Operational Quantum Average-Case Distances

Filip B. Maciejewski[1,2], Zbigniew Puchała[3,4], and Michał Oszmaniec[1]

[1]Center for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warszawa, Poland
[2]Research Institute for Advanced Computer Science (RIACS), USRA, Moffett Field, CA
[3]Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, 44-100 Gliwice, Poland
[4]Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, 30-348 Kraków, Poland

We introduce distance measures between quantum states, measurements, and channels based on their statistical distinguishability in generic experiments. Specifically, we analyze the average Total Variation Distance (TVD) between output statistics of protocols in which quantum objects are intertwined with random circuits and measured in a standard basis. We show that for circuits forming approximate 4-designs, the average TVDs can be approximated by simple explicit functions of the underlying objects – the average-case (AC) distances. We apply AC distances to analyze the effects of noise in quantum advantage experiments and for efficient discrimination of high-dimensional states and channels without quantum memory. We argue that AC distances are better suited for assessing the quality of NISQ devices than common distance measures such as trace distance or the diamond norm.

Introduction. In the era of Noisy Intermediate Scale Quantum (NISQ) devices [48], it is instrumental to have figures of merit that quantify how close two quantum protocols are. The distance measures commonly used for this purpose, for example, in the context of quantum error correction [27], such as trace distance or diamond norm, have an operational interpretation in terms of *optimal statistical distinguishability* between two quantum states, measurements, or channels [6, 18, 43, 49]. While it is natural to consider the optimal protocols when one wishes to distinguish between two objects, alas, in reality, such protocols might be not practical. For example, in general, they require high-depth, complicated quantum circuits [10]. From a complementary perspective, quantum distances are often used to compare an *ideal* implementation (of a state, measurement, or channel) with its *noisy* experimental version. In this context, using the distances based on optimal distinguishability gives information about the worst-case performance of a device in question. This may be impractical as well – it is not expected that the performance of typical experiments on a quantum device will be comparable to the worst-case scenario.

In this work, we consider the average Total-Variation (TV) distance between output statistics of two protocols in which random circuits interlace quantum objects of interest (see Figure 1). This can be thought to mimic the typical circumstances in which quantum states, measurements, or channels appear as parts of quantum-information protocols. We show that for a broad class of easy-to-implement random circuits (forming approximate 4-designs), the average TV distance is approximated by simple explicit functions expressible by degree 2 polynomials in objects in question. We use these functions to define distance measures between states, measurements, and channels. The so-defined average-case (AC) distances are thus distance measures that approximate average-case total variation distance. Contrary to conventional distances such as the trace distance or the diamond norm, the AC distances capture the generic behavior of quantum objects in experiments involving only moderate-depth quantum circuits. This feature can be especially relevant in the context of near-term algorithms, such as the Quantum Approximate Optimization Algorithm (QAOA) [13, 14, 23] and Variational Quantum Eigensolver (VQE) [33, 46, 47], as it is expected that generic variational circuits will, on average, have properties of unitary designs [41]. We present numerical results suggesting that AC distances are more suitable for quantifying the impact of imperfections on variational algorithms than the conventional distance measures.

Multiple recent quantum advantage proposals are based on random circuits sampling [5, 51]. We apply AC distances to understand the effects of noise on such protocols. We approach the problem from two sides. First, the AC distances allow to easily *lower* bound the average-case TV distance between the noisy distribution and the ideal distribution, thus

Michał Oszmaniec: oszmaniec@cft.edu.pl

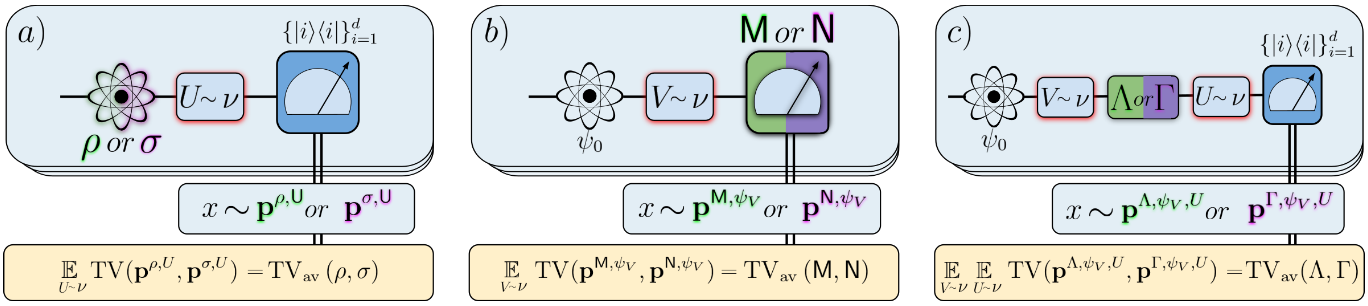

$a)$
$\{|i\rangle\langle i|\}_{i=1}^{d}$
$U\sim\nu$
ρ or σ
$x\sim \mathbf{p}^{\rho,\mathsf{U}}$ or $\mathbf{p}^{\sigma,\mathsf{U}}$
$\underset{U\sim\nu}{\mathbb{E}}\,\mathrm{TV}(\mathbf{p}^{\rho,U},\mathbf{p}^{\sigma,U}) = \mathrm{TV}_{\mathrm{av}}(\rho,\sigma)$
$b)$
M or N
$V\sim\nu$
ψ_0
$x\sim \mathbf{p}^{\mathsf{M},\psi_V}$ or $\mathbf{p}^{\mathsf{N},\psi_V}$
$\underset{V\sim\nu}{\mathbb{E}}\,\mathrm{TV}(\mathbf{p}^{\mathsf{M},\psi_V},\mathbf{p}^{\mathsf{N},\psi_V}) = \mathrm{TV}_{\mathrm{av}}(\mathsf{M},\mathsf{N})$
$c)$
$\{|i\rangle\langle i|\}_{i=1}^{d}$
$V\sim\nu$
Λ or Γ
$U\sim\nu$
ψ_0
$x\sim \mathbf{p}^{\Lambda,\psi_V,U}$ or $\mathbf{p}^{\Gamma,\psi_V,U}$
$\underset{V\sim\nu}{\mathbb{E}}\,\underset{U\sim\nu}{\mathbb{E}}\,\mathrm{TV}(\mathbf{p}^{\Lambda,\psi_V,U},\mathbf{p}^{\Gamma,\psi_V,U}) = \mathrm{TV}_{\mathrm{av}}(\Lambda,\Gamma)$

Figure 1: Measures of the distance between quantum objects based on *average* statistical distinguishability. For quantum states a), we take the average over random unitaries applied to the state, followed by measurement in the standard basis. For quantum measurements b), we take the average over random pure states measured on the detector. Finally, for quantum channels c) we take the average over independent random unitaries applied *before* and *after* the application of the channel.

giving insight into how well separated, on average, are noisy distributions from target distributions. Second, AC distances allow to *upper* bound the average-case TV distance between a noisy distribution and a (trivial) uniform distribution. This allows to study how quickly the noise makes the average distribution useless. For example, we show that even in the absence of gate and state-preparation noise, the local, symmetric bitflip error in measurements causes noisy distribution to approach trivial one exponentially quickly in system size.

Recently there has been a lot of interest in algorithms that use randomized quantum circuits, such as shadow tomography [1, 11, 19, 20, 28] and randomized-benchmarking [12, 15, 16, 26, 39, 40]. Our results can be employed to quantify the performance of randomized algorithms in the task of statistical distinguishability of quantum objects. Namely, if the average-case distance between a pair of quantum objects on N qubit systems is large, then they can be (statistically) distinguished almost perfectly using a randomized protocol with just a few implementations of local random circuits of depth $O(N)$. We observe that such behavior takes place in two scenarios related to those recently analyzed in the context of so-called Quantum Algorithmic Measurement [2] and complexity growth of quantum circuits [10]: (i) distinguishing Haar random N qubit pure state from maximally mixed state and (ii) distinguishing N qubit Haar random unitary from maximally depolarizing channel. This shows that protocols employing random circuits can be used to efficiently discriminate quantum objects. Since they do not depend on the objects to be distinguished, randomized measurement schemes can be interpreted as "universal discriminators", analogous to the SWAP test but not requiring the usage of entanglement or coherent access to copies of quantum systems.

The manuscript is accompanied by a complementary work [37] that contains proofs of theorems, a thorough analysis of the properties of average-case quantum distances, and further examples. In contrast, the following work focuses on providing intuition behind AC distances and demonstrating how they can be applied to understand the power of random quantum circuits in practically relevant scenarios, which is followed by numerical demonstrations.

Notation and basic concepts. Our result concern quantum systems on finite-dimensional Hilbert space $\mathcal{H}_d \approx \mathbb{C}^d$. General quantum measurements, also known as POVMs, are described by tuples $\mathsf{M} = (M_i)_{i=1}^{n}$ of operators on $\mathcal{H}_d$ which satisfy $M_i \geq 0$ and $\sum_{i=1}^{n} M_i = \mathbb{I}_d$, where $\mathbb{I}_d$ is the identity on $\mathcal{H}_d$. General quantum operations on $\mathcal{H}_d$ is described by a quantum channel, i.e., a completely-positive trace-preserving map Λ : $\mathrm{Herm}(\mathcal{H}_d) \rightarrow \mathrm{Herm}(\mathcal{H}_d)$. We will use the notation $\tau_d = \mathbb{I}/d$ to denote maximally mixed state on $\mathcal{H}_d$.

We will consider general protocols consisting of three stages (i) state preparation, in which quantum system is initialized in state ρ, (ii) evolution given by a quantum channel Λ and (iii) measurement of the resulting state $\Lambda(\rho)$ by a POVM M. The outcome statistics of such a protocol are given by the Born rule: $p_i^{\rho,\Lambda,\mathsf{M}} = \mathrm{tr}(M_i\Lambda(\rho))$. Total Variation (TV) distance between distributions $\mathbf{p} = (p_i)_{i=1}^{n}$ and $\mathbf{q} = (q_i)_{i=1}^{n}$ is defined as $\mathrm{TV}(\mathbf{p},\mathbf{q}) = \frac{1}{2}\sum_{i=1}^{n}|p_i - q_i|$. TV distance defines the 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 equal probability), the optimal probability of correctly guessing the answer is $p_{\mathrm{succ}} = \frac{1}{2}(1+\mathrm{TV}(\mathbf{p},\mathbf{q}))$. The related distance between quantum objects is constructed by considering the optimal success probability of distinguishing between pairs of relevant quantum objects, where the optimization is carried out not only over classical post-processing strategies but also over *quantum* strategies that produce classical outcomes given the objects in question (see Supplementary Material (SM) for details).

