Shadow Tomography on General Measurement Frames

L. Innocenti,[1,*] S. Lorenzo,[1] I. Palmisano,[2] F. Albarelli,[3,4] A. Ferraro,[2,3] M. Paternostro,[1,2] and G. M. Palma[1,5]

[1] *Università degli Studi di Palermo, Dipartimento di Fisica e Chimica—Emilio Segrè, via Archirafi 36, Palermo I-90123, Italy*

[2] *Centre for Quantum Materials and Technologies, School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, United Kingdom*

[3] *Quantum Technology Lab, Dipartimento di Fisica Aldo Pontremoli, Università degli Studi di Milano, Milano I-20133, Italy*

[4] *Istituto Nazionale di Fisica Nucleare, Sezione di Milano, via Celoria 16, Milan 20133, Italy*

[5] *NEST, Istituto Nanoscienze-CNR, Piazza S. Silvestro 12, Pisa 56127, Italy*

(Received 13 February 2023; revised 24 July 2023; accepted 2 October 2023; published 20 November 2023)

We provide a new perspective on shadow tomography by demonstrating its deep connections with the general theory of measurement frames. By showing that the formalism of measurement frames offers a natural framework for shadow tomography—in which "classical shadows" correspond to unbiased estimators derived from a suitable dual frame associated with the given measurement—we highlight the intrinsic connection between standard state tomography and shadow tomography. Such a perspective allows us to examine the interplay between measurements, reconstructed observables, and the estimators used to process measurement outcomes, while paving the way to assessing the influence of the input state and the dimension of the underlying space on estimation errors. Our approach generalizes the method described by Huang *et al.* [H.-Y. Huang *et al.*, Nat. Phys. **16**, 1050 (2020)], whose results are recovered in the special case of covariant measurement frames. As an application, we demonstrate that a sought-after target of shadow tomography can be achieved for the entire class of tight rank-1 measurement frames—namely, that it is possible to accurately estimate a finite set of generic rank-1 bounded observables while avoiding the growth of the number of the required samples with the state dimension.

DOI: 10.1103/PRXQuantum.4.040328

I. INTRODUCTION

The reliable reconstruction of the information encoded in a quantum register is one of the stepping stones of any quantum information processing device. In this respect, quantum state tomography (QST), i.e., the task of estimating quantum states from a measured data set, is the gold standard for verification and benchmarking of quantum devices [1–3]. QST has been performed in countless experiments by measuring a complete set of observables the expectation values of which determine the quantum state.

As the typical representation of density matrices implies a number of coefficients exponential in the number of constituent subsystems, the standard formulation of tomography [4] of a generic state requires an exponential time in the system size. Alternative methods based on efficient representations of multiparty quantum states—such as matrix product states [5]—have led to improved schemes for state tomography. Such an advantage, however, is achieved only for those states that are efficiently represented in the ansatz that is chosen. On the other hand, performing QST of d-dimensional quantum states, within error ϵ (in trace distance), requires a number of copies of the unknown state that scales polynomially with d [2,4]. In this context, tight lower bounds to single-copy nonadaptive state reconstruction have been proven [6–9].

However, the reconstruction of specific features of a state, rather than performing full tomographic reconstruction, is achievable with a much smaller amount of resources [10,11]. In particular, the number of measurements required to estimate the expectation value of M observables within error ϵ scales logarithmically with M and does not depend explicitly on the state's

*luca.innocenti@unipa.it

dimension—the associated task is referred to as "shadow tomography" [12]. An explicit way to implement shadow tomography via random Clifford circuits has recently been proposed [13–15]. A review discussing some of the relations between state tomography and shadow tomography can be found in Ref. [16]. In particular, a generalization of shadow tomography to general quantum measurements has recently been proposed in Ref. [17,18].

Here, we further the grounding of shadow tomography for agile property reconstruction by highlighting its deep connection with the approach of state tomography via measurement frames [19–25]. Our formalism reduces to the standard approach of Ref. [14] in special cases, and is compatible with its generalizations presented in Refs. [17,18]. We demonstrate that this general formalism provides a simple framework to understand the relationship between measurement, target observable, and estimator used to postprocess measurement outcomes, as well as how the input state and the dimension of the underlying space affect the estimation error. This approach also connects directly with general metrological considerations, showing how classical shadows can be seen as minimum-variance unbiased linear estimators. This formalism can also potentially be of great use to study the efficiency of state-estimation schemes involving generalized measurements and single-setting measurement schemes, which have recently attracted significant attention [26,27].

More specifically, we take the analysis of measurement frames developed for state tomography and specialize it to analyze estimation errors for shadow-tomography tasks. We discuss how the mean-squared-error (MSE) matrix, a quantity defined to study state tomography the trace of which gives the estimation error, also reveals a powerful tool to study errors in shadow tomography. We show how, for any choice of measurement, multiple possible unbiased estimators can be used to postprocess the measurement data to recover the target observables and we discuss how to find the unbiased estimator that minimizes the variance with respect to any given input state, as well as the one with minimum averaged variance—with the average taken with respect to uniformly random input states. We also demonstrate that the notion of *shadow norm* of an observable, introduced in Ref. [14], emerges naturally in this more general formalism. Furthermore, we examine the behavior of errors for different choices of measurement, given a fixed optimal estimator. A crucial feature of shadow tomography is the favorable scaling of estimation errors with the dimension of the state. Focusing on this aspect, we derive explicit bounds for best- and worst-case estimation errors corresponding to different measurement choices and find a wide class of measurements that allow us to estimate properties as efficiently as the protocol used in Ref. [14].

The remainder of this paper is organized as follows. In Sec. II, we present a reformulation of shadow tomography using the formalism of measurement frames. In Sec. III, we introduce the notion of the canonical estimator, review standard results for linear tomography in the measurement-frames formalism, and highlight the strong analogy between shadow and linear tomography. In Sec. IV, we derive general bounds for the variance of the introduced estimators, both in the averaged and best- and worst-case settings, and establish general results connecting the symmetry of the measurement with the associated variances. In Sec. VI, we show explicitly how the formalism introduced in Ref. [14] can be viewed as a special instance of our approach, specifically when employing covariant measurements and canonical estimators. Our conclusions and an outlook are finally given in Sec. VII. Additional in-depth discussions about the derivations and formalism used throughout the paper can be found in the appendixes.

II. SHADOW TOMOGRAPHY ON MEASUREMENT FRAMES

In this section, we demonstrate explicitly how the formalism of measurement frames provides a natural framework for discussing shadow tomography on general quantum measurements. The approach to shadow tomography [12] introduced in Refs. [13,14] relies on the idea of *classical shadows*, which are functions of the measurement outcomes that can be used to derive good estimates for target observables. These classical shadows can be understood as a way to construct unbiased estimators for the input state that operate on individual measurement outcomes. Unbiased estimators for target observables are then easily obtained via these classical shadows. By not requiring us to recover a tomographically complete description of the states, such specialized estimators allow us to efficiently estimate desired features of input states. An explicit protocol to perform shadow tomography with Clifford circuits has recently been proposed in Refs. [13,14] and some generalizations to general measurements have been proposed in Refs. [17,18]. Here, we demonstrate that frame theory [28,29]—and, in particular, the formalism of measurement frames [19,30–32]—provide a remarkably simple conceptual framework to think about shadow tomography and allow us to directly view the "classical shadows" as the unbiased estimators that constitute the elements of the dual measurement frame.

A. Notation

We will restrict our attention to finite-dimensional states and measurements with a finite number of outcomes. This constraint allows a more concise presentation and can relaxed later without significantly changing the formalism or the results. Following the notation of Ref. [33], we will denote the real vector space of Hermitian operators acting on a d-dimensional complex vector space $\mathbb{C}^d$ by $\mathrm{Herm}(\mathbb{C}^d)$, the set of positive semidefinite operators acting on the same space by $\mathrm{Pos}(\mathbb{C}^d)$, and the subset of density matrices by $\mathrm{D}(\mathbb{C}^d) \subset \mathrm{Pos}(\mathbb{C}^d)$. To focus on the linear algebraic properties involved in the calculations, we will use the notation $\langle X, Y\rangle \equiv \mathrm{tr}(X^\dagger Y)$ to denote the Hilbert-Schmidt inner product between operators X and Y and $\|X\|_2 \equiv \sqrt{\mathrm{tr}(X^2)}$ for the corresponding operator norm. We will denote a positive operator-valued measure (POVM) with ℓ outcomes by $\boldsymbol{\mu} \equiv (\mu_a)_{a=1}^{\ell}$, where $\mu_a \in \mathrm{Pos}(\mathbb{C}^d)$ and $\sum_a \mu_a = I$. Given a state $\rho \in \mathrm{D}(\mathbb{C}^d)$, the associated outcome probabilities are thus given by $p_a(\rho) = \langle \mu_a, \rho\rangle$. Any procedure involving an arbitrary evolution followed by a measurement in some basis can be concisely modeled via one such POVM.

B. Frame theory

In linear algebra, a *frame* [28,29,34] for a vector space V is a collection of vectors $v_k \in V$ such that, for all $v \in V$, $A\|v\|^2 \le \sum_k |\langle v_k, v\rangle|^2 \le B\|v\|^2$, for some $0 < A \le B < \infty$. These can informally be thought of as overcomplete bases: sets of vectors spanning the space, thus providing a linear decomposition for all other vectors. For finite frames in finite-dimensional spaces, a set $(v_k)_k$ is a frame if and only if it spans V [28]. Given a frame $(v_k)_k$, any $v \in V$ can be linearly decomposed as

$$v = \sum_k \langle v_k, v\rangle \tilde{v}_k = \sum_k \langle \tilde{v}_k, v\rangle v_k, \tag{1}$$

where $(\tilde{v}_k)_k$ is another frame, referred to as a *dual frame* of $(v_k)_k$. A frame $(v_k)_k$ admits infinitely many possible dual frames if and only if it is not linearly independent—i.e., if it is "overcomplete."

If we want to estimate a given unknown state ρ from measurement outcomes, a natural class of objects to study are *unbiased estimators*. These are functions $\hat{f} : \Sigma \to \mathrm{Herm}(\mathbb{C}^d)$, which map the set of measurement outcomes Σ into Hermitian operators that, on average, reproduce the measured state. That is, more precisely,

$$\mathbb{E}[\hat{f}\,|\rho] \equiv \sum_a \langle \mu_a, \rho\rangle \hat{f}(a) = \rho. \tag{2}$$

The elements of a POVM $\boldsymbol{\mu} \equiv (\mu_a)_{a\in\Sigma}$ are vectors in $\mathrm{Herm}(\mathbb{C}^d)$ and span linearly the space of Hermitian operators if and only if they are informationally complete (IC) [33]. We can therefore think of $\boldsymbol{\mu}$ as a frame of operators in the real space $\mathrm{Herm}(\mathbb{C}^d)$ equipped with the Hilbert-Schmidt inner product. Such frames of operators are referred to as *measurement frames* [19–21,23,35,36]. The task of finding unbiased estimators is thus equivalent to that of finding dual measurement frames for a given IC POVM $\boldsymbol{\mu}$. A natural choice of dual frame is the *canonical dual frame* $(\mu_a^\star)_{a\in\Sigma}$, defined via the *frame superoperator* $\mathcal{F} \in \mathrm{Lin}(\mathrm{Herm}(\mathbb{C}^d))$ as

$$\mu_a^\star \equiv \mathcal{F}^{-1}(\mu_a), \quad \mathcal{F}(X) \equiv \sum_a \langle \mu_a, X\rangle \mu_a. \tag{3}$$

This definition of a canonical dual frame is a direct application of the standard procedure used in frame theory for generic frames of vectors, where one can define a *frame operator* that, acting on frame elements, gives the corresponding canonical dual frame elements. Here, the vectors making up the frame are operators themselves. Therefore, in our context, such frame operators are linear operators acting on operators. We will refer to this type of linear transformation as a frame *super*operator in order to highlight such technical aspects. Equivalently, $\mathcal{F}$ and $\mathcal{F}^{-1}$ can be thought of as quantum maps, which linearly transform operators into other operators. The frame superoperator can also be concisely written as $\mathcal{F} = \sum_a \mathbb{P}(\mu_a)$, where $\mathbb{P}(Y) \in \mathrm{Pos}(\mathrm{Herm}(\mathbb{C}^d))$ denotes the outer product of $Y \in \mathrm{Herm}(\mathbb{C}^d)$ with itself, i.e., the superoperator acting as $\mathbb{P}(Y) : \rho \mapsto \langle Y, \rho\rangle Y$ on any $\rho \in \mathrm{Herm}(\mathbb{C}^d)$. In vectorized bra-ket notation, this is also often denoted by $\mathbb{P}(Y) \equiv |Y\rangle\rangle\langle\langle Y|$. Note that $\mathbb{P}(Y)$ is therefore again a quantum map and its action on an operator ρ would thus read explicitly $\mathbb{P}(Y)(\rho) = \langle Y, \rho\rangle Y \equiv \mathrm{tr}(Y^\dagger \rho) Y$. There are, in general, infinitely many dual frames associated with any given $\boldsymbol{\mu}$, each one corresponding to a different unbiased estimator. These estimators are not generally equivalent and can result in different reconstruction efficiencies. This will be discussed in detail in Sec. III. In particular, while $(\mu_a^\star)_{a\in\Sigma}$ is a standard choice of dual in the context of frame theory, we will show that it is not in fact the optimal choice to estimate properties of input states.

C. Estimators from measurement frames

In summary, for any IC POVM $\boldsymbol{\mu}$ and dual measurement frame $\tilde{\boldsymbol{\mu}}$, we have an unbiased estimator $\hat{f}(b) \equiv \tilde{\mu}_b$ for the unknown input state ρ and, vice versa, any such unbiased estimator can be obtained from a dual measurement frame of $\boldsymbol{\mu}$. If the goal is estimating the expectation value of an observable $\mathcal{O}$, we use the estimator $\hat{o}(b) \equiv \langle \mathcal{O}, \hat{f}(b)\rangle$. With this formalism, we can understand the main scaling results of shadow tomography as the observation that by carefully choosing the measurement $\boldsymbol{\mu}$ and the associated dual measurement frame $\tilde{\boldsymbol{\mu}}$, we obtain favorable scalings to estimate (finite sets of) target observables. The connection with the standard framing of shadow tomography is that the *classical shadows* are precisely a particular—in some sense optimal—choice of the state estimators $\hat{f}$. If a finite set of outcomes $\{b_1, \ldots, b_N\}$ is collected, we compute and store the values of the single-outcome estimators $\hat{f}(b_k)$ and then build from these an estimator for the expectation value—typically via the sample mean $1/N \sum_{k=1}^N \hat{f}(b_k)$ or the median of means. To estimate the expectation value of $\mathcal{O}$, the average is instead computed on the values $\langle \mathcal{O}, \hat{f}(b)\rangle$.

D. Variance of the estimators

A standard way to assess the magnitude of the statistical fluctuations in the estimator is to consider its variance. For state estimators, considering the errors in the L_2 distance, the variance reads

$$\mathrm{Var}[\hat{f}\,] = \mathbb{E}[\|\hat{f} - \rho\|_2^2] = \sum_b \langle \mu_b, \rho\rangle \|\hat{f}(b) - \rho\|_2^2. \quad (4)$$

Similarly, for observable estimators, the variance reads

$$\mathrm{Var}[\hat{o}] = \mathbb{E}[(\hat{o} - \langle \mathcal{O}, \rho\rangle)^2] = \sum_b \langle \mu_b, \rho\rangle (\hat{o}(b) - \langle \mathcal{O}, \rho\rangle)^2. \quad (5)$$

These variances depend on the input state ρ, the measurement $\boldsymbol{\mu}$, the estimator $\hat{f}$, and the target observable $\mathcal{O}$. For the sake of conciseness, the dependence on some or all of these will often not be made explicit, using the shorthand $\mathrm{Var}[\hat{o}] \equiv \mathrm{Var}[\hat{o}|\rho, \boldsymbol{\mu}, \hat{f}, \mathcal{O}]$. Knowledge of the variance grants performance guarantees for the additive estimation error, via standard statistical bounds such as Chebyshev's, Hoeffding's, or Bernstein's inequalities or by employing median-of-means estimators. A recent discussion of these statistical bounds and their applications to quantum state estimation is given in Ref. [16]. As will be shown in detail in the following sections, for the entire class of so-called "tight measurement frames," we can derive the unbiased estimator that minimizes the averaged variance and show that its averaged variance does not depend explicitly on the state dimension. Furthermore, for any measurement frame that forms a 3-design, we will prove that the worst-case-scenario variance can also be similarly upper bounded. This generalizes some of the results reported in Ref. [14] for random measurements.

E. Nonpositivity of state estimators

It is worth noting that the state estimators $\hat{f}(b)$ obtained with this scheme are Hermitian matrices but do not necessarily have unit trace and are not necessarily positive semidefinite. This means that if the goal is to estimate the state itself, the estimated state might not be a valid density matrix. This is precisely what happens in the context of linear state tomography and it is also the defining setting of shadow tomography. This feature of the scheme is particularly unproblematic in the shadow-tomography setting because the focus is on reconstructing expectation values of observables, rather than on the density matrix itself.

F. Mean versus median-of-means estimators

The median-of-means estimator, which has been used in, e.g., Ref. [14], has recently been found to not provide an advantage over the standard mean estimator in some situations [18,37]. More generally, Hoeffding-like bounds provide the same scaling-performance guarantees for any sub-Gaussian distribution and thus, in particular, for bounded ones [38]. All the estimators for finite-dimensional observables that we study are bounded by construction: for any IC POVM $\boldsymbol{\mu}$, estimator $\tilde{\boldsymbol{\mu}}$, and observable $\mathcal{O}$, we have

$$\begin{aligned}|\langle \mathcal{O}, \tilde{\mu}_b\rangle| &= |\langle \mathcal{O}, \mathcal{F}^{-1}(\mu_b)\rangle| \le \|\mathcal{O}\|_2 \|\mathcal{F}^{-1}(\mu_b)\|_2 \\ &= \|\mathcal{O}\|_2 \sqrt{\langle \mu_b, \mathcal{F}^{-2}(\mu_b)\rangle} \le \|\mathcal{O}\|_2 \|\mathcal{F}^{-2}\|_{\mathrm{op}}^{1/2},\end{aligned} \quad (6)$$

where $\mathcal{F}$ is the rescaled frame operator, $\|\cdot\|_{\mathrm{op}}$ is the operator norm, and $\|X\|_2 \equiv \sqrt{\mathrm{tr}(X^\dagger X)}$ is the L_2 operator norm of X. For the second identity, we have used the self-adjoint nature of the linear operator $\mathcal{F}^{-1}$ to move it across the inner product, thus obtaining

$$\|\mathcal{F}^{-1}(\mu_b)\|_2^2 = \langle \mathcal{F}^{-1}(\mu_b), \mathcal{F}^{-1}(\mu_b)\rangle = \langle \mu_b, \mathcal{F}^{-2}(\mu_b)\rangle. \quad (7)$$

Moreover, we have used the shorthand notation $\mathcal{F}^{-2} \equiv \mathcal{F}^{-1} \circ \mathcal{F}^{-1}$. The last step in the chain of relations in Eq. (6) then follows from

$$\langle \mu_b, \mathcal{F}^{-2}(\mu_b)\rangle \le \|\mu_b\|_2^2 \|\mathcal{F}^{-2}\|_{\mathrm{op}} \le \|\mathcal{F}^{-2}\|_{\mathrm{op}}. \quad (8)$$

As this holds for all b, $b \mapsto \hat{o}(b)$ is a bounded estimator. This implies that Hoeffding-like performance guarantees can always be used, i.e., that to have $\Pr(|\overline{o}_N - \mathbb{E}[\hat{o}]| \ge \epsilon) \le \delta$, where $\overline{o}_N$ is the sample mean taken over N independently drawn samples, it is sufficient to use $N \ge C/\epsilon^2 \log(2/\delta)$, where C is a constant that is independent of ϵ and δ. This matches the type of performance guarantees provided by the median-of-means estimator, explaining why in many practical scenarios the standard mean can perform better than the median-of-means estimator. Nonetheless, it is worth remarking that the constant C will depend on the interval of values taken by the estimator $\hat{o}$, which, as shown above, are only upper bounded by $\|\mathcal{F}^{-2}\|_{\mathrm{op}}^{1/2}$. This quantity can increase with the state dimension d. Consequently, while the median of means is never useful from the perspective of the scaling of N with respect to ϵ and δ, it might provide advantages in higher-dimensional spaces, as has been found to be the case in the analytical derivation for Clifford circuits in Ref. [37]. It is worth stressing that the results that we present in this paper are completely agnostic to the choice of between means and the median of means, as our analysis is performed at the level of the single-shot estimator. It is therefore entirely possible to apply the estimators that we propose using either the standard mean or the median of means, or possibly estimators that provide even more advantageous bounds [39].

