling that in basis we consider Pauli eigenstate is simply $|0\rangle\langle 0|^{\otimes N}$, we thus need to simply take value of the first matrix element of ρ, obtaining $\mathrm{tr}\,(\rho\,\tilde{\rho}) = \Pi_i q^{(i)}$. Summing up and inserting into definition of average-case distance yields Eq. (122). □

We now consider scenario where our target POVM is computational-basis measurement P, and we wish to calculate its distance from some other POVM M. This choice is motivated by the fact that in quantum computing the computational-basis measurement is often a model for ideal detector [2], and M can be thought of as its noisy implementation. In particular, we considered situation in which $\mathsf{M} = \mathrm{T}\,\mathsf{P}$, where T is a left-stochastic map, i.e., its columns' are probability distributions. Such noise is equivalent to classical post-processing of ideal statistics (i.e., probabilities one would have obtained on P), hence we call it classical noise. This is practically relevant scenario, as it has been experimentally observed that classical noise is a dominant type of readout noise in contemporary quantum devices based on superconducting qubits [69].

We now define, in analogy to average-case quantum distance, the *average-case classical distance* between POVMs M and N

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{classical}}(\mathsf{M},\mathsf{N}) := \mathop{\mathbb{E}}_{|k\rangle\langle k|}\mathrm{TV}\left(\mathbf{p}\left(|k\rangle\langle k|,\mathsf{M}\right),\mathbf{p}\left(|k\rangle\langle k|,\mathsf{N}\right)\right)\,, \tag{124}$$

where by $\mathbb{E}_{|k\rangle\langle k|}$ we denote average over all *classical* deterministic states $|k\rangle\langle k|$. The above distance turns out to be a helpful tool in investigating some of the properties of average-case quantum distances for quantum measurements. In considerations about distances between measurements, the following Lemma 26 and Lemma 27 proved useful.

Lemma 26 (Average-case quantum vs classical distance)**.** *Let* P *be measurement in computational basis, and* T *arbitrary stochastic map, i.e.,* $\sum_i T_{ij} = 1$*. Define POVM* $\mathrm{T}\mathsf{P}$ *via* $(\mathrm{T}\,\mathsf{P})_i = \sum_j \mathrm{T}_{ij} P_j$*. Then we have*

$$\frac{1}{2}\,\mathrm{d}_{\mathrm{av}}^{\mathrm{classical}}(\mathrm{T}\mathsf{P},\mathsf{P}) \leq \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathrm{T}\,\mathsf{P},\mathsf{P})\,. \tag{125}$$

Proof. We start by directly computing classical distance from Eq. (124)

$$\begin{aligned}\mathrm{d}_{\mathrm{av}}^{\mathrm{classical}}(\mathrm{T}\mathsf{P},\mathsf{P}) &= \frac{1}{2d}\sum_{k=1}^{d}\sum_{i=1}^{d}|\,\mathrm{tr}(|k\rangle\langle k|\,(\sum_j T_{ij}\,|j\rangle\langle j| - |i\rangle\langle i|))| = \frac{1}{2d}\sum_{k=1}^{d}\sum_{i=1}^{d}|T_{ik}-\delta_{k,i}| \\ &= \frac{1}{2d}\sum_{k=1}^{d}(1-T_{kk}+\sum_{i\neq k}T_{ik}) = \frac{1}{2d}\sum_{k=1}^{d}2(1-T_{kk}) = 1-\frac{\mathrm{tr}(T)}{d}\,,\end{aligned} \tag{126}$$

where we used the fact that T is left-stochastic, hence $\sum_{i\neq k} T_{ik} = 1 - T_{kk}$. Now we notice that

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathrm{T}\mathsf{P},\mathsf{P}) = \frac{1}{2d}\sum_{i=1}^{d}\sqrt{(1-\mathrm{T}_{ii})^2+(1-\sum_j T_{ij})^2+\sum_{k\neq i}T_{ik}^2} \geq \frac{1}{2}\,\frac{1}{d}\sum_{i=1}^{d}\sqrt{(1-T_{ii})^2} = \frac{1}{2}\,(1-\frac{\mathrm{tr}(T)}{d})\,, \tag{127}$$

thus the expression on the RHS of Eq. (127) is exactly equal to $\frac{1}{2}\mathrm{d}_{\mathrm{av}}^{\mathrm{classical}}(\mathrm{T}\mathsf{P},\mathsf{P})$, which concludes the proof. □

Lemma 27 (Distance of classical part of the measurement noise)**.** *Let* M *be an arbitrary* d*-outcome POVM, and* P *measurement in computational basis. Decompose* M *as* $\mathsf{M} = \mathrm{T}\,\mathsf{P} + \Delta$*, where* $\mathrm{T}\,\mathsf{P}$ *is a POVM obtained by taking only diagonal elements of operators* M*, i.e.,* $(\mathrm{T}\,\mathsf{P})_i := \mathrm{diag}(M_i)$ *. Then we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathrm{T}\,\mathsf{P},\mathsf{P}) \leq \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathsf{M},\mathsf{P})\,. \tag{128}$$

Proof. Consider action of (the dual of) completely dephasing noise $\Lambda_{\mathrm{deph}}^{\dagger}$ on POVMs' effects, namely $\Lambda_{\mathrm{deph}}^{\dagger}(M_i) = (\mathrm{T}\mathsf{P})_i$ (this is because to begin with we defined $\mathrm{T}\,\mathsf{P}$ as diagonal part of POVM M). Since dephasing noise is unital and it preserves computational-basis measurement P, the above property follows directly from data-processing inequality for unital pre-processing of quantum measurements proved in Lemma 14. □

Remark 11. *We note that while decomposition of POVM* $\mathsf{M} = \mathrm{T}\,\mathsf{P} + \Delta$ *into diagonal and off-diagonal parts may seem arbitrary, it has been in fact previously used in the context of measurement error-mitigation. In particular, the* $\mathrm{T}\mathsf{P}$ *can be interpreted as a "classical part" of the noise and if we are able to reconstruct* T *(for example, using Diagonal Detector Tomography), we can use it to reduce the noise via classical post-processing of the statistics estimated on faulty detector* M *[13, 70, 71].*

Corollary 2. *By combining Lemma 26 with Lemma 27, we immediately get that for any POVM decomposed into diagonal and off-diagonal part as* $\mathsf{M} = \mathrm{T}\,\mathsf{P} + \Delta$*, its distance from standard measurement can be bounded from below via*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathsf{M},\mathsf{P}) \geq \frac{1}{2}\,\mathrm{d}_{\mathrm{av}}^{\mathit{classical}}(\mathrm{T}\mathsf{P},\mathsf{P})\,. \tag{129}$$

Let us now consider a simplified scenario where target POVM is computational-basis measurement, and its noisy version corresponds to local, symmetric classical noise.

Example 10 (Computational basis and local symmetric bitflip)**.** *Let* P *denote measurement in computational basis and its noisy version* $\mathrm{T}^{sym}\mathsf{P}$ *affected by noise* $\mathrm{T} = \otimes_{i=1}^{N}\Lambda_i^{(\mathrm{sym})}$*, where* $\Lambda_i^{(\mathrm{sym})}$ *denotes local stochastic noise describing symmetric bitflip specified by parameter* p_i *(bitflip error probability). In this case we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}\,(\mathrm{T}^{sym}\mathsf{P},\mathsf{P}) = \frac{1}{2}\sqrt{1-2\Pi_{i=1}^{N}\,(1-p_i)+\Pi_{i=1}^{N}(1-2p_i(1-p_i))}, \tag{130}$$

$$\mathrm{d}_{\mathrm{op}}\,(\mathrm{T}^{sym}\mathsf{P},\mathsf{P}) = 1-\Pi_{i=1}^{N}\,(1-p_i)\,, \tag{131}$$

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathrm{T}^{sym}\mathsf{P},\mathsf{M}^{\mathcal{I}}) = \frac{1}{2}\sqrt{\Pi_{i=1}^{N}(1-2p_i(1-p_i))-\frac{1}{d}}\,, \tag{132}$$

where N *is the number of qubits.*

Proof. To obtain (130) we calculate explicitly

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathsf{P},\mathsf{TP}) = \frac{1}{2d}\sum_{i=1}^{d}\sqrt{(1-\mathrm{T}_{ii})^2+(1-\sum_j T_{ij})^2+\sum_{k\neq i}T_{ik}^2} = \frac{1}{2d}\sum_{i=1}^{d}\sqrt{(1-\mathrm{T}_{ii})^2+\sum_{k\neq i}T_{ik}^2}\,, \tag{133}$$

where first equality follows from the fact that T is bistochastic. Then we notice that for identical symmetric bitflip each term on RHS is the same, namely for each i we have

$$(1-\mathrm{T}_{ii}^{\mathrm{sym}})^2+\sum_{k\neq i}(\mathrm{T}_{ik}^{\mathrm{sym}})^2 = 1-2\mathrm{T}_{ii}^{\mathrm{sym}}+\sum_k(\mathrm{T}_{ik}^{\mathrm{sym}})^2\,. \tag{134}$$

Furthermore, from product structure of T it follows that

$$\mathrm{T}_{kk}^{\mathrm{sym}} = \Pi_{i=1}^{N}(1-p_i) \tag{135}$$

$$,\sum_k(\mathrm{T}_{lk}^{\mathrm{sym}})^2 = \Pi_{i=1}^{N}((1-p_i)^2+p_i^2)\,. \tag{136}$$

Summing over $i=1,\ \ldots,\ d$ yields Eq. (130).