Here we propose alternative distance measures based on scenarios where the strategy of discrimination of quantum objects is based on intertwining them with random quantum circuits and then comparing their outcome statistics [37]. Specifically, consider output statistics $\mathbf{p}^{\alpha,\beta}$ of a quantum protocol where α is a fixed quantum object while β is taken to be a random variable (specifying a quantum circuit) distributed according to probability distribution ν. The average statistical distinguishability of two objects α_1, α_2 is quantified by

$$\mathrm{TV}_{\mathrm{av}}(\alpha_1, \alpha_2) = \mathop{\mathbb{E}}_{\beta \sim \nu} \mathrm{TV}(\mathbf{p}^{\alpha_1,\beta}, \mathbf{p}^{\alpha_2,\beta}) \ . \qquad (1)$$

Explicit computation of $\mathrm{TV}_{\mathrm{av}}(\alpha_1, \alpha_2)$ is difficult because $\mathrm{TV}(\mathbf{p}, \mathbf{q})$ is not a polynomial function of the involved probabilities. However, if ν forms an approximate 4-design, it is possible to find simple estimates to $\mathrm{TV}_{\mathrm{av}}$. Unitary k-designs are measures on $\mathrm{U}(\mathcal{H}_d)$ that reproduce averages of Haar measure μ on balanced polynomials of degree k in U [3]. For approximate k-designs these averages agree only approximately. Measure ν on $\mathrm{U}(\mathcal{H}_d)$ is δ-approximate k-design if $\|\mathcal{T}_{k,\nu} - \mathcal{T}_{k,\mu}\|_\diamond \leq \delta$, where $\mathcal{T}_{k,\nu}(\rho) = \int_{\mathrm{U}(\mathcal{H})} d\nu(U) U^{\otimes k} \rho (U^\dagger)^{\otimes k}$. Importantly, random quantum circuits in the 1D architecture formed from *arbitrary* universal gates that randomly couple neighboring qubits, generate approximate k-designs efficiently with the number of qubits N [9, 21, 25, 45]. Specifically, δ-approximate 4-designs are generated by the 1D random brickwork architecture in depth $O(N + \log(1/\delta))$, with moderate numerical constants [21].

Quantum average-case distances between states, measurements, and channels. We are now ready to formulate our main technical results - dimension independent relative error estimates on average TV distances between three types of quantum objects depicted in Figure 1. To simplify the formulation of the Theorems, we will use the symbol $\approx$ to denote equality up to a dimension-independent relative error. The specific constants are given in [37]. In Appendix B we provide simplified proofs of the following theorems in the setting of exact unitary designs. The proofs for approximate unitary designs can be found in Appendix B of [37].

Quantum states. Let $\mathbf{p}^{\rho,U}$ denote the probability distribution of a quantum process in which ρ undergoes a unitary transformation U and is then 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 $\mathcal{H}_d$.

Theorem 1 (Average-case distinguishability of quantum sates – Theorem 1 from [37])**.** *Let ρ, σ be quantum states in $\mathcal{H}_d$ and let ν be a distribution in the unitary group $\mathrm{U}(\mathcal{H}_d)$ forming δ-approximate 4-design for $\delta = \frac{\delta'}{2d^4}$, for $\delta' \in (0, \frac{1}{3})$. We then have*

$$\mathop{\mathbb{E}}_{U \sim \nu} \mathrm{TV}(\mathbf{p}^{\rho,U}, \mathbf{p}^{\sigma,U}) \approx \mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_{\mathrm{HS}} \ , \qquad (2)$$

where $\|X\|_{\mathrm{HS}} = \sqrt{\mathrm{tr}(X^2)}$ denotes Hilbert-Schmidt norm.

The proof of Theorem 1 (and also theorems 2 and 3 stated below) is inspired by the proof of Theorem 4 from [3] where Berger inequality (stating that for every random variable X with well-defined 2nd and 4th moments we have $(\mathbb{E}[X^2])^{\frac{3}{2}} (\mathbb{E}[X^4])^{-\frac{1}{2}} \leq \mathbb{E}|X|$) 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 1. *We can interpret the above average statistical distinguishability as* TV*-distance of output statistics resulting from a measurement of a* single *POVM with effects $M_{i,V_j} = \nu_j U_j^\dagger |i\rangle\langle i| U_j$, where ν_j is the probability of occurence of circuit U_j in the ensemble ν (for simplicity of presentation we assumed that ensemble ν is discrete). This POVM can be interpreted as a convex combination [44] of projective measurements M^{U_j} with effects $\mathsf{M}_i^{U_j} = U_j^\dagger |i\rangle\langle i| U_j$. Lower bound on average TV distance implies that such randomized protocol distinguishes between quantum states with high probability. It immediately follows that there also exists a deterministic (not randomized)* optimal *distinguishability protocol that achieves the same success probability. Such a measurement can be implemented, for example, via Naimark's dilation using an ancillary system [43]. Analogous interpretation holds also for the average* TV*-distances from Theorems 2 and 3 below.*

Remark 2. *We note that while the dependence of δ on the dimension of the system d is very high in Theorem 1 (as well as in Theorems 2 and 3), it does not pose a practical problem. Indeed, exponentially accurate δ-approximate unitary designs can be implemented already with linear-depth quantum circuits [21].*

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 is evolved according by unitary V and is subsequently measured via a n-outcome POVM $\mathsf{M} = (M_1, M_2, \ldots, M_n)$. In other words $p_i^{\mathsf{M},\psi_V} = \mathrm{tr}(V \psi_0 V^\dagger M_i)$.

Theorem 2 (Average-case distinguishability of quantum measurements – Theorem 2 from [37])**.** *Let* M,N *be* n*-outcome POVMs on* $\mathcal{H}_d$ *and let* ν *be a distribution on on* $\mathrm{U}(\mathcal{H}_d)$ *forming* δ*-approximate* 4*-design for* $\delta = \frac{\delta'}{(2d)^8}$*, for* $\delta' \in (0, \frac{1}{3})$*. We then have*

$$\mathop{\mathbb{E}}_{V\sim\nu} \mathrm{TV}(\mathbf{p}^{\mathsf{M},\psi_V}, \mathbf{p}^{\mathsf{N},\psi_V}) \approx \mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{M},\mathsf{N}) \quad , \textit{ where}$$

$$\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{M},\mathsf{N}) = \frac{1}{2d}\sum_{i=1}^{n}\sqrt{\|M_i - N_i\|^2_{\mathrm{HS}} + \mathrm{tr}(M_i - N_i)^2} \ . \tag{3}$$

Quantum channels. Let $\mathbf{p}^{\Lambda,\psi_V,U}$ by the probability distribution associated to a quantum process in in which a fixed pure quantum state ψ_0 is subsequently acted on by unitary V, channel Λ and unitary U, and is subsequently measured in the computational basis of $\mathcal{H}$. In other words we have $p_i^{\Lambda,\psi_V,U} = \mathrm{tr}(|i\rangle\langle i|U\Lambda(V\psi_0 V^\dagger)U^\dagger)$.

Theorem 3 (Average-case distinguishability of quantum channels – Theorem 3 from [37])**.** *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* $\delta = \frac{\delta'}{(2d)^8}$*, for* $\delta' \in (0, \frac{1}{9})$*. Then we have*

$$\mathop{\mathbb{E}}_{V\sim\nu}\mathop{\mathbb{E}}_{U\sim\nu} \mathrm{TV}(\mathbf{p}^{\Lambda,\psi_V,U}, \mathbf{p}^{\Gamma,\psi_V,U}) \approx \mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma) \quad , \textit{ where}$$

$$\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma) = \frac{1}{2}\sqrt{\|\mathcal{J}_\Lambda - \mathcal{J}_\Gamma\|^2_{\mathrm{HS}} + \mathrm{tr}\left((\Lambda-\Gamma)[\tau_d]^2\right)} \tag{4}$$

and $\mathcal{J}_\Lambda$ *denotes Jamiołkowski-Choi state of* Λ*.*

Remark 3. *Having defined randomized distinguishability strategies, it is natural to ask how they compare to optimal protocols on a d-dimensional Hilbert space* $\mathcal{H}_d$*. We give upper bounds on the maximal ratio between worst-case and average-case distances to answer this. It turns out that this ratio is at most* $d^{\frac{1}{2}}$*,* d*,* $d^{\frac{3}{2}}$ *for quantum states, measurements, and channels, respectively. This implies that there exist scenarios where the optimal protocol for distinguishing two quantum objects performs exponentially better than protocol using random quantum circuits. Indeed, in the technical version of the manuscript, [37] we construct examples that saturate those bounds.*

The above theorems suggest to define average-case distances between quantum states, measurements, and channels via formulas $\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}$, $\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}$, $\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}$ appearing in approximations (2), (3), and (4). This approach has several pleasant consequences. First, functions describing these distances can be expressed via simple, degree-two polynomials in underlying objects and can be easily explicitly computed for objects acting on systems of moderate dimension (no optimization is needed as in the case of the diamond norm [50]). Second, all average-case distances utilize in some way the Hilbert-Schmidt norm. This gives this norm an operational interpretation it did not possess before (especially for quantum states for which $\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho,\sigma) = \frac{1}{2}\|\rho-\sigma\|_{\mathrm{HS}}$). Third, it turns out that so-defined distances satisfy plethora of natural properties such as subadditivity: $\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho_1\otimes\rho_2, \sigma_1\otimes\sigma_2) \leq \mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho_1,\sigma_1) + \mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho_2,\sigma_2)$, joint convexity: $\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\sum_\alpha p_\alpha\rho_\alpha, \sum_\alpha p_\alpha\sigma_\alpha) \leq \sum_\alpha p_\alpha \mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho_\alpha,\sigma_\alpha)$, or restricted data-processing inequalities (typically various distances d_{av} are non-increasing under application of unital quantum channels). See [37] for details and proofs of various properties of average-case distances. Fourth, while it may seem that condition of being (approximate) 4-design is quite stringent, from a recent paper [21] it follows that ensembles of quantum circuits required by Theorems 1-3 can be realized by random circuits in the 1D brickwork architecture in depth $O(N)$ (with moderate prefactors) [21]. Finally, we expect that our average-case distances will more accurately capture the behavior of errors in the performance of quantum objects in generic moderate size quantum algorithms (note that many architectures of variational circuits used in NISQ algorithms are expected to exhibit, on average, design-like behavior [41]). We back up this last claim numerically by testing the usefulness of our distance measures on families of random quantum circuits originating from random instances of variational quantum algorithms on few-qubit systems.