III. CANONICAL ESTIMATORS

A. Minimum-variance unbiased estimators for tomography

It has been shown [19,23,30], in the context of state tomography, that the operators $\tilde{\mu}_b^{(\rho)}$, defined as

$$\tilde{\mu}_b^{(\rho)} \equiv \frac{\mathcal{F}_\rho^{-1}(\mu_b)}{\langle \mu_b, \rho \rangle}, \quad \mathcal{F}_\rho \equiv \sum_b \frac{\mathbb{P}(\mu_b)}{\langle \mu_b, \rho \rangle}, \tag{9}$$

give an unbiased estimator that minimizes the L_2 state-estimation error if the input state is ρ and the measurement is $\boldsymbol{\mu}$. Here, $\mathcal{F}_\rho$ is the frame superoperator associated with the rescaled measurement frame with elements $\mu_b/\sqrt{\langle \mu_b, \rho \rangle}$. Note that $\tilde{\boldsymbol{\mu}}^{(\rho)} \equiv (\tilde{\mu}_b^{(\rho)})_b$ is a dual measurement frame for $\boldsymbol{\mu}$ but not its canonical dual measurement frame. It is a suitably rescaled version of the canonical dual to the rescaled measurement frame with elements $\mu_b/\sqrt{\langle \mu_b, \rho \rangle}$. To use $\tilde{\boldsymbol{\mu}}^{(\rho)}$, one needs to already have a good guess about the underlying state ρ that is being measured and we thus interpret ρ as the prior information on the input state [40]. Thus, $\tilde{\boldsymbol{\mu}}^{(\rho)}$ is the minimum-variance unbiased estimator when the input state is ρ. A convenient tool to study the precision of an estimator is the *MSE matrix*. Following Ref. [23], this is defined with respect to a generic dual frame $\tilde{\boldsymbol{\mu}}$ and state ρ as

$$\mathcal{C}_\rho \equiv \sum_b \langle \mu_b, \rho \rangle \mathbb{P}(\tilde{\mu}_b) - \mathbb{P}(\rho). \tag{10}$$

While we do not write the functional relationship explicitly, $\mathcal{C}_\rho$ depends on the choice of $\boldsymbol{\mu}$, $\tilde{\boldsymbol{\mu}}$, and ρ. The expected L_2 state-estimation error associated with the estimator $\hat{f}(b) = \tilde{\mu}_b$ can be written concisely using the MSE matrix as

$$\mathcal{E}_\rho \equiv \mathbb{E}[\|\hat{f} - \rho\|_2^2] = \mathrm{tr}(\mathcal{C}_\rho). \tag{11}$$

When using the estimator $\tilde{\mu}_b = \tilde{\mu}_b^{(\rho)}$, the MSE matrix simplifies to

$$\mathcal{E}_\rho = \mathrm{tr}(\mathcal{F}_\rho^{-1}) - \mathrm{tr}(\rho^2), \tag{12}$$

which is the expected mean-squared error when using the estimator with minimum variance when the input state is ρ [41]. In the expression $\mathrm{tr}(\mathcal{F}^{-1})$, the argument $\mathcal{F}^{-1}$ is a superoperator but its trace is defined as in linear algebra for a standard trace. However, it is often the case that the trace of a superoperator is referred to as a "superoperator trace." Explicitly, the (superoperator) trace of a generic superoperator Φ can be defined as $\mathrm{tr}(\Phi) = \sum_k \langle \sigma_k, \Phi(\sigma_k) \rangle$ for any orthonormal basis of operators $\{\sigma_k\}_k$. In our case, $\mathcal{F}_\rho^{-1}$ is considered as an operator acting in the subspace of Hermitian operators and its trace is thus

$$\mathrm{tr}(\mathcal{F}_\rho^{-1}) = \sum_{k=1}^{d^2} \langle \sigma_k, \mathcal{F}_\rho^{-1}(\sigma_k) \rangle, \tag{13}$$

where $\{\sigma_k\}_{k=1}^{d^2}$ a generic orthonormal basis of Hermitian operators and d is the dimension of the underlying space. It is also often convenient to pick an orthonormal basis of the form $\{I/\sqrt{d}\} \cup \{\tilde{\sigma}_k\}_{k=1}^{d^2-1}$, where $I/\sqrt{d}$ is the (normalized) identity and $\{\tilde{\sigma}_k\}_{k=1}^{d^2-1}$ forms an orthonormal basis for the subspace of *traceless* Hermitian operators. This can always be done and is very useful in our calculations for a twofold reason. On one hand, it provides the following decomposition for the (superoperator) trace:

$$\mathrm{tr}(\mathcal{F}_\rho^{-1}) = \frac{\mathrm{tr}(\mathcal{F}_\rho^{-1}(I))}{d} + \sum_{k=1}^{d^2-1} \langle \tilde{\sigma}_k, \mathcal{F}_\rho^{-1}(\tilde{\sigma}_k) \rangle. \tag{14}$$

On the other, as $\mathcal{F}_\rho^{-1}(I) = \rho$—which follows directly from the readily verifiable relation $\mathcal{F}_\rho(\rho) = I$—we reduce the calculation of the trace to the calculation of the trace on the subspace of traceless Hermitian operators.

B. Canonical estimator

A standard scenario is the lack of any prior information about the input state. In such cases, because the error will generally depend on the input state, it is common to consider as "optimal" the estimator that minimizes the *average* L_2 estimation error, which corresponds to the optimal estimator with respect to the reference state $\rho = I/d$. Following Ref. [23], we will refer to this as the *canonical estimator*, denoted by $\tilde{\boldsymbol{\mu}}^{\mathrm{can}} \equiv \tilde{\boldsymbol{\mu}}^{(I/d)}$, which is thus written explicitly as

$$\tilde{\mu}_b^{\mathrm{can}} \equiv \frac{d\mathcal{F}_{I/d}^{-1}(\mu_b)}{\mathrm{tr}(\mu_b)}, \quad \mathcal{F}_{I/d} \equiv d \sum_b \frac{\mathbb{P}(\mu_b)}{\mathrm{tr}(\mu_b)}. \tag{15}$$

It is worth noting that this is not the same as the canonical dual with respect to the measurement frame $\boldsymbol{\mu}$ [42]. The canonical estimator thus minimizes the L_2 error averaged over unitarily equivalent input states [19,30,43]. This average L_2 error turns out to depend only on the purity $P \equiv \mathrm{tr}(\rho^2)$ of the input state and will be denoted by $\overline{\mathcal{E}_P}$. As discussed in Ref. [23], this quantity is lower bounded by

$$\overline{\mathcal{E}_P} \geq d^2 + d - 1 - P, \tag{16}$$

with the lower bound saturated if and only if the measurement is composed of projectors onto subnormalized pure states that form a weighted 2-design. Such measurements are referred to as *tight rank-1 IC POVMs* and have elements $\mu_b = w_b\mathbb{P}(\psi_b)$, with the weights satisfying $\sum_b w_b = d$ and

$$\frac{1}{d}\sum_b w_b\mathbb{P}(\psi_b)^{\otimes 2} = \binom{d+1}{2}^{-1}\Pi_{\text{sym}}, \tag{17}$$

where Π_{sym} is the projection onto the symmetric subspace, which can be written explicitly as $\Pi_{\text{sym}} = (I + W)/2$, in which W is the SWAP operator. For all tight rank-1 IC POVMs, the canonical estimator has the form

$$\tilde{\mu}_b^{\text{can}} = (d+1)\mathbb{P}(\psi_b) - I \tag{18}$$

and the MSE matrix equals

$$\mathcal{C}_{I/d} = \frac{d+1}{d}\Pi_{H_0}, \tag{19}$$

where $\Pi_{H_0} \equiv \text{Id} - \mathbb{P}(I/\sqrt{d})$ is the superoperator that projects onto the subspace of traceless linear operators. A more in-depth discussion of these results and, more generally, of the connection between weighted 2-designs and tight IC POVMs, is given in Appendix D.

C. Estimation of observables

The usefulness of shadow tomography lies in the potentially favorable scalings of the associated estimation errors with respect to the state dimension d. More specifically, we are interested in the variance of $\hat{o}$ for different choices of ρ, $\boldsymbol{\mu}$, $\tilde{\boldsymbol{\mu}}$, and $\mathcal{O}$. For notational convenience, we indicate explicitly only the dependence of the variance on ρ:

$$\text{Var}[\hat{o}|\rho] = \mathbb{E}[|\hat{o} - \langle\mathcal{O},\rho\rangle|^2] = \sum_b \langle\mu_b,\rho\rangle\langle\mathcal{O},\tilde{\mu}_b\rangle^2 - \langle\mathcal{O},\rho\rangle^2, \tag{20}$$

for different choices of ρ, $\boldsymbol{\mu}$, $\tilde{\boldsymbol{\mu}}$, and $\mathcal{O}$. This can be conveniently written using the MSE matrix $\mathcal{C}_\rho$ as

$$\text{Var}[\hat{o}|\rho] = \langle\mathbb{P}(\mathcal{O}),\mathcal{C}_\rho\rangle \equiv \langle\mathcal{O},\mathcal{C}_\rho(\mathcal{O})\rangle. \tag{21}$$

As discussed in detail in Appendix C, we can derive a general expression for the minimum-variance unbiased estimator for a given target observable and input state and this is found to match the corresponding estimator for state tomography on the support of the observable. More precisely, if $\tilde{\boldsymbol{\mu}}^{(\rho)}$ is a minimum-variance unbiased estimator for state tomography with respect to the state ρ, then any $\tilde{\boldsymbol{\mu}}$ such that $\langle\mathcal{O},\tilde{\mu}_b\rangle = \langle\mathcal{O},\tilde{\mu}_b^{(\rho)}\rangle$ is a minimum-variance unbiased estimator for $\mathcal{O}$. Although derived using different methods and notation, this result is similar to some of the results reported in Refs. [20,21,44]. If we want an estimator that gives small errors for arbitrary target observables, the natural candidate to use is the one that minimizes the variance averaged over the observables. In this case, the minimum-variance unbiased estimator is again the one that we have found for state tomography. Given that in shadow tomography we do not generally want to fix the observables to be estimated beforehand, we can safely fix as optimal estimators the $\tilde{\boldsymbol{\mu}}^{(\rho)}$ derived for state tomography. We will furthermore focus on the scenario in which only the purity of the input state is known beforehand and thus, in the following, we will always use the canonical estimator $\tilde{\boldsymbol{\mu}}^{\text{can}}$ given in Eq. (15). This has the added advantage of being independent of both ρ and $\mathcal{O}$, though the estimation variance will still, in general, depend on these quantities. In summary, if there is prior information suggesting that the input state is or is close to ρ, the minimum-variance estimator is given by $\tilde{\boldsymbol{\mu}}^{(\rho)}$, as discussed in this section and proved explicitly in Appendixes B and C. If no prior knowledge is assumed about the input state, the canonical estimator $\tilde{\boldsymbol{\mu}}^{\text{can}}$ can be used and provides the minimal averaged estimation variance.

D. Numerical examples

We illustrate explicitly how different choices of dual frames provide nonequivalent estimators in Figs. 1 and 2. In particular, the canonical estimator $\tilde{\boldsymbol{\mu}}^{\text{can}}$ has, on average, the lowest variance, albeit the estimator $\tilde{\boldsymbol{\mu}}^{(\rho)}$ can give even lower variances if ρ matches the true input state. The nonrescaled estimator $\boldsymbol{\mu}^\star$ tends to perform worse than $\tilde{\boldsymbol{\mu}}^{\text{can}}$, consistently with the latter having a smaller averaged variance. On the other hand, using the estimator $\tilde{\boldsymbol{\mu}}^{(\sigma)}$—which has minimum variance when the input is σ—to estimate properties of $\rho \neq \sigma$ will still, on average, reproduce the correct expectation values but will result in a generally larger estimation error.

E. Shadow tomography versus state tomography

It is worth stressing the tight relation between shadow and state tomography emerging from the above discussion. The general formalism of measurement frames clarifies how these can be viewed as one and the same experimental protocol, with the only difference being how estimation errors are evaluated. Both linear state tomography and our formalism for shadow tomography can be performed for arbitrary IC POVMs—albeit, as discussed previously, not always with favorable error scalings—and the postprocessing procedure is the same in both cases. The core difference is in the problem setting: whether the target is recovering an approximation of the full density matrix or just recovering the expectation values of finitely many observables.

IV. BOUNDS ON AVERAGED VARIANCE

In this section, we will derive useful bounds for the averaged estimation variance of an observable in terms

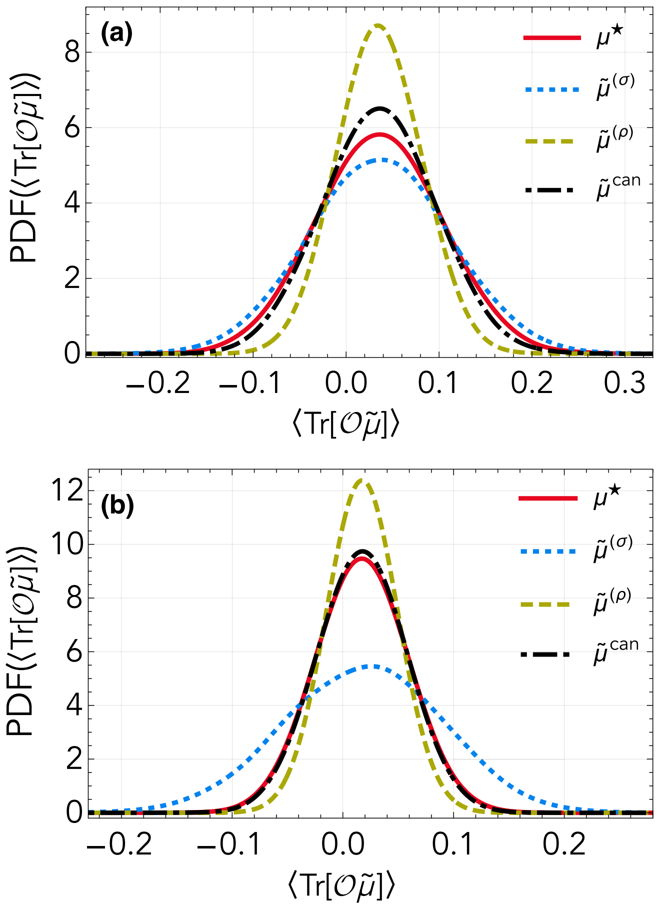

FIG. 1. The probability distributions of the sample means. Histograms of the probability distribution of the sample mean $\overline{o}_N$ with $N = 10^3$, obtained taking the average of $\hat{o}(b) \equiv \langle \hat{f}(b), \mathcal{O}\rangle$ over N randomly sampled outcomes b, for different choices of estimator $\hat{f}$. The histograms are computed using 10^4 realizations of the sample mean. The input state is $\rho \equiv \mathbb{P}_0$ in all cases and the measurements are random rank-1 POVMs built as $\mu_b = V\mathbb{P}_b V^\dagger$ with V random isometries. In each case, we show the distribution of the sample mean for the nonrescaled estimator $\mu^\star$ [cf. Eq. (3)]; the estimators $\tilde{\mu}^{(\rho)}$ and $\tilde{\mu}^{(\sigma)}$ [cf. Eq. (9)] with $\sigma \equiv \mathbb{P}_1$; and the canonical estimator $\tilde{\mu}^{\mathrm{can}}$ [cf. Eq. (15)]. We show the data for (a) two-dimensional states with ten-outcome measurements and (b) five-dimensional states and 100-outcome measurements.

of the eigenvalues of the frame superoperator associated with the measurement. These eigenvalues will then be bounded in terms of a quantity that measures how far a given IC POVM is from being tight. Finally, we will show that, for tight measurements and any suitable normalized observable, the resources needed to estimate the expectation value of the observable via the shadow-tomography apparatus do not scale with the dimension of the state.

A. Bounds via eigenvalues of frame superoperator

The variance averaged over unitarily equivalent input states is

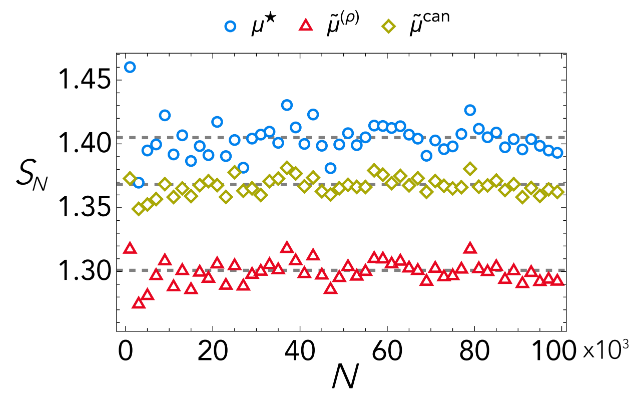

FIG. 2. The sample variance for different estimators. Examples of the behavior of the sample variance $\hat{S}_N$ of the estimator $\overline{o}_N \equiv 1/N \sum_{k=1}^N \hat{o}(b_k)$ as a function of N, computed with respect to the estimators $\boldsymbol{\mu}^\star$, $\tilde{\boldsymbol{\mu}}^{(\rho)}$, and $\tilde{\boldsymbol{\mu}}^{\mathrm{can}}$. The sample variance is defined as $\hat{S}_N \equiv 1/(N-1)\sum_{k=1}^N (\hat{o}(b_k) - \overline{o}_N)^2$. The dashed lines give the values of the variance $\mathrm{Var}[\hat{o}|\rho]$ in each case, as computed via Eq. (20). The data are obtained using $d = 2$-dimensional systems, with fixed input state $\rho = \mathbb{P}_0$, random rank-1 POVMs with ten outcomes, and random target observables with $\mathrm{tr}(\mathcal{O}) = 0$ and $\mathrm{tr}(\mathcal{O}^2) = 1$.

$$\overline{\mathrm{Var}[\hat{o}|P, \mathcal{O}, \boldsymbol{\mu}]} \equiv \int_{\mathbf{U}(d)} dU \, \mathrm{Var}[\hat{o}|U\rho U^\dagger]$$

$$= \underbrace{\sum_b \langle \mu_b, I/d\rangle \langle \mathcal{O}, \tilde{\mu}_b^{\mathrm{can}}\rangle^2}_{=\langle \mathcal{O}, \mathcal{F}_{I/d}^{-1}(\mathcal{O})\rangle} - \underbrace{\int_{\mathbf{U}(d)} dU \langle \mathcal{O}, U\rho U^\dagger\rangle^2}_{\equiv \beta}. \quad (22)$$