To compute (131) we notice that in case of (any) stochastic noise T affecting standard measurement we have

$$\mathrm{d}_{\mathrm{op}}(\mathsf{TP},\mathsf{P}) = \max_j(1-T_{jj})\,, \tag{137}$$

which after substituting T_{jj} from Eq. (135) yields Eq. (131).

To obtain Eq. (132), we first notice that multiqubit symmetric bitflip is represented by bistochastic map that does not change trace – thus second term in Eq. (118) vanishes. Then we calculate explicitly purity of ith effect as

$$\mathrm{tr}\left(((\mathrm{T}^{\mathrm{sym}}\mathsf{P})_i)^2\right) = \sum_k(\mathrm{T}_{ik}^{\mathrm{sym}})^2\,. \tag{138}$$

Combining the above observations with Eq. (135) yields Eq. (132). □

Lemma 28 (Computational basis and local asymmetric bitflip)**.** *Consider a noisy version* $\mathrm{T}^{asym}\mathsf{P}$ *of computational basis measurement* P*, where* $\mathrm{T}^{asym} = \otimes_{i=1}^{N}\mathrm{T}_i^{asym}$ *is a separable,* asymmetric *stochastic map. For each qubit* i*, such map is characterized by two parameters,* $p_i(1|0)$ *and* $p_i(0|1)$*, specifying probability of erroneously measuring 1 (0) if the input state was* $|0\rangle$ *(*$|1\rangle$*). Define average error probability*

$$q_i^{av} = \frac{p_i(1|0)+p_i(0|1)}{2}\,, \tag{139}$$

and corresponding symmetric bitflip map $\mathrm{T}_i^{av} = (1-q_i^{av})\mathbb{I}+q_i^{av}\sigma_x$*, together with global map* $\mathrm{T}^{av} = \bigotimes_{i=1}^{N}\mathrm{T}_i^{av}$*. Then we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathrm{T}^{av}\mathsf{P},\mathsf{P}) \leq \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathrm{T}^{asym}\mathsf{P},\mathsf{P})\,, \tag{140}$$

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathrm{T}^{av}\mathsf{P},\mathsf{M}^{\mathcal{I}}) \leq \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathrm{T}^{asym}\mathsf{P},\mathsf{M}^{\mathcal{I}})\,. \tag{141}$$

Proof. The proof uses data-processing inequality for unital channels and stochastic post-processing proved in Lemma 14. The idea is to present a strategy that "symmetrizes" stochastic noise on each qubit via randomized measurements and post-processing (while not changing computational basis measurement P or trivial POVM $\mathsf{M}^{\mathcal{I}}$). Consider a strategy that applies combinations of X and $\mathbb{I}$ gates uniformly at random just before measurement, and then applies a post-processing strategy that combines the outcomes of measurements to "undo" the effects of the unital channel. Namely, for each qubit, if the applied gate was $\mathbb{I}$ do not do anything, and if it was X then swap the outcomes. Working out Kraus operators for this process shows that it corresponds to CPTP, unital map. Finally, it follows from direct calculation that if the initial POVM was $\mathrm{T}^{\mathrm{asym}}\mathsf{P}$, now the implemented POVM is exactly $\mathrm{T}^{\mathrm{av}}\mathsf{P}$. Clearly, such strategy does not affect the computational basis measurement (nor a trivial POVM $\mathsf{M}^{\mathcal{I}}$). Recalling Lemma‘14 concludes the proof.

We note that the above strategy was used for single-qubit error mitigation in Ref. [72], and more general multi-qubit versions were considered in context of noise characterization and mitigation Refs. [73–75]. □

From the Lemma 28 it follows that when studying separation between asymmetric stochastic noise and ideal measurement in computational basis, one can instead study symmetric noise with "average" error probability (Eq. (139)), which is easier to handle computationally. The same holds for studying separation from uniform distribution. The usefulness of this comes from the fact that asymmetric bitflip is more realistic model of measurement noise than symmetric bitflip, (see, e.g., [13, 70]).

We now consider a few interesting scenarios for distances between channels.

Example 11 (Two arbitrary state preparation channels)**.** *Denote by Λ_ρ and Λ_σ the state preparation channels that regardless of the input state always prepare state $\rho \in \mathrm{D}(\mathcal{H}_d)$ or $\sigma \in \mathrm{D}(\mathcal{H}_d)$, respectively. Then we have*

$$\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda_\rho, \Lambda_\sigma) = \sqrt{1 + \frac{1}{d}}\ \frac{1}{2}||\rho - \sigma||_{\mathrm{HS}} \,. \tag{142}$$

Example 12 (Two arbitrary unitary channels)**.** *Denote by Λ_U and Λ_V the unitary channels associated with unitaries U and V, i.e., $\Lambda_U(\rho) = U\rho U^\dagger$ for any state $\rho \in \mathrm{D}(\mathcal{H}_d)$. Then we have*

$$\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda_U, \Lambda_V) = \sqrt{\frac{1}{2}\left(1 - \frac{|\operatorname{tr}(U^\dagger V)|^2}{d^2}\right)} \,. \tag{143}$$

Example 13 (Identity channel and separable unitary rotations)**.** *Let $\mathcal{I}$ denote identity channel, and Λ_V be unitary channel corresponding to separable rotation $V = \bigotimes_{j=1}^N \exp(i\ \mathbf{n}_j \cdot \boldsymbol{\sigma}\ \frac{\phi_j}{2})$, where $|\mathbf{n}_j| = 1$ and $\phi_j > 0$. Assume that $\sum_{j=1}^N \phi_j \leq \frac{\pi}{2}$. Define $\phi_{\max} = \max_j \phi_j$ and $\phi_{\min} = \min_j \phi_j$. Then we have*

$$\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\mathcal{I}, \Lambda_V) \leq \sqrt{N}\ \frac{\phi_{\max}}{\sqrt{8}} \,, \tag{144}$$

$$\mathrm{d}_{\diamond}(\mathcal{I}, \Lambda_V) \geq \frac{1}{\sqrt{2}} N \phi_{\min} \,. \tag{145}$$

To obtain the above, we first note that since distances are unitarily invariant, we can rotate each unitary so it is a phase shift gate with with angle ϕ_j. To get first inequality, we calculate explicitly (see Example 12) $\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\mathcal{I}, \Lambda_V) = \sqrt{\frac{1}{2}(1 - \prod_{j=1}^N \cos^2(\frac{\phi_j}{2}))}$. Then we use inequality $\cos^2(\frac{\phi_j}{2}) \leq \cos^2(\frac{\phi_{\max}}{2})$ for $\phi_j \in [0, \pi]$, and employ inequalities $\cos(x)^2 \geq 1 - x^2$ and $(1-x)^N \geq 1 - Nx$. To get second inequality we calculate diamond norm explicitly $\mathrm{d}_{\diamond}(\mathcal{I}, \Lambda_V) = 2|\sin(\sum_{j=1}^N \frac{\phi_j}{2})|$ and employ inequality $|\sin(x)| \geq \frac{x}{2\sqrt{2}}$ for $x \in \left[0, \frac{\pi}{2}\right]$.