Applications. For all the reasons mentioned above, we believe that introduced distances will prove useful in analyzing the practical performance of near-term quantum processors. We expect that they can also be useful in other branches of quantum information requiring the usage of randomized protocols like quantum communication, quantum complexity theory, or quantum machine learning. The following simple examples illustrate potential usefulness of our results.

Application 1: Noise in quantum advantage experiments.

Here we consider examples which help to understand how noise affects average probability distributions in experiments with random circuits sampling. First, AC distances between noisy and ideal state allow to lower-bound average TVDs between target and noisy distributions. Second, AC distances allow to upper-bound average-case TVD between noisy distribution and trivial (uniform) one.

Indeed, to bound average TVD between uniform and noisy distribution, one calculates AC distance to maximally mixed state $\frac{\mathbb{I}}{d}$ (states), trivial POVM $\mathsf{M}^{\mathcal{I}} = \left(\frac{\mathbb{I}}{d}, \ldots, \frac{\mathbb{I}}{d}\right)$ (measurements), or maximally depolarizing channel Λ_{dep} that acts as $\Lambda_{\mathrm{dep}}(\rho) = \frac{\mathbb{I}}{d}$ for any state ρ (channels). This follows directly from definitions of AC distances – see Lemmas 23, 24 and 25 in [37].

In what follows, most of the examples make use of some average noise parameter q^{av} (with different meaning for each example) that describes an average (over qubits) probability of errors of considered type *not* occurring. In most of them, we make an assumption that $q^{av} \leq \sqrt[N]{\frac{1}{2}}$. This is done solely to achieve a particularly appealing form of lower bounds. One can derive expressions that are more complicated and do not require this assumption (see SM for details and proofs of the following examples). In general, since $\sqrt[N]{\frac{1}{2}} \xrightarrow{N\to\infty} 1$, the assumption becomes less restrictive for higher-dimensional systems and the presented bounds are intended for use in such cases.

Example 1 (Pauli eigenstates and tensor product 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 tensor product 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). Define average properties of noise as* $q^{av} = \frac{1}{N}\sum_{i=1}^{N} q^{(i)}$ *and* $f^{av} = \frac{1}{N}\sum_{i=1}^{N} q^{(i)}(1 - q^{(i)})$*. Assume* $q^{(i)} \geq \frac{1}{2}$ *for each qubit and that* $q^{av} \leq \sqrt[N]{\frac{1}{2}}$*. Then we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\Lambda^{pauli}(\psi^{pauli}), \frac{\mathbb{I}}{d}) < \frac{1}{2}\,\exp\left(-2f^{av}\,N\right), \qquad (5)$$

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\Lambda^{pauli}(\psi^{pauli}), \psi^{pauli}) > \frac{1}{2}\sqrt{1 - 2(q^{av})^{N}}, \qquad (6)$$

The above example might be relevant, for example, in QAOA algorithms where input state is often indeed a tensor product Pauli state [13], or can be useful for estimating effects of state-preparation errors for standard setting where input state is $|0\rangle\langle 0|^{\otimes N}$. We see that with growing system size, the average noisy distribution approaches uniform distribution exponentially quickly (while moving away from target distribution).

This demonstrates that even in the absence of noise in random unitaries, the state-preparation errors will quickly aggregate. Exactly the same behaviour is demonstrated for the following simplified measurement noise model.

Example 2 (Symmetric bitflip measurement noise)**.** *Consider a noisy version* $\mathrm{T}^{sym}\mathsf{P}$ *of computational basis measurement* P*, where* $\mathrm{T}^{sym} = \otimes_{i=1}^{N}\mathrm{T}_i^{sym}$ *and* k*th effect of noisy measurement is given by* $\left(\mathrm{T}^{sym}\mathsf{P}\right)_k = \sum_l \mathrm{T}_{kl}^{sym}\,|l\rangle\langle l|$*. Here for each qubit we have* $\mathrm{T}_i^{sym} = p^{(i)}\mathbb{I} + (1 - p^{(i)})\sigma_x$*, where* $(1 - p^{(i)})$ *is a bitflip error probability on* i*th qubit. Define* $f^{av} = \frac{1}{N}\sum_{i=1}^{N} p^{(i)}(1 - p^{(i)})$*. Assume* $p^{(i)} \geq \frac{1}{2}$ *for each qubit. Then we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathrm{T}^{sym}\mathsf{P}, \mathsf{M}^{\mathcal{I}}) < \frac{1}{2}\exp\left(-2f^{av}\,N\right)\ , \qquad (7)$$

The above means that even in the absence of state-preparation and gate errors, for symmetric bitflip noise the resulting average distribution exponentially quickly converges to uniform. We now consider a distance from ideal measurement for more realistic case of generic tensor product measurement noise.

Example 3 (Generic tensor product measurement noise)**.** *Let* $\mathsf{P} = (|\mathbf{x}\rangle\langle\mathbf{x}|)_{\mathbf{x}\in\{0,1\}^N}$ *be a computational basis measurement on* N *qubit system. Let* $\mathsf{M} = (M_{\mathbf{x}})_{\mathbf{x}\in\{0,1\}^N}$ *be a POVM specified by effects* $M_{\mathbf{x}} = \Lambda_1^{\dagger}(|x_1\rangle\langle x_1|) \otimes \ldots \otimes \Lambda_N^{\dagger}(|x_N\rangle\langle x_N|)$*, where* Λ_i *are quantum channels affecting* i*'th qubit, and* $\Lambda_i^{\dagger}$ *is the conjugate of* Λ_i*. Define classical success probability as* $p^{(i)}(k|k) = \mathrm{tr}\left(\Lambda_i^{\dagger}\left(|x_i\rangle\langle x_i|\right)|x_i\rangle\langle x_i|\right)$ *and corresponding average* $q_{av}^{(i)} = \frac{p^{i}(0|0)+p^{(i)}(1|1)}{2}$*. Let* $q^{av} := \frac{1}{N}\sum_{i=1}^{N} q_{av}^{(i)}$*. Assume that for each qubit* $q_{av}^{(i)} \geq \frac{1}{2}$ *and that* $q^{av} \leq \sqrt[N]{\frac{1}{2}}$*. Then we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathsf{M}, \mathsf{P}) > \frac{1}{2}\sqrt{1 - 2(q^{av})^{N}}\ . \qquad (8)$$

The quantity q^{av} is the survival probability of classical single-qubit state $|x_i\rangle\langle x_i|$ that goes through a channel Λ_i, averaged over all qubits and input states. We note that those quantities are routinely reported in experimental works, which makes the above bound particularly useful. Indeed, data from recent quantum advantage experiments [5, 51] suggests that q^{av} is around 97% (we take average of values reported in both papers). Assume perfect gates, no state preparation errors and $q^{av} = 0.97$. Furthermore, assume that random circuits used in experiments form approximate 4-designs (this assumption is consistent with results of [25]). Then from Theorem 2 it follows that if readout errors remain constant with scaling of the system, for a 54-qubit quantum computer, on average (over realizations of random quantum circuits) output distributions $\mathbf{p}^{\mathsf{M},\psi_V}$ will have a constant ≈ 0.13 TV-distance from the ideal probability distributions $\mathbf{p}^{\mathsf{P},\psi_V}$ solely due to effects of readout noise.

Example 4 (Tensor product Pauli noise in the middle of the circuit)**.** *Consider tensor product Pauli channel Λ^{pauli} defined in Example 1. For each qubit i define $||\mathbf{p}^{(i)}||_2^2 = \sum_j \left(p_j^{(i)}\right)^2$, and corresponding average $p_2^{av} = \frac{1}{N}\sum_{i=1}^N ||\mathbf{p}^{(i)}||_2^2$, as well as average probability of application of identity channel $p_1^{av} = \frac{1}{N}\sum_{i=1} p_1^{(i)}$. Assume $p_1^{av} \leq \sqrt[N]{\frac{1}{2}}$. Then we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda^{pauli}, \Lambda_{dep}) < \frac{1}{2}\exp\left(-p_2^{av}\ N\right)\ , \tag{9}$$

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda^{pauli}, \mathcal{I}) > \frac{1}{\sqrt{2}}\sqrt{1 - 2\left(p_1^{av}\right)^N}\ . \tag{10}$$

Recall that the above scenario corresponds to inserting local Pauli noise "between" two random circuits (two averages in Eq. (4)). Similarly to previous cases, whenever there is non-zero noise, we will observe an exponential convergence to the trivial distribution and high separation from ideal distribution corresponding to identity channel $\mathcal{I}$.

Example 5 (Single Pauli error the middle of the circuit)**.** *Consider tensor product channel $\Lambda_\sigma^{(i)}$ that applies some traceless unitary σ on qubit i (and identity to all other qubits). Then we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_\sigma^{(i)}, \mathcal{I}) = \frac{1}{\sqrt{2}}\ . \tag{11}$$

Physically, the above may correspond to a unitary noise applying one of Pauli matrices on qubit i somewhere in the circuit. We then observe a constant separation (value of $\frac{1}{\sqrt{2}}$) between ideal distribution and the noisy distribution. Such significant average distance between noisy and target distribution suggests that local strong coherent errors can dramatically affect the performance of a given device in typical circumstances. This result is in agreement with empirical observations made in Refs. [5, 8] where single-qubit errors were causing "speckle pattern" of output bitstrings probabilities to break, resulting in very low cross-entropy benchmarking fidelity.