As mentioned previously, the explicit dependence on $\mathcal{O}$ and $\boldsymbol{\mu}$ will be left implicit in the following for notational conciseness and we will write this averaged variance as simply $\overline{\mathrm{Var}[\hat{o}|P]}$. The coefficient β, the explicit expression for which is reported in Appendix E, is computed explicitly using known formulas to integrate polynomials in the components of unitary matrices over the uniform Haar measure [45,46] and does not depend on $\boldsymbol{\mu}$. Furthermore, as shown in Appendix A, the canonical superoperator decomposes as

$$\mathcal{F}_{I/d} = d\mathbb{P}(I/\sqrt{d}) + \tilde{\mathcal{F}}_{I/d}, \quad (23)$$

where $\tilde{\mathcal{F}}_{I/d} \equiv \Pi_{H_0}\mathcal{F}_{I/d}\Pi_{H_0}$ is the projection of $\mathcal{F}_{I/d}$ onto the subspace of traceless operators. Using such a decomposition, we rewrite

$$\langle \mathcal{O}, \mathcal{F}_{I/d}^{-1}(\mathcal{O})\rangle = \frac{\mathrm{tr}(\mathcal{O})^2}{d^2} + \langle \mathcal{O}, \tilde{\mathcal{F}}_{I/d}^{-1}(\mathcal{O})\rangle. \quad (24)$$

The second term can then be bounded in terms of the eigenvalues of $\tilde{\mathcal{F}}_{I/d}$, as

$$\frac{Vd}{\lambda_+(\tilde{\mathcal{F}}_{I/d})} \le \langle \mathcal{O}, \tilde{\mathcal{F}}_{I/d}^{-1}(\mathcal{O})\rangle \le \frac{Vd}{\lambda_-(\tilde{\mathcal{F}}_{I/d})}, \tag{25}$$

where $\lambda_-(\tilde{\mathcal{F}}_{I/d})$ and $\lambda_+(\tilde{\mathcal{F}}_{I/d})$ denote the smallest and largest eigenvalues of $\tilde{\mathcal{F}}_{I/d}$, respectively, and $V \equiv \mathrm{tr}(\mathcal{O}^2)/d - \mathrm{tr}(\mathcal{O})^2/d^2$ is the variance of $\mathcal{O}$ with respect to the totally mixed state I/d. As further explained in Appendix E, this expression is obtained observing that $\tilde{\mathcal{F}}_{I/d}^{-1}$ is a Hermitian linear (super)operator that only acts nontrivially on the subspace of traceless Hermitian operators. Since $\tilde{\mathcal{F}}_{I/d}$ is positive definite as an operator whenever $\boldsymbol{\mu}$ is informationally complete, we ensure that $\lambda_\pm(\tilde{\mathcal{F}}_{I/d}) > 0$. For any $\boldsymbol{\mu}$, as again shown in Appendix E, the eigenvalues can be bounded as a function of $a \equiv \mathrm{tr}(\tilde{\mathcal{F}}_{I/d})$ and $b \equiv \mathrm{tr}(\tilde{\mathcal{F}}_{I/d}^2)$. Focusing on the variance for the hardest-to-estimate observable, we find that the smallest such variance compatible with a and b reads

$$\max_{\mathcal{O}} \frac{\overline{\mathrm{Var}[\hat{o}|P]}}{Vd} \ge \frac{1}{\lambda_1^*} - \frac{P - 1/d}{d^2-1}, \qquad \lambda_1^* \equiv \frac{a}{d^2-1} - \frac{\sqrt{(d^2-2)((d^2-1)b-a^2)}}{(d^2-1)(d^2-2)}. \tag{26}$$

This relation tells us that if $\boldsymbol{\mu}$ gives a frame superoperator such that $a = \mathrm{tr}(\tilde{\mathcal{F}}_{I/d})$ and $b = \mathrm{tr}(\tilde{\mathcal{F}}_{I/d}^2)$, then the worst-case average variance is lower bounded as in Eq. (26). In other words, a and b define a bound on the best possible performance of the canonical estimator (in the scenario in which we average over input states and take the worst-case scenario with respect to the observables).

B. Performance for tight measurements

In the case of tight measurements, $\tilde{\mathcal{F}}_{I/d}$ is a multiple of the identity, $(d^2-1)b = a^2$, and Eq. (26) simplifies to

$$\overline{\mathrm{Var}[\hat{o}|P]} = Vd\left(\frac{d^2+d-1-P}{d^2-1}\right), \tag{27}$$

where the maximum over the observables no longer applies, because all observables give the same expression for the averaged average. We recognize, in particular, the term $d^2+d-1-P$, which is the optimal state-estimation L_2 error discussed in Appendix D. Equation (27) shows that for tight rank-1 measurements, the variance increases with the state dimension only due to the variance V of the observable calculated with respect to the totally mixed state. Note that for rank-1 observables of the form $\mathcal{O} = \mathbb{P}_\psi$ for any $|\psi\rangle$, we have $Vd = 1 - 1/d$, while for observables normalized as $\mathrm{tr}(\mathcal{O}) = 0$ and $\mathrm{tr}(\mathcal{O}^2) = 1$, we have $Vd = 1$. It immediately follows that for all such cases, $Vd \to 1$ for large d and thus the variance does not increase with d, converging asymptotically to $V \to 1$. On top of estimating best- and worst-case scenarios for the variance, we also show in Appendix F how to compute the variance averaging with respect to unitarily equivalent observables.

We have thus shown that for the entire class of tight rank-1 measurement frames, which includes but is not limited to covariant measurements, the sampling statistics required to estimate arbitrary rank-1 observables with a bounded norm do not increase with the state dimension, in direct contrast to the corresponding results about state tomography. More generally, we can explicitly characterize the class of observables that correspond to such favorable scalings. This directly implies that all these measurements can be used to implement shadow-tomography schemes. While not all such measurements will allow an efficient circuit decomposition like the one presented in Ref. [14], this will depend on the experimental context that is being considered. Having a good characterization of the general class of viable measurements can greatly help to find measurement schemes to efficiently implement shadow tomography in different experimental scenarios.

V. BEST- AND WORST-CASE SCENARIO VARIANCES

In Sec. IV, we have derived bounds for the variance averaged over input states. In this section, we focus instead on the derivation of bounds for minimum and maximum variance with respect to the input states. This is particularly relevant for comparing with the results of Ref. [14] because, as will be discussed in detail in Sec. VI, the often-used "shadow norm" is precisely the variance maximized over the input states.

A. Concise expression for variance via A operator

We first observe that the general expression for the variance in Eq. (20), for a generic input state ρ, can be rewritten as

$$\mathrm{Var}[\hat{o}|\rho] + \langle \mathcal{O}, \rho\rangle^2 = \sum_b \langle \mu_b, \rho\rangle \langle \mathcal{O}, \tilde{\mu}_b\rangle^2 = \langle A, \rho\rangle, \tag{28}$$

where we have defined the operator

$$A \equiv \sum_b \langle \mathcal{O}, \tilde{\mu}_b\rangle^2 \mu_b. \tag{29}$$

Notably, the only part of Eq. (28) that is nonlinear with respect to ρ is $\langle \mathcal{O}, \rho\rangle^2$, which does not depend on the measurement choice and is bounded as $\langle \mathcal{O}, \rho\rangle^2 \le \mathrm{tr}(\mathcal{O}^2)$. Furthermore, the linearity of $\langle A, \rho\rangle$ with respect to ρ

means that for any choice of measurement, estimator, and observable, we can write the general bounds

$$\lambda_{\min}(A) \leq \langle A, \rho \rangle \leq \lambda_{\max}(A) \equiv \|A\|_{\mathrm{op}}, \tag{30}$$

where $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ are the smallest and largest eigenvalues of A, respectively. In particular, we have the following upper bound for the worst-case (with respect to input states) variance:

$$\max_{\rho} \mathrm{Var}[\hat{o}|\rho] \leq \|A\|_{\mathrm{op}}. \tag{31}$$

As will be further discussed in more detail in Sec. VI, the right-hand side of this expression corresponds to the so-called "shadow norm" $\|\mathcal{O}\|_{\mathrm{sh}}^2 = \|A\|_{\mathrm{op}}$ introduced in Ref. [14].

B. Explicit expression for 3-designs

In the case of rank-1 measurements that also give a weighted 3-design, we can find a remarkably simple expression for the state- and observable-dependent variance even in the nonaveraged scenario. To see this, we start by observing that

$$\sum_b \langle \mu_b, \rho \rangle \langle \mathcal{O}, \tilde{\mu}_b \rangle^2 = \left\langle \rho \otimes \mathcal{O} \otimes \mathcal{O}, \sum_b \mu_b \otimes \tilde{\mu}_b \otimes \tilde{\mu}_b \right\rangle. \tag{32}$$

For any tight rank-1 POVM with elements $\mu_b = w_b \mathbb{P}(\psi_b)$, using the canonical estimators $\tilde{\mu}_b^{\mathrm{can}}$ given in Eq. (18) we can also write

$$\sum_b \mu_b \otimes \tilde{\mu}_b^{\mathrm{can}} \otimes \tilde{\mu}_b^{\mathrm{can}} = (d+1)^2 S_3 - (d+1) S_2 + I, \tag{33}$$

where $S_3 \equiv \sum_b w_b \mathbb{P}(\psi_b)^{\otimes 3}$ and

$$S_2 \equiv \sum_b w_b \mathbb{P}(\psi_b)^{\otimes 2} \otimes I + \sum_b w_b \mathbb{P}(\psi_b) \otimes I \otimes \mathbb{P}(\psi_b). \tag{34}$$

If the states $|\psi_b\rangle$ form a complex projective 3-design, then $S_3 = d\Pi_{\mathrm{sym},3}/\binom{d+2}{3}$ where $\Pi_{\mathrm{sym},3} \in \mathrm{Lin}((\mathbb{C}^d)^{\otimes 3})$ is the projection onto the completely symmetric subspace of $(\mathbb{C}^d)^{\otimes 3}$, and $S_2\binom{d+1}{2} = d\Pi_{\mathrm{sym},2}^{(1,2)} + d\Pi_{\mathrm{sym},2}^{(1,3)}$ is a sum of the projections on the symmetric subspace of $(\mathbb{C}^d)^{\otimes 2}$ on first and second and first and third qubits, respectively. These projections can be written more explicitly as $\Pi_{\mathrm{sym},2} = (I \otimes I + W)/2$, where W is the SWAP operator, $\Pi_{\mathrm{sym},3} = \frac{1}{3!}\sum_{\pi \in \mathcal{S}_3} W_\pi$, with $\mathcal{S}_3$ denoting the symmetric group over three elements, and W_π is the unitary operator, defined as [33]

$$W_\pi = \sum_{i_1,i_2,i_3} \left| i_{\pi(1)}, i_{\pi(2)}, i_{\pi(3)} \right\rangle \langle i_1, i_2, i_3 | . \tag{35}$$

With these and Eq. (32), we can work out the explicit expressions for state- and observable-dependent variances and obtain

$$\mathrm{Var}[\hat{o}|\rho] = -\frac{\mathrm{tr}(\mathcal{O})^2 + 2\mathrm{tr}(\mathcal{O})\mathrm{tr}(\rho\mathcal{O})}{d+2} + \frac{d+1}{d+2}\left[\mathrm{tr}(\mathcal{O}^2) + 2\mathrm{tr}(\mathcal{O}^2\rho)\right] - \mathrm{tr}(\mathcal{O}\rho)^2. \tag{36}$$

This expression shows explicitly that for any rank-1 measurement that forms a 3-design, we obtain an explicit expression for the variance even in the nonaveraged regime. This dramatically simplifies the study of the relations between best, worst, and average cases with respect to both the input state and the target observable. Random Clifford circuits and Haar-random unitaries, considered in Ref. [14], as well as single-qubit mutually unbiased bases, are examples of rank-1 measurements that form a 3-design [48,49].

C. Worst-case variance bounds for 3-designs

The explicit expression for the variance for 3-designs allows us to also derive general bounds for the variance maximized over the input states: given any rescaled observable, $\mathrm{tr}(\mathcal{O}) = 0$, we obtain, from Eq. (36),

$$\max_{\rho} \mathrm{Var}[\hat{o}|\rho] \leq \mathrm{tr}(\mathcal{O}^2) + 2\|\mathcal{O}^2\|_{\mathrm{op}} \leq 3\mathrm{tr}(\mathcal{O}^2), \tag{37}$$

which shows that by increasing the dimension d, even in the worst-case scenario, the variance only increases with d via the observable. Thus, for any rescaled observable for which $\mathrm{tr}(\mathcal{O}) = 0$, $\mathrm{tr}(\mathcal{O}^2) = 1$, we obtain a dimension-independent upper bound. Note that the $3\mathrm{tr}(\mathcal{O}^2)$ upper bound is identical to the one derived in Ref. [14] for random Clifford and unitary measurements.

D. Numerical examples with MUBs

In Fig. 3, we report numerical results obtained for the average, minimum, and maximum variance, in the case of mutually unbiased basis (MUB) measurements in prime dimensions [47], calculated via Eq. (30). We note in particular how even the worst-case variance does not increase with the state dimension. This is compatible with the general expression for the variance that we will obtain for 3-designs, although MUBs do not correspond to a 3-design, indicating that these favorable scaling results might hold even more generally.

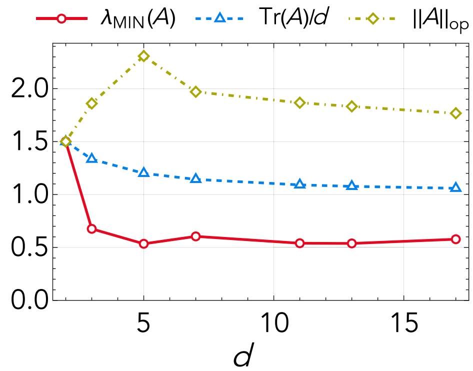

FIG. 3. The average, minimum, and maximum variance for mutually unbiased basis (MUB) POVMs. We plot the values of $\lambda_{\min}(A)$, $\|A\|_{\rm op}$, and $\operatorname{tr}(A)/d$, as a function of the state dimension d, for the case of canonical estimators, with a random target observable for each d. The data are shown for prime d because these are the values corresponding to which explicit constructions for MUBs are known [47]. These results give the range of possible values of $\langle A,\rho\rangle$ varying over the input states ρ, for the case of MUB measurements. These values are then tightly connected with the estimation variance via Eq. (28). The data shown correspond to a random target observable with $\operatorname{tr}(\mathcal{O})=0$ and $\operatorname{tr}(\mathcal{O}^2)=1$.

VI. RELATION WITH CONSTRUCTION OF REF. [13]

We now specialize our discussion in Sec. II to the formalism presented in Ref. [14]. The goal is to show that the latter can be viewed and studied from the general perspective of measurement frames and corresponds to the special case where the employed IC POVM is a *covariant measurement* [23,31,50,51].

A. Description of the formalism

The procedure to build the classical shadows introduced in Ref. [14] involves the following steps:

(1) Perform a random unitary rotation $\rho\mapsto U\rho U^\dagger$ on the state and then measure the evolved state in the computational basis $|b\rangle$.
(2) Define the operator

$$\begin{aligned}\mathcal{M}(\rho) &\equiv \mathbb{E}\left[U^\dagger|\hat b\rangle\langle\hat b|U\right]\\ &\equiv \mathbb{E}_{U\sim\mathcal{U}}\sum_b \left\langle U\rho U^\dagger\middle|b\right\rangle\, U^\dagger\,|b\rangle\langle b|\,U,\end{aligned}\tag{38}$$

where $|\hat b\rangle$ is a random variable associating, with each outcome b, the corresponding state $|b\rangle$. The expectation value is taken with respect to some distribution $\mathcal{U}$ in the group of unitary matrices and with respect to the possible outcomes b for each choice of unitary.
(3) Compute and store the operators $\hat\rho\equiv\mathcal{M}^{-1}(U^\dagger|\hat b\rangle\langle\hat b|U)$. These are referred to as the "classical shadows" of the state.

To estimate the expectation values of an observable $\mathcal{O}$, one then uses the estimator $\hat o\equiv\langle\mathcal{O},\hat\rho\rangle$ built from the classical shadows. We will focus here on the task of estimating expectation values, although in Ref. [14] the estimation of other kinds of quantities is also discussed. Another important aspect discussed in Ref. [14] is the efficiency of computing and storing the classical shadows for large many-qubit Hilbert spaces, which can be solved by leveraging Clifford circuits and the formalism of stabilizer states. We will not focus on these aspects here but, rather, on the general structure of the shadow-tomography protocol.

B. Equivalence: Step 1

The equivalence between the formalism thus outlined and our approach is seen by observing that a measurement in the computational basis $\{|b\rangle\}$ after evolving the state through a random unitary rotation U amounts to a direct measurement, with the POVM having elements

$$\mu_{U,b}\equiv U^\dagger\,|b\rangle\langle b|\,U.\tag{39}$$

As such a measurement has (uncountably) infinitely many outcomes, its normalization reads

$$\int_{\mathbf{U}(d)} dU\sum_b \mu_{U,b}=I,\tag{40}$$

where the integral is performed with respect to the Haar measure over the unitary group of suitable dimension and thus $\int_{\mathbf{U}(d)}dU=1$.

C. Equivalence: Step 2

The introduced map $\mathcal{M}$ is precisely the frame operator corresponding to the measurement frame $\{\mu_{U,b}\}_{U,b}$. This becomes more evident upon rewriting Eq. (38) in the form:

$$\mathcal{M}(\rho)=\int_{\mathbf{U}(d)}dU\sum_b\langle\mu_{U,b},\rho\rangle\mu_{U,b},\tag{41}$$

which matches the structure of the frame superoperator defined in Eq. (3).

D. Equivalence: Step 3

From the above considerations, it is now clear that the classical shadows, which read in terms of the POVM $(\mu_{U,b})$ as $\hat\rho=\mathcal{M}^{-1}(\mu_{U,b})$, are the elements of the canonical dual frame of the measurement frame. This shows that the formalism to compute classical shadows with random unitary rotations and projective measurement follows as a special case of the general procedure for measurement frames outlined in Sec. II.

E. Equivalence of the formalisms

At first glance, this procedure might still appear different from the one discussed in Sec. II, as we did not explicitly use rescaled measurement frames in this section (as opposed to the discussion in Sec. II). This is due to the covariant measurements being such that $\mathrm{tr}(\mu_{U,b}) = 1$ for all U, b, making the rescaling factors used in the definition $\tilde{\mathcal{F}}_{I/d}$ unnecessary in these cases. It follows that $\mathcal{M}$ and $\tilde{\mathcal{F}}_{I/d}$ only differ by the proportionality constant d. These observations show that the formalism of shadow tomography via random unitary rotations can be seen as a direct application of the general formalism that we present to rank-1 POVMs of the form $\mu_\psi = p_\psi \mathbb{P}_\psi$ for some distribution over the states $|\psi\rangle$. A direct numerical comparison between the results of applying our formalism to estimate observables from MUB measurements and the approach with uniformly random unitaries is presented in Fig. 4. As clearly shown in the figures, while the distribution of the estimators differs considerably in the two cases, the induced sample means have similar distributions and both converge to the same Gaussian in the limit of infinite statistics.