From derivations in the above example it follows that if we adopt perspective of average-case statistical distinguishability, any local coherent noise (when target operation is identity) can be viewed simply as a phase shift error. Furthermore, for angles such that $\frac{\phi_{\max}}{\phi_{\min}} = O(1)$, we see that worst-case distance grows quadratically faster than average-case.

Example 14. *[Separable Pauli noise in the middle of the circuit] Consider separable Pauli channel Λ^{pauli} defined in Example 9. Then we have*

$$\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda^{pauli}, \Lambda_{dep}) = \frac{1}{2}\sqrt{\Pi_{i=1}^N ||\mathbf{p}^i||_2^2 - \frac{1}{d^2}} \,, \tag{146}$$

$$\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda^{pauli}, \mathcal{I}) = \frac{1}{2}\sqrt{1 + \Pi_{i=1}^N ||\mathbf{p}^i||_2^2 - 2\Pi_{i=1}^N p_1^i} \,, \tag{147}$$

where $||\mathbf{p}^i||_2^2 = \sum_j (p_j^i)^2$ is a Euclidean norm of the vector of noise coefficients on ith qubit.

Proof. To begin the proof, we notice that the Pauli noise is mixed unitary channel and is thus unital. Since both completely depolarizing and identity channels are unital as well, in both average-case distances the terms that relate to action on maximally-mixed state equal 0. We are therefore left with the task of calculating Hilbert-Schmidt norms of relevant Choi matrices.

To show that Eq. (146) holds, we note that purity of separable Choi state is product of purities – this follows from the fact that any Choi matrix of product channel is permutationally similar to a tensor product of Choi matrices of those channels. We thus need to consider only single-qubit purity (note that this is analogous to proof for states in Example 9). Denote by $\mathcal{J}^{(i)}_{\mathrm{pauli}}$ a Choi matrix of Pauli channel on qubit i. By directly evaluating action of that channel on operators of the form $|k\rangle\langle l|$ (recall definition of Choi matrix) we explicitly write down matrix representation of $\mathcal{J}^{(i)}_{\mathrm{pauli}}$ and calculate

$$\operatorname{tr}\left(\mathcal{J}^{(i)}_{\mathrm{pauli}}\right)^2 = \frac{1}{4}\left(\operatorname{tr}\left(\Lambda^{(i)}_{\mathrm{pauli}}(|0\rangle\langle 0|)\right)^2 + \operatorname{tr}\left(\Lambda^{(i)}_{\mathrm{pauli}}(|1\rangle\langle 1|)\right)^2 + 2\operatorname{tr}\left(\Lambda^{(i)}_{\mathrm{pauli}}(|0\rangle\langle 1|)\left(\Lambda^{(i)}_{\mathrm{pauli}}(|0\rangle\langle 1|)\right)^\dagger\right)\right) \tag{148}$$

From direct evaluation we get that

$$\operatorname{tr}\left(\Lambda^{(i)}_{\mathrm{pauli}}(|k\rangle\langle k|)\right)^2 = (p_1 + p_{z_i})^2 + (p_{x_i} + p_{y_i})^2 \tag{149}$$

and

$$\mathrm{tr}\left(\Lambda^{(i)}_{\mathrm{pauli}}(|k\rangle\langle l|)\left(\Lambda^{(i)}_{\mathrm{pauli}}(|k\rangle\langle l|)\right)^{\dagger}\right) = (p_1 - p_{z_i})^2 + (p_{x_i} - p_{y_i})^2 \tag{150}$$

for $k \neq l$. Summing up everything we get that cross-terms cancel and $||\mathcal{J}^{(i)}_{\mathrm{pauli}}||^2_{HS} = \sum_j p_j^{(i)} = ||\mathbf{p^{(i)}}||_2^2$ which combined with Lemma 25 yields Eq. (146).

To get Eq. (147) we follow identical strategy as for Example 9. Namely, we recall the fact that for any states ρ and $\tilde{\rho}$, the HS distance can be written as $||\rho - \tilde{\rho}||^2_{HS} = \mathrm{tr}\,\rho^2 + \mathrm{tr}\,\tilde{\rho}^2 - 2\,\mathrm{tr}\,(\rho\,\tilde{\rho})$. Now in our case $\rho = \mathcal{J}_{\Lambda^{\mathrm{pauli}}}$ and $\tilde{\rho} = \mathcal{J}_{\mathcal{I}}$. Choi of identity channel is a maximally-entangled state, its purity is thus equal to 1, while purity of the Choi of the noisy channel was already calculated above. To evaluate cross-term, we note that it factorizes into product of single-qubit terms (as for purity, it follows from permutational equivalence between Choi matrix of product channel and tensor product of Choi matrices), each of them being equal to

$$\mathrm{tr}\left(\mathcal{J}^{(i)}_{\mathrm{pauli}}\mathcal{J}^{(i)}_{\mathcal{I}}\right) = \frac{1}{4}\sum_{k,l\in\{0,1\}} \mathrm{tr}\left(\Lambda\left(|k\rangle\langle l|\right)|l\rangle\langle k|\right). \tag{151}$$

This evaluates to

$$\mathrm{tr}\left(\Lambda^{(i)}_{\mathrm{pauli}}\left(|k\rangle\langle k|\right)|k\rangle\langle k|\right) = p_1^{(i)} + p_{z_i}^{(i)}, \tag{152}$$

and

$$\mathrm{tr}\left(\Lambda^{(i)}_{\mathrm{pauli}}\left(|k\rangle\langle l|\right)|k\rangle\langle l|\right) = p_1^{(i)} - p_{z_i}^{(i)}, \tag{153}$$

for $k \neq l$. Summing up we obtain $\mathrm{tr}\left(\mathcal{J}^{(i)}_{\mathrm{pauli}}\mathcal{J}_{\mathcal{I}}\right) = p_1^{(i)}$. Combining all of the above with definition of average-case distance yields Eq. (147). □

VII. OPEN PROBLEMS

Our work leaves many interesting problems left for future research. First important question one can ask is how to estimate average-case quantum distances in easy-to-implement setting. Natural candidate seems to be randomized-benchmarking types of experiments, as they also employ unitary designs [29, 41]. It would be also very interesting to connect quantum average-case distances with commonly used figures of merit used to assess quality of quantum devices. Those include measures such as average fidelity [41] (which is perhaps the most widely used quality measure), unitarity of quantum channels [29, 76, 77], or partitioned trace distances [78]. We note that the partitioned trace distances share with the average-case distances the property of being non-increasing under unital channels, which might suggest a deeper connection between the two. Furthermore, one can ask whether similar results can be obtained for different functions of output probability distributions such us classical fidelity or f-divergences [79, 80]. Another natural direction to pursue is to obtain better constants that appear in bounds (for example by considering higher moments) relating the average TV distance with quantum average-case distance. Another straightforward research direction is to check how the average-case quantum distances compare with worst-case distances for small subsets of qubits in actual quantum devices. For example, for pairs of qubits full Quantum Process Tomography [2] (or even Gate Set Tomography [81]) is possible, therefore one would be able to calculate the distances directly from objects in question. Such studies could provide some insight into what to expect from existing devices in worst and average-case scenarios.

Acknowledgements We would like to thank Richard Kueng, Victor Albert and Ingo Roth for interesting discussions and comments. We sincerely thank Susane Calegari for help with tables formatting and proofreading the manuscript. The authors acknowledge the financial support by TEAM-NET project co-financed by EU within the Smart Growth Operational Programme (contract no. POIR.04.04.00-00-17C1/18-00).

[1] F. B. Maciejewski, Z. Puchała, and M. Oszmaniec, Operational quantum average-case distances (2022), arXiv:2112.14283 [quant-ph].
[2] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010).
[3] F. G. Brandão, W. Chemissany, N. Hunter-Jones, R. Kueng, and J. Preskill, PRX Quantum **2**, 030316 (2021).
[4] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and et al., Nature Reviews Physics **3**, 625–644 (2021).