Application 2: Sample efficient distinguishability of quantum objects with incoherent access

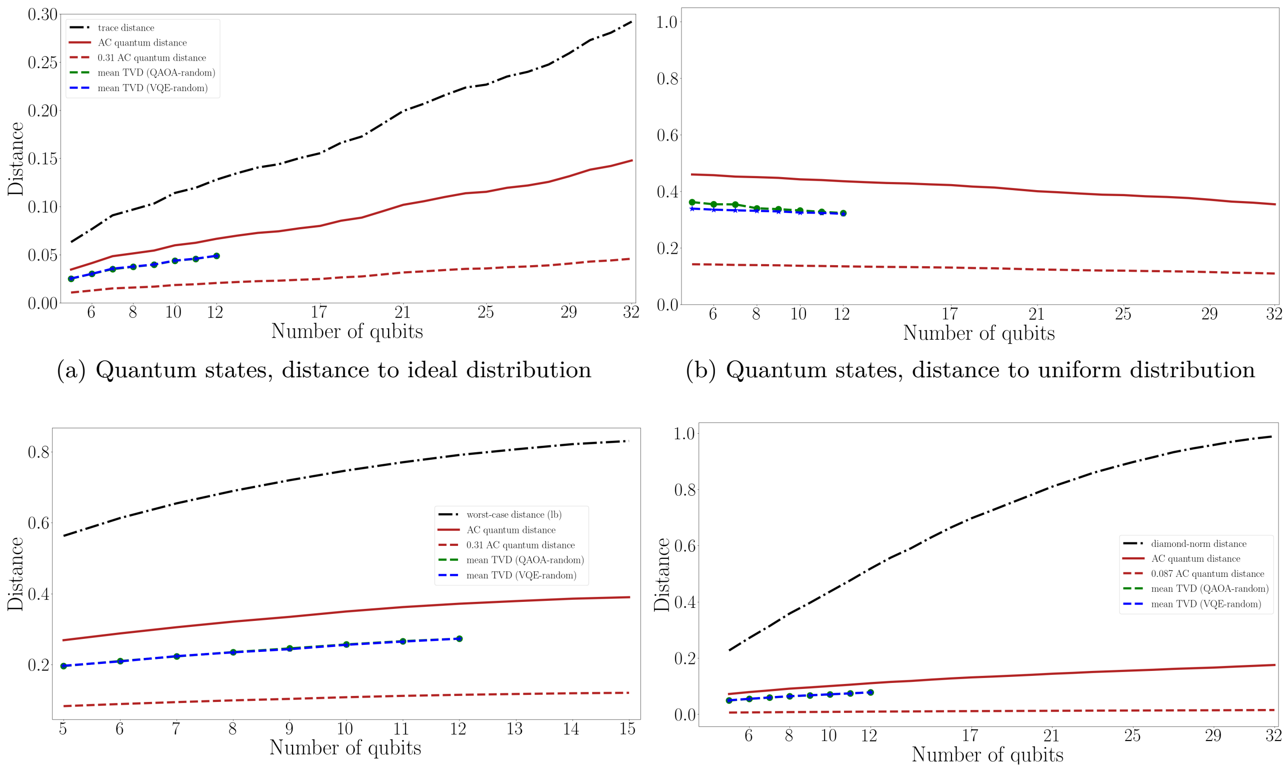
Example 6. *For any pure state ψ on $\mathcal{H}_d$ we have $\mathrm{d}_{\mathrm{av}}^{\mathrm{s}}\left(\psi, \tau_d\right) = \frac{1}{2}\sqrt{1 - \frac{1}{d}}$.*

It follows that a single round of a randomized protocol implicit in the definition of $\mathrm{d}_{\mathrm{av}}^{\mathrm{s}}$ (cf. Remark 1), realized via approximate 4-design and computational basis measurements, gives a constant bias in distinguishing *any* pure N qubit state ψ from the maximally mixed state: $p_{\mathrm{succ}}^{\mathrm{av}} \gtrsim 0.57$. This probability can be made arbitrarily close 1 by repeating the protocol and using the majority-vote strategy. Importantly, this method *does not* utilize the coherent access or a quantum memory (in a sense defined, e. g., in [2, 29]). We note that a related but distinct scenario is considered in Ref. [2]. There, the authors introduced the task of `PurityTesting` corresponding to discrimination between *unknown* Haar-random pure random state and maximally mixed state. For N qubit systems, Theorem 4 of [2] implies exponential lower bound for the query complexity k (number of usages of unknown quantum state) needed to succeed in this task, given incoherent access to objects in question. In contrast, our randomized measurement protocol gives high statistical distinguishability already for a single query *for all* states ψ. The difference comes from the fact that in the scenario considered in Example 6 the random state is arbitrary but known.

Example 7. *Let Λ_U be a a unitary channel corresponding to a unitary U on $\mathcal{H}_d$ and let Λ_{dep} be a depolarizing channel i.e. $\Lambda_{\mathrm{dep}}(\rho) = \tau_d$ for any ρ. Then we have $\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}\left(\Lambda_U, \Lambda_{\mathrm{dep}}\right) = \frac{1}{2}\sqrt{1 - \frac{1}{d^2}}$.*

In related task `FixedUnitary` studied in [2], one is asked to distinguish *unknown* Haar-random unitary channel Λ_U from Λ_{dep}. Exponential query complexity lower bound incoherent protocols was shown in [2]. By repeating analogous reasoning as for states, we get that when Λ_U is arbitrary but known, randomized, non-adaptive, and incoherent protocol, utilizing two realizations of approximate 4-designs, gives constant bias in success probability of discrimination of Λ_U from Λ_{dep} using just a single query.

Application 3: Strong complexity of quantum states and unitaries. The above o examples have interesting consequences for the notion of a strong state and unitary complexity investigated in [10]. There, the authors defined complexity C_Δ of N-qubit pure state ψ (resp. unitary circuit Λ_U) as the number of elementary gates needed to construct a circuit necessary to implement a *two-outcome* measurement discriminating between ψ (resp. depolarizing channel Λ_{dep}) with success probability $p_{\mathrm{succ}} = \frac{1}{2} + \Delta$. Our results imply that if the requirement of two-outcome measurement is relaxed, then measurements realizable with circuit depths $r = \mathrm{poly}(N)$



(a) Quantum states, distance to ideal distribution

(b) Quantum states, distance to uniform distribution

(c) Quantum measurements, distance to ideal distribution

(d) Quantum channels, distance to ideal distribution

Figure 2: Results of numerical studies for comparison between worst-case distance, average-case quantum distance and numerically calculated mean TVD. Plots 2a, 2c and 2d correspond to distance to ideal (noiseless) distribution. For states, we additionally plot distance to uniform (trivial) distribution on plot 2b. For average-case distance, we also plot value corresponding to lower bound on average-case TVD (following from Eqs. (2), (3), (4)). In case of worst-case distance, "lb" indicates lower-bound. Average-case quantum distances were calculated explicitly. Mean TVDs were calculated between (exact numerical) probability distributions over 1000 random instances of random unitaries.

can succeed in these discrimination tasks with a constant bias Δ_* *for all* states ψ and unitary channels Λ_U. This renders the so-defined notion of complexity trivial - all states and unitaries will have complexity $C_\Delta \leq \mathrm{poly}(N)$, unless bias δ satisfies $\Delta > \Delta_*$.

We note that large average-case distance d_{av} implies only information-theoretic distinguishability of quantum objects. The cost of classical post-processing needed to distinguish the probability distributions resulting from randomized protocols can be very large since they operate on exponentially large sample space.

Numerical results. Here we present the results of numerical studies of small-size quantum systems. We compare scaling with the system size for worst-case distance, average-case distance, and a mean TVD taken over an ensemble of random unitaries. The mean Total-Variation distance is calculated numerically over two types of ensembles of unitaries with a structure of variational circuits. One ensemble has a QAOA-like structure, while the other is a standard hardware-efficient VQE ansatz [47], both initialized with random parameters (see SM for exact form). Based on recent results [41], we expect them to form (approximate) unitary 4-designs.

We consider the following scenarios.

1. (States) We compare a randomly chosen Pauli eigenstate affected by random local Pauli noise with its ideal version (Fig. 2a) and with maximally mixed state $\frac{\mathbb{I}}{d}$ (Fig. 2b). This is the scenario considered in Example 1. The error probabilities are chosen randomly from range $[0.001, 0.01]$.

2. (Measurements) The noisy measurement is a tensor product POVM constructed from single-qubit measurements obtained via Quantum Detector Tomography [35] of IBM's 15-qubit Melbourne device. We compare it to ideal computational-basis measurement (Fig. 2c). Since the measurement noise in superconducting devices is usually highly asymmetric [36], we do not expect it to converge to the uniform distribution.

3. (Channels) We compare channel corresponding to random tensor product 1-qubit rotations around a random axis with ideal identity channel $\mathcal{I}$ (Fig 2d). Explicitly, the unitary corresponding to the channel has a form $\bigotimes_{k=1}^{N}\exp(-i\gamma_k V^{(k)})$, where $V^{(k)}$ is chosen randomly to be X, Y or Z gate, and $\gamma_k \in [0.025\pi, 0.0313\pi]$. Similarly to POVMs, we do not expect coherent errors to bring noisy distributions close to the uniform distribution.

In each case, the number of circuit layers is $\lfloor 1.5N \rfloor$. In Fig. 2 we collectively present the results of all simulations. Recall that both ensembles presented in Fig. 2 consist of circuits that are variational QAOA and VQE circuits with random parameters. From the plots, it is clear that in all studied cases for those ensembles, the average-case quantum distance is both significantly closer and more similar in scaling to the mean Total Variation distance between distributions in question, as compared to worst-case distance.

Acknowledgments

Acknowledgements We would like to thank Richard Kueng, Victor Albert, and Ingo Roth for interesting discussions and comments. We are grateful to Jordan Cotler for clarifying tasks of `PurityTesting` and `FixedUnitary` considered in Ref. [2]. We sincerely thank Susane Calegari for proofreading the manuscript. We used Qiskit [4] and QuTiP [31, 32] to perform some of the simulations.

The authors acknowledge the financial support by the TEAM-NET project co-financed by the EU within the Smart Growth Operational Programme (contract no. POIR.04.04.00-00-17C1/18-00). We are grateful to Jordan Cotler for improving our understanding of Purity Testing and Fixed Unitary tasks considered in Ref. [2]. We used Qiskit [4] and QuTiP [32] to perform some of the simulations.

References

[1] Scott Aaronson. Shadow tomography of quantum states. *SIAM Journal on Computing*, 49(5):STOC18–368–STOC18–394, 2020. doi:10.1137/18M120275X.

[2] Dorit Aharonov, Jordan Cotler, and Xiao-Liang Qi. Quantum algorithmic measurement. *Nature Communications*, 13(1), feb 2022. doi:10.1038/s41467-021-27922-0.