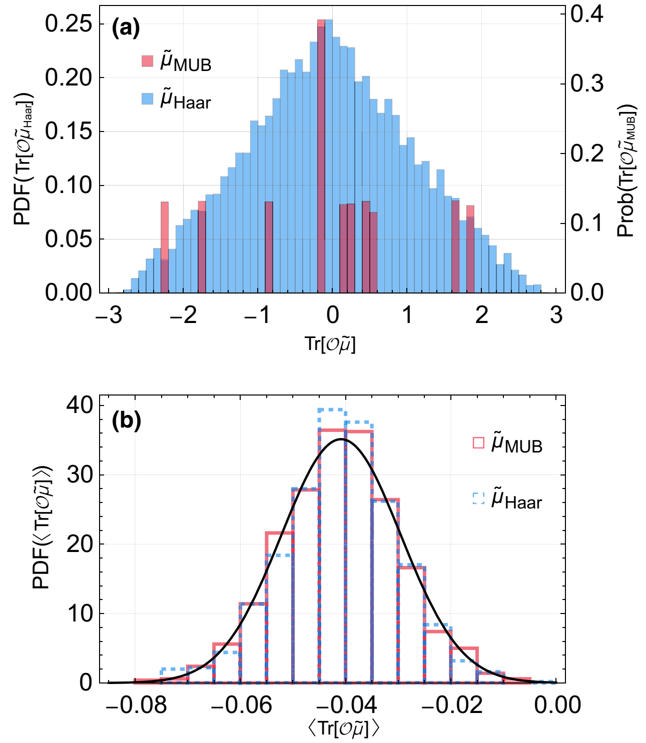

FIG. 4. The distributions of the estimators and their sample mean, corresponding to MUBs and Haar-random unitary POVMs. (a) Histograms of the probability distributions for the estimator $\langle \mathcal{O}, \tilde{\mu}\rangle$ for a random (fixed) observable $\mathcal{O}$ with $\mathrm{tr}(\mathcal{O}) = 0, \mathrm{tr}(\mathcal{O}^2) = 1$, and fixed qutrit state $\rho = \mathbb{P}_0$. The reported results correspond to MUBs, $\boldsymbol{\mu}_{\mathrm{MUB}}$ (red), and random measurements $\boldsymbol{\mu}_{\mathrm{Haar}}$, which have elements $\mu_{U,b} = U\mathbb{P}_b U^\dagger$ with Haar-random unitaries U (blue). For $\boldsymbol{\mu}_{\mathrm{MUB}}$, there is a finite number of outcomes and we plot the probability associated with each outcome directly. For $\boldsymbol{\mu}_{\mathrm{Haar}}$, owing to the infinitely many outcomes, we uniformly draw a number of random unitaries U and plot a histogram of the observed estimator values $\langle \tilde{\boldsymbol{\mu}}_{\mathrm{Haar}}, \mathcal{O}\rangle$. We show two different scales on the vertical axis: in the presence of a continuum of possible outcomes, as we have for $\langle \mathcal{O}, \tilde{\boldsymbol{\mu}}_{\mathrm{Haar}}\rangle$, we plot the probability density function (PDF); while for finitely many outcomes, we show the probability mass function. (b) The histogram of possible outcomes of the sample mean $\overline{o}_N \equiv \frac{1}{N}\sum_{k=1}^N \hat{o}(b_k)$ of $\hat{o}(b) \equiv \langle \mathcal{O}, \tilde{\mu}_b\rangle$, estimated with statistics of $N = 10^3$ samples. The histogram is drawn sampling 10^4 realizations of this sample mean, in the same condition as the other histogram. The black solid line is a Gaussian with the same mean and variance as both estimators, $\tilde{\boldsymbol{\mu}}_{\mathrm{MUB}}$ and $\tilde{\boldsymbol{\mu}}_{\mathrm{Haar}}$—which have the same variance, both being tight measurement frames. Both histograms approach this Gaussian for $N \to \infty$, due to the central-limit theorem.

F. Variance and shadow norm

In Ref. [14], the variance of the estimators for observables is bounded in terms of their so-called "shadow norm," which is defined there as

$$\|\mathcal{O}\|_{\mathrm{sh}} \equiv \max_\sigma \left(\mathbb{E}_{U\sim\mathcal{U}} \sum_b \langle b|U\sigma U^\dagger|b\rangle\langle b|U \times \mathcal{M}^{-1}(\mathcal{O})U^\dagger|b\rangle^2 \right)^{1/2}, \tag{42}$$

where the maximization is performed with respect to all possible states σ. This expression is equivalent to

$$\|\mathcal{O}\|_{\mathrm{sh}}^2 = \max_\sigma \sum_b \langle \mu_b, \sigma\rangle \langle \mathcal{O}, \tilde{\mu}_b^{\mathrm{can}}\rangle^2, \tag{43}$$

in the special case of $\boldsymbol{\mu}$ being the covariant measurement, i.e., mapping $b \to (U, b)$ and $\mu_b \to \mu_{U,b} \equiv U^\dagger \mathbb{P}_b U$, and with $\tilde{\mu}_b^{\mathrm{can}}$ being the canonical estimator associated with this measurement, as given in Eq. (15), which in this case reads $\tilde{\mu}_{U,b}^{\mathrm{can}} = d\mathcal{F}_{I/d}^{-1}(\mu_{U,b})$. Note that the explicit expression for $\mathcal{F}_{I/d}$ for this POVM is

$$\mathcal{F}_{I/d} = d \int_{\mathbf{U}(d)} dU \sum_b \mathbb{P}(\mu_{U,b}), \tag{44}$$

where we have used $\mathrm{tr}(\mu_{U,b}) = 1$ for all U, b. Therefore, in terms of the operator $\mathcal{M}$ defined in Eq. (38), we have $\mathcal{F}_{I/d} = d\mathcal{M}$ and $\tilde{\mu}_{U,b}^{\mathrm{can}} = \mathcal{M}^{-1}(\mu_{U,b})$. We finally recover Eq. (42), observing that $\mathcal{M}$ or, equivalently, $\mathcal{F}_{I/d}$, is Hermitian as a superoperator and thus

$$\langle b|U\mathcal{M}^{-1}(\mathcal{O})U^\dagger|b\rangle = \mathrm{tr}(\mu_{U,b}\mathcal{M}^{-1}(\mathcal{O})) = \mathrm{tr}(\mathcal{O}\tilde{\mu}_{U,b}^{\mathrm{can}}). \tag{45}$$

Rewriting the shadow norm as in Eq. (43) clearly shows that it corresponds to the nontrivial part of the variance, maximized over the input states, and that the definition of shadow norm is thus applicable for any choice of measurement and estimator. In fact, we have the general result $\|\mathcal{O}\|_{\rm sh}^2 = \|A\|_{\rm op}$, where A is the operator defined in Eq. (29), and thus the scaling results derived for random (Haar or Clifford) unitaries can be viewed as a particular instance of the more general results presented in Sec. IV.

VII. CONCLUSIONS AND OUTLOOK

We have demonstrated how the general theory of measurement frames embodies a natural framework for shadow tomography. In doing so, we have assessed thoroughly the interplay between general measurements and associated optimal estimators to recover expectation values of target observables. Our results push the current knowledge in this context, recovering previously reported seminal results (cf. Ref. [14]) as special cases of our general framework, providing a natural understanding of the notion of shadow norm often used in the topical literature, and allowing estimation of finite sets of rank-1 bounded observables with a number of samples that does not grow with the dimension of the underlying space. We have provided analytical bounds for the estimation variance in several cases of interest, including the variance averaged over input states and the variance averaged over both input states and target observables. Among other things, we have provided explicit results for the averaged variance in the case of tight measurement frames and general bounds tying the average variance to how close a POVM is to being a tight measurement frame and we have also found an explicit expression for the nonaveraged variance for rank-1 POVMs that form 3-designs. In Table I, we provide a useful summary of some of the main expressions for frame operators and variances discussed throughout the paper. To further ease the understanding of the different notions introduced in this paper, we have also included in Appendix H several toy examples in which we explicitly work out frame superoperators and other relevant quantities.

Besides improving our understanding of general shadow-tomography protocols, our results help the analysis and assessment of estimation errors in general measurement protocols, providing a unifying framework to understand both linear state tomography and shadow tomography. Our work thus contributes to the design of optimal strategies for single-setting quantum state tomography, which has recently attracted significant attention [26,27,52,53], as well as more general experimental protocols relying on learning properties of input states from measurement outcomes [44,54–57]. Another context where our results will prove useful is the analysis of quantum reservoir-computing architectures, which have

TABLE I. A summary of the introduced quantities. A schematic review of the expressions provided in the text for frame operator ($\mathcal{F}$), state estimator ($\tilde\mu$), and associated variances. The first two rows summarize some of the quantities associated with the tomographic estimation of input states. Similarly, the next three rows refer to the case of recovering the expectation value of some target observable $\mathcal{O}$. The first and third rows summarize the quantities associated with the estimators that have minimum variance when the true input state is ρ. The second and fourth rows summarize the quantities associated with estimators that have minimum variance on average over the possible input states—or, equivalently, that have minimum variance when the true input state is I/d. The fifth row contains quantities associated with the estimator with minimum variance on average over both input states and target observables. Finally, the last row gives the explicit expressions for the canonical frame operator, the canonical estimator, the MSE matrix, and the averaged variance, in the special case of tight rank-1 measurement frames.

		Frame operator	Estimator	Variance of estimator
State estimation	(with prior ρ)	$\mathcal{F}_\rho \equiv \sum_b \dfrac{\mathbb{P}(\mu_b)}{\langle\mu_b,\rho\rangle}$	$\tilde\mu_b^{(\rho)} \equiv \dfrac{\mathcal{F}_\rho^{-1}(\mu_b)}{\langle\mu_b,\rho\rangle}$	$\mathcal{E}_\rho = \mathrm{tr}(\mathcal{F}_\rho^{-1}) - \mathrm{tr}(\rho^2)$
	(average over ρ)	$\mathcal{F}_{I/d} \equiv d\sum_b \dfrac{\mathbb{P}(\mu_b)}{\mathrm{tr}(\mu_b)}$	$\tilde\mu_b^{\rm can} \equiv \dfrac{d\mathcal{F}_{I/d}^{-1}(\mu_b)}{\mathrm{tr}(\mu_b)}$	$\overline{\mathcal{E}}_P = \mathrm{tr}(\mathcal{F}_{I/d}^{-1}) - P$
Observable estimation	(with prior ρ)	$\mathcal{F}_\rho$	$\langle\mathcal{O},\tilde\mu_b\rangle = \langle\mathcal{O},\tilde\mu_b^{(\rho)}\rangle$	$\mathrm{Var}[\hat o\|\rho] = \langle\mathcal{O},\mathcal{C}_\rho(\mathcal{O})\rangle$
	(average over ρ)	$\mathcal{F}_{I/d}$	$\langle\mathcal{O},\tilde\mu_b\rangle = \langle\mathcal{O},\tilde\mu_b^{\rm can}\rangle$	$\overline{\mathrm{Var}[\hat o\|P]} = \langle\mathcal{O},\overline{\mathcal{C}}_\rho(\mathcal{O})\rangle$
	(average over ρ, $\mathcal{O}$)	$\mathcal{F}_{I/d}$	$\tilde\mu^{\rm can}$	$\overline{\overline{\mathrm{Var}[\hat o\|P]}} = \dfrac{Vd}{d^2-1}\left[\mathrm{tr}(\mathcal{F}_{I/d}^{-1}) - P\right]$
Tight rank-1 POVM $\mu_b = w_b\mathbb{P}(\psi_b)$		$\mathcal{F}_{I/d} = d\dfrac{\mathbb{P}(I)+\mathrm{Id}}{d+1}$	$\tilde\mu_b^{\rm can} = (d+1)\mathbb{P}(\psi_b) - I$	$\overline{\mathcal{E}}_P = d^2+d-1-P$ $\overline{\mathrm{Var}[\hat o\|P]} = \dfrac{Vd\left(d^2+d-1-P\right)}{d^2-1}$

been recently shown to be representable via generalized measurements summarizing the properties of the reservoir and to be applicable for quantum state-estimation tasks [58]. More generally, our formalism can be applied to any scenario where the goal is to extract properties of states from measurement outcomes, especially (although not exclusively) when the goal is to efficiently extract few properties from high-dimensional states. Other potential avenues for research in this context include a more thorough exploration of the performance guarantees for 2-designs that are not also 3-designs, which would significantly expand the class of experimental situations where efficient dimension-independent estimation is possible.

By demonstrating the connection between the computation of classical shadows and the associated unbiased linear estimators, our approach establishes useful connections with metrology and estimation theory. In particular, estimating only certain properties of an unknown quantum state is formally a quantum semiparametric estimation problem [59]—also known in the finite-dimensional case as estimation with nuisance parameters [60]. While quantum estimation is most commonly studied in a local and/or asymptotic scenario, we hope that our approach will lead to further connections between shadow tomography and semiparametric estimation in the nonasymptotic regime. Another intriguing area where an approach based on infinite-dimensional measurement frames could provide useful insights is continuous-variable shadow tomography, which has only very recently been proposed [61,62]. Finally, the agility of the framework that we put forward holds the promise to inform experimental efforts aimed at demonstrating a resource-inexpensive route to quantum state and property reconstruction. The code used to generate the figures in the paper can be found at Ref. [63].

ACKNOWLEDGMENTS

L.I. acknowledges support from the Italian Ministry for Universities and Research (MUR) and Amazon Web Services (AWS) under the Programma Operativo Nazionale (PON) project "Ricerca e Innovazione" 2014–2020, "Calcolo quantistico in dispositivi quantistici rumorosi nel regime di scala intermedia" (NISQ—Noisy, Intermediate-Scale Quantum). I.P. is grateful to the Marie Skłodowska-Curie Actions (MSCA) COFUND project Collaboration In Training and Innovation for Growing, Evolving and Networked Societies (CITI-GENS) (Grant No. 945231). F.A. acknowledges financial support from MUR under the PON Ricerca e Innovazione 2014–2020 project Efficienza Energetica con risorse QUantistiche (EEQU). M.P. acknowledges the support by the European Union Horizon 2020 FET-Open project "Testing the Large-Scale Limit of Quantum Mechanics" (TEQ) (Grant Agreement No. 766900), the Horizon Europe EIC Pathfinder project "Quantum Control of Gravity with Levitated Mechanics" (QuCoM) (Grant Agreement No. 101046973), the Leverhulme Trust Research Project Grant Ultracold Quantum Thermo-machines (UltraQuTe) (Grant No. RPG-2018-266), the Royal Society Wolfson Fellowship (Grant No. RSWF/R3/183013), the United Kingdom Engineering and Physical Sciences Research Council (EPSRC) (Grant No. EP/T028424/1), and the Department for the Economy Northern Ireland under the U.S.-Ireland R&D Partnership Programme (Grants No. USI 175 and No. USI 194).

APPENDIX A: PROPERTIES OF FRAME SUPEROPERATORS

In this appendix, we briefly review some important properties of the frame superoperators used in the paper.

1. Definition

The frame superoperator that provides the minimum-variance state estimator when the true input is some reference state ρ is

$$\mathcal{F}_\rho = \sum_b \frac{\mathbb{P}(\mu_b)}{\langle \mu_b, \rho\rangle}, \tag{A1}$$

where we denote by $\mathbb{P}(\mu_b)$ the quantum map sending operators X to $\mu_b\langle\mu_b, X\rangle$. If $\rho \in \mathrm{D}(\mathbb{C}^d)$ is a d-dimensional state, then $\mu_b \in \mathrm{Pos}(\mathbb{C}^d)$, $\mathcal{F}_\rho : \mathrm{Lin}(\mathbb{C}^d) \to \mathrm{Lin}(\mathbb{C}^d)$, and $\mathcal{F}_\rho \in \mathrm{Lin}(\mathrm{Lin}(\mathbb{C}^d))$. Being a linear function defined on linear operators, $\mathcal{F}_\rho$ is a quantum map. To connect to the more general theory of frames in linear algebra, this map is the frame operator corresponding to the rescaled frame of operators with elements $\{\mu_b/\sqrt{\langle\mu_b,\rho\rangle}\}_b$.

2. General properties

Thinking of $\mathcal{F}_\rho$ as a linear operator, we define its trace in the standard way, i.e.,

$$\mathrm{tr}(\mathcal{F}_\rho) = \sum_\alpha \langle \sigma_\alpha, \mathcal{F}_\rho(\sigma_\alpha)\rangle, \tag{A2}$$

for an arbitrary orthonormal basis of Hermitian operators $\{\sigma_\alpha\}_{\alpha=1}^{d^2}$. In particular, $\mathrm{tr}(\mathbb{P}(\mu_b)) = \langle\mu_b,\mu_b\rangle = \mathrm{tr}(\mu_b^2)$, and thus

$$\mathrm{tr}(\mathcal{F}_\rho) = \sum_b \frac{\mathrm{tr}(\mu_b^2)}{\langle\mu_b,\rho\rangle}. \tag{A3}$$

We can furthermore verify by direct substitution that

$$\mathcal{F}_\rho(\rho) = I, \quad \mathcal{F}_\rho^{-1}(I) = \rho. \tag{A4}$$

3. Properties of the inverse

As discussed in the main text and derived in Appendix B, the minimum-variance unbiased estimator provided by $\mathcal{F}_\rho$ is $\hat{f}(b) \equiv \tilde{\mu}_b^{(\rho)}$, with

$$\tilde{\mu}_b^{(\rho)} \equiv \frac{1}{\langle \mu_b, \rho\rangle} \mathcal{F}_\rho^{-1}(\mu_b). \tag{A5}$$

In particular, this means that the canonical dual frame corresponding to this frame operator has elements $\{\sqrt{\langle \mu_b, \rho\rangle}\tilde{\mu}_b^{(\rho)}\}_b$ and

$$\mathcal{F}_\rho^{-1} = \sum_b \mathbb{P}(\sqrt{\langle \mu_b, \rho\rangle}\tilde{\mu}_b^{(\rho)}) = \sum_b \langle \mu_b, \rho\rangle \mathbb{P}(\tilde{\mu}_b^{(\rho)}). \tag{A6}$$

Taking the trace, we obtain

$$\operatorname{tr}(\mathcal{F}_\rho^{-1}) = \sum_b \langle \mu_b, \rho\rangle \operatorname{tr}((\tilde{\mu}_b^{(\rho)})^2). \tag{A7}$$

This expression is particularly useful in that it directly enters the corresponding MSE matrix.

4. Canonical estimator

The minimum-variance unbiased estimator when no prior knowledge about the true input state is assumed is obtained by setting $\rho = I/d$ in the frame superoperator. We show in Appendix D that the unbiased state estimator that minimizes the L_2 error averaged over unitarily equivalent states is $\hat{f}(b) \equiv \tilde{\mu}_b^{\mathrm{can}}$, with

$$\tilde{\mu}_b^{\mathrm{can}} = \frac{d\mathcal{F}_{I/d}^{-1}(\mu_b)}{\operatorname{tr}(\mu_b)}. \tag{A8}$$

The map $\mathcal{F}_{I/d}$ has some further properties compared with its general counterpart. In particular, we have $\mathcal{F}_{I/d}(I) = dI$, which means that I is an eigenvector of $\mathcal{F}_{I/d}$. This observation can be exploited to write the general decomposition

$$\mathcal{F}_{I/d} = d\mathbb{P}(I/\sqrt{d}) + \tilde{\mathcal{F}}_{I/d}, \tag{A9}$$

where $\tilde{\mathcal{F}}_{I/d}$ is defined as the projection of $\mathcal{F}_{I/d}$ on the subspace of traceless operators, i.e.,

$$\tilde{\mathcal{F}}_{I/d} = \Pi_{H_0}\mathcal{F}_{I/d}\Pi_{H_0} = \Pi_{H_0}\tilde{\mathcal{F}}_{I/d}\Pi_{H_0}, \tag{A10}$$

where $\Pi_{H_0} \equiv \mathrm{Id} - \mathbb{P}(I/\sqrt{d})$ is the (superoperator) projector onto the subspace of traceless operators. We employ the rescaled identity operator $I/\sqrt{d}$ in these expressions to ensure the normalization of the corresponding operator with respect to the Hilbert-Schmidt inner product: $\|I/\sqrt{d}\|_2 \equiv \operatorname{tr}((I/\sqrt{d})^2) = 1$. This decomposition also translates into corresponding simplified expressions for the inverse and trace:

$$\begin{aligned}\mathcal{F}_{I/d}^{-1} &= \frac{1}{d}\mathbb{P}(I/\sqrt{d}) + \tilde{\mathcal{F}}_{I/d}^{-1},\\ \operatorname{tr}(\mathcal{F}_{I/d}^{-1}) &= \frac{1}{d} + \operatorname{tr}(\tilde{\mathcal{F}}_{I/d}^{-1}).\end{aligned} \tag{A11}$$

As discussed in more detail in Appendix D, these expressions simplify even further in the special case of tight rank-1 measurement frames.