[5] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm (2014), arXiv:1411.4028 [quant-ph].
[6] E. Farhi and A. W. Harrow, Quantum supremacy through the quantum approximate optimization algorithm (2019), arXiv:1602.07674 [quant-ph].
[7] M. P. Harrigan *et al.*, Nature Physics **17**, 332 (2021).
[8] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, Nature Communications **5**, 10.1038/ncomms5213 (2014).
[9] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Nature **549**, 242 (2017).
[10] R. M. Parrish, E. G. Hohenstein, P. L. McMahon, and T. J. Martínez, Phys. Rev. Lett. **122**, 230401 (2019).
[11] C. Xue, Z.-Y. Chen, Y.-C. Wu, and G.-P. Guo, Chinese Physics Letters **38**, 030302 (2021).
[12] J. Marshall, F. Wudarski, S. Hadfield, and T. Hogg, IOP SciNotes **1**, 025208 (2020).
[13] F. B. Maciejewski, F. Baccari, Z. Zimborás, and M. Oszmaniec, Quantum **5**, 464 (2021).
[14] D. Stilck França and R. García-Patrón, Nature Physics **17**, 1221 (2021).
[15] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Nature Communications **9**, 4812 (2018), arXiv:1803.11173 [quant-ph].
[16] J. Watrous, Theory of Computing **5**, 217 (2009).
[17] Z. Puchała, L. Pawela, A. Krawiec, and R. Kukulski, Physical Review A **98**, 10.1103/physreva.98.042103 (2018).
[18] C. Fuchs and J. van de Graaf, IEEE Transactions on Information Theory **45**, 1216 (1999).
[19] I. Bengtsson and K. Zyczkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, 2006).
[20] W. Matthews, S. Wehner, and A. Winter, Communications in Mathematical Physics **291**, 813 (2009).
[21] C. Lancien and A. Winter, Communications in Mathematical Physics **323**, 555 (2013).
[22] G. De Palma, M. Marvian, D. Trevisan, and S. Lloyd, IEEE Transactions on Information Theory **67**, 6627–6643 (2021).
[23] S. Luo and Q. Zhang, Phys. Rev. A **69**, 032106 (2004).
[24] A. Acín, Physical Review Letters **87**, 10.1103/physrevlett.87.177901 (2001).
[25] M. D. Bowdrey, D. K. Oi, A. Short, K. Banaszek, and J. Jones, Physics Letters A **294**, 258–260 (2002).
[26] M. A. Nielsen, Physics Letters A **303**, 249–252 (2002).
[27] B. Schumacher, Phys. Rev. A **54**, 2614 (1996).
[28] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. A **60**, 1888 (1999).
[29] Y. Nakata, D. Zhao, T. Okuda, E. Bannai, Y. Suzuki, S. Tamiya, K. Heya, Z. Yan, K. Zuo, S. Tamate, Y. Tabuchi, and Y. Nakamura, PRX Quantum **2**, 030339 (2021).
[30] Z. Puchała, L. Pawela, A. Krawiec, R. Kukulski, and M. Oszmaniec, Quantum **5**, 425 (2021).
[31] J. S. Lundeen, A. Feito, H. Coldenstrodt-Ronge, K. L. Pregnell, C. Silberhorn, T. C. Ralph, J. Eisert, M. B. Plenio, and I. A. Walmsley, Nature Physics **5**, 27 (2008).
[32] L. Zhang, A. Datta, H. B. Coldenstrodt-Ronge, X.-M. Jin, J. Eisert, M. B. Plenio, and I. A. Walmsley, New Journal of Physics **14**, 115005 (2012).
[33] M. Endo, T. Sonoyama, M. Matsuyama, F. Okamoto, S. Miki, M. Yabuno, F. China, H. Terai, and A. Furusawa, Opt. Express **29**, 11728 (2021).
[34] F. Arute *et al.*, Nature **574**, 505 (2019).
[35] S. Aaronson, Shadow tomography of quantum states (2018), arXiv:1711.01053 [quant-ph].
[36] H.-Y. Huang, R. Kueng, and J. Preskill, Nature Physics **16**, 1050–1057 (2020).
[37] C. Hadfield, S. Bravyi, R. Raymond, and A. Mezzacapo, Measurements of quantum hamiltonians with locally-biased classical shadows (2020), arXiv:2006.15788 [quant-ph].
[38] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, PRX Quantum **2**, 10.1103/prxquantum.2.030348 (2021).
[39] C. Hadfield, Adaptive pauli shadows for energy estimation (2021), arXiv:2105.12207 [quant-ph].
[40] J. Emerson, R. Alicki, and K. Życzkowski, Journal of Optics B: Quantum and Semiclassical Optics **7**, S347–S352 (2005).
[41] E. Magesan, J. M. Gambetta, and J. Emerson, Physical Review Letters **106**, 10.1103/physrevlett.106.180504 (2011).
[42] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, and et al., Physical Review Letters **109**, 10.1103/physrevlett.109.080505 (2012).
[43] J. M. Gambetta, A. D. Córcoles, S. T. Merkel, B. R. Johnson, J. A. Smolin, J. M. Chow, C. A. Ryan, C. Rigetti, S. Poletto, T. A. Ohki, and et al., Physical Review Letters **109**, 10.1103/physrevlett.109.240504 (2012).
[44] J. Helsen, X. Xue, L. M. K. Vandersypen, and S. Wehner, A new class of efficient randomized benchmarking protocols (2019), arXiv:1806.02048 [quant-ph].
[45] S. T. Flammia, Averaged circuit eigenvalue sampling (2021), arXiv:2108.05803 [quant-ph].
[46] J. Helsen, M. Ioannou, I. Roth, J. Kitzinger, E. Onorati, A. H. Werner, and J. Eisert, arXiv e-prints , arXiv:2110.13178 (2021), arXiv:2110.13178 [quant-ph].
[47] A. Ambainis and J. Emerson, in *Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity*, CCC '07 (IEEE Computer Society, USA, 2007) p. 129–140.
[48] J. Radhakrishnan, M. Rötteler, and P. Sen, Algorithmica **55**, 490 (2009).
[49] J. Lee, M. S. Kim, and v. Brukner, Physical Review Letters **91**, 10.1103/physrevlett.91.087902 (2003).
[50] A. Peres, *Quantum Theory: Concepts and Methods* (Springer Netherlands, 2002).
[51] C. W. Helstrom, *Quantum detection and estimation theory*, Vol. 1 (1969) pp. 231–252.
[52] M. Navascués and S. Popescu, Phys. Rev. Lett. **112**, 140502 (2014).
[53] R. A. Low, *Pseudo-randomness and Learning in Quantum Computation*, Ph.D. thesis, - (2010).
[54] F. G. S. L. Brandão, A. W. Harrow, and M. Horodecki, Communications in Mathematical Physics **346**, 397 (2016), arXiv:1208.0692 [quant-ph].

[55] M. Oszmaniec, A. Sawicki, and M. Horodecki, IEEE Transactions on Information Theory **68**, 989 (2022).
[56] J. Haferkamp and N. Hunter-Jones, Phys. Rev. A **104**, 022417 (2021).
[57] J. L. W. V. Jensen, Acta Mathematica **30**, 175 (1906).
[58] B. Berger, SIAM Journal on Computing **26**, 1188 (1997), https://doi.org/10.1137/S0097539792240005.
[59] A. W. Harrow, The church of the symmetric subspace (2013), arXiv:1308.6595 [quant-ph].
[60] R. Kueng, H. Zhu, and D. Gross, arXiv e-prints , arXiv:1609.08595 (2016), arXiv:1609.08595 [quant-ph].
[61] M. Oszmaniec, L. Guerini, P. Wittek, and A. Acín, Phys. Rev. Lett. **119**, 190501 (2017).
[62] A. Gilchrist, N. K. Langford, and M. A. Nielsen, Physical Review A **71**, 10.1103/physreva.71.062310 (2005).
[63] P. M. Alberti and A. Uhlmann, *Stochasticity and partial order* (Deutscher Verlag der Wissenschaften Berlin, 1982).
[64] F. Buscemi, M. Keyl, G. M. D'Ariano, P. Perinotti, and R. F. Werner, Journal of Mathematical Physics **46**, 082109 (2005), arXiv:quant-ph/0505095 [quant-ph].
[65] W. Roga, Z. Puchała, Ł. Rudnicki, and K. Życzkowski, Physical Review A **87**, 032308 (2013).
[66] A. Jenčová and M. Plávala, Journal of Mathematical Physics **57**, 122203 (2016).
[67] I. Nechita, Z. Puchała, Ł. Pawela, and K. Życzkowski, Journal of Mathematical Physics **59**, 052201 (2018).
[68] Z. Puchała, L. Pawela, A. Krawiec, and R. Kukulski, Phys. Rev. A **98**, 042103 (2018).
[69] F. B. Maciejewski, Z. Zimborás, and M. Oszmaniec, Quantum **4**, 257 (2020).
[70] M. R. Geller and M. Sun, Quantum Science and Technology **6**, 025009 (2021).
[71] S. Bravyi, S. Sheldon, A. Kandala, D. C. Mckay, and J. M. Gambetta, Physical Review A **103**, 10.1103/physreva.103.042605 (2021).
[72] M. Oszmaniec, F. B. Maciejewski, and Z. Puchała, Phys. Rev. A **100**, 012351 (2019).
[73] E. van den Berg, Z. K. Minev, and K. Temme, Physical Review A **105**, 10.1103/physreva.105.032620 (2022).
[74] A. W. R. Smith, K. E. Khosla, C. N. Self, and M. S. Kim, Science Advances **7**, 10.1126/sciadv.abi8009 (2021).
[75] S. Tang, C. Zheng, and K. Wang, Detecting and eliminating quantum noise of quantum measurements (2022).
[76] B. Dirkse, J. Helsen, and S. Wehner, Physical Review A **99**, 10.1103/physreva.99.012315 (2019).
[77] M. Girling, C. Cirstoiu, and D. Jennings, arXiv e-prints , arXiv:2104.04352 (2021), arXiv:2104.04352 [quant-ph].
[78] A. E. Rastegin, Quantum Information Processing **9**, 61 (2010).
[79] D. Petz, Reports on Mathematical Physics **23**, 57 (1986).
[80] M. Jarzyna and J. Kolodynski, IEEE Journal on Selected Areas in Information Theory **1**, 367–386 (2020).
[81] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Quantum **5**, 557 (2021).

