

Optimal Fidelity Estimation from Binary Measurements for Discrete and Continuous Variable Systems

Omar Fawzi¹, Aadil Oufkir², and Robert Salzmann^{1,*}

¹*Inria, ENS de Lyon, Université Claude Bernard Lyon 1, LIP, 69342, Lyon cedex 07, France*

²*Institute for Quantum Information, RWTH Aachen University, Aachen, Germany*



(Received 7 April 2025; accepted 22 September 2025; published 15 January 2026)

Estimating the fidelity between a desired target quantum state and an actual prepared state is essential for assessing the success of experiments. For pure target states, we use functional representations that can be measured directly and determine the number of copies of the prepared state needed for fidelity estimation. In continuous variable (CV) systems, we use the Wigner function, which can be measured via displaced parity measurements. We provide upper and lower bounds on the sample complexity required for fidelity estimation, considering the worst-case scenario across all possible prepared states. For target states of particular interest, such as Fock and Gaussian states, we find that this sample complexity is characterized by the L^1 -norm of the Wigner function, a measure of Wigner negativity widely studied in the literature, in particular in resource theories of quantum computation. For discrete variable systems consisting of n qubits, we explore fidelity estimation protocols using Pauli string measurements. Similarly as for the CV approach, the sample complexity is shown to be characterized by the L^1 -norm of the characteristic function of the target state for both Haar random states and stabilizer states. Furthermore, in a general black box model, we prove that, for any target state, the optimal sample complexity for fidelity estimation is characterized by the smoothed L^1 -norm of the target state. To the best of our knowledge, this is the first time the L^1 -norm of the Wigner function provides a lower bound on the cost of some information processing task.

DOI: [10.1103/qd1c-1fk9](https://doi.org/10.1103/qd1c-1fk9)

I. INTRODUCTION

With the possibility to produce a variety of complex and highly entangled quantum states in experiments, it is essential to develop reliable methods for certifying that a desired target state has been prepared successfully. For a pure target state $\rho = |\psi\rangle\langle\psi|$ and an actually prepared state σ , an operationally meaningful way to achieve such certification is by estimating the fidelity $F(\rho, \sigma)$ through measurements on the state σ . For example, performing the projective measurement $\{|\psi\rangle\langle\psi|, \mathbb{I} - |\psi\rangle\langle\psi|\}$ allows fidelity estimation using a number of copies of σ that depends only on the desired accuracy and success probability, and that is independent of the target state or system size. However, for most states $|\psi\rangle\langle\psi|$ of interest, performing this measurement is not possible in practice in an efficient and reliable manner. Therefore, it is sensible to express the fidelity

as the overlap of certain functions representing the quantum states, which can be directly measured with current experimental techniques [1,2].

For continuous variable (CV) quantum systems with $m \in \mathbb{N}$ degrees of freedom, e.g., an m -mode photonic system, we use the phase space representation of the fidelity between a pure state ρ and a general mixed state σ in terms of their *Wigner functions* W_ρ and W_σ given by

$$F(\rho, \sigma) = \pi^m \int_{\mathbb{C}^m} W_\rho(\alpha) W_\sigma(\alpha) d\alpha. \quad (1)$$

Assuming full knowledge of the pure target state ρ , the fidelity can hence be estimated by directly measuring the Wigner function $W_\sigma(\alpha)$ at different points in phase space $\alpha \in \mathbb{C}^m$ through *Wigner tomography*. This can be achieved by using the fact that the Wigner function of σ is given by the expectation value of the *displaced parity operator* [3], i.e.,

$$W_\sigma(\alpha) = \left(\frac{2}{\pi}\right)^m \text{Tr}(D(\alpha)\Pi D^*(\alpha)\sigma), \quad (2)$$

with $D(\alpha)$ denoting the *displacement operator* and $\Pi = (-1)^N$ the *parity operator*, with N being the m -mode

*Contact author: robert.als.salzmann@gmail.com

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

number operator (see Sec. IID for the definition). The observable $(2/\pi)^m D(\alpha) \Pi D^*(\alpha)$ has measurement outcomes $\{(2/\pi)^m, -(2/\pi)^m\}$, showing that we can estimate $W_\sigma(\alpha)$ by some binary measurement on the state σ . A practical scheme for performing these displaced parity measurements relying on Ramsey interferometry was proposed in Ref. [4] and was experimentally first implemented by Nogues *et al.* [5], Bertet *et al.* [6] and Vlastakis *et al.* [7]. Directly measuring the Wigner function in this way and performing fidelity estimation using Eq. (1) has become conventional in practice; see Refs. [8–12]. In the following, we refer to the above-described paradigm as the “Wigner model of fidelity estimation.”