5. MSE matrix

Following Ref. [23], we define the *MSE matrix* corresponding to a state ρ, measurement μ, and estimator $\tilde{\mu}$, as

$$\mathcal{C}_\rho = \sum_b \langle \mu_b, \rho\rangle \mathbb{P}(\tilde{\mu}_b) - \mathbb{P}(\rho). \tag{A12}$$

Using the *minimum-variance dual estimator* given in Eq. (A5), the MSE matrix takes the simplified form

$$\mathcal{C}_\rho^{\mathrm{opt}} = \mathcal{F}_\rho^{-1} - \mathbb{P}(\rho). \tag{A13}$$

For an arbitrary choice of possibly suboptimal estimator, we have the inequality $\mathcal{C}_\rho \geq \mathcal{C}_\rho^{\mathrm{opt}}$. A remarkable property of the MSE matrix is that its trace equals the average L_2 state-estimation error, as will be further discussed in the following appendixes. The optimal MSE matrix can also be regarded as the (classical) Fisher information matrix, when the states are considered parametrized via their coefficients in some orthonormal basis.

APPENDIX B: MINIMUM-VARIANCE STATE ESTIMATORS

Let us consider a generic unbiased estimator—or, equivalently, as discussed before, a generic dual measurement frame—and ask what is the associated average estimation error. Measuring the error in the Hilbert-Schmidt distance, we find that

$$\begin{aligned}\mathbb{E}[\|\hat{f} - \rho\|_2^2] &\equiv \sum_b \langle \mu_b, \rho\rangle \|\hat{f}(b) - \rho\|_2^2\\ &= \mathbb{E}\operatorname{tr}(\hat{f}^2) - \operatorname{tr}(\rho^2),\\ \mathbb{E}\operatorname{tr}(\hat{f}^2) \equiv \Delta^2(\rho,\mu,\tilde{\mu}) &\equiv \sum_b \langle \mu_b, \rho\rangle \operatorname{tr}(\tilde{\mu}_b^2),\end{aligned} \tag{B1}$$

where we have introduced the notation $\Delta^2 \equiv \Delta^2(\rho,\mu,\tilde{\mu})$ to denote the component of the average error that depends on the choice of measurement μ and dual $\tilde{\mu}$. The dependence of this quantity on these choices will not the explicitly shown in the following in order to ease the notation.

1. Minimum-variance dual frame

As previously mentioned, different dual frames generally exist and from Eq. (B1) we can see that the choice of dual frame $\tilde{\mu}$ can affect the associated average estimation error. It is then natural to ask what is the choice of dual frame that minimizes the estimation variance? This issue is addressed in Refs. [19–21,30–32]. We include here a different approach to deriving the minimum-variance unbiased estimators from the rescaled frame superoperator, using the method of Lagrange multipliers to directly perform the optimization with respect to all possible linear unbiased estimators.

2. Problem definition in vectorized notation

To find the minimum-variance estimator $\tilde{\boldsymbol{\mu}}$, we observe that the task involves optimizing a quadratic function under linear constraints. To see this more clearly, we temporarily neglect the fact that the various objects in Eq. (B1) are operators and simply think of them as vectors, upon some choice of orthonormal basis for the underlying Hilbert space. The error term Δ^2, which is what we need to minimize, can be written in vectorized notation as

$$\sum_b \langle\mu_b,\rho\rangle \operatorname{tr}(\tilde{\mu}_b^2) = \sum_b \langle\mu_b,\rho\rangle \|\tilde{\mu}_b\|^2 = \sum_{b,i,j} \mu_{bi}\rho_i\tilde{\mu}_{bj}^2, \tag{B2}$$

and the minimization must be performed with respect to the real parameters $\tilde{\mu}_{bj}$. More explicitly, this notation amounts to decomposing the operators as

$$\tilde{\mu}_{bj} \equiv \langle\sigma_j,\tilde{\mu}_b\rangle, \quad \mu_{bj} \equiv \langle\sigma_j,\mu_b\rangle, \quad \rho_i \equiv \langle\sigma_i,\rho\rangle, \tag{B3}$$

for some fixed choice of orthonormal operatorial basis $\{\sigma_i\}$.

We need to take into consideration that not all sets of parameters $\tilde{\mu}_{bj}$ correspond to a valid dual frame of μ. The definition of a dual frame can be written in vectorized notation as

$$\sum_{b,i} \mu_{bi}\rho_i\tilde{\mu}_{bj} = \rho_j \tag{B4}$$

and this must hold for all possible choices of ρ. Although there are, in principle, an infinite amount of constraints, they can be thought of as equivalent to the finite set of constraints corresponding to using as ρ the elements of the considered operatorial basis $\{\sigma_i\}$. These constraints read

$$\sum_b \mu_{bi}\tilde{\mu}_{bj} = \delta_{ij}, \quad \forall i,j. \tag{B5}$$

Let us denote this set of constraints by $\phi_{ij} \equiv \phi_{ij}(\mu,\tilde{\mu}) = 0$, having defined

$$\phi_{ij} \equiv \sum_b \mu_{bi}\tilde{\mu}_{bj} - \delta_{ij}. \tag{B6}$$

3. Lagrange multipliers to find stationary points

To find the minimum of Eq. (B3) under the constraints in Eq. (B5), we can use the general method of Lagrange multipliers. For there to be a stationary point for the cost function under the given constraints, the gradient of the cost must be in the linear span of the gradients of the constraints. More explicitly, this means that there must be a set of coefficients λ_{ij} such that, for all b and k, we have

$$\frac{\partial\Delta^2}{\partial\tilde{\mu}_{bk}} = \sum_{ij}\lambda_{ij}\frac{\partial\phi_{ij}}{\partial\tilde{\mu}_{bk}}. \tag{B7}$$

Computing the derivatives explicitly, we find that

$$\frac{\partial\Delta^2}{\partial\tilde{\mu}_{bk}} = 2\sum_i \mu_{bi}\rho_i\tilde{\mu}_{bk}, \qquad \frac{\partial\phi_{ij}}{\partial\tilde{\mu}_{bk}} = \mu_{bi}\delta_{jk}, \tag{B8}$$

and thus Eq. (B7) becomes

$$2\sum_i \mu_{bi}\rho_i\tilde{\mu}_{bk} = \sum_i \lambda_{ik}\mu_{bi}. \tag{B9}$$

Thinking of λ, μ, and $\tilde{\mu}$ as matrices and defining the diagonal matrix Λ with components $\Lambda_{ab} \equiv \delta_{ab}\langle\mu_b,\rho\rangle$, Eqs. (B9) and (B5) can be written concisely as

$$2\Lambda\tilde{\mu} = \mu\lambda, \quad \mu^T\tilde{\mu} = I. \tag{B10}$$

Putting these together, and assuming Λ to be invertible—which amounts to using ρ such that $\langle\mu_b,\rho\rangle > 0$ for all b—we obtain $2I = 2\mu^T\tilde{\mu} = \mu^T\Lambda^{-1}\mu\lambda$. We thus conclude that the set of coefficients λ_{ij} must have the form

$$\lambda = 2(\mu^T\Lambda^{-1}\mu)^{-1}. \tag{B11}$$

In writing this, we are interpreting λ as a matrix, i.e., as a linear operator in the underlying Hilbert space of Hermitian operators. In other words, we can in this context interpret the set of Lagrange multipliers as a quantum map satisfying the given relations. We can safely talk about the inverse of $\mu^T\Lambda^{-1}\mu$ because the corresponding map is invertible provided that μ is an IC POVM. This is because $\mu^T\Lambda^{-1}\mu$, going back to the original formalism in terms of operators, corresponds to the map

$$\mathcal{F}_\rho \equiv \sum_b \frac{\mathbb{P}(\mu_b)}{\langle\mu_b,\rho\rangle} \tag{B12}$$

and if $\{\mu_b\}$ is an IC POVM, then its elements span the space and the quantum map thus defined is invertible.

With this solution for λ, we can now find the minimum-variance dual frame $\tilde{\mu}$ using Eq. (B10) as

$$\tilde{\mu} = \Lambda^{-1}\mu(\mu^T\Lambda^{-1}\mu)^{-1}. \quad (B13)$$

Note that μ is not in general an invertible, nor a squared, matrix and thus we cannot simplify the inverse $(\mu^T\Lambda^{-1}\mu)^{-1}$ using the inverse of its elements.

Going back to the notation with operators, the minimum-variance dual frame that we have just found corresponds to the operators

$$\tilde{\mu}_b = \frac{1}{\langle\mu_b, \rho\rangle}\mathcal{F}_\rho^{-1}(\mu_b), \quad (B14)$$

where we have denoted by $\mathcal{F}_\rho$ the map corresponding to the Lagrange multipliers, which can also be seen as the frame operator of the rescaled frame with elements $\mu_b/\sqrt{\langle\mu_b, \rho\rangle}$. An explicit expression of $\mathcal{F}_\rho^{-1}$ in terms of $\tilde{\mu}$ can be obtained using Eq. (B10) again: we get that $\lambda = 2\tilde{\mu}^T\Lambda\tilde{\mu}$ and therefore

$$\mathcal{F}_\rho^{-1} \equiv \sum_b \langle\mu_b, \rho\rangle\mathbb{P}(\tilde{\mu}_b), \quad (B15)$$

to be compared with $\mathcal{F}_\rho$ of Eq. (B12).

It is worth stressing the precise kind of "optimality" that we have just derived. While the above optimal dual frame $\tilde{\mu}_b$ is an unbiased estimator with respect to all states, meaning that $\sum_b\langle\mu_b, \rho\rangle\tilde{\mu}_b = \rho$ for all ρ, the associated estimation error and its optimality depend upon the specific state ρ that is being examined. Different choices of ρ will correspond to different minimum-variance estimators, although all of these estimators are unbiased with respect to all states. To find the estimator that has minimum variance *on average with respect to all possible input states*—sampled uniformly from the Haar measure—we just need to set $\rho = I/d$, obtaining

$$\tilde{\mu}_b = \frac{d}{\mathrm{tr}(\mu_b)}\mathcal{F}_{I/d}^{-1}(\mu_b), \quad (B16)$$

where

$$\mathcal{F}_{I/d} \equiv d\sum_b \frac{\mathbb{P}(\mu_b)}{\mathrm{tr}(\mu_b)}, \quad \mathcal{F}_{I/d}^{-1} \equiv \frac{1}{d}\sum_b \mathrm{tr}(\mu_b)\mathbb{P}(\tilde{\mu}_b). \quad (B17)$$

This can be deduced from the linearity of Δ^2 in Eq. (B1) with respect to ρ. Therefore, integrating it over Haar-distributed states is equivalent to evaluating it at the maximally mixed state $\rho = I/d$.

APPENDIX C: MINIMUM-VARIANCE OBSERVABLE ESTIMATORS

In Appendix B, we have derived the form of the unbiased state estimator that minimizes the averaged L_2 estimation error. The focus of shadow-tomography protocols is, however, the estimation of observables, not retrieving tomographically complete descriptions of the state itself. It would stand to reason that if the goal was to estimate some target observable $\mathcal{O}$, this might be possible with a different strategy that does not pass through state estimators and gives even lower variance. In this appendix, we will show that this is in fact not the case: any unbiased estimator $\hat{o}$ for an observable $\mathcal{O}$, assuming that it is unbiased for all possible input states, is bound to have the form $\hat{o}(b) = \langle\mathcal{O}, \tilde{\mu}_b\rangle$ for some dual measurement frame $\tilde{\boldsymbol{\mu}}$.

1. All observable estimators pass through dual frames

Let $\hat{o}$ be an unbiased estimator for a target observable $\mathcal{O}$. By definition, this means that we have the relation

$$\sum_b \hat{o}(b)\langle\mu_b, \rho\rangle = \langle\mathcal{O}, \rho\rangle \quad (C1)$$

for all states ρ. But by linearity of the inner product, this implies $\sum_b \hat{o}(b)\mu_b = \mathcal{O}$, which tells us that $\hat{o}(b) \in \mathbb{R}$ can be interpreted as the coefficients appearing in the expansion of $\mathcal{O}$ as a linear combination of the frame elements $(\mu_b)_b$. From the general theory of frames, we then conclude that there must be some dual frame $(\tilde{\mu}_b)_b$ such that $\hat{o}(b) = \langle\tilde{\mu}_b, \mathcal{O}\rangle$. The opposite direction is immediate: if $\tilde{\mu}_b$ is a dual frame and thus gives an unbiased state estimator, it is clear that $\langle\mathcal{O}, \tilde{\mu}_b\rangle$ is an unbiased estimator for $\mathcal{O}$. We conclude that unbiased observable estimators always pass through some state estimator $\tilde{\boldsymbol{\mu}}$.

2. Minimum-variance observable estimators

The above considerations tell us that we can restrict our attention to estimators of the form $\hat{o}(b) = \langle\tilde{\mu}_b, \mathcal{O}\rangle$. The question remains as to what choice of estimator is best—in the sense of having minimum variance—to recover $\mathcal{O}$ specifically. To answer this question, we follow a reasoning similar to that in Appendix B. If $\tilde{\mu}_b$ is a generic dual frame, with corresponding estimator $\hat{o}(b) \equiv \tilde{\mu}_b$, and ρ is the true state, the variance reads

$$\mathrm{Var}[\hat{o}|\mathcal{O}] = \sum_b \langle\mu_b, \rho\rangle\langle\mathcal{O}, \tilde{\mu}_b\rangle^2 - \langle\mathcal{O}, \rho\rangle^2. \quad (C2)$$

We focus on minimizing the first term with respect to $\tilde{\mu}$, as the second term only depends on ρ and $\mathcal{O}$. In vectorized notation, the first term can be rewritten as

$$\sum_b \langle\mu_b, \rho\rangle\langle\mathcal{O}, \tilde{\mu}_b\rangle^2 = \mathcal{O}^T\tilde{\mu}^T\Lambda\tilde{\mu}\mathcal{O}. \tag{C3}$$

In this notation, $\tilde{\mu}$ and μ are matrices, Λ is a diagonal matrix, and $\mathcal{O}$ is a vector. The constraints on the estimators remain $\tilde{\mu}^T\mu = \mu^T\tilde{\mu} = I$, which amounts to the set of constraints $\phi_{ij} = \sum_b \mu_{bi}\tilde{\mu}_{bj} - \delta_{ij}$. Taking the derivative with respect to $\tilde{\mu}_{bk}$ on the cost function given in Eq. (C3), and using the constraints on $\tilde{\mu}$, we conclude that there must be coefficients λ_{ij} such that

$$2\sum_j \Lambda_{bb}\mathcal{O}_k\tilde{\mu}_{bj}\mathcal{O}_j = \sum_{ij}\lambda_{ij}\,\mu_{bi}\delta_{jk}. \tag{C4}$$

In more compact matrix notation, denoting by λ the matrix with components λ_{ij}, we obtain the condition

$$2\Lambda\tilde{\mu}\mathcal{O}\mathcal{O}^T = \mu\lambda. \tag{C5}$$

Multiplying both sides from the left first by Λ^{-1} and then by μ^T, and observing that $\mu^T\Lambda^{-1}\mu$ is the matrix representation of $\mathcal{F}_\rho$, which is invertible for IC POVMs, we find that

$$\lambda = 2(\mu^T\Lambda^{-1}\mu)^{-1}\mathcal{O}\mathcal{O}^T. \tag{C6}$$

We thus conclude that the minimum-variance estimators are given by

$$\tilde{\mu}\mathcal{O}\mathcal{O}^T = \Lambda^{-1}\mu(\mu^T\Lambda^{-1}\mu)^{-1}\mathcal{O}\mathcal{O}^T. \tag{C7}$$

More explicitly, this amounts to

$$\sum_k \tilde{\mu}_{bk}\mathcal{O}_k = \sum_{ik}\Lambda_{bb}^{-1}\mu_{bi}(\mathcal{F}_\rho^{-1})_{ik}\mathcal{O}_k. \tag{C8}$$

In operator notation, this reads

$$\langle\mathcal{O}, \tilde{\mu}_b\rangle = \frac{\langle\mathcal{O}, \mathcal{F}_\rho^{-1}(\mu_b)\rangle}{\langle\mu_b, \rho\rangle}. \tag{C9}$$

We conclude that the estimator $\tilde{\mu}_b$ that minimizes $\mathrm{Var}[\hat{o}|\mathcal{O}]$ when the input state is ρ always has the form:

$$\langle\mathcal{O}, \tilde{\mu}_b\rangle = \langle\mathcal{O}, \tilde{\mu}_b^{(\rho)}\rangle, \tag{C10}$$

in other words, the estimators equal to the minimum-variance state estimator $\tilde{\mu}_b^{(\rho)}$ on the span of $\mathcal{O}$. The associated variance can be written in terms of the MSE matrix as

$$\langle\mathbb{P}(\mathcal{O}), \mathcal{C}_\rho\rangle \equiv \langle\mathcal{O}, \mathcal{C}_\rho(\mathcal{O})\rangle = \sum_b \langle\mu_b, \rho\rangle\langle\mathcal{O}, \tilde{\mu}_b\rangle^2 - \langle\mathcal{O}, \rho\rangle^2, \tag{C11}$$

where $\mathbb{P}(\mathcal{O})$ denotes the map $X \mapsto \langle\mathcal{O}, X\rangle\mathcal{O}$ for all $X \in \mathrm{Lin}(\mathbb{C}^d)$. We thus conclude that finding the state estimator giving an observable estimator with the smallest variance amounts to finding an estimator that acts like the overall minimum-variance state estimator on the support of the observable. In other words, the minimum-variance state estimator also provides the minimum-variance observable estimator for any observable (under the same assumptions on the input state). As in Appendix B, all these results also hold in the averaged scenario: the estimators minimizing the variance on average over input states are obtained with the choice $\rho = I/d$, i.e., using $\tilde{\boldsymbol{\mu}}^{\mathrm{can}}$.

APPENDIX D: TIGHT MEASUREMENTS AND WEIGHTED 2-DESIGNS

In this appendix, we prove the equivalence between weighted complex projective 2-designs and tight measurement frames, discuss the general property of tight measurement frames, and prove the known lower bounds on L_2 average estimation error corresponding to canonical state estimators. Although using a slightly different formalism, the idea behind the proof reported here is analogous to that reported in Ref. [43].