[3] Andris Ambainis and Joseph Emerson. Quantum T-designs: T-wise independence in the quantum world. In *Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity*, CCC '07, page 129–140, USA, 2007. IEEE Computer Society. doi:10.1109/CCC.2007.26.

[4] MD SAJID ANIS et al. Qiskit: An open-source framework for quantum computing, 2021. doi:10.5281/zenodo.2573505.

[5] Frank Arute et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 10 2019. doi:10.1038/s41586-019-1666-5.

[6] Ingemar Bengtsson and Karol Zyczkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press, 2006. doi:10.1017/CBO9780511535048.

[7] Bonnie Berger. The fourth moment method. *SIAM Journal on Computing*, 26(4):1188–1207, 1997. doi:10.1137/S0097539792240005.

[8] Sergio Boixo, Sergei V. Isakov, Vadim N. Smelyanskiy, Ryan Babbush, Nan Ding, Zhang Jiang, Michael J. Bremner, John M. Martinis, and Hartmut Neven. Characterizing quantum supremacy in near-term devices. *Nature Physics*, 14(6):595–600, 6 2018. doi:10.1038/s41567-018-0124-x.

[9] Fernando G. S. L. Brandão, Aram W. Harrow, and Michał Horodecki. Local random quantum circuits are approximate polynomial-designs. *Communications in Mathematical Physics*, 346(2):397–434, 9 2016. doi:10.1007/s00220-016-2706-8.

[10] Fernando G.S.L. Brandão, Wissam Chemissany, Nicholas Hunter-Jones, Richard Kueng, and John Preskill. Models of quantum complexity growth. *PRX Quantum*, 2:030316, 7 2021. doi:10.1103/PRXQuantum.2.030316.

[11] Senrui Chen, Wenjun Yu, Pei Zeng, and Steven T. Flammia. Robust shadow estimation. *PRX Quantum*, 2(3), 9 2021. doi:10.1103/prxquantum.2.030348.

[12] Joseph Emerson, Robert Alicki, and Karol Życzkowski. Scalable noise estimation with random unitary operators. *Journal of Optics B: Quantum and Semiclassical Optics*, 7(10):S347–S352, 9 2005. doi:10.1088/1464-4266/7/10/021.

[13] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm, 2014. arXiv:1411.4028.

[14] Edward Farhi and Aram W Harrow. Quantum supremacy through the quantum approx-

imate optimization algorithm, 2019. arXiv:1602.07674.

[15] Steven T. Flammia. Averaged circuit eigenvalue sampling, 2021. arXiv:2108.05803.

[16] Jay M. Gambetta, A. D. Córcoles, S. T. Merkel, B. R. Johnson, John A. Smolin, Jerry M. Chow, Colm A. Ryan, Chad Rigetti, S. Poletto, Thomas A. Ohki, and et al. Characterization of addressability by simultaneous randomized benchmarking. *Physical Review Letters*, 109(24), 12 2012. doi:10.1103/physrevlett.109.240504.

[17] Guillermo García-Pérez, Matteo A.C. Rossi, Boris Sokolov, Francesco Tacchino, Panagiotis Kl. Barkoutsos, Guglielmo Mazzola, Ivano Tavernelli, and Sabrina Maniscalco. Learning to measure: Adaptive informationally complete generalized measurements for quantum algorithms. *PRX Quantum*, 2(4), 11 2021. doi:10.1103/prxquantum.2.040342.

[18] Alexei Gilchrist, Nathan K. Langford, and Michael A. Nielsen. Distance measures to compare real and ideal quantum processes. *Phys. Rev. A*, 71(6):062310, 6 2005. doi:10.1103/PhysRevA.71.062310.

[19] Charles Hadfield. Adaptive pauli shadows for energy estimation, 2021. arXiv:2105.12207.

[20] Charles Hadfield, Sergey Bravyi, Rudy Raymond, and Antonio Mezzacapo. Measurements of quantum hamiltonians with locally-biased classical shadows. *Communications in Mathematical Physics*, 391(3):951–967, May 2022. doi:10.1007/s00220-022-04343-8.

[21] Jonas Haferkamp and Nicholas Hunter-Jones. Improved spectral gaps for random quantum circuits: Large local dimensions and all-to-all interactions. *Phys. Rev. A*, 104:022417, 8 2021. doi:10.1103/PhysRevA.104.022417.

[22] Pierre Hansen and Brigitte Jaumard. Algorithms for the maximum satisfiability problem. *Computing*, 44(4):279–303, 12 1990. doi:10.1007/BF02241270.

[23] Matthew P. Harrigan et al. Quantum approximate optimization of non-planar graph problems on a planar superconducting processor. *Nature Physics*, 17(3):332–336, feb 2021. doi:10.1038/s41567-020-01105-y.

[24] Aram W. Harrow. The church of the symmetric subspace, 2013. arXiv:1308.6595.

[25] Aram W. Harrow and Saeed Mehraban. Approximate unitary t-designs by short random quantum circuits using nearest-neighbor and long-range gates. *Communications in Mathematical Physics*, 401(2):1531–1626, may 2023. doi:10.1007/s00220-023-04675-z.

[26] Jonas Helsen, Xiao Xue, Lieven M. K. Vandersypen, and Stephanie Wehner. A new class of efficient randomized benchmarking protocols. *npj Quantum Information*, 5(1):71, Aug 2019. doi:10.1038/s41534-019-0182-7.

[27] Eric Huang, Andrew C. Doherty, and Steven Flammia. Performance of quantum error correction with coherent errors. *Phys. Rev. A*, 99(2):022313, 2 2019. doi:10.1103/PhysRevA.99.022313.

[28] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 6 2020. doi:10.1038/s41567-020-0932-7.

[29] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Information-theoretic bounds on quantum advantage in machine learning. *Phys. Rev. Lett.*, 126(19):190505, 5 2021. doi:10.1103/PhysRevLett.126.190505.

[30] J. L. W. V. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30(none):175 – 193, 1906. doi:10.1007/BF02418571.

[31] J. R. Johansson, P. D. Nation, and Franco Nori. QuTiP: An open-source python framework for the dynamics of open quantum systems. *Computer Physics Communications*, 183(8):1760–1772, 2012. doi:https://doi.org/10.1016/j.cpc.2012.02.021.

[32] J. R. Johansson, P. D. Nation, and Franco Nori. QuTiP 2: A python framework for the dynamics of open quantum systems. *Computer Physics Communications*, 184(4):1234–1240, 2013. doi:https://doi.org/10.1016/j.cpc.2012.11.019.

[33] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671):242–246, sep 2017. doi:10.1038/nature23879.

[34] Richard Kueng, Huangjun Zhu, and David Gross. Distinguishing quantum states using clifford orbits. *arXiv e-prints*, 9 2016. arXiv:1609.08595.

[35] J. S. Lundeen, A. Feito, H. Coldenstrodt-Ronge, K. L. Pregnell, Ch. Silberhorn, T. C. Ralph, J. Eisert, M. B. Plenio, and I. A. Walmsley. Tomography of quantum detectors. *Na-

ture Physics, 5:27, 11 2008. doi:10.1038/nphys1133.

[36] Filip B. Maciejewski, Flavio Baccari, Zoltán Zimborás, and Michał Oszmaniec. Modeling and mitigation of cross-talk effects in readout noise with applications to the quantum approximate optimization algorithm. *Quantum*, 5:464, 6 2021. doi:10.22331/q-2021-06-01-464.

[37] Filip B. Maciejewski, Zbigniew Puchała, and Michał Oszmaniec. Exploring quantum average-case distances: Proofs, properties, and examples. *IEEE Transactions on Information Theory*, 69(7):4600–4619, 2023. doi:10.1109/TIT.2023.3250100.

[38] Filip B. Maciejewski, Zoltán Zimborás, and Michał Oszmaniec. Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography. *Quantum*, 4:257, 4 2020. doi:10.22331/q-2020-04-24-257.

[39] Easwar Magesan, J. M. Gambetta, and Joseph Emerson. Scalable and robust randomized benchmarking of quantum processes. *Physical Review Letters*, 106(18), 5 2011. doi:10.1103/physrevlett.106.180504.

[40] Easwar Magesan, Jay M. Gambetta, B. R. Johnson, Colm A. Ryan, Jerry M. Chow, Seth T. Merkel, Marcus P. da Silva, George A. Keefe, Mary B. Rothwell, Thomas A. Ohki, and et al. Efficient measurement of quantum gate error by interleaved randomized benchmarking. *Physical Review Letters*, 109(8), 8 2012. doi:10.1103/physrevlett.109.080505.

[41] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9:4812, 11 2018. doi:10.1038/s41467-018-07090-4.

[42] Miguel Navascués and Sandu Popescu. How energy conservation limits our measurements. *Phys. Rev. Lett.*, 112:140502, 4 2014. doi:10.1103/PhysRevLett.112.140502.

[43] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. doi:10.1017/CBO9780511976667.

[44] Michał Oszmaniec, Leonardo Guerini, Peter Wittek, and Antonio Acín. Simulating positive-operator-valued measures with projective measurements. *Phys. Rev. Lett.*, 119:190501, 11 2017. doi:10.1103/PhysRevLett.119.190501.

[45] Michał Oszmaniec, Adam Sawicki, and Michał Horodecki. Epsilon-nets, unitary designs and random quantum circuits. *IEEE Transactions on Information Theory*, pages 1–1, 2021. doi:10.1109/TIT.2021.3128110.

[46] Robert M. Parrish, Edward G. Hohenstein, Peter L. McMahon, and Todd J. Martínez. Quantum computation of electronic transitions using a variational quantum eigensolver. *Phys. Rev. Lett.*, 122:230401, 6 2019. doi:10.1103/PhysRevLett.122.230401.

[47] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5(1), 7 2014. doi:10.1038/ncomms5213.

[48] John Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, 2:79, 8 2018. doi:10.22331/q-2018-08-06-79.

[49] Zbigniew Puchała, Łukasz Pawela, Aleksandra Krawiec, and Ryszard Kukulski. Strategies for optimal single-shot discrimination of quantum measurements. *Physical Review A*, 98(4), 10 2018. doi:10.1103/physreva.98.042103.

[50] John Watrous. Semidefinite programs for completely bounded norms. *Theory of Computing*, 5(11):217–238, 2009. doi:10.4086/toc.2009.v005a011.