Supplementary Material: Exploring Quantum Average-Case Distances: Proofs, properties, and examples

Filip B. Maciejewski,[1] Zbigniew Puchała,[2,3] and Michał Oszmaniec[1]

[1]*Center for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warszawa, Poland*

[2]*Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, 44-100 Gliwice, Poland*

[3]*Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, 30-348 Kraków, Poland*

Here we provide proofs of more technical results from the main part – proof of Lemma 5 and proofs of main Theorems 1, 2, 3 for *approximate* 4-designs.

A. PROOF OF LEMMA 5

In what follows we prove Lemma 5, which we repeat here for Reader's convenience.

Lemma 1 (Repeated Lemma 5)**.** *Let $X, Y \in \mathrm{Herm}(\mathcal{H})$ be Hermitian operators acting on $\mathcal{H} \simeq \mathcal{H}$. Let $\mathbb{P}^{(k)}_{\mathrm{sym}}$ denotes the orthogonal projector onto k-fold symmetrization of $\mathcal{H}^{(k)}_{\mathrm{sym}} \subset \mathcal{H}^{\otimes k}$. We then have the following inequality*

$$\mathrm{tr}\left(X^{\otimes 2} \otimes Y^{\otimes 2}\, \mathbb{P}^{(4)}_{\mathrm{sym}}\right) \leq C\, \mathrm{tr}\left(X^{\otimes 2}\, \mathbb{P}^{(2)}_{\mathrm{sym}}\right) \mathrm{tr}\left(Y^{\otimes 2}\, \mathbb{P}^{(2)}_{\mathrm{sym}}\right)\ ,\ \textit{where } C = \frac{13}{6}\ . \tag{A.1}$$

Proof. We begin by noting that, for Hermitian matrices A and B we have

$$\mathrm{tr}\left((A^{\otimes 2} \otimes B^{\otimes 2})(\mathbb{P}^{(2)}_{\mathrm{sym}} \otimes \mathbb{P}^{(2)}_{\mathrm{sym}})\right) = \frac{1}{4}(\mathrm{tr}(A^2) + (\mathrm{tr}(A))^2)(\mathrm{tr}(B^2) + (\mathrm{tr}(B))^2). \tag{A.2}$$

We also have

$$\begin{aligned} 4!\ \mathrm{tr}\left((A^{\otimes 2}B^{\otimes 2})\mathbb{P}^{(4)}_{\mathrm{sym}}\right) &= ((\mathrm{tr}(A))^2 + \mathrm{tr}(A^2))((\mathrm{tr}(B))^2 + \mathrm{tr}(B^2)) \\ &\quad + 4\,\mathrm{tr}(A)\,\mathrm{tr}(B)\,\mathrm{tr}(AB) + 4\,\mathrm{tr}(A)\,\mathrm{tr}(AB^2) + 4\,\mathrm{tr}(B)\,\mathrm{tr}(A^2B) \\ &\quad + 2(\mathrm{tr}(AB))^2 + 2\,\mathrm{tr}(A^2B^2) + 2\,\mathrm{tr}(ABAB). \end{aligned} \tag{A.3}$$

Now we consider the following difference for, with arbitrary scalar parameter c

$$\begin{aligned} &c\ \mathrm{tr}\left((A^{\otimes 2} \otimes B^{\otimes 2})(\mathbb{P}^{(2)}_{\mathrm{sym}} \otimes \mathbb{P}^{(2)}_{\mathrm{sym}})\right) - \mathrm{tr}\left((A^{\otimes 2} \otimes B^{\otimes 2})\mathbb{P}^{(4)}_{\mathrm{sym}}\right) = \\ &= \frac{1}{4!}\Big((6c-1)(\mathrm{tr}(A^2) + (\mathrm{tr}(A))^2)(\mathrm{tr}(B^2) + (\mathrm{tr}(B))^2) \\ &- 4\,\mathrm{tr}(A)\,\mathrm{tr}(B)\,\mathrm{tr}(AB) - 4\,\mathrm{tr}(A)\,\mathrm{tr}(AB^2) - 4\,\mathrm{tr}(B)\,\mathrm{tr}(A^2B) \\ &- 2(\mathrm{tr}(AB))^2 - 2\,\mathrm{tr}(A^2B^2) - 2\,\mathrm{tr}(ABAB)\Big). \end{aligned} \tag{A.4}$$

Now we will bound the terms which occur above using standard inequalities, to get

$$\begin{aligned} -4\,\mathrm{tr}(A)\,\mathrm{tr}(B)\,\mathrm{tr}(AB) &\geq -4|\,\mathrm{tr}(A)||\,\mathrm{tr}(B)|\sqrt{\mathrm{tr}(A^2)}\sqrt{\mathrm{tr}(B^2)} \geq -2(|\,\mathrm{tr}(A)|^2|\,\mathrm{tr}(B)|^2 + \mathrm{tr}(A^2)\,\mathrm{tr}(B^2)); \\ -4\,\mathrm{tr}(A)\,\mathrm{tr}(AB^2) &\geq -4|\,\mathrm{tr}(A)|\sqrt{\mathrm{tr}(A^2)}\sqrt{\mathrm{tr}(B^4)} \geq -4|\,\mathrm{tr}(A)|\sqrt{\mathrm{tr}(A^2)}\,\mathrm{tr}(B^2) \geq -2(|\,\mathrm{tr}(A)|^2 + \mathrm{tr}(A^2))\,\mathrm{tr}(B^2); \\ -4\,\mathrm{tr}(B)\,\mathrm{tr}(A^2B) &\geq -4|\,\mathrm{tr}(B)|\sqrt{\mathrm{tr}(B^2)}\sqrt{\mathrm{tr}(A^4)} \geq -4|\,\mathrm{tr}(B)|\sqrt{\mathrm{tr}(B^2)}\,\mathrm{tr}(A^2) \geq -2(|\,\mathrm{tr}(B)|^2 + \mathrm{tr}(B^2))\,\mathrm{tr}(A^2); \\ -2(\mathrm{tr}(AB))^2 &\geq -2\,\mathrm{tr}(A^2)\,\mathrm{tr}(B^2); \\ -2\,\mathrm{tr}(A^2B^2) &\geq -2\,\mathrm{tr}(A^2)\,\mathrm{tr}(B^2); \\ -2\,\mathrm{tr}(ABAB) &\geq -2\,\mathrm{tr}(A^2)\,\mathrm{tr}(B^2). \end{aligned} \tag{A.5}$$