1. Weighted 2-designs and tight measurement frames

Consider a rank-1 measurement with elements $\mu_b = w_b\mathbb{P}(\psi_b)$, $b = 1, \ldots, m$, for some set of weights $w_b \in \mathbb{R}$ such that $\sum_b w_b = d$ and some set of vectors $|\psi_b\rangle \in \mathbb{C}^d$. The corresponding canonical frame superoperator is, by definition, equal to

$$\mathcal{F}_{I/d} = d\sum_b \frac{\mathbb{P}(\mu_b)}{\mathrm{tr}(\mu_b)} = d\sum_b w_b\mathbb{P}(\mathbb{P}(\psi_b)), \tag{D1}$$

where we have used $\mathrm{tr}(\mu_b) = w_b$ and we have denoted by $\mathbb{P}(\mathbb{P}(\psi_b))$ the projector onto the projector $\mathbb{P}(\psi_b)$. Here, $\psi_b \in \mathbb{C}^d$ is a vector, $\mathbb{P}(\psi_b) \equiv |\psi_b\rangle\langle\psi_b| \in \mathrm{Herm}(\mathbb{C}^d)$ is a linear operator projecting onto $|\psi_b\rangle$, and thus $\mathbb{P}(\mathbb{P}(\psi_b))$ is a linear operator acting in the space of linear operators, which projects onto the linear operator $\mathbb{P}(\psi_b)$. This object is a quantum map, which acts on any $X \in \mathrm{Lin}(\mathbb{C}^d)$ as follows:

$$\mathbb{P}(\mathbb{P}(\psi_b))(X) = \mathbb{P}(\psi_b)\langle\mathbb{P}(\psi_b), X\rangle \equiv \mathbb{P}(\psi_b)\langle\psi_b, X\psi_b\rangle. \tag{D2}$$

Being this a quantum map, we can consider its Choi representation. Given any map $\Phi : \mathrm{Lin}(\mathcal{H}_A) \to \mathrm{Lin}(\mathcal{H}_B)$, we define its Choi representation as the operator $J(\Phi) \in \mathrm{Lin}(\mathcal{H}_B \otimes \mathcal{H}_A)$ such that

$$J(\Phi) = \sum_{ij}\Phi(|i\rangle\langle j|) \otimes |i\rangle\langle j|. \tag{D3}$$

For an arbitrary map of the form $\Phi(X) = \langle \bar{A}, X\rangle B$, the Choi is $J(\Phi) = B \otimes \bar{A}$. It follows that

$$J(\mathbb{P}(\mathbb{P}(\psi_b))) = \mathbb{P}(\psi_b) \otimes \mathbb{P}(\psi_b)^T, \tag{D4}$$

and thus, for the frame superoperator,

$$J(\mathcal{F}_{I/d}) = d\left[\sum_b w_b \mathbb{P}(\psi_b)^{\otimes 2}\right]^{T_B}, \tag{D5}$$

where T_B denotes the partial transpose of the second space. This expression is useful because it provides a direct connection with the defining property of weighted 2-designs. The vectors $|\psi_b\rangle$ form a complex projective 2-design with weights w_b if and only if we have

$$\sum_b w_b \mathbb{P}(\psi_b)^{\otimes 2} = d\frac{\Pi_{\rm sym}}{\binom{d+1}{2}}. \tag{D6}$$

The d normalization factor on the right-hand side of this equation comes from $\sum_b w_b = d$, whereas in the standard definition of weighted 2-designs the weights are normalized to 1. Using this relation, we obtain

$$J(\mathcal{F}_{I/d})^{T_B} = d^2\frac{\Pi_{\rm sym}}{\binom{d+1}{2}} = d\frac{I\otimes I + W}{d+1}, \tag{D7}$$

where we have expressed the projector in terms of the SWAP operator W via $\Pi_{\rm sym} = (I + W)/2$. Observing that $W^{T_B} = \sum_{ij} |ii\rangle\langle jj|$, $J(\mathbb{P}(I)) = I\otimes I$, and $J(\mathrm{Id}) = W^{T_B}$, together with the fact that the Choi is a linear isomorphism between maps and operators, we conclude that

$$\mathcal{F}_{I/d} = d\frac{\mathbb{P}(I) + \mathrm{Id}}{d+1}. \tag{D8}$$

This derivation shows that, for any rank-1 IC POVM with elements $\mu_b = w_b\mathbb{P}(\psi_b)$, the frame superoperator $\mathcal{F}_{I/d}$ has this form *if and only if* the vectors $|\psi_b\rangle$ and weights w_b form a weighted 2-design. Equation (D8) differs by a factor of d from the expressions for tight frames found in, e.g., Ref. [19] but that is simply due to the definitions of the frame superoperator differing by a factor of d and will not affect our results.

2. Properties of tight-frame superoperators

Now suppose that μ is a tight rank-1 IC POVM and thus the frame superoperator satisfies Eq. (D8). In light of the decomposition of Eq. (A9), we can rewrite the frame operator as

$$\mathcal{F}_{I/d} = d\mathbb{P}(I/\sqrt{d}) + \frac{d}{d+1}\left(\mathrm{Id} - \mathbb{P}(I/\sqrt{d})\right). \tag{D9}$$

This expression is useful because it splits the action of $\mathcal{F}_{I/d}$ into two invariant orthogonal subspaces. The superoperators $\mathbb{P}(I)/d$ and $\mathrm{Id} - \mathbb{P}(I)/d$ project onto the one-dimensional subspace spanned by I and the (d^2-1)-dimensional subspace of traceless Hermitian matrices, respectively. It follows that the inverse has the form

$$\begin{aligned}\mathcal{F}_{I/d}^{-1} &= \frac{1}{d}\mathbb{P}(I/\sqrt{d}) + \frac{d+1}{d}\left(\mathrm{Id} - \mathbb{P}(I/\sqrt{d})\right)\\ &= \frac{(d+1)\,\mathrm{Id} - \mathbb{P}(I)}{d}.\end{aligned} \tag{D10}$$

Using $\tilde{\mathcal{F}}_{I/d}$, defined as in Eq. (A9), we then also obtain, for tight measurement frames, the expression

$$\mathrm{tr}(\tilde{\mathcal{F}}_{I/d}) = \frac{d(d^2-1)}{d+1} = d(d-1). \tag{D11}$$

3. Estimators for tight measurement frames

With knowledge of the general structure of the optimal frame corresponding to a tight measurement with elements $\mu_b = w_b\mathbb{P}(\psi_b)$, we can compute explicitly the structure of the corresponding estimator $\hat{f}(b) \equiv \tilde{\mu}_b^{\rm can}$, which gives

$$\tilde{\mu}_b^{\rm can} = \frac{d}{\mathrm{tr}(\mu_b)}\mathcal{F}_{I/d}^{-1}(\mu_b) = (d+1)\mathbb{P}(\psi_b) - I. \tag{D12}$$

4. Lower error bounds for tight measurement frames

We will show here that the L_2 estimation error averaged over unitarily invariant input states, when using any unbiased estimator, is lower bounded by $d^2 + d - 1 - \mathrm{tr}(\rho^2)$, with the inequality saturated for rank-1 tight measurements. This has first been proven in Refs. [19,30]. To estimate the average state-estimation errors, we use the MSE matrix $\mathcal{C}_\rho$ discussed in Eq. (A12). If we assume that the estimators $\tilde{\mu}_b$ do not depend on the input state—as is the case for the canonical estimator but not for the optimal ones—then taking the uniform average with respect to states unitarily equivalent to ρ, we obtain

$$\begin{aligned}\overline{\mathcal{C}}_\rho &= \sum_b \langle \mu_b, I/d\rangle \mathbb{P}(\tilde{\mu}_b) - \int_{\mathbf{U}(d)} dU\,\mathbb{P}(U\rho U^\dagger)\\ &= \mathcal{F}_{I/d}^{-1} - \int_{\mathbf{U}(d)} dU\,\mathbb{P}(U\rho U^\dagger),\end{aligned} \tag{D13}$$

where the integral is taken with respect to the uniform Haar measure in the group of unitary matrices. Taking the trace, we obtain the average error as

$$\overline{\mathcal{E}}_\rho = \mathrm{tr}(\overline{\mathcal{C}}_\rho) = \mathrm{tr}(\mathcal{F}_{I/d}^{-1}) - \mathrm{tr}(\rho^2). \tag{D14}$$

For tight rank-1 measurement frames we know from Eq. (D10) that

$$\mathrm{tr}(\mathcal{F}_{I/d}^{-1}) = d^2 + d - 1. \tag{D15}$$

Let us now show that this is also the lower bound for an arbitrary measurement. From Eq. (A3), we see that for any $\boldsymbol{\mu}$,

$$\mathrm{tr}(\mathcal{F}_{I/d}) = d\sum_b \frac{\mathrm{tr}(\mu_b^2)}{\mathrm{tr}(\mu_b)} \le d\sum_b \mathrm{tr}(\mu_b) = d^2, \quad \text{(D16)}$$

where we have used the inequality $\mathrm{tr}(X^2) \le \mathrm{tr}(X)^2$ for $X \ge 0$, which is saturated if and only if $\mathrm{rank}(X) = 1$. Thus $\mathrm{tr}(\mathcal{F}_{I/d}) \le d^2$ and $\mathrm{tr}(\tilde{\mathcal{F}}_{I/d}) = \mathrm{tr}(\mathcal{F}_{I/d}) - d \le d(d-1)$, with equality for rank-1 measurements. But also, being $\tilde{\mathcal{F}}_{I/d}$ Hermitian and nonsingular as a linear (super)operator, we have

$$\mathrm{tr}(\tilde{\mathcal{F}}_{I/d}) = \sum_{k=1}^{d^2-1}\lambda_k, \quad \mathrm{tr}(\tilde{\mathcal{F}}_{I/d}^{-1}) = \sum_{k=1}^{d^2-1}\frac{1}{\lambda_k}, \quad \text{(D17)}$$

where the λ_k are the eigenvalues of $\tilde{\mathcal{F}}_{I/d}$ and there are d^2-1 terms in the sum because $\mathrm{rank}(\tilde{\mathcal{F}}_{I/d}) = d^2-1$. A direct application of Lagrange's multipliers then allows us to find the minimum value of $\mathrm{tr}(\tilde{\mathcal{F}}_{I/d}^{-1})$ under the constraint of $\lambda_k \ge 0$ and $\mathrm{tr}(\tilde{\mathcal{F}}_{I/d}) = d(d-1)$, which reads

$$\mathrm{tr}(\tilde{\mathcal{F}}_{I/d}^{-1}) \ge \frac{(d^2-1)(d+1)}{d}, \quad \text{(D18)}$$

with equality holding if and only if all the eigenvalues have the same value, i.e., if and only if $\tilde{\mathcal{F}}_{I/d}$ is a multiple of the identity (when acting on the (d^2-1)-dimensional subspace of traceless Hermitian matrices). We conclude that for *any* measurement, we have the lower bound

$$\mathrm{tr}(\mathcal{F}_{I/d}^{-1}) = \frac{1}{d} + \mathrm{tr}(\tilde{\mathcal{F}}_{I/d}^{-1}) \ge d^2 + d - 1, \quad \text{(D19)}$$

with the inequality saturated for tight rank-1 measurements. We have therefore just proved that for any measurement, the average L_2 estimation error when using the canonical estimator is lower bounded as

$$\overline{\mathcal{E}}_\rho \ge d^2 + d - 1 - \mathrm{tr}(\rho^2). \quad \text{(D20)}$$

It is also possible to study the errors corresponding to more general—not necessarily rank-1—tight IC POVMs. This analysis can be found in Ref. [32] and the smallest possible average L_2 estimation error, when the POVM elements have average purity $\wp$, works out to be

$$\overline{\mathcal{E}}_\rho = \frac{(d^2-1)^2}{d^2\wp - d} - \left[\mathrm{tr}(\rho^2) - \frac{1}{d}\right], \quad \text{(D21)}$$

where

$$\wp \equiv \frac{1}{d}\sum_b \frac{\mathrm{tr}(\mu_b^2)}{\mathrm{tr}(\mu_b)} = \frac{\mathrm{tr}(\mathcal{F}_{I/d})}{d^2} \in [1/d, 1]. \quad \text{(D22)}$$

APPENDIX E: ERRORS TO ESTIMATE SINGLE OBSERVABLES

As discussed in Appendix D, to study the estimation errors associated with a given state estimator, it is useful to introduce the MSE matrix $\mathcal{C}_\rho$. Now suppose that we want to estimate the expectation value of some observable $\mathcal{O}$ on a state ρ, using the unbiased estimator $\hat{o}(b) \equiv \langle \mathcal{O}, \hat{f}(b)\rangle = \langle \mathcal{O}, \tilde{\mu}_b\rangle$. The associated variance is

$$\mathrm{Var}[\hat{o}|\rho, \mathcal{O}, \boldsymbol{\mu}, \tilde{\boldsymbol{\mu}}] = \sum_b \langle \mu_b, \rho\rangle\langle \mathcal{O}, \tilde{\mu}_b\rangle^2 - \langle \mathcal{O}, \rho\rangle^2. \quad \text{(E1)}$$

As in the main text, the functional dependence on $\mathcal{O}$, $\boldsymbol{\mu}$, and $\tilde{\boldsymbol{\mu}}$ will be left implicit for notational conciseness. This variance can be expressed via the MSE matrix as

$$\mathrm{Var}[\hat{o}|\rho] = \langle \mathbb{P}(\mathcal{O}), \mathcal{C}_\rho\rangle \equiv \langle \mathcal{O}, \mathcal{C}_\rho(\mathcal{O})\rangle. \quad \text{(E2)}$$

1. Expression for averaged variance

Let us focus on the behavior of the variance when using the canonical state-independent estimator $\tilde{\mu}_b^{\mathrm{can}}$. With this choice, taking the average over input states with purity $P \equiv \mathrm{tr}(\rho^2)$, we have

$$\begin{aligned}\overline{\mathrm{Var}[\hat{o}|P]} &\equiv \int_{\mathbf{U}(d)} dU\, \mathrm{Var}[\hat{o}|U\rho U^\dagger, \mathcal{O}] \\ &= \sum_b \langle \mu_b, I/d\rangle\langle \mathcal{O}, \tilde{\mu}_b^{\mathrm{can}}\rangle^2 - \int_{\mathbf{U}(d)} dU \langle \mathcal{O}, U\rho U^\dagger\rangle^2 \\ &= \langle \mathcal{O}, \mathcal{F}_{I/d}^{-1}(\mathcal{O})\rangle - \beta, \end{aligned} \quad \text{(E3)}$$

where β is the expectation value of $\langle \mathcal{O}, \rho\rangle^2$ over states with purity P. This quantity is computed using the known formulas to integrate polynomials in the components of unitary matrices over the uniform Haar measure [45] and equals

$$\beta = \frac{\mathrm{tr}(\mathcal{O})^2}{d^2} + \frac{dP-1}{d^2-1}V, \quad \text{(E4)}$$

where $V \equiv \langle \mathcal{O}^2\rangle - \langle \mathcal{O}\rangle^2$ is the variance of the observable computed on the maximally mixed state, with $\langle \mathcal{O}\rangle \equiv \mathrm{tr}(\mathcal{O})/d$ and $\langle \mathcal{O}^2\rangle \equiv \mathrm{tr}(\mathcal{O}^2)/d$. Note that the averaged variance depends on P but not on the specific choice of ρ. Let us now focus on the term $\langle \mathcal{O}, \mathcal{F}_{I/d}^{-1}(\mathcal{O})\rangle$, which is the one that depends on the POVM. Using the decomposition in Eq. (A9), for $\mathcal{F}_{I/d}$, we have

$$\langle \mathcal{O}, \mathcal{F}_{I/d}^{-1}(\mathcal{O})\rangle = \frac{\mathrm{tr}(\mathcal{O})^2}{d^2} + \langle \mathcal{O}, \tilde{\mathcal{F}}_{I/d}^{-1}(\mathcal{O})\rangle. \quad \text{(E5)}$$

Putting together Eqs. (E3)–(E5), we obtain the general expression for the averaged variance corresponding to the canonical estimator:

$$\overline{\mathrm{Var}[\hat{o}|P]} = \langle \mathcal{O}, \tilde{\mathcal{F}}_{I/d}^{-1}(\mathcal{O})\rangle - \frac{dP-1}{d^2-1}V. \tag{E6}$$

2. Bounds for the averaged variance

The first term can be bounded in terms of the eigenvalues of $\tilde{\mathcal{F}}_{I/d}^{-1}$, as

$$\frac{Vd}{\lambda_{\max}(\tilde{\mathcal{F}}_{I/d})} \le \langle \mathcal{O}, \tilde{\mathcal{F}}_{I/d}^{-1}(\mathcal{O})\rangle \le \frac{Vd}{\lambda_{\min}(\tilde{\mathcal{F}}_{I/d})}, \tag{E7}$$

where $\lambda_{\min}(\tilde{\mathcal{F}}_{I/d}), \lambda_{\max}(\tilde{\mathcal{F}}_{I/d})$ are the smallest and largest eigenvalues of $\tilde{\mathcal{F}}_{I/d}$ (which is positive definite as an operator whenever $\boldsymbol{\mu}$ is IC). This bound is obtained by observing that $\tilde{\mathcal{F}}_{I/d}$, and therefore also $\tilde{\mathcal{F}}_{I/d}^{-1}$, is a (Hermitian) linear operator acting on the space of $\mathrm{Herm}(\mathbb{C}^d)$ spanned by traceless Hermitian operators. In general, if $H \in \mathrm{Lin}(V)$ is a Hermitian operator acting on some vector space V, with support $W \equiv \mathrm{supp}(H) \subseteq V$, then for any $v \in W$, we have

$$\lambda_{\min}(H)\|v_W\|^2 \le \langle v, Hv\rangle \le \lambda_{\max}(H)\|v_W\|^2, \tag{E8}$$

where v_W is the projection of v on W and $\lambda_{\min}(H), \lambda_{\max}(H)$ are the smallest and largest *nonzero* eigenvalues of H. Applying this with $H = \tilde{\mathcal{F}}_{I/d}^{-1}$ and $v = \mathcal{O}$, we obtain Eq. (E7), because the orthogonal projection of $\mathcal{O}$ on the subspace of traceless Hermitian operators is $\mathcal{O} - \mathrm{tr}(\mathcal{O})I/d$ and $\|\mathcal{O} - \mathrm{tr}(\mathcal{O})I/d\|^2 = Vd$.

3. General bounds for worst-case variance

From Eq. (E7), we obtain a general upper bound for the variance in the form

$$\overline{\mathrm{Var}[\hat{o}|P]} \le Vd\left[\frac{1}{\lambda_{\min}(\tilde{\mathcal{F}}_{I/d})} - \frac{P-1/d}{d^2-1}\right]. \tag{E9}$$

This upper bound still depends on $\mathcal{O}$ via V but this dependence is intrinsic to the observable—it is the average variance that one would obtain estimating $\langle\mathcal{O}\rangle$ from projective measurements in its eigenbasis and it is thus the absolute lower bound achievable for $\overline{\mathrm{Var}[\hat{o}|P]}$. We can thus interpret Eq. (E9) as the average variance corresponding to the hardest-to-estimate observable. We will now attempt to provide more precise bounds for this quantity in terms of general symmetry properties of the POVM. In particular, remembering that a POVM is tight *if and only if* its frame superoperator satisfies $(d^2-1)\mathrm{tr}(\tilde{\mathcal{F}}_{I/d}^2) = \mathrm{tr}(\tilde{\mathcal{F}}_{I/d})^2$, a natural choice is to explore the set of IC POVMs under the constraints $\mathrm{tr}(\tilde{\mathcal{F}}_{I/d}) = a$ and $\mathrm{tr}(\tilde{\mathcal{F}}_{I/d}^2) = b$ for some given $a, b > 0$.

We then analyze what is the smallest possible value of the average variance for the hardest-to-estimate observable, as a function of a and b. More formally, we therefore consider the following question: *What is the POVM that gives the smallest $1/\lambda_{min}(\tilde{\mathcal{F}}_{I/d})$, under the above constraints?* This is equivalent to asking for the largest possible $\lambda_{\min}(\tilde{\mathcal{F}}_{I/d})$ under the same constraints. In turn, focusing on the eigenvalues, this question is equivalent to: *Within the set of tuples $\lambda_1, \ldots, \lambda_{d^2-1} > 0$ such that $\sum_k \lambda_k = a$ and $\sum_k \lambda_k^2 = b$, what is the largest possible value of* $\min(\lambda_k)$? For consistency, the coefficients a and b need to satisfy $0 < b \le a^2 \le b(d^2-1)$, which follows directly from the AM–GM inequality.