On the other hand, for discrete variable (DV) quantum systems consisting of n qubits, we consider the representation of the fidelity given by

$$F(\rho, \sigma) = \frac{1}{d} \sum_{P \in \mathcal{P}_n} \chi_\rho(P) \chi_\sigma(P), \quad (3)$$

with dimension of the overall system $d = 2^n$, $\mathcal{P}_n = \{1, X, Y, Z\}^{\otimes n}$ denoting the set of Pauli strings of length n and

$$\chi_\sigma(P) = \text{Tr}(P\sigma) \quad (4)$$

denoting the *characteristic function* of the state σ . Again, assuming full knowledge of the pure target state ρ , we can estimate the fidelity by measuring Pauli strings with binary outcome in $\{1, -1\}$ on the state σ . We refer to this paradigm as the “Pauli model of fidelity estimation” in the following.

For both these settings, efficient fidelity estimation algorithms are known for some special families of states, e.g., well-conditioned states including stabilizer states [2]. However, as far as we are aware, no lower bound for fidelity estimation is known. This raises the following natural question: *Are there states for which fidelity estimation is hard?* More specifically, in the Wigner model of fidelity estimation, are there states for which the sample complexity is arbitrarily large? In the Pauli model, are there states for which the number of samples required is exponential in n ? Our results give an affirmative answer to both these questions for Fock states [Eq. (10)] and Haar-random n -qubit states [Eq. (17)]. Going beyond such hard states, we can ask the following question: *Which property of ρ quantifies the sample complexity of fidelity estimation?* We identify the smoothed L^1 -norm as quantifying the optimal sample complexity for fidelity estimation in a general black box model [Eq. (7)].

A. Main results

We now describe more precisely the setup for our results. We analyze for pure target states ρ the *sample complexity of fidelity estimation in the Wigner model and the*

Pauli model, respectively denoted by $N_{\varepsilon, \delta}^W(\rho)$ and $N_{\varepsilon, \delta}^P(\rho)$. They are defined as the minimal number of copies of the state σ needed to estimate the fidelity $F(\rho, \sigma)$ with accuracy $\varepsilon > 0$ and probability of failure at most $\delta > 0$ in the worst case over all states σ with the use of the paradigms described around Eqs. (1) and (3) respectively.

For this, we unify the discussion for continuous and discrete variable systems by considering general overlaps of functions f and g on some measurable space Ω with measure μ . Inspired by Eqs. (1) and (3), we choose for CV systems $\Omega_W = \mathbb{C}^m$ with rescaled Lebesgue measure $d\mu_W(\alpha) = \pi^m d\alpha$ and for DV systems $\Omega_P = \mathcal{P}_n$ [13] with $\mu_P = (1/d)\mu_{\text{count}}$, with μ_{count} denoting the counting measure.

The function f corresponds to the Wigner function or characteristic function of the target state ρ in the Wigner model or the Pauli model respectively and hence full knowledge of this function is assumed. On the other hand, for the function g , which plays the role of the Wigner or characteristic function of the prepared state σ in the Wigner model or the Pauli model respectively, we assume we have *black box access*: that is, for each $\lambda \in \Omega$ we can obtain samples of a random variable Y_λ with binary outcome space and that satisfies $\mathbb{E}[Y_\lambda] = g(\lambda)$.

The objective is then to estimate the overlap

$$\int_{\Omega} f(\lambda) g(\lambda) d\mu(\lambda) = \int_{\Omega} f(\lambda) \mathbb{E}[Y_\lambda] d\mu(\lambda) \quad (5)$$

from as few samples of Y_λ at different points λ as possible. We call this task “black box overlap estimation” in the following.

Here, f is assumed to be fixed and the function g , for which black box access is provided, is allowed to be any element of some fixed set of functions $\mathcal{S}$. The *sample complexity of black box overlap estimation*, denoted by $N_{\varepsilon, \delta}(f, \mathcal{S})$, is then given by the minimal number of samples of Y_λ needed to estimate the overlap (5) with accuracy $\varepsilon > 0$ and probability of failure at most $\delta > 0$ in the worst case over all $g \in \mathcal{S}$. In particular, choosing Ω and μ for CV and DV quantum systems as outlined above and denoting the set of possible Wigner and characteristic functions by $\mathcal{S}_W$ and $\mathcal{S}_P$ respectively, we can make the following identifications:

$$\begin{aligned} \text{CV systems: } N_{\varepsilon, \delta}^W(\rho) &= N_{\varepsilon, \delta}(W_\rho, \mathcal{S}_W), \\ \text{DV systems: } N_{\varepsilon, \delta}^P(\rho) &= N_{\varepsilon, \delta}(\chi_\rho, \mathcal{S}_P). \end{aligned} \quad (6)$$

We provide upper and lower bounds on the sample complexity of black box overlap estimation $N_{\varepsilon, \delta}(f, \mathcal{S})$ for a large class of sets of functions $\mathcal{S}$.