[51] Qingling Zhu et al. Quantum computational advantage via 60-qubit 24-cycle random circuit sampling. *Science Bulletin*, 67(3):240–245, 2022. doi:10.1016/j.scib.2021.10.017.

A Worst-case quantum distances

As mentioned in the main text, commonly used distance measures are based on optimal statistical distinguishability of the objects in question. We have the following statistical interpretations of trace distance d_{tr} between quantum states [43], operational distance d_{op} [42, 49] between quantum measurements, and the diamond norm distance $\mathrm{d}_\diamond$ [43] between quantum channels

$$\mathrm{d}_{\mathrm{tr}}(\rho,\sigma) = \max_{\mathsf{M}} \mathrm{TV}(\mathbf{p}^{\rho,\mathsf{M}},\mathbf{p}^{\sigma,\mathsf{M}}) \;, \tag{12}$$

$$\mathrm{d}_{\mathrm{op}}(\mathsf{M},\mathsf{N}) = \max_{\rho} \mathrm{TV}(\mathbf{p}^{\rho,\mathsf{M}},\mathbf{p}^{\rho,\mathsf{N}}) \;, \tag{13}$$

$$\mathrm{d}_\diamond(\Lambda,\Gamma) = \max_{\rho,\ \mathsf{M}} \mathrm{TV}(\mathbf{p}^{\rho,\Lambda,\mathsf{M}},\mathbf{p}^{\rho,\Gamma,\mathsf{M}}) \;. \tag{14}$$

For the case of states, the maximization is over POVMs M used to distinguish them. We have a dual situation for measurements, the maximization is over input quantum states used to differentiate between one POVM and another. Finally, for the case of quantum channels and the diamond norm – the maximization is over both input states (on a possibly extended system) and over POVMs applied after a channel is implemented.

B Simplified proofs of main Theorems

Here we present simplified versions of proofs of Theorems 1, 2, and 3 from the main text. We refer the Reader to [37] for detailed calculations. Since in the main text we omitted dependence on δ in δ-approximate unitary designs, we consider here proofs only for exact (not approximate) unitary designs. The functional dependence for approximate designs, as well as proofs for approximate designs, can be found in [37].

B.1 Lower and upper bounds on absolute values

In scenarios we consider, we aim to find bounds on a random variable that is a Total-Variation distance (TVD) between two probability distributions. Note that since the expectation value is linear, it suffices to focus attention on a single outcome probability, and then add resulting bounds to obtain bounds on TVD.

Let us thus denote by $X_i = p_i - q_i$ the value of a difference of probabilities of measurement outcome i taken from probability distributions p and q that correspond to two quantum-mechanical protocols. This is a shorthand notation – the protocols are described in the main text and correspond to discrimination between two states, measurements, or general channels. Conveniently, it turns out that for considered scenarios and probability measures (Haar measure and unitary designs), one can find real parameters a such that the following holds.

Lemma 1. *(Lower bound on absolute value)*

$$a\ \sqrt{(\mathbb{E}[X_i^2])} \leq \mathbb{E}|X_i| \;, \tag{15}$$

where the value of a depends on whether we discriminate between states, measurements, or channels.

Proof. From Lemmas 4, and 5 in [37] it follows that one can find constants a such that

$$\mathbb{E}[X_i^4] \leq \frac{1}{\sqrt{a}}\ \left(\mathbb{E}[X_i^2]\right)^2 \;. \tag{16}$$

We note that Lemma 4 from [37] is Lemma 2 from [34], while Lemma 5 from [37] is one of the results in the accompanying technical manuscript [37]. Recall that Berger's inequality [7] states that for random variable Y with well-defined 2nd and 4th moments, we have

$$\frac{(\mathbb{E}[Y^2])^{\frac{3}{2}}}{(\mathbb{E}[Y^4])^{\frac{1}{2}}} \leq \mathbb{E}|Y| \;, \tag{17}$$

Then the proof follows from combining Eq. (16) with Berger's inequality. □

At the same time, we have that the following holds for any random variable Y.

Lemma 2. *(Upper bound on absolute value)*

$$\mathbb{E}[|Y|] = \mathbb{E}[\sqrt{Y^2}] \leq \sqrt{\mathbb{E}[Y^2]} \;.$$

Proof. The above is a special case of Jensen's inequality [30] which states that for a concave function f we have $\mathbb{E}[f(Y)] \leq f\left(\mathbb{E}[Y]\right)$. □

From the above one can see that to obtain both lower and upper bound on TVD it suffices to calculate the 2nd moment of $|X_i|$. To do so, the following Lemma will be useful.

Lemma 3 (Ancillary integral for 2nd moment)**.** *Let A be a Hermitian operator on $(\mathcal{H}_d)$ and μ be a Haar measure. Then we have*

$$\underset{U\sim\mu}{\mathbb{E}}\left[\mathrm{tr}(|i\rangle\langle i|UAU^\dagger)^2\right] = \frac{1}{d(d+1)}\left(\mathrm{tr}(A^2)+\mathrm{tr}(A)^2\right) . \tag{18}$$

Proof. We first write simple manipulation

$$\underset{U\sim\mu}{\mathbb{E}}\left[\mathrm{tr}(|i\rangle\langle i|UAU^\dagger)^2\right] = \underset{U\sim\mu}{\mathbb{E}}\left[\mathrm{tr}\left((U^\dagger)^{\otimes 2}|i\rangle\langle i|^{\otimes 2}U^{\otimes 2}A^{\otimes 2}\right)\right] . \tag{19}$$

This allows us to evaluate the RHS using standard techniques of Haar measure integration (see, e.g., [24, Prop. 6]), and obtain that it is proportional to $\mathbb{P}_{\mathrm{sym}}^{(2)}$, i.e., projector onto 2-fold symmetric subspace of $\mathcal{H}_{\mathrm{d}}^{\otimes 2}$. Then the proof follows from applying identities $\mathbb{P}_{\mathrm{sym}}^{(2)} = \frac{1}{2}(\mathbb{I}+\mathbb{S})$ and $\mathrm{tr}\left(\mathbb{S}A^{\otimes 2}\right) = \mathrm{tr}\left(A^2\right)$, where $\mathbb{S} = \sum_{i,j=1}^{d}|ij\rangle\langle ji|$ is a generalized SWAP operator. □

B.2 Proofs of Theorems 1 and 2

For states and measurements, the proofs are essentially identical, thus we consider them together. As stated above, obtaining both bounds reduces to calculating second moments of $|X_i|$, which we will now outline.

Consider discrimination of states ρ and σ. We calculate the second moment by applying Lemma 3 to operator $\Delta_i = \rho - \sigma$, which yields

$$\underset{U\sim\mu}{\mathbb{E}}[X_i^2] = \underset{U\sim\mu}{\mathbb{E}}\,\mathrm{tr}(U^\dagger|i\rangle\langle i|U\Delta_i)^2 = \frac{1}{d(d+1)}\,\mathrm{tr}(\Delta^2) \;= \frac{1}{d(d+1)}||\rho-\sigma||_{\mathrm{HS}}^2. \tag{20}$$

Note that the RHS does not depend on index i. The proof concludes by taking a square root of the RHS and summing over i.

Consider discrimination of measurements M and N. In analogy to states, we calculate the 2nd moment by applying Lemma 3 to operator $\tilde{\Delta}_i = M_i - N_i$, and obtain

$$\underset{U\sim\mu}{\mathbb{E}}[X_i^2] = \underset{U\sim\mu}{\mathbb{E}}\,\mathrm{tr}(U^\dagger\psi_0 U\tilde{\Delta}_i)^2 = \frac{1}{d(d+1)}\left(\mathrm{tr}(\tilde{\Delta}_i^2)+\mathrm{tr}(\tilde{\Delta}_i)^2\right) . \tag{21}$$

B.3 Proof of Theorem 3

In the case of states and measurements, there was only a single average (over projective measurements for states and over pure states for measurements). However, for quantum channels we have both quantum inputs and outputs, thus we need to calculate two averages. Consider discrimination between two channels Λ and Γ. Denote $\Delta = \Lambda - \Gamma$.

To proceed, we first apply Theorem 1 to perform averaging over projective measurements after the application of the channel (or, equivalently, averaging over unitaries acting on the output of channels followed by fixed measurement in a standard basis). In this way, we remove one integral and reduce the problem to finding bounds on the expected value of $\underset{\psi\sim\nu_S}{\mathbb{E}}\|\Delta[\psi]\|_{\mathrm{HS}} = \underset{\psi\sim\nu_S}{\mathbb{E}}\left[\sqrt{\mathrm{tr}\left(\Delta[\psi]^2\right)}\right]$. Using the same line of arguments as before, this quantity can be lower and upper bounded by evaluating $\underset{\psi\sim\nu_S}{\mathbb{E}}\mathrm{tr}\left(\Delta[\psi]^2\right)$. This is done by first performing simple manipulation

$$\underset{\psi\sim\nu_S}{\mathbb{E}}\left[\mathrm{tr}\left(\Delta[\psi]^2\right)\right] = \underset{\psi\sim\nu_S}{\mathbb{E}}\left[\mathrm{tr}\left(\mathbb{S}\Delta[\psi]^{\otimes 2}\right)\right] = \mathrm{tr}\left(\mathbb{S}\underset{\psi\sim\nu_S}{\mathbb{E}}\left[\Delta[\psi]^{\otimes 2}\right]\right). \tag{22}$$

The last term in the above can then be evaluated using standard techniques of Haar measure integration (see, e.g., [24, Prop. 6], and recall the proof of Lemma 3). The computation yields

$$\mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}}\left[\mathrm{tr}(\varDelta[\psi]^2)\right] = \frac{d^2}{d(d+1)}\left(\mathrm{tr}\left(\varDelta\left[\frac{\mathbb{I}}{d}\right]^2\right) + \mathrm{tr}\left(\mathbb{S}\varDelta^{\otimes 2}\left[\frac{\mathbb{S}}{d^2}\right]\right)\right) . \tag{23}$$

Noticing that $\mathrm{tr}\left(\mathbb{S}\varDelta^{\otimes 2}\left[\frac{\mathbb{S}}{d^2}\right]\right) = \|\mathcal{J}_{\varDelta}\|^2_{\mathrm{HS}}$ concludes the proof.

C Proofs of claims in Examples 1-4

As mentioned in the main text, Examples 1-5 follow directly from more general expressions in examples in technical manuscript [37]. Specifically, the Example 1 follows from Example 9, Examples 2 and 3 follow from Example 10 (in case of Example 3 arguments are slightly more involved, as presented below), while Examples 4 and 5 follow from Example 14.