Solving this optimization problem is made somewhat more difficult by the cost function $\min(\lambda_1, \ldots, \lambda_{d^2-1})$ being nondifferentiable. We can nonetheless convert it into a differentiable cost by introducing additional slack variables. For notational conciseness, let us define $m \equiv d^2-1$. Our problem can be restated as that of maximizing λ_1, with respect to the $2m-1$ variables $\lambda_1, \ldots, \lambda_m, s_2, \ldots, s_m$, subject to the constraints

$$\begin{gathered}\lambda_k \ge 0, \quad \sum_{k=1}^m \lambda_k = a, \quad \sum_k \lambda_k^2 = b, \\ \lambda_1 + s_k^2 = \lambda_k, \quad \forall k = 2, \ldots, m.\end{gathered} \tag{E10}$$

The constraints $\lambda_1 + s_k^2 = \lambda_k$ are introduced to enforce $\lambda_1 \le \lambda_k$, and thus ensure that the solution to this problem corresponds to the solution of the original one. Using the method of Lagrange multipliers [64], define the Lagrangian function

$$\begin{aligned}L = {} & \lambda_1 + \alpha\left(\sum_{k=1}^m \lambda_k - a\right) + \beta\left(\sum_{k=1}^m \lambda_k^2 - b\right) \\ & + \sum_{k=2}^m \gamma_k(\lambda_k - \lambda_1 - s_k^2).\end{aligned} \tag{E11}$$

Imposing $\nabla L = 0$ gives the conditions

$$\begin{aligned}&1 + \alpha + 2\beta\lambda_1 - \sum_{k=2}^m \gamma_k = 0, \\ &\alpha + 2\beta\lambda_k + \gamma_k = 0, \\ &\forall k \ge 2, \gamma_k s_k = 0, \quad \forall k \ge 2.\end{aligned} \tag{E12}$$

We can explore the different sets of solutions compatible with these constraints by taking into account the number of coefficients s_k that equal 0:

(1) Suppose that $s_2, \ldots, s_m \neq 0$. This implies that $\gamma_2 = \cdots = \gamma_m = 0$, which in turns implies that $\lambda_1 < \lambda_2$ and $\lambda_2 = \cdots = \lambda_m$. The constraints in terms of a

and b simplify to $\lambda_1 + (m-1)\lambda_2 = a$ and $\lambda_1^2 + (m-1)\lambda_2^2 = b$. These two equations give two solutions for λ_1, one of which is unfeasible because it corresponds to $\lambda_1 > \lambda_2$. The other one is feasible and is a possible solution:

$$\lambda_1 = \frac{a}{m} - \frac{\sqrt{(m-1)(bm-a^2)}}{m}. \tag{E13}$$

(2) More generally, suppose that $s_2 = \cdots = s_\ell = 0$ and $s_{\ell+1}, \ldots, s_m \neq 0$ for some $2 \leq \ell \leq m$. This implies that $\gamma_{\ell+1} = \cdots = \gamma_m = 0$, which in turn implies that $\lambda_{\ell+1} = \cdots = \lambda_m$. Furthermore, $s_2 = \cdots = s_\ell = 0$ means that $\lambda_1 = \cdots = \lambda_\ell$. We therefore reduce again to a situation with only two distinct values for the coefficients λ_k and the constraints again simplify to $\ell\lambda_1 + (m-\ell)\lambda_m = a$ and $\ell\lambda_1^2 + (m-\ell)\lambda_m^2 = b$. Solving this and keeping the solution consistent with the constraints gives

$$\lambda_1 = \frac{a}{m} - \frac{\sqrt{\ell(m-\ell)(bm-a^2)}}{m\ell}. \tag{E14}$$

The above covers all possible scenarios, up to a permutation of the vanishing coefficients s_k (any such permutation does not affect the resulting solution for λ_1 due to the problem symmetry). The final solution is thus the maximum of Eq. (E14) for $\ell = 1, \ldots, m$. Observing that $\sqrt{\ell(m-\ell)}/\ell$ decreases monotonically with $\ell = 1, \ldots, m$, we conclude that the largest λ_1 is obtained when $\ell = m$. This case, however, corresponds to having $\lambda_1 = \cdots = \lambda_m = a/m$, which is only compatible with the constraints if $bm = a^2$. The more general scenario is obtained for $\ell = m-1$, corresponding to having $\lambda_1 = \cdots = \lambda_{m-1} < \lambda_m$, and is possible for all $a, b > 0$ with $a^2 \leq bm$.

To summarize, we have concluded that the largest $\min(\lambda_1, \ldots, \lambda_m)$, $m \equiv d^2 - 1$ compatible with given values of $a = \text{tr}(\tilde{\mathcal{F}}_{I/d})$ and $b = \text{tr}(\tilde{\mathcal{F}}_{I/d}^2)$ is

$$\lambda_1^* \equiv \frac{a}{m} - \frac{\sqrt{(m-1)(bm-a^2)}}{m(m-1)}, \tag{E15}$$

which, in the special case where $bm = a^2$, corresponding to $\tilde{\mathcal{F}}_{I/d}$ being a multiple of the identity and thus $\boldsymbol{\mu}$ being a tight measurement frame, reduces to $\lambda_1^* = a/m$. Reformulating this in terms of the variance, we have concluded that, if $\boldsymbol{\mu}$ is such that $\text{tr}(\tilde{\mathcal{F}}) = a$ and $\text{tr}(\tilde{\mathcal{F}}^2) = b$, then

$$\max_{\mathcal{O}} \frac{\overline{\text{Var}[\hat{o}|P, \mathcal{O}]}}{Vd} \geq \frac{1}{\lambda_1^*} - \frac{P - 1/d}{d^2 - 1}, \tag{E16}$$

with the inequality saturated by some POVM the canonical estimator of which gives equal average variance for all observables (in some orthonormal basis of Hermitian operators) but one. Furthermore, for tight measurements, $\tilde{\mathcal{F}}_{I/d}$ is a multiple of the identity, $\text{tr}(\tilde{\mathcal{F}}_{I/d}) = d(d-1)$ as per Eq. (D11), $\lambda_1^* = \text{tr}(\tilde{\mathcal{F}}_{I/d})/(d^2-1)$, and thus

$$\overline{\text{Var}[\hat{o}|P, \mathcal{O}]} = Vd\left(\frac{d^2 + d - 1 - P}{d^2 - 1}\right), \tag{E17}$$

where we have made the dependence of $\mathcal{O}$ explicit to point out that all observables give the same expression for the variance. We recognize in particular the term $d^2 + d - 1 - P$, which is the optimal state-estimation L_2 error discussed in Appendix D. From Eq. (E17), we see that for tight rank-1 measurements, the asymptotic growth of the variance with the state dimension can be canceled out by the choice of observable, since it only depends on the factor Vd. For example, for any observable that is a projection onto a pure state, $\mathcal{O} = \mathbb{P}_\psi$ for some $|\psi\rangle$, we have $\text{tr}(\mathcal{O}^2) = \text{tr}(\mathcal{O}) = 1$, $Vd = (d-1)/d$, and therefore

$$\overline{\text{Var}[\hat{o}|P, \mathbb{P}_\psi]} = \frac{d^2 + d - 1 - P}{d(d+1)}, \tag{E18}$$

where we have now included the explicit dependence of the variance on the observable $\mathcal{O} = \mathbb{P}_\psi$. This gives $\overline{\text{Var}[\hat{o}|P, \mathbb{P}_\psi]} \to 1$ for large d, regardless of $|\psi\rangle$, meaning that the estimation errors to estimate such observables do not increase with the dimension of the space. Similarly, for normalized observables $\mathcal{O}_N$ with $\text{tr}(\mathcal{O}_N) = 0$ and $\text{tr}(\mathcal{O}_N^2) = 1$, we have $Vd = 1$ and thus

$$\overline{\text{Var}[\hat{o}|P, \mathcal{O}_N]} = \frac{d^2 + d - 1 - P}{d^2 - 1}. \tag{E19}$$

As a counterexample, if one studies the variance associated with estimating $\mathcal{O}_n = \sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n}$ defined as a product of n Pauli matrices, then $\text{tr}(\mathcal{O}_n) = 0$, $\text{tr}(\mathcal{O}_n^2) = 2^n$, $Vd = d = 2^n$, and

$$\overline{\text{Var}[\hat{o}|P, \mathcal{O}_n]} = d\frac{d^2 + d - 1 - P}{d^2 - 1} = O(d), \quad d \to \infty \tag{E20}$$

meaning that the average variance increases linearly with d.

APPENDIX F: AVERAGED ERROR FOR THE ESTIMATION OF OBSERVABLES

We focus in this appendix on the variance averaged over both the input states at fixed purity and over unitarily equivalent random target observables.

We have already derived in Eq. (E3) the expression for the variance averaged over unitarily equivalent input states, for any given fixed observable $\mathcal{O}$. Also performing an average over unitarily equivalent observables, we obtain

$$\overline{\overline{\overline{\mathrm{Var}[\hat{o}]}}} \equiv \int_{\mathbf{U}(d)} dU\, \overline{\mathrm{Var}[\hat{o}|U\mathcal{O}U^\dagger]} \equiv \alpha' - \beta, \tag{F1}$$

where

$$\begin{aligned}\alpha' &= \frac{1}{d(d^2-1)} \sum_b \mathrm{tr}\mu_b \left\{ (\mathrm{tr}\mathcal{O})^2 \left[(\mathrm{tr}\tilde{\mu}_b)^2 - \frac{\mathrm{tr}\tilde{\mu}_b^2}{d} \right] \right. \\ &\quad \left. + \,\mathrm{tr}\mathcal{O}^2 \left[\mathrm{tr}\tilde{\mu}_b^2 - \frac{(\mathrm{tr}\tilde{\mu}_b)^2}{d} \right] \right\} \\ &= \frac{\mathrm{tr}(\mathcal{F}_{I/d}^{-1}) \left(d\mathrm{tr}\mathcal{O}^2 - (\mathrm{Tr}\mathcal{O})^2 \right) + \left(d(\mathrm{tr}\mathcal{O})^2 - \mathrm{tr}\mathcal{O}^2 \right)}{d(d^2-1)} \\ &= V\frac{d\mathrm{tr}(\mathcal{F}_{I/d}^{-1}) - 1}{d^2-1} + \frac{(\mathrm{tr}\mathcal{O})^2}{d^2}\end{aligned} \tag{F2}$$

and where β, given in Eq. (E4), does not change in performing this second average since it only depends on $\mathcal{O}$ via $\mathrm{tr}(\mathcal{O})$ and $\mathrm{tr}(\mathcal{O}^2)$. These expressions further simplify to

$$\begin{aligned}\overline{\overline{\overline{\mathrm{Var}[\hat{o}|\mathcal{O}]}}} &= \frac{Vd}{d^2-1} \left(\mathrm{tr}(\mathcal{F}_{I/d}^{-1}) - P \right) \\ &= \frac{Vd}{d^2-1} \left(\mathrm{tr}(\tilde{\mathcal{F}}_{I/d}^{-1}) - P + \frac{1}{d} \right),\end{aligned} \tag{F3}$$

using the expression for β given in Eq. (E4). It is instructive to compare this equation with the results of Appendix E and with, e.g., the upper bound of Eq. (E9). Using the lower bound on the trace given by Eq. (D18), we obtain

$$\overline{\overline{\overline{\mathrm{Var}[\hat{o}|\mathcal{O}]}}} \geq \frac{Vd\left(d^2+d-1-P\right)}{d^2-1}, \tag{F4}$$

with equality *if and only if* $\boldsymbol{\mu}$ is a tight rank-1 measurement.

To provide some examples, let us consider observables of the form $\mathcal{O} = \mathbb{P}_\psi$ for some $|\psi\rangle$, for which we have $Vd = (d-1)/d$ and thus, from Eq. (F3),

$$\overline{\overline{\overline{\mathrm{Var}[\hat{o}|\mathbb{P}_\psi]}}} = \frac{1}{d(d+1)} \left[\mathrm{tr}\left(\tilde{\mathcal{F}}_{I/d}^{-1} \right) - P + \frac{1}{d} \right]. \tag{F5}$$

Equation (F4) now reads

$$\frac{2}{3} \leq \min_{\boldsymbol{\mu}} \left\{ \overline{\overline{\overline{\mathrm{Var}[\hat{o}|\mathbb{P}_\psi]}}} \right\} = 1 - \frac{1+P}{d(d+1)} \leq 1, \tag{F6}$$

which is an increasing function of d but is bounded from above, as expected for this type of observable.

Similarly, for Pauli observables of the form $\mathcal{O}_n = \sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n}$, acting on n qubits ($d = 2^n$), we have $Vd = d$ and therefore Eq. (F3) becomes

$$\overline{\overline{\overline{\mathrm{Var}[\hat{o}|\mathcal{O}_n]}}} = \frac{d}{d^2-1} \left[\mathrm{tr}\left(\tilde{\mathcal{F}}_{I/d}^{-1} \right) - P + \frac{1}{d} \right]. \tag{F7}$$

As before, this quantity is bounded from below by

$$\overline{\overline{\overline{\mathrm{Var}[\hat{o}|\mathcal{O}_n]}}} \geq \min_{\boldsymbol{\mu}} \left\{ \overline{\overline{\overline{\mathrm{Var}[\hat{o}|\mathcal{O}_n]}}} \right\} = d + 1 - \frac{dP-1}{d^2-1}, \tag{F8}$$

with equality for tight rank-1 IC POVMs.

In contrast to projectorlike observables, this lower bound is not bounded from above by a constant that is independent on the dimension d and, indeed, one has that, $\forall \rho$,

$$\min_{\boldsymbol{\mu}} \left\{ \overline{\overline{\overline{\mathrm{Var}[\hat{o}|\mathcal{O}_n]}}} \right\} \sim \mathcal{O}(d), \quad d \to \infty. \tag{F9}$$

Comparing the average variances of Eqs. (F5) and (F7), we can also write, for general measurement frames $\boldsymbol{\mu}$,

$$\overline{\overline{\overline{\mathrm{Var}[\hat{o}|\mathcal{O}_n]}}} = \frac{d^2}{d-1} \overline{\overline{\overline{\mathrm{Var}[\hat{o}|\mathbb{P}_\psi]}}}. \tag{F10}$$

APPENDIX G: OPTIMAL DUAL FRAME AND RESCALED FRAMES

As discussed in Sec. II, the nonrescaled canonical dual frame $\mathcal{F} = \sum_b \mathbb{P}(\mu_b)$ is not, in general, the optimal choice of unbiased estimator. Nonetheless, it can be interesting to note that we can see the optimal dual frame as corresponding to the canonical dual frame computed with respect to a rescaled frame. More precisely, the estimator $\tilde{\mu}^{(\rho)}$, introduced in Sec. III, can be derived considering the rescaled frame with elements

$$\mu_b^N \equiv \frac{\mu_b}{\sqrt{\langle \mu_b, \rho \rangle}}. \tag{G1}$$

The set of operators $\{\mu_b^N\}_b$ is a frame if and only if $\{\mu_b\}_b$ is also a frame and the nonrescaled frame operator corresponding to $\{\mu_b^N\}_b$ is precisely the rescaled frame operator corresponding to $\{\mu_b\}_b$.

1. Frame operators for arbitrary rescalings

We briefly show in this section the consequences of considering frames of operators defined in terms of rescaled POVM elements, for arbitrary rescalings. In particular, we provide the explicit expansion of a generic state obtained using this formalism, and the associated unbiased estimators corresponding to each choice of rescaling. These observations are not pivotal to the main results of the paper but are presented here for the sake of completeness.

2. Definition of general rescaled measurement frames

Consider a *rescaled measurement frames* with elements $\mu_b/\sqrt{\alpha_b}$ for some set of positive real coefficients α_b. The associated nonrescaled frame operator is

$$\mathcal{F}_\alpha \equiv \sum_b \frac{\mathbb{P}(\mu_b)}{\alpha_b}. \tag{G2}$$

If $\mu_b^{(\alpha)\star} = \mathcal{F}_\alpha^{-1}(\mu_b/\sqrt{\alpha_b})$ denotes the corresponding (nonrescaled) canonical dual frame, the associated decomposition of a state ρ reads

$$\rho = \sum_b \frac{\langle \mu_b, \rho\rangle}{\sqrt{\alpha_b}} \mu_b^{(\alpha)\star} = \sum_b \frac{\langle \mu_b, \rho\rangle}{\alpha_b} \mathcal{F}_\alpha^{-1}(\mu_b). \tag{G3}$$

Recognizing that $\langle \mu_b, \rho\rangle$ is a probability, we then define an unbiased estimator for the state as

$$\hat{f}(b) \equiv \frac{1}{\sqrt{\alpha_b}} \mu_b^{(\alpha)\star}, \tag{G4}$$

which thus satisfies $\mathbb{E}[\hat{f}\,] = \rho$. Note that, in general, $\mu_b^{(\alpha)\star} \neq \sqrt{\alpha_b}\mu_b^\star$ and thus different frame scalings provide nontrivially different canonical estimators, albeit Eq. (G3) means that each set of operators $\{(1/\sqrt{\alpha_b})\mu_b^{(\alpha)\star}\}_b$ is a generally noncanonical valid dual frame of the nonrescaled measurement frame $\{\mu_b\}_b$.

3. Average error with rescaled frames

The main usefulness of considering rescaled measurement frames is that the associated average L_2 error now reads

$$\mathbb{E}\|\hat{f} - \rho\|_2^2 = \mathbb{E}\,\mathrm{tr}(\hat{f}^{\,2}) - \mathrm{tr}(\rho^2), \tag{G5}$$

where

$$\mathbb{E}\,\mathrm{tr}(\hat{f}^{\,2}) \equiv \sum_b \frac{\langle \mu_b, \rho\rangle}{\alpha_b} \mathrm{tr}((\mu_b^{(\alpha)\star})^2). \tag{G6}$$

Therefore, if we rescale the operators via $\alpha_b = \langle \mu_b, \rho\rangle$, we can write

$$\mathbb{E}\,\mathrm{tr}(\hat{f}^{\,2}) = \sum_b \mathrm{tr}((\mu_b^{(\alpha)\star})^2) = \mathrm{tr}(\mathcal{F}_\alpha^{-1}). \tag{G7}$$

This simplifies the problem to searching for the measurement $\boldsymbol{\mu}$ that minimizes $\mathrm{tr}(\mathcal{F}_\alpha^{-1})$ using a given set of coefficients $\{\alpha_b\}_b$.

APPENDIX H: TOY EXAMPLES

In this appendix, we present a number of toy examples to better illustrate how the techniques set forth in the main text would be used in practice.

1. Projective measurement

Consider a simple single-qubit projective measurement: $\mu_0 \equiv \mathbb{P}_0$ and $\mu \equiv \mathbb{P}_1$. The corresponding canonical frame superoperator, as per Eq. (15), is

$$\mathcal{F}_{I/2} = 2[\mathbb{P}(\mathbb{P}_0) + \mathbb{P}(\mathbb{P}_1)] = 2\begin{pmatrix}1&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&1\end{pmatrix}, \tag{H1}$$

where we have represented the superoperator in the standard vectorized notation. This $\mathcal{F}_{I/2}$ is clearly singular, which corresponds to the POVM not being informationally complete. The associated estimator is not well defined, correspondingly to the POVM not being a frame. Nonetheless, the general property $\mathcal{F}_{I/2}(I/2) = I$, as per Eq. (A4), still holds, as directly verified by observing that upon vectorization, the identity operator I becomes $\mathrm{vec}(I) = (1,0,0,1)^T$. Similarly, the decomposition given in Eq. (A9) applies: we can write

$$\mathcal{F}_{I/2} = 2\mathbb{P}(I/\sqrt{2}) + \tilde{\mathcal{F}}_{I/2}, \quad \tilde{\mathcal{F}}_{I/2} = \begin{pmatrix}1&0&0&-1\\0&0&0&0\\0&0&0&0\\-1&0&0&1\end{pmatrix} \tag{H2}$$

and we can directly verify that $\mathbb{P}(I/\sqrt{2})$ and $\tilde{\mathcal{F}}_{I/2}$ act on orthogonal spaces and that $\tilde{\mathcal{F}}_{I/2}$ acts on the space of traceless operators, as $\tilde{\mathcal{F}}_{I/2}(Z) = 2Z$, where $Z \equiv \mathbb{P}_0 - \mathbb{P}_1$.