1. Black box overlap estimation

For general measurable spaces Ω with measure μ we prove upper and lower bounds on the sample complexity $N_{\varepsilon,\delta}(f, \mathcal{S})$ for $f \in L^2(\Omega, \mu)$ and all sets of functions $\mathcal{S} \subset \mathcal{S}_{\max}$ where for some fixed number $r > 0$ the set $\mathcal{S}_{\max}$ consists of all functions g satisfying $\|g\|_{L^2(\Omega, \mu)} \leq 1$ and $\|g\|_{L^\infty(\Omega, \mu)} \leq r$. These conditions on the sets $\mathcal{S}$ and $\mathcal{S}_{\max}$ are motivated by the Wigner and Pauli models as for all quantum states ρ of a continuous (or discrete) variable system the corresponding Wigner (or characteristic) function satisfies $\|W_\rho\|_{L^2(\mathbb{C}^m, \mu_W)} \leq 1$ and $\|W_\rho\|_{L^\infty(\mathbb{C}^m, \mu_W)} \leq (2/\pi)^m$ (or $\|\chi_\rho\|_{L^2(\mathcal{P}_n, \mu_P)} \leq 1$ and $\|\chi_\rho\|_{L^\infty(\mathcal{P}_n, \mu_P)} \leq 1$). Hence, under the appropriate choice of Ω and μ as above, the sets of Wigner and characteristic functions for continuous and discrete variable quantum systems respectively provide examples of such sets $\mathcal{S}$, i.e., $\mathcal{S}_W \subset \mathcal{S}_{\max}$ and $\mathcal{S}_P \subset \mathcal{S}_{\max}$.

For all such sets $\mathcal{S}$, we provide an explicit algorithm for estimating the overlap (5) yielding an upper bound on $N_{\varepsilon,\delta}(f, \mathcal{S})$ (see Theorem 1) and furthermore, through some information-theoretic argument, the corresponding lower bound (see Theorem 2). For the particular choice of $\mathcal{S} = \mathcal{S}_{\max}$, these upper and lower bounds can be shown to be essentially matching in their scaling in the relevant variables f, ε, δ since they are given as

$$\begin{aligned} \left(\frac{\|f\|_1^{(2\varepsilon)}}{2\varepsilon} \right)^2 \ln \left(\frac{1}{3\delta} \right) &\lesssim N_{\varepsilon,\delta}(f, \mathcal{S}_{\max}) \\ &\lesssim \inf_{\varepsilon' \in [0, \varepsilon]} \left(\frac{\|f\|_1^{(\varepsilon')}}{\varepsilon - \varepsilon'} \right)^2 \ln \left(\frac{1}{\delta} \right). \end{aligned} \quad (7)$$

Here, we have denoted the *smoothed L^1 -norm* of the function f by

$$\|f\|_1^{(\varepsilon)} = \inf_{\tilde{f} \in L^1(\Omega) \cap L^2(\Omega), \|\tilde{f} - f\|_2 \leq \varepsilon} \|\tilde{f}\|_1. \quad (8)$$

Note that Eq. (7) provides an *instance-optimal* characterization for the task of black box overlap estimation, i.e., it identifies precisely the property of f that governs the sample complexity as the smoothed L^1 -norm. The study of instance-optimal bounds is well-developed in the learning literature; for example for the problem of identity testing, one of the flagship results is that classical identity testing is characterized by a smoothed version of the $L^{2/3}$ -quasinorm of the distribution to be tested [14] (see also Ref. [15] for results for quantum identity testing in this direction).

For functions f with a rapidly decaying tail as defined in Sec. III and discussed in detail in Appendix A 1, which include for example exponentially decaying functions on Euclidean space, we find that the sample complexity is

characterized by the usual L^1 -norm as

$$N_{\varepsilon,\delta}(f, \mathcal{S}_{\max}) = \Theta \left(\left(\frac{\|f\|_1}{\varepsilon} \right)^2 \ln \left(\frac{1}{\delta} \right) \right), \quad (9)$$

with the precise statement given in Corollary 1.