Combining above inequalities, we will determine the value of parameter c, for which (A.4) is non-negative

$$\begin{aligned} & c\ \mathrm{tr}(A^{\otimes 2}\otimes B^{\otimes 2})(\mathbb{P}^{(2)}_{\mathrm{sym}}\otimes\mathbb{P}^{(2)}_{\mathrm{sym}}) - \mathrm{tr}(A^{\otimes 2}\otimes B^{\otimes 2})\mathbb{P}^{(4)}_{\mathrm{sym}} \geq \\ & = \frac{1}{4!}\Big((6c-1)(\mathrm{tr}\,A^2+(\mathrm{tr}\,A)^2)(\mathrm{tr}\,B^2+(\mathrm{tr}\,B)^2) - 12\,\mathrm{tr}\,A^2\,\mathrm{tr}\,B^2 \\ & - 2|\,\mathrm{tr}\,A|^2|\,\mathrm{tr}\,B|^2 - 2|\,\mathrm{tr}\,A|^2\,\mathrm{tr}\,B^2 - 2|\,\mathrm{tr}\,B|^2\,\mathrm{tr}\,A^2\Big) \\ & = \frac{1}{4!}\Big((6c-13)\,\mathrm{tr}\,A^2\,\mathrm{tr}\,B^2 + (6c-3)(\mathrm{tr}\,A^2(\mathrm{tr}\,B)^2 + (\mathrm{tr}\,A)^2\,\mathrm{tr}\,B^2 + (\mathrm{tr}\,A)^2(\mathrm{tr}\,B)^2)\Big). \end{aligned} \tag{A.6}$$

Note, that the above is larger than 0 for $c \geq 13/6$. □

B. PROOFS OF MAIN THEOREMS FOR δ-APPROXIMATE 4-DESIGNS

Here we outline the extension of proofs of Theorems 1, 2, and 3 for approximate 4-designs.

1. Quantum states and measurements

We will start with quantum states. Let us consider δ-approximate 4-design ν (recall Section II C), i.e., we have

$$\|\mathcal{T}_{4,\nu} - \mathcal{T}_{4,\mu}\|_\diamond \leq \delta\ , \tag{B.1}$$

where μ is the Haar measure in $\mathrm{U}(\mathcal{H}_d)$ and $\mathcal{T}_{4,\nu}$ is the quantum channel acting on $\mathcal{H}_d^{\otimes 4}$ defined as $\mathcal{T}_{4,\nu}(A) = \int_{\mathrm{U}(\mathcal{H}_d)} d\nu(U) U^{\otimes 4} A (U^\dagger)^{\otimes 4}$. For a measure $\nu = \{\nu_\alpha, U_\alpha\}$ on $\mathrm{U}(\mathcal{H}_d)$ let $\tilde{\nu}$ denote a measure supported on 'inverted gates' i.e. $\nu = \{\nu_\alpha, U_\alpha^\dagger\}$ (the generalization to non-discrete measures is straightforward). From the definition of the diamond norm and the identity $\mu = \tilde{\mu}$ it follows that

$$\|\mathcal{T}_{k,\nu} - \mathcal{T}_{k,\mu}\|_\diamond = \|\mathcal{T}_{k,\tilde{\nu}} - \mathcal{T}_{k,\mu}\|_\diamond\ . \tag{B.2}$$

Denote $X_{i,U} = \mathrm{tr}(|i\rangle\langle i|\, U\Delta U^\dagger)$, where $\Delta = \rho - \sigma$ for two quantum states $\rho, \sigma \in \mathrm{D}(\mathcal{H}_d)$ which we wish to compare. Using Berger's inequality (cf. Lemma 4) for every summand in the expression for the TV distance $\mathrm{TV}(\mathbf{p}^{\rho,U}, \mathbf{p}^{\sigma,U})$, where $U \sim \nu$ we get

$$\mathop{\mathbb{E}}_{U\sim\nu}|X_{i,U}| \geq \frac{\left(\mathop{\mathbb{E}}_{U\sim\nu} X_{i,U}^2\right)^{3/2}}{\left(\mathop{\mathbb{E}}_{U\sim\nu} X_{i,U}^4\right)^{1/2}}\ . \tag{B.3}$$

Our goal is to compare the right-hand side of the above expression with its counterpart evaluated using the Haar measure μ i.e. : $\left(\mathop{\mathbb{E}}_{U\sim\mu} X_{i,U}^2\right)^{\frac{3}{2}}\left(\mathop{\mathbb{E}}_{U\sim\mu} X_{i,U}^4\right)^{-\frac{1}{2}}$. We begin with the lower bound for the numerator of (B.3).

$$\begin{aligned} \mathop{\mathbb{E}}_{U\sim\nu} X_{i,U}^2 &\geq \mathop{\mathbb{E}}_{U\sim\mu} X_{i,U}^2 - \left|\mathrm{tr}(\mathcal{T}_{2,\mu} - \mathcal{T}_{2,\tilde{\nu}})[|i\rangle\langle i|^{\otimes 2}]\Delta^{\otimes 2}\right| \\ &\geq \frac{1}{d(d+1)}\,\mathrm{tr}(\Delta^{\otimes 2}) - \left\|\mathrm{tr}(\mathcal{T}_{2,\mu} - \mathcal{T}_{2,\tilde{\nu}})[|i\rangle\langle i|^{\otimes 2}]\right\|_1 \|\Delta\|_\infty^2 \\ &\geq \frac{1}{d(d+1)}\,\mathrm{tr}(\Delta^{\otimes 2}) - \delta\|\Delta\|_\infty^2 \\ &\geq \frac{1}{d(d+1)}\,\mathrm{tr}(\Delta^{\otimes 2})(1 - d(d+1)\delta). \end{aligned} \tag{B.4}$$

where we used standard inequalities $|\,\mathrm{tr}(AB)| \leq ||A||_1\ ||B||_\infty$, $||A||_1 \leq ||A||_\diamond$, $||A||_\infty^2 \leq ||A||_{\mathrm{HS}}^2$, the definition of the diamond norm and (B.1).

Next, we bound denominator from above using Lemma 4 and reasoning analogous as before

$$\begin{aligned} \mathop{\mathbb{E}}_{U\sim\nu} X_{i,U}^4 &\leq \mathop{\mathbb{E}}_{U\sim\mu} X_{i,U}^4 + \left|\mathrm{tr}(\mathcal{T}_{4,\mu} - \mathcal{T}_{4,\tilde{\nu}})[|i\rangle\langle i|^{\otimes 4}]\Delta^{\otimes 4}\right| \\ &\leq C\left(\mathop{\mathbb{E}}_{U\sim\mu} X_{i,U}^2\right)^2 + \delta\|\Delta\|_\infty^4 \\ &\leq C\frac{\mathrm{tr}(\Delta^2)^2}{(d(d+1))^2}\left(1 + \frac{(d(d+1))^4\delta}{C}\right)\ , \end{aligned} \tag{B.5}$$

with $C = 10.1$.

Combining above inequalities, we obtain that for δ approximate 4-desing, we have

$$\frac{\left(\mathop{\mathbb{E}}_{U\sim\nu} X_{i,U}^2\right)^{3/2}}{\left(\mathop{\mathbb{E}}_{U\sim\nu} X_{i,U}^4\right)^{1/2}} \geq \tilde{\ell}(\delta)\frac{\left(\mathop{\mathbb{E}}_{U\sim\mu} X_{i,U}^2\right)^{3/2}}{\left(\mathop{\mathbb{E}}_{U\sim\mu} X_{i,U}^4\right)^{1/2}} \tag{B.6}$$

with

$$\tilde{\ell}(\delta) = \frac{(1-d(d+1)\delta)^{3/2}}{\left(1+\frac{\delta(d(d+1))^2}{C}\right)^{1/2}} \geq \frac{(1-2d^2\delta)^{3/2}}{\left(1+\frac{4d^4\delta}{C}\right)^{1/2}} \geq \frac{(1-2d^2\delta)^{3/2}}{(1+2d^4\delta)^{1/2}} \tag{B.7}$$

where we used the fact that $x(x+1) \leq 2x^2$ and $x^2(x+1)^2 \leq 4x^4$ for any $x \geq 1$, and $\frac{2}{C} = \frac{2}{10.1} < 1$.