2. Simple non-IC POVM

Consider the following single-qubit POVM:

$$\mu_1 = \frac{1}{2}\mathbb{P}_0, \quad \mu_2 = \frac{1}{2}\mathbb{P}_1, \quad \mu_3 = \frac{1}{2}\mathbb{P}_+, \quad \mu_4 = \frac{1}{2}\mathbb{P}_-. \tag{H3}$$

Note that in vectorized notation we have $\mu_1 = \frac{1}{2}(1,0,0,0)$, $\mu_3 = \frac{1}{4}(1,1,1,1)^T$, etc. The corresponding canonical frame superoperator is then

$$\mathcal{F}_{I/2} = \frac{1}{2}\begin{pmatrix}3&0&0&1\\0&1&1&0\\0&1&1&0\\1&0&0&3\end{pmatrix}. \tag{H4}$$

This has eigenvalues $\{2,1,1,0\}$ and is therefore again singular, consistently with the POVM again not being informationally complete. Note how the number of nonzero eigenvalues reflects the dimension of the span of the POVM, which is in this case larger than for the simple projective case. The eigenvectors corresponding to the nonzero eigenvalues are $(1,0,0,1)^T$, $(1,0,0,-1)^T$, and

$(0,1,1,0)^T$, respectively—which, upon devectorizing, correspond to the Pauli operators I, Z, and X. This is again consistent with the general statement that $\mathcal{F}_{I/2}(I/2) = I$. Note that in this example, by defining the frame superoperator directly via Eq. (3) and thus not introducing the rescaling factors used in Eq. (15), the frame operator would have been $\mathcal{F}_{I/2}/4$ instead.

3. Example of IC POVM

As an example of a single-qubit IC POVM, consider

$$\mu_1 = \frac{1}{3}\mathbb{P}_0, \quad \mu_2 = \frac{1}{3}\mathbb{P}_+, \quad \mu_3 \equiv \frac{1}{3}\mathbb{P}_R,$$
$$\mu_4 = I - \mu_1 - \mu_2 - \mu_3, \tag{H5}$$

with $\mathbb{P}_R = |R\rangle\langle R|$ and $|R\rangle = (|0\rangle + i\,|1\rangle)/\sqrt{2}$.

4. Frame operators and canonical estimator

This POVM is informationally complete and its corresponding frame operator is

$$\mathcal{F}_{I/2} = \frac{1}{18}\begin{pmatrix} 22 & 1+i & 1-i & 14 \\ 1-i & 8 & -2i & -1+i \\ 1+i & 2i & 8 & -1-i \\ 14 & -1-i & -1+i & 22 \end{pmatrix}, \tag{H6}$$

the eigenvalues of which are 2, 2/3, 1/3, and 1/3. The actual matrix representation of the frame operator depends on the choice of operator basis. The above representation corresponds to a standard choice of operatorial basis with elements $\{|i\rangle\langle j|\}_{i,j}$. Another possibility is to represent the operator in a basis of Hermitian operators, such as $\{I/\sqrt{2}, X/\sqrt{2}, Y/\sqrt{2}, Z/\sqrt{2}\}$. With this choice, we obtain instead

$$\mathcal{F}_{I/2} = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 4/9 & 1/9 & 1/9 \\ 0 & 1/9 & 4/9 & 1/9 \\ 0 & 1/9 & 1/9 & 4/9 \end{pmatrix}, \tag{H7}$$

which makes some of the underlying structure more transparent. As always, the first eigenvalue corresponds to the I eigenvector, i.e., the general property $\mathcal{F}_{I/2}(I) = 2I$. The remaining eigenvalues are eigenvalues of $\tilde{\mathcal{F}}_{I/2}$. In particular, the eigenvectors corresponding to the eigenvalues 2/3, 1/3, and 1/3, and thus also the eigenvectors of $\tilde{\mathcal{F}}_{I/2}$, are the operators $X+Y+Z$, $X-Z$, and $X+Z-2Y$, respectively. We can now compute the canonical-estimator elements $\tilde{\mu}_b^{\text{can}}$, which work out to be

$$\tilde{\mu}_1^{\text{can}} = \frac{1}{2}(I - X - Y + 5Z), \quad \tilde{\mu}_2^{\text{can}} = \frac{1}{2}(I + 5X - Y - Z),$$
$$\tilde{\mu}_3^{\text{can}} = \frac{1}{2}(I - X + 5Y - Z), \quad \tilde{\mu}_4^{\text{can}} = \frac{1}{2}(I - X - Y - Z). \tag{H8}$$

These then provide unbiased estimators to estimate arbitrary observables. For example, if the target observable is the Pauli matrix, $\mathcal{O} = Z$, then the observable estimator would be $\hat{o}$ such that $\hat{o}(b) = \langle \mathcal{O}, \tilde{\mu}_b^{\text{can}}\rangle$, the values of which are

$$\hat{o}(1) = 5, \quad \hat{o}(2) = \hat{o}(3) = \hat{o}(4) = -1. \tag{H9}$$

Because the POVM is minimal — which means that the number of outcomes equals d^2 — the POVM elements are also linearly independent. This implies that there is a single possible choice of dual frame and therefore a single choice of estimator. In other words, performing similar calculations using the nonrescaled frame operators will produce the same exact estimators in this case.

5. Assessment of estimator variances

We can then use Eq. (21) to compute the variances in different scenarios. For example, if $\rho = \mathbb{P}_0$ and $\mathcal{O} = Z$, then

$$\begin{aligned}\text{Var}[\hat{o}|\mathbb{P}_0, Z] &= \langle \mathcal{O}, \mathcal{C}_{\mathbb{P}_0}(\mathcal{O})\rangle \\ &= \left[5^2\frac{1}{3} + (-1)^2\left(1 - \frac{1}{3}\right)\right] - 1 = 8,\end{aligned} \tag{H10}$$

where $\mathcal{C}_{\mathbb{P}_0}$ is the MSE matrix, as defined in Eq. (10), computed using the canonical estimator $\tilde{\mu}_b^{\text{can}}$. If, on the other hand, we have $\rho = \mathbb{P}_1$, then

$$\text{Var}[\hat{o}|\mathbb{P}_1, X] = \text{Var}[\hat{o}|\mathbb{P}_1, Y] = 5 \tag{H11}$$

but $\text{Var}[\hat{o}|\mathbb{P}_1, Z] = 0$, consistently with the first outcome being the only one that gives $\hat{o}(1) = 5$ and this outcome having zero probability due to $\langle \mu_1, \mathbb{P}_1\rangle = 0$. We can also gain a more general understanding of how the variance has changed with the input state using the A operator defined in Eq. (29). For example, if $\mathcal{O} = X$, this equals

$$A = \sum_b \langle X, \tilde{\mu}_b^{\text{can}}\rangle \mu_b = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}. \tag{H12}$$

This operator has eigenvalues 9 and 1, which immediately tells us that $1 \le \mathbb{E}[\hat{o}^2] \le 9$ and thus $0 \le \text{Var}[\hat{o}|\rho, X] \le 9$. In particular, the eigenvector of A corresponding to the eigenvalue $+1$ is $(1,-1)^T$, which tells us that the state $\rho = \mathbb{P}_-$ is such that $\langle A, \mathbb{P}_-\rangle = 1$ and because $\langle X, \mathbb{P}_-\rangle^2 = 1$, we conclude that $\text{Var}[\hat{o}|\mathbb{P}_-, X] = 0$.

6. Bounds on the average variance

To work with the averaged variance, we can use Eqs. (22), (25), and (E6), which immediately tell us that the possible values of the averaged variance depend on the eigenvalues of $\tilde{\mathcal{F}}_{I/2}$. As shown above, in the case we are studying, these eigenvalues are 3, 3, and 3/2. Sticking to pure states for simplicity, we thus obtain the general bounds for the averaged variance in this example, as

$$2.67 \simeq \frac{8}{3} \leq \frac{\overline{\mathrm{Var}[\hat{o}|\mathcal{O}]}}{V} \leq \frac{17}{3} \simeq 5.67. \tag{H13}$$

7. Consistency with general bounds on average variance

Finally, we can also attempt to directly verify the consistency of the general bounds provided in Eq. (26). Working out explicitly the various terms for our canonical frame operator, we find $a = 10/3$, $b = 14/3$, and

$$\lambda_1^* = \frac{10 - \sqrt{13}}{9} \simeq 0.71, \tag{H14}$$

and thus the bound reads, considering pure states for simplicity,

$$\max_{\mathcal{O}} \frac{\overline{\mathrm{Var}[\hat{o}|\mathcal{O}]}}{V} \geq \frac{2}{\lambda_1^*} - \frac{1}{3} \simeq 2.48. \tag{H15}$$

This is consistent with Eq. (H13), because $2.48 < 17/3$. This tells us that there are better choices of measurement, which produce frame superoperators compatible with the given values of a and b, that give a much better worst-case average variance.

[1] M. Paris and J. Řeháček, eds., *Quantum State Estimation*, Lecture Notes in Physics, Vol. 649 (Springer-Verlag, Berlin, 2004).
[2] Y. S. Teo, *Introduction to Quantum-State Estimation* (World Scientific, Singapore, 2015).
[3] G. M. D'Ariano, M. G. Paris, and M. F. Sacchi, Quantum tomography, Adv. Imaging Electron Phys. **128**, 206 (2003).
[4] D. F. James, P. G. Kwiat, W. J. Munro, and A. G. White, Measurement of qubits, Phys. Rev. A **64**, 052312 (2001).
[5] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, Matrix product state representations, Quantum Inf. Comput. **7**, 401 (2007).
[6] A. Lowe and A. Nayak, Lower bounds for learning quantum states with single-copy measurements, ArXiv:2207.14438 (2022).
[7] M. Guţă, J. Kahn, R. Kueng, and J. A. Tropp, Fast state tomography with optimal error bounds, J. Phys. A: Math. Theor. **53**, 204001 (2020).
[8] R. Kueng, H. Rauhut, and U. Terstiege, Low rank matrix recovery from rank one measurements, ArXiv:1410.6913 (2014).
[9] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, in *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing* (IEEE, Cambridge, MA (USA), 2016), p. 913.
[10] J. Morris, V. Saggio, A. Gočanin, and B. Dakić, Quantum verification and estimation with few copies, Adv. Quantum Technol. **5**, 2100118 (2022).
[11] V. Gebhart, R. Santagati, A. A. Gentile, E. M. Gauger, D. Craig, N. Ares, L. Banchi, F. Marquardt, L. Pezzè, and C. Bonato, Learning quantum systems, Nat. Rev. Phys. **5**, 141 (2023).
[12] S. Aaronson, in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing* (Association for Computing Machinery, New York, 2018), p. 325.
[13] H.-Y. Huang and R. Kueng, Predicting features of quantum systems from very few measurements, ArXiv:1908.08909 (2019).
[14] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nat. Phys. **16**, 1050 (2020).
[15] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, The randomized measurement toolbox, Nat. Rev. Phys. **5**, 9 (2023).
[16] M. Kliesch and I. Roth, Theory of quantum system certification, PRX Quantum **2**, 010201 (2021).
[17] H. C. Nguyen, J. L. Bönsel, J. Steinberg, and O. Gühne, Optimizing shadow tomography with generalized measurements, Phys. Rev. Lett. **129**, 220502 (2022).
[18] A. Acharya, S. Saha, and A. M. Sengupta, Shadow tomography based on informationally complete positive operator-valued measure, Phys. Rev. A **104**, 052418 (2021).
[19] A. J. Scott, Tight informationally complete quantum measurements, J. Phys. A: Math. General **39**, 13507 (2006).
[20] G. M. D'Ariano and P. Perinotti, Optimal data processing for quantum measurements, Phys. Rev. Lett. **98**, 020403 (2007).
[21] P. Perinotti and G. M. D'Ariano, Optimal estimation of ensemble averages from a quantum measurement, ArXiv:quant-ph/0701231 (2007).
[22] C. Ferrie and J. Emerson, Frame representations of quantum mechanics and the necessity of negativity in quasi-probability representations, J. Phys. A: Math. Theor. **41**, 352001 (2008).
[23] H. Zhu, Quantum state estimation with informationally overcomplete measurements, Phys. Rev. A **90**, 012115 (2014).
[24] F. C. Pérez, V. G. Avella, and D. Goyeneche, Mutually unbiased frames, Quantum **6**, 851 (2022).
[25] C. A. Fuchs, M. C. Hoang, and B. C. Stacey, The SIC question: History and state of play, Axioms **6**, 21 (2017).
[26] R. Stricker, M. Meth, L. Postler, C. Edmunds, C. Ferrie, R. Blatt, P. Schindler, T. Monz, R. Kueng, and M. Ringbauer, Experimental single-setting quantum state tomography, PRX Quantum **3**, 040310 (2022).
[27] G. García-Pérez, M. A. Rossi, B. Sokolov, F. Tacchino, P. K. Barkoutsos, G. Mazzola, I. Tavernelli, and S. Maniscalco, Learning to measure: Adaptive informationally complete generalized measurements for quantum algorithms, PRX Quantum **2**, 040342 (2021).

[28] P. G. Casazza, G. Kutyniok, and F. Philipp, in *Finite Frames: Theory and Applications*, edited by P. G. Casazza and G. Kutyniok (Birkhäuser, Boston, 2013), p. 1.

[29] P. G. Casazza and R. G. Lynch, A brief introduction to Hilbert space frame theory and its applications, ArXiv:1509.07347 (2015).

[30] H. Zhu and B.-G. Englert, Quantum state tomography with fully symmetric measurements and product measurements, Phys. Rev. A **84**, 022327 (2011).

[31] Z. Huangjun, Ph.D. thesis, NUS Graduate School for Integrative Sciences and Engineering Centre for Quantum Technologies, National University of Singapore (2012).

[32] H. Zhu, Tomographic and lie algebraic significance of generalized symmetric informationally complete measurements, Phys. Rev. A **90**, 032309 (2014).

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

[34] O. Christensen, *An Introduction to Frames and Riesz Bases* (Springer International Publishing, Cham, 2016).

[35] Y. C. Eldar and G. D. Forney, Optimal tight frames and quantum measurement, IEEE Trans. Inf. Theory **48**, 599 (2002).

[36] A. Bisio, G. Chiribella, G. D'Ariano, S. Facchini, and P. Perinotti, Optimal quantum tomography, IEEE J. Sel. Top. Quantum Electron. **15**, 1646 (2009).

[37] J. Helsen and M. Walter, Thrifty shadow estimation: Re-using quantum circuits and bounding tails, ArXiv:2212.06240 (2022).

[38] L. Devroye, M. Lerasle, G. Lugosi, and R. I. Oliveira, Sub-Gaussian mean estimators, Ann. Stat. **44**, 2695 (2016).

[39] S. Minsker, Efficient median of means estimator, ArXiv: 2305.18681 (2023).

[40] Note that we use the term "prior information" informally, without introducing a prior distribution explicitly, since we do not adopt a Bayesian framework. However, we mention that shadow tomography has recently been studied from a Bayesian perspective [65].

[41] Note that these expressions can be written more rigorously as $\mathcal{E}_\rho(\boldsymbol{\mu}, \tilde{\boldsymbol{\mu}}) = \operatorname{tr}(\mathcal{C}_\rho(\boldsymbol{\mu}, \tilde{\boldsymbol{\mu}}))$ and $\mathcal{E}_\rho(\boldsymbol{\mu}, \tilde{\boldsymbol{\mu}}^{(\rho)}) = \operatorname{tr}(\mathcal{F}_\rho(\boldsymbol{\mu})^{-1}) - \operatorname{tr}(\rho^2)$, making explicit the dependence of $\mathcal{F}_\rho$ on the measurement $\boldsymbol{\mu}$.

[42] The notational mismatch is analogous to how a "tight measurement frame" is not a "tight frame" built from the POVM elements. These differences can be traced back to the fact that any POVM satisfies the normalization condition $\sum_b \mu_b = I$, which imposes some structure on the associated frame operator and derived quantities, which is easier to handle by introducing the specific notions of "tight measurement frame." On the other hand, the mismatch in the use of the term "canonical dual" is due to the fact that if the goal is to minimize the variance corresponding to the estimator, the optimal choice is to use the canonical dual (in the standard sense) of a *rescaled* measurement frame. These aspects are discussed in greater depth in Appendix G.

[43] A. Roy and A. Scott, Weighted complex projective 2-designs from bases: optimal state determination by orthogonal measurements, J. Math. Phys. **48**, 072110 (2007).

[44] M. C. Tran, D. K. Mark, W. W. Ho, and S. Choi, Measuring arbitrary physical properties in analog quantum simulation, Phys. Rev. X **13**, 011049 (2023).

[45] B. Collins, Moments and cumulants of polynomial random variables on unitary groups, the Itzykson-Zuber integral, and free probability, International Mathematics Research Notices **2003**, 953 (2003).

[46] B. Collins, S. Matsumoto, and J. Novak, The Weingarten calculus, Not. Am. Math. Soc. **69**, 1 (2022).

[47] T. Durt, B.-G. Englert, I. Bengtsson, and K. Życzkowski, On mutually unbiased bases, Int. J. Quantum Inf. **8**, 535 (2010).

[48] R. Kueng and D. Gross, Qubit stabilizer states are complex projective 3-designs, ArXiv:1510.02767 (2015).

[49] Z. Webb, The Clifford group forms a unitary 3-design, Quantum Inf. Comput. **16**, 1379 (2016).

[50] G. M. D'Ariano, P. Perinotti, and M. F. Sacchi, Informationally complete measurements and group representation, J. Opt. B: Quantum Semiclassical Opt. **6**, S487 (2004).

[51] W. Matthews, S. Wehner, and A. Winter, Distinguishability of quantum states under restricted families of measurements with an application to quantum data hiding, Commun. Math. Phys. **291**, 813 (2009).

[52] L. Guerini, R. Wiersema, J. F. Carrasquilla, and L. Aolita, Quasiprobabilistic state-overlap estimator for NISQ devices, ArXiv:2112.11618 (2021).

[53] G. García-Pérez, E.-M. Borrelli, M. Leahy, J. Malmi, S. Maniscalco, M. A. Rossi, B. Sokolov, and D. Cavalcanti, Virtual linear map algorithm for classical boost in near-term quantum computing, ArXiv:2207.01360 (2022).

[54] A. Suprano, D. Zia, E. Polino, T. Giordani, L. Innocenti, A. Ferraro, M. Paternostro, N. Spagnolo, and F. Sciarrino, Dynamical learning of a photonics quantum-state engineering process, Adv. Photonics **3**, 066002 (2021).

[55] A. Suprano, D. Zia, E. Polino, T. Giordani, L. Innocenti, M. Paternostro, A. Ferraro, N. Spagnolo, and F. Sciarrino, Enhanced detection techniques of orbital angular momentum states in the classical and quantum regimes, New J. Phys. **23**, 073014 (2021).

[56] D. Zia, R. Checchinato, A. Suprano, T. Giordani, E. Polino, L. Innocenti, A. Ferraro, M. Paternostro, N. Spagnolo, and F. Sciarrino, Regression of high-dimensional angular momentum states of light, Phys. Rev. Res. **5**, 013142 (2023).

[57] A. Suprano, D. Zia, L. Innocenti, S. Lorenzo, V. Cimini, T. Giordani, I. Palmisano, E. Polino, N. Spagnolo, and F. Sciarrino, *et al.*, Experimental property-reconstruction in a photonic quantum extreme learning machine, ArXiv:2308.04543 (2023).

[58] L. Innocenti, S. Lorenzo, I. Palmisano, A. Ferraro, M. Paternostro, and G. M. Palma, Potential and limitations of quantum extreme learning machines, Commun. Phys. **6**, 118 (2023).

[59] M. Tsang, F. Albarelli, and A. Datta, Quantum semiparametric estimation, Phys. Rev. X **10**, 031023 (2020).

[60] J. Suzuki, Y. Yang, and M. Hayashi, Quantum state estimation with nuisance parameters, J. Phys. A: Math. Theor. **53**, 453001 (2020).

[61] S. Gandhari, V. V. Albert, T. Gerrits, J. M. Taylor, and M. J. Gullans, Continuous-variable shadow tomography, ArXiv:2211.05149 (2022).

[62] S. Becker, N. Datta, L. Lami, and C. Rouzé, Classical shadow tomography for continuous variables quantum systems, ArXiv:2211.07578 (2022).
[63] https://github.com/lucainnocenti/Shadow-tomography-on-general-measurement-frames-PRXQ.
[64] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, New York, 2004).
[65] J. M. Lukens, K. J. H. Law, and R. S. Bennink, A Bayesian analysis of classical shadows, npj Quantum Inf. **7**, 1 (2021).