We now recall statements of Example 9 for Reader's convenience.

Example 8. *[Example 9 from [37]] 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 tensor product Pauli channel* $\varLambda^{pauli} = \otimes_{i=1}^{N}\varLambda_i^{pauli}$, *where single-qubit channel is* $\varLambda_i^{pauli}(\rho) = \sum_{j=1} p_j^{(i)}\sigma_j\rho\sigma_j$ *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*

$$\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\varLambda^{pauli}(\psi^{pauli}), \frac{\mathbb{I}}{d}) = \frac{1}{2}\sqrt{\varPi_{i=1}^{N}\left(1 - 2q^{(i)}(1-q^{(i)})\right) - \frac{1}{d}} , \tag{24}$$

$$\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\varLambda^{pauli}(\psi^{pauli}), \psi^{pauli}) = \frac{1}{2}\sqrt{1 - 2\varPi_{i=1}^{N}q^{(i)} + \varPi_{i=1}^{N}(1 - 2q^{(i)}(1-q^{(i)}))} , \tag{25}$$

We start by defining function $f^{(i)} = q^{(i)}(1-q^{(i)})$, as well as average noise properties $q^{av} = \frac{1}{N}\sum_{i=1}^{N} q^{(i)}$ and $f^{av} = \frac{1}{N}\sum_{i=1}^{N} f^{(i)}$. We then bound Eq. (24) from above as

$$\sqrt{\varPi_{i=1}^{N}\left(1 - 2f^{(i)}\right) - \frac{1}{d}} \leq \sqrt{\varPi_{i=1}^{N}\left(1 - 2f^{(i)}\right)} , \tag{26}$$

and continue with bounding (positive) expression inside square root as

$$\varPi_{i=1}^{N}\left(1 - 2f^{(i)}\right) = \left(\sqrt[N]{\varPi_{i=1}^{N}\left(1 - 2f^{(i)}\right)}\right)^N \leq \left(\frac{\sum_{i=1}^{N}(1-2f^{(i)})}{N}\right)^N = (1 - 2f^{av})^N \leq \exp(-2f^{av}N) , \tag{27}$$

where in first inequality we used inequality between geometric and arithmetic means together with a fact that $x^N \geq y^N$ for $x > y > 0$. In second inequality we used that for $0 \leq x \leq 1$ and $N \geq 1$, we have $(1-x)^N \leq \exp(-xN)$. Note that each term $2f^{(i)}$ lies in interval $2f^{(i)} \in \left[0, \frac{1}{2}\right]$. Combining everything we obtain

$$\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\varLambda^{\mathrm{pauli}}(\psi^{\mathrm{pauli}}), \frac{\mathbb{I}}{d}) \leq \frac{1}{2}\exp(-f^{av}N) , \tag{28}$$

which concludes the proof of first bound.

To bound Eq. (25) from below, we start by again employing inequality between geometric and arithmetic mean, namely

$$1 - 2\varPi_{i=1}^{N}q^{(i)} = 1 - 2\left(\sqrt[N]{\varPi_{i=1}^{N}q^{(i)}}\right)^N \geq 1 - 2\left(\frac{\sum_{i=1}q^{(i)}}{N}\right)^N = 1 - 2\left(q^{av}\right)^N , \tag{29}$$

which after combining with Eq. (25) yields

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\mathit{\Lambda}^{\mathrm{pauli}}(\psi^{\mathrm{pauli}}),\psi^{\mathrm{pauli}}) \geq \frac{1}{2}\sqrt{1-2\left(q^{av}\right)^{N}+\mathit{\Pi}_{i=1}^{N}(1-2q^{(i)}(1-q^{(i)}))} \;\geq \frac{1}{2}\sqrt{1-2\left(q^{av}\right)^{N}}\;. \tag{30}$$

The above bound is valid provided that argument is still contained in the domain of square root, i.e., we need to impose

$$1-2\left(q^{av}\right)^{N} \geq 0 \implies q^{av} \leq \sqrt[N]{\frac{1}{2}}\;. \tag{31}$$

Note that $\sqrt[N]{\frac{1}{2}} \xrightarrow{N\rightarrow\infty} 1$, and since q^{av} is by definition lower than 1, the bound becomes less restrictive for higher system sizes. For small systems it is valid only for high noise (small q^{av}), but in such cases one can simply use the exact expressions from Eqs. (24) and (25).

The exactly same reasoning is applied for Examples 2 and 4, for which all expressions have almost the same functional forms (see [37]). We now consider bound from Example 3 from the main text, for which the first part of the proof is slightly more involved due to more general noise model considered.

Example 9 (Example 3 from the main text)**.** *Let* $\mathsf{P} = (|\mathbf{x}\rangle\!\langle\mathbf{x}|)_{\mathbf{x}\in\{0,1\}^N}$ *be a computational basis measurement on* N *qubit system. Let* $\mathsf{M} = (M_{\mathbf{x}})_{\mathbf{x}\in\{0,1\}^N}$ *be a POVM specified by effects* $M_{\mathbf{x}} = \mathit{\Lambda}_1^{\dagger}(|x_1\rangle\!\langle x_1|)\otimes \ldots \otimes \mathit{\Lambda}_N^{\dagger}(|x_N\rangle\!\langle x_N|)$*, where* $\mathit{\Lambda}_i$ *are quantum channels affecting* i*'th qubit, and* $\mathit{\Lambda}_i^{\dagger}$ *is the conjugate of* $\mathit{\Lambda}_i$*. Define classical success probability as* $p^{(i)}(k|k) = \mathrm{tr}\left(\mathit{\Lambda}_i^{\dagger}\left(|x_i\rangle\langle x_i|\right)|x_i\rangle\langle x_i|\right)$ *and corresponding average* $q_{av}^{(i)} = \frac{p^{i}(0|0)+p^{(i)}(1|1)}{2}$*. Let* $q^{av} := \frac{1}{N}\sum_{i=1}^{N} q_{av}^{(i)}$*. Assume* $q_{av}^{(i)} \geq \frac{1}{2}$ *for each qubit* i *and that* $q^{av} \leq \sqrt[N]{\frac{1}{2}}$*. Then we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathsf{M},\mathsf{P}) > \frac{1}{2}\sqrt{1-2(q^{av})^{N}}\;. \tag{32}$$

To prove the above, first one applies maximally-dephasing channel to both measurements and uses data-processing inequality for average-case distance to bound the distance from below by the diagonal part of the POVM M. Specifically, define dephased POVM $\mathit{\Phi}_{\mathrm{dep}}(\mathsf{M})$ via its effects $\mathit{\Phi}_{\mathrm{dep}}(\mathsf{M})_i = \mathit{\Phi}_{\mathrm{dep}}(M_i)$, where maximally dephasing channel acts on any operator A as $\mathit{\Phi}_{\mathrm{dep}}(A) = \mathrm{diag}(A)$, with $\mathrm{diag}(A)$ denoting diagonal part of A. Note that for compuational basis measurement P we have $\mathit{\Phi}_{\mathrm{dep}}(\mathsf{P}) = \mathsf{P}$. Thus we have

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathit{\Phi}_{\mathrm{dep}}(\mathsf{M}),\mathit{\Phi}_{\mathrm{dep}}(\mathsf{P})) \geq \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathit{\Phi}_{\mathrm{dep}}(\mathsf{M}),\mathsf{P})\;. \tag{33}$$

The above allows to treat noise as classical and look only on assignment infidelities for classical states (i.e., error probabilites when measured states are computational-basis states). Note that, importantly, maximally dephasing channel does not change the product structure of M. Thus we can treat this dephased POVM $\mathit{\Phi}_{\mathrm{dep}}(\mathsf{M})$ as related to computational basis measurement via some tensor product stochastic map $\mathrm{T} = \bigotimes_{i=1}^{N}\mathrm{T}^{(i)}$, where $\mathrm{T}^{(i)}$ acts on ith qubit and is specified by two success probabilities $p^{(i)}(0|0)$ and $p^{(i)}(1|1)$ (see, for example, Ref. [38] for more details on stochastic readout noise). Thus we have

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathsf{M},\mathsf{P}) \geq \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathrm{T}\mathsf{P},\mathsf{P})\;, \tag{34}$$

where $\mathrm{T}\mathsf{P}$ is a POVM with ith effect given by $(\mathrm{T}\mathsf{P})_i = \sum_i \mathrm{T}_{ij}\,|j\rangle\langle j|$ and stochastic map T is defined via diagonal elements of original POVM M (as in discussion above).

Now one applies Lemma 28 from technical version of the work [37] that lower bounds the distance via symmetrized version of T, where now both error probabilities are the same and equal to $q_{av}^{(i)} = \frac{p^{(i)}(0|0)+p^{(i)}(1|1)}{2}$ (note that this is equivalent to Pauli bitflip channel applied with probability $q_{av}^{(i)}$). Denote such symmetrized version of T as $\mathrm{T}^{\mathrm{sym}}$. This gives

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathrm{T}\mathsf{P},\mathsf{P}) \geq \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathrm{T}^{\mathrm{sym}}\mathsf{P},\mathsf{P}). \tag{35}$$

Therefore we reduced the lower bound to scenario considered in Example 2 from the main text, for which the bound was proved above.

D Details on numerical simulations

In the main text, we presented numerical results of calculating mean Total-Variation distances over ensembles of random unitaries. Here we describe how those ensembles were constructed. In each case, the p-layer circuit can be written as

$$U_p = \prod_{j=1}^{p} U_{\mathrm{rot},j}\; U_{\mathrm{ent},j}\ . \tag{36}$$

where $U_{\mathrm{rot},j}$ is a "rotation block" and U_{ent} is an "entangling block". Exact form of the evolution, as well as the initial state depend on the ensemble. We consider two such ensembles:

1. Circuits that originate from QAOA instance for fixed Hamiltonian $\mathrm{H}_{2\mathrm{SAT}}$ encoding fixed (random) instance of random MAX-2-SAT problem [22]. In this case, the initial state is of the form $|+\rangle^{\otimes N}$ with $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, while unitary evolution is given by $U_{\mathrm{rot,j}} := U_{\alpha_j} = \exp\left(-i\alpha_j \sum_{k=1}^{N} \sigma_x^{(k)}\right)$, and $U_{\mathrm{ent},j} := U_{\beta_j} = \exp\left(-i\beta_j \mathrm{H}_{2\mathrm{SAT}}\right)$, with $\sigma_x^{(k)}$ being X gate on kth qubit. For each j, α_j and β_j are N-dimensional vectors of parameters chosen randomly from range $[-\pi, \pi]$.