By setting $\delta = \frac{\delta'}{2d^4}$, we obtain

$$\tilde{l}(\delta') \geq \sqrt{\frac{(1-\frac{\delta'}{d^2})^3}{1+\delta'}} =: \ell(\delta') \,. \tag{B.8}$$

Using analogous reasoning for bounding $\mathop{\mathbb{E}}_{U\sim\nu} X_{i,U}^2$ from above, we obtain upper bound

$$\mathop{\mathbb{E}}_{U\sim\nu} |X_{i,U}| \leq \tilde{u}(\delta) \mathop{\mathbb{E}}_{U\sim\mu} |X_{i,U}| \,, \tag{B.9}$$

with

$$\tilde{u}(\delta) = (1+d(d+1)\delta)^{1/2} \leq (1+2\delta d^2)^{1/2} = (1+\frac{\delta'}{d^2})^{1/2} =: u(\delta'), \tag{B.10}$$

where we used the fact that $x(x+1) \leq 2x^2$ for any $x \geq 1$. This concludes the proof for quantum states.

For quantum measurements, we follow the analogous technique of proof. For POVMs M and N, each $\Delta_i = M_i - N_i$ will play a role of previous Δ. The only difference will be that the second moment is equal to

$$\mathop{\mathbb{E}}_{U\sim\nu} X_{i,U}^2 = \frac{1}{d(d+1)}(\mathrm{tr}(\Delta_i)^2 + \mathrm{tr}(\Delta_i^2)), \tag{B.11}$$

because operators Δ_i are generally not traceless.

2. Quantum channels

Let us now proceed to the proof of Theorem 3 for channels. Denote $X_{i,V,U} = \mathrm{tr}(|i\rangle\langle i|\, U\Delta(V\psi_0 V^\dagger)U^\dagger)$ where $\Delta = \Lambda - \Gamma$ with two quantum channels $\Lambda, \Gamma \in \mathrm{CPTP}(\mathcal{H}_d)$ that we are comparing. From the reasoning given in the preceding section (i.e. the proof of Theorem 1 for approximate 4-designs) we have that for $\delta = \delta'/(2d^4)$

$$\ell(\delta')\,\frac{a}{2}\|\Delta[\psi_V]\|_{\mathrm{HS}} \leq \mathop{\mathbb{E}}_{U\sim\nu} \mathrm{TV}(\mathbf{p}^{\Lambda,\psi_V,U}, \mathbf{p}^{\Gamma,\psi_V,U}) \leq u(\delta')\,\frac{A}{2}\|\Delta[\psi_V]\|_{\mathrm{HS}} \,. \tag{B.12}$$

We will use the above bounds together with Jensen's and Berger's inequality (applied for the function $Y_V := \|\Delta[\psi_V]\|_{\mathrm{HS}}$ and $V \sim \nu$) to establish the desired result. We start by re-expressing the second and fourth moment of Y_V in a convenient form :

$$\mathop{\mathbb{E}}_{V\sim\nu} Y_V^2 = \mathop{\mathbb{E}}_{V\sim\nu} \mathrm{tr}\left(\mathbb{S}\Delta^{\otimes 2}[\psi_V^{\otimes 2}]\right) = 2\,\mathrm{tr}\left(\mathbb{P}_{\mathrm{sym}}^{(2)}\Delta^{\otimes 2}[\mathcal{T}_{2,\nu}(\psi_0^{\otimes 2})]\right) \,, \tag{B.13}$$

$$\mathop{\mathbb{E}}_{V\sim\nu} Y_V^4 = \mathop{\mathbb{E}}_{V\sim\nu} \mathrm{tr}\left(\mathbb{S}\Delta^{\otimes 2}[\psi_V^{\otimes 2}]\right)^2 = 4\,\mathrm{tr}\left(\mathbb{P}_{\mathrm{sym}}^{(2)} \otimes \mathbb{P}_{\mathrm{sym}}^{(2)}\Delta^{\otimes 4}[\mathcal{T}_{4,\nu}(\psi_0^{\otimes 4})]\right) \,, \tag{B.14}$$

where we have used the 'swap trick': $\mathrm{tr}(AB) = \mathrm{tr}(A \otimes B\mathbb{S})$, and the fact that $\mathrm{tr}(\Delta[\psi_V]) = 0$. We start by using Eq. (B.13) to derive an *upper bound* on $\mathop{\mathbb{E}}_{V\sim\nu} Y_V^2$. From above for δ-approximate 4-design ν

$$\mathop{\mathbb{E}}_{V\sim\nu} Y_V^2 \leq \mathop{\mathbb{E}}_{V\sim\mu} Y_V^2 + 2\left|\mathrm{tr}\left(\mathbb{P}^{(2)}_{\mathrm{sym}}\Delta^{\otimes 2}[(\mathcal{T}_{2,\nu} - \mathcal{T}_{2,\mu})(\psi_0^{\otimes 2})]\right)\right| \tag{B.15}$$

$$= \mathop{\mathbb{E}}_{V\sim\mu} Y_V^2 + 2\left|\mathrm{tr}\left((\Delta^\dagger)^{\otimes 2}\left[\mathbb{P}^{(2)}_{\mathrm{sym}}\right](\mathcal{T}_{2,\nu} - \mathcal{T}_{2,\mu})(\psi_0^{\otimes 2})\right)\right| \tag{B.16}$$

$$\leq \mathop{\mathbb{E}}_{V\sim\mu} Y_V^2 + 2\left\|(\Delta^\dagger)^{\otimes 2}\left[\mathbb{P}^{(2)}_{\mathrm{sym}}\right]\right\|_\infty \delta\,, \tag{B.17}$$

where we have used the definition of the dual of a super operator and utilized that δ-approximate 4-design ν is also δ-approximate 2-design. We proceed with bounding the operator norm of $(\Delta^\dagger)^{\otimes 2}\left[\mathbb{P}^{(2)}_{\mathrm{sym}}\right]$ in terms of HS norm of the Jamiołkowski-Choi state $\mathcal{J}_\Delta$:

$$\left\|(\Delta^\dagger)^{\otimes 2}\left[\mathbb{P}^{(2)}_{\mathrm{sym}}\right]\right\|_\infty = \binom{d+1}{2}\left\|(\Delta^\dagger)^{\otimes 2}\left[\mathop{\mathbb{E}}_{U\sim\mu}\psi_U^{\otimes 2}\right]\right\|_\infty \leq \tag{B.18}$$

$$\leq \binom{d+1}{2}\max_{\psi\in\mathcal{S}(\mathcal{H})}\left\|(\Delta^\dagger)^{\otimes 2}\left[\psi^{\otimes 2}\right]\right\|_\infty = \binom{d+1}{2}\left(\max_{\psi,\phi\in\mathcal{S}(\mathcal{H})}\mathrm{tr}(\phi\Delta(\psi))\right)^2. \tag{B.19}$$

The result of double maximization can be upper bounded as follows:

$$\max_{\psi,\phi\in\mathcal{S}(\mathcal{H}_d)}\mathrm{tr}(\phi\Delta(\psi)) = \max_{\psi,\phi\in\mathcal{S}(\mathcal{H}_d)} d\,\mathrm{tr}(\mathcal{J}_\Delta\phi\otimes\psi^T) \leq d\|\mathcal{J}_\Delta\|_{\mathrm{HS}}\,. \tag{B.20}$$

Inserting this to (B.17) and recalling that $\mathop{\mathbb{E}}_{V\sim\mu} Y_V^2 = (A^{\mathrm{ch}})^2(\|\mathcal{J}_\Delta\|_{\mathrm{HS}}^2 + \mathrm{tr}(\Delta(\tau_d)^2)$ with $A^{\mathrm{ch}} = \frac{d}{d+1}$(cf. Eq. (B.12)) we get:

$$\mathop{\mathbb{E}}_{V\sim\nu} Y_V^2 \leq (A^{\mathrm{ch}})^2(\|\mathcal{J}_\Delta\|_{\mathrm{HS}}^2 + \mathrm{tr}(\Delta(\tau_d)^2) + d^5(d+1)\|\mathcal{J}_\Delta\|_{\mathrm{HS}}^2\delta \leq \mathop{\mathbb{E}}_{V\sim\mu} Y_V^2\left(1 + \frac{d^5(d+1)}{(A^{\mathrm{ch}})^2}\delta\right)\,. \tag{B.21}$$

Integrating both sides of the upper bound in Eq. (B.12) , using Jensen's inequality, and noting that $\frac{1}{2}\sqrt{\mathop{\mathbb{E}}_{V\sim\mu} Y_V^2} = \mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma)$ yields

$$\mathop{\mathbb{E}}_{V\sim\nu}\mathop{\mathbb{E}}_{U\sim\nu}\mathrm{TV}(\mathbf{p}^{\Lambda,\psi_V,U},\mathbf{p}^{\Gamma,\psi_V,U}) \leq u(2d^4\,\delta)\sqrt{1+\frac{d^5(d+1)}{(A^{\mathrm{ch}})^2}\delta}\,\frac{A^{\mathrm{ch}}}{2}\sqrt{\mathop{\mathbb{E}}_{V\sim\mu} Y_V^2}\,, \tag{B.22}$$

$$= u(2d^4\,\delta)\sqrt{1+d^3(d+1)^3\delta}\,A^{\mathrm{ch}}\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma)\,, \tag{B.23}$$

$$= \tilde{u}^{\mathrm{ch}}(\delta)\,A^{\mathrm{ch}}\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma)\,, \tag{B.24}$$

where we defined

$$\tilde{u}^{\mathrm{ch}}(\delta) := u(2d^4\,\delta)\,\sqrt{1+d^3(d+1)^3\delta} \leq u(2d^4\,\delta)\,\sqrt{1+8d^6\delta}\,. \tag{B.25}$$

where in second step we used inequality $(x+a) \leq x(1+a)$ for $x, a \geq 1$,

To get the lower bound, we will integrate LHS of Eq. (B.12) and apply berger inequality. Proceeding analogously as before we obtain

$$\mathop{\mathbb{E}}_{V\sim\nu} Y_V^2 \geq \mathop{\mathbb{E}}_{V\sim\mu} Y_V^2 - 2\left\|(\Delta^\dagger)^{\otimes 2}\left[\mathbb{P}^{(2)}_{\mathrm{sym}}\right]\right\|_\infty\delta \;\geq\; \mathop{\mathbb{E}}_{V\sim\mu} Y_V^2(1-d^3(d+1)^3\delta)\,. \tag{B.26}$$

$$\mathop{\mathbb{E}}_{V\sim\nu} Y_V^4 \leq \mathop{\mathbb{E}}_{V\sim\mu} Y_V^4 + 4\left\|(\Delta^\dagger)^{\otimes 2}\left[\mathbb{P}^{(2)}_{\mathrm{sym}}\right]\right\|_\infty^2\delta \;\leq \mathop{\mathbb{E}}_{V\sim\mu} Y_V^4 + d^4(d+1)^4\|\mathcal{J}_\Delta\|_{\mathrm{HS}}^4\delta \tag{B.27}$$

We now recall that fourth moment w.r.p. to Haar measure is bounded by (see Eq. 47)

$$\mathop{\mathbb{E}}_{V\sim\mu} Y_V^4 \leq v\cdot\left(\mathop{\mathbb{E}}_{V\sim\mu} Y_V^2\right)^2 = v\cdot(A^{\mathrm{ch}})^4\,(2\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma))^4\,, \tag{B.28}$$

with $v = \frac{\frac{13}{6}\binom{d+1}{2}^2}{\binom{d+3}{4}}$, where we used the fact that $\mathop{\mathbb{E}}_{V\sim\mu} Y_V^2 = (A^{\mathrm{ch}})^2\,(\|\mathcal{J}_\Delta\|_{\mathrm{HS}}^2 + \mathrm{tr}(\Delta(\tau_d)^2) = (A^{\mathrm{ch}})^2\,(2\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma))^2$.

We now note that $\|\mathcal{J}_\Delta\|_{\rm HS}^4 \le \left(\|\mathcal{J}_\Delta\|_{\rm HS}^2 + \mathrm{tr}\left(\Delta\left(\tau_d\right)^2\right)\right)^2 = (2\mathrm{d}_{\rm av}^{\rm ch}(\Lambda,\Gamma))^4$, and combine it with Eq. (B.28) and Eq. (B.27) to obtain

$$\mathop{\mathbb{E}}_{V\sim\nu} Y_V^4 \le v\,(A^{\rm ch})^4\,(2\mathrm{d}_{\rm av}^{\rm ch}(\Lambda,\Gamma))^4\,\left(1-\delta\,\frac{d^4(d+1)^4}{v\,(A^{\rm ch})^4}\right) \tag{B.29}$$

$$= v\,(A^{\rm ch})^4\,(2\mathrm{d}_{\rm av}^{\rm ch}(\Lambda,\Gamma))^4\,\left(1+\delta\,\frac{(d+1)^7(d+2)(d+3)}{13d}\right)\,. \tag{B.30}$$

Similarly, we get

$$\mathop{\mathbb{E}}_{V\sim\nu} Y_V^2 \ge (A^{\rm ch})^2\,(2\mathrm{d}_{\rm av}^{\rm ch}(\Lambda,\Gamma))^2\,\left(1-\delta\,d^3(d+1)^3\right) \tag{B.31}$$

Inserting the above to Berger's inequality yields

$$\frac{\left(\mathop{\mathbb{E}}_{V\sim\nu} Y_V^2\right)^{3/2}}{\left(\mathop{\mathbb{E}}_{V\sim\nu} Y_V^4\right)^{1/2}} \ge 2\,\tilde{b}_d\,\mathrm{d}_{\rm av}^{\rm ch}(\Lambda,\Gamma)\,\frac{\left(1-\delta\,d^3(d+1)^3\right)^{3/2}}{\left(1+\delta\,\frac{(d+1)^7(d+2)(d+3)}{13d}\right)^{1/2}} \tag{B.32}$$

where $\tilde{b}_d = \frac{d}{d+1}\sqrt{\frac{(d+2)(d+3)}{13d(d+1)}}$.

Now we integrate Eq. (B.12) and combine with the above to obtain

$$\mathop{\mathbb{E}}_{U\sim\nu} \mathrm{TV}(\mathbf{p}^{\Lambda,\psi_V,U},\mathbf{p}^{\Gamma,\psi_V,U}) \ge a^{\rm ch}\,\mathrm{d}_{\rm av}^{\rm ch}(\Lambda,\Gamma)\,\tilde{l}^{\rm ch}(\delta)\,, \tag{B.33}$$

where

$$\tilde{l}^{\rm ch}(\delta) = \ell(2d^4\,\delta)\,\frac{\left(1-\delta\,d^3(d+1)^3\right)^{3/2}}{\left(1+\delta\,\frac{(d+1)^7(d+2)(d+3)}{13d}\right)^{1/2}} \tag{B.34}$$

$$\ge \ell(2d^4\,\delta)\,\frac{\left(1-\delta\,8\,d^6\right)^{3/2}}{\left(1+\delta\,\frac{d^8\cdot 2^7\cdot 12}{13}\right)^{1/2}} \tag{B.35}$$

$$\ge \ell(2d^4\,\delta)\,\frac{\left(1-\delta\,2^8\,d^6\right)^{3/2}}{\left(1+\delta\,(2d)^8\right)^{1/2}}\,, \tag{B.36}$$

where we used inequality $(x+a)\le x(1+a)$ for $x,a\ge 1$.

Now we set $\delta := \frac{\delta'}{(2d)^8}$ and get

$$\tilde{l}^{\rm ch}(\delta') \ge \ell(\frac{\delta'}{2^7d^4})\,\frac{\left(1-\frac{\delta'}{d^2}\right)^{3/2}}{(1+\delta')^{1/2}} \ge \ell(\delta')\,\frac{\left(1-\frac{\delta'}{d^2}\right)^{3/2}}{(1+\delta')^{1/2}} \ge \frac{\left(1-\frac{\delta'}{d^2}\right)^{3}}{1+\delta'} =: \ell^{\rm ch}(\delta')\,, \tag{B.37}$$

where in second inequality we used the fact that $\ell(x)$ is a decreasing function of $x\ge 0$.

With set δ, we also rewrite upper bound from Eq. (B.25) as

$$\tilde{u}^{\rm ch}(\delta') \le u(\frac{\delta'}{2^7d^4})\sqrt{1+\frac{\delta'}{2^5d^2}} \le u(\delta)\sqrt{1+\frac{\delta'}{d^2}} \le 1+\frac{\delta'}{d^2} =: u^{\rm ch}(\delta')\,, \tag{B.38}$$

where we used the fact that $u(x)$ is an increasing function of $x\ge 0$.