2. Circuits of a form of generic Hamiltonian-independent VQE ansatz with initial state being $|0\rangle^{\otimes N}$. We choose the rotation block to be of the form $U_{\mathrm{rot,j}} := U_{\alpha_j} = \bigotimes_{k=1}^{N} \exp\left(-i\alpha_{2j}\ \sigma_Z^{(k)}\right) \circ \exp\left(-i\alpha_{2j+1}\ \sigma_Y^{(k)}\right)$, where σ_Y, σ_Z are Y and Z gates. The entangling block is $U_{\mathrm{ent},j} = U_{\mathrm{ent}} := \prod_{k=1}^{N-1} \mathrm{CX}_{k,k+1}$ with $\mathrm{CX}_{k,l}$ denoting CX gate between qubits k and l. For each j, α_j is a $2N$-dimensional vector of parameters chosen randomly from range $[-\pi, \pi]$.

E Additional numerical results

Here we provide some additional plots with numerical results.

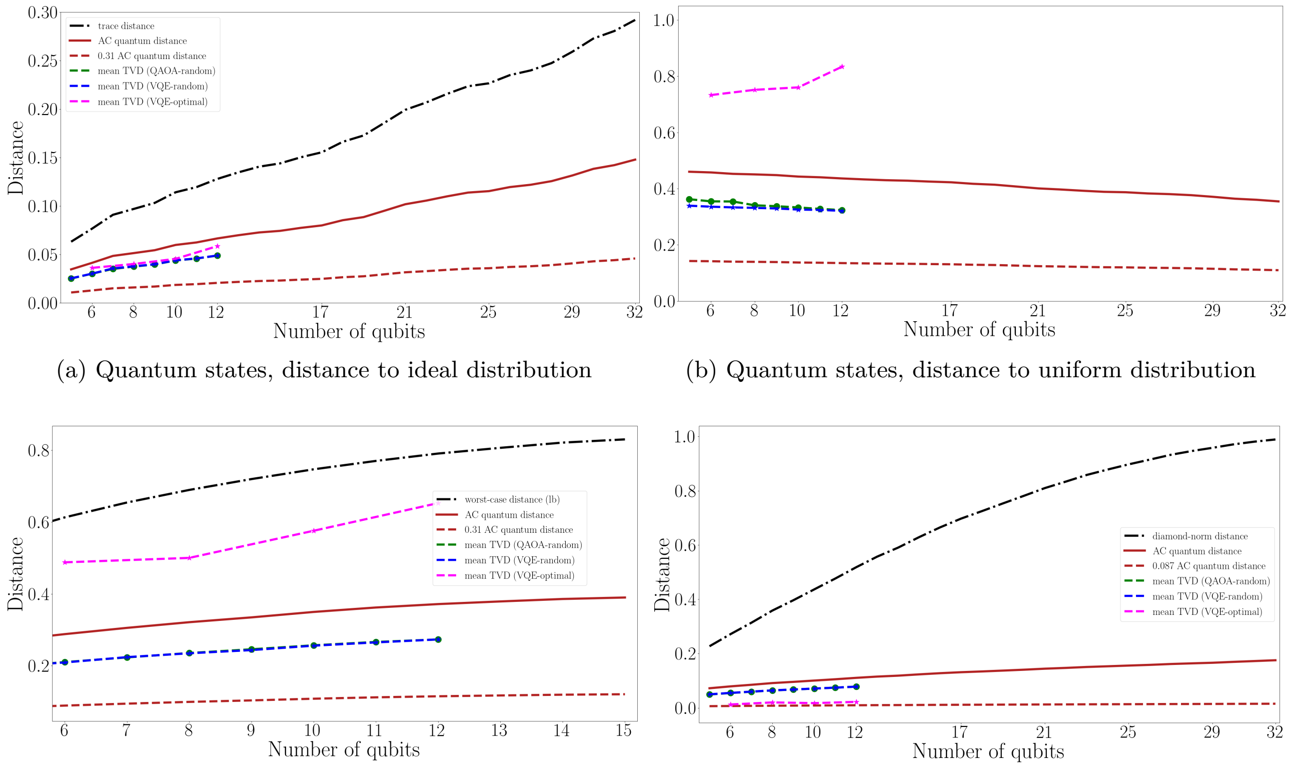
In Fig 3 we present the same plots as for Fig 2 in the main text, but with additional, third ensemble of unitaries considered (see previous section for description of two ensembles used in the main text).

3. The third ensemble is similar to the second VQE-like (see previous section), but now rotation block contains only Y rotations. Furthermore, the angles are *not random*, but they are chosen from a fixed set of parameters that come from solutions of variational optimization. In other words, each used unitary corresponds to a circuit that was found to be optimal in a VQE optimization (as opposed to uniformly random angles taken for both previous ensembles). We use datasets from Ref. [17] where authors developed an adaptive measurement scheme that improves performance of VQE.

Ensemble of type 3, due to limited computational resources, consist of only $7-12$ unitaries (recall that generating each unitary requires performing full VQE optimization). This implies that this ensemble *does not* form even unitary 2-design. It is nevertheless still interesting to investigate its behaviour, since those are circuits of particular practical importance.

From Fig. 3 we see that for distances between ideal and noisy distributions, the results are qualitatively similar to random ensembles in case of states and channels, but significantly different for quantum measurements. Recall that POVMs used to generate plot 3c are results of detector tomography of actual quantum device from IBM. In this case, the noise affects results so much, that empirical TVDs are closer to worst-case than to average-case bounds. In case of distance between noisy and uniform distribution for states (Fig. 3b) we also observe that average-case distances do not capture well the behaviour of the distributions for unitaries obtained in VQE optimization.

In Fig 4 we present histograms of TVDs over random unitaries. The data-points correspond to simulations presented in Fig. 2 in the main text. The plots show how the TVDs concentrate for small system sizes and demonstrate that all random points lied well within bounds provided by average-case distances.



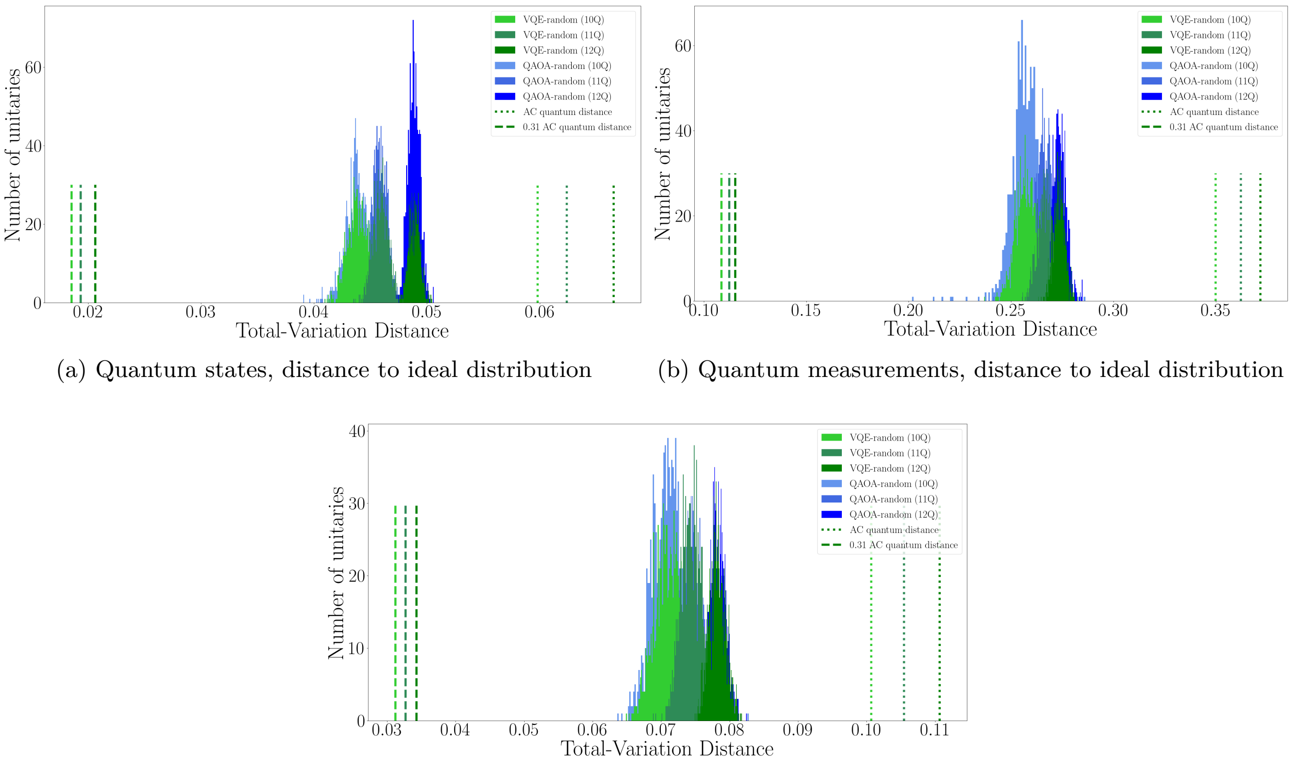
(a) Quantum states, distance to ideal distribution

(b) Quantum states, distance to uniform distribution

(c) Quantum measurements, distance to ideal distribution

(d) Quantum channels, distance to ideal distribution

Figure 3: Results of numerical studies for comparison between worst-case distance, average-case quantum distance and numerically calculated mean TVD. The plot is exactly the same as Fig 2 in the main text, but with additional ensemble of unitaries considered (see text description).



(a) Quantum states, distance to ideal distribution

(b) Quantum measurements, distance to ideal distribution

(c) Quantum channels, distance to ideal distribution

Figure 4: Histograms of TVDs obtained for random ensembles considered in numerical simulations corresponding to Fig 2 in the main text. Different shades of a given color (blue or green) correspond to different system sizes for a given ensemble (QAOA or VQE). Bounds from average-case distances are indicated via dashed lines and for each dimension are the same for both ensembles (they depend only on quantum objects in question, not on the choice of random ensemble).