2. Fidelity estimation in the Wigner model

In Theorems 3 and 4 we give upper and lower bounds on the sample complexity of fidelity estimation in the Wigner model, $N_{\varepsilon,\delta}^W(\rho)$, for pure states ρ on $L^2(\mathbb{R}^m)$ from the ones established for the abstract framework of black box overlap estimation. In particular, the upper bound involves the smoothed L^1 -norm of W_ρ , giving that for every fixed pure state ρ and in the worst case over all states σ , the fidelity $F(\rho, \sigma)$ can be estimated by performing displaced parity measurements on a finite number of copies of σ , i.e., $N_{\varepsilon,\delta}^W(\rho) < \infty$.

Showing for general pure states ρ that these upper and lower bounds on $N_{\varepsilon,\delta}^W(\rho)$ are essentially matching similarly to Eqs. (7) and (9) is, however, hard as the set of Wigner functions, $\mathcal{S}_W$, is a proper and complicated subset of $\mathcal{S}_{\max} = B_1^2(0) \cap B_{(2/\pi)^m}^\infty(0)$. Hence, we focus on interesting examples of states of continuous variable systems and establish for those that the sample complexity can, up to constants, be expressed in terms of the L^1 -norm of the corresponding Wigner function:

Firstly, we consider the Fock states $(|n\rangle\langle n|)_{n \in \mathbb{N}}$, i.e., the eigenstates of the harmonic oscillator or number operator. In this case, we find in Proposition 1 that their sample complexity of fidelity estimation in the Wigner model is fully characterized as

$$N_{\varepsilon,\delta}^W(|n\rangle\langle n|) = \Theta \left(\left(\frac{\|W_{|n\rangle\langle n|}\|_1}{\varepsilon} \right)^2 \ln \left(\frac{1}{\delta} \right) \right). \quad (10)$$

Here, the L^1 -norm of the Wigner functions of the Fock states satisfies the scaling behavior

$$\|W_{|n\rangle\langle n|}\|_1 \sim \sqrt{n} \quad (11)$$

as shown in Eq. (74) by straightforward use of known asymptotic expressions of the L^p -norms of the Laguerre polynomials in Ref. [16]. Note this shows that there exist states with arbitrarily large sample complexity.

As another example, we construct in Sec. IV A 2 the so-called *spike states* for any m -mode continuous variable quantum system. These are pure states $(\rho_n)_{n \in \mathbb{N}} = (|\psi_n\rangle\langle\psi_n|)_{n \in \mathbb{N}}$ on the Hilbert space $L^2(\mathbb{R}^m)$ whose Wigner functions are uniformly vanishing as $n \rightarrow \infty$, i.e.,

$$\|W_{\rho_n}\|_\infty \xrightarrow{n \rightarrow \infty} 0. \quad (12)$$

Furthermore, we show that $W_{\rho_n} \in L^1(\mathbb{C}^m)$ for all $n \in \mathbb{N}$ with L^1 -norms blowing up and scaling reciprocally to their

L^∞ -norms, i.e.,

$$\|W_{\rho_n}\|_1 \sim \|W_{\rho_n}\|_\infty^{-1}. \quad (13)$$

Using Eqs. (12) and (13) together with the established bounds on $N_{\varepsilon,\delta}^W(\rho)$, we show in Proposition 2 that also for the spike states the corresponding L^1 -norm characterizes the sample complexity of fidelity estimation, i.e., precisely

$$N_{\varepsilon,\delta}^W(\rho_n) = \Theta\left(\left(\frac{\|W_{\rho_n}\|_1}{\varepsilon}\right)^2 \ln\left(\frac{1}{\delta}\right)\right). \quad (14)$$

We believe that the existence of a sequence of pure states satisfying Eqs. (12) and (13) can be of independent interest.

For ρ being a Gaussian state, we know that $W_\rho \geq 0$ and hence by normalization of the Wigner function that $\|W_\rho\|_1 = 1$. From this, we see in Proposition 3 that the sample complexity of fidelity estimation in the Wigner model satisfies for all pure Gaussian states ρ

$$N_{\varepsilon,\delta}^W(\rho) = \Theta\left(\frac{1}{\varepsilon^2} \ln\left(\frac{1}{\delta}\right)\right) \quad (15)$$

and hence that for all states σ , the fidelity $F(\rho, \sigma)$ can efficiently be estimated with the use of displaced parity measurements on a few copies of σ .

In addition, looking at bounds that depend only on the L^1 -norm of the state, we find in Theorem 5 a characterization of the sample complexity in the worst case over all pure states ρ whose Wigner function has L^1 -norm less than or equal to some fixed threshold value. In particular, we see for all $t \geq 1$

$$\sup_{\substack{\rho \text{ pure state,} \\ \|W_\rho\|_1 \leq t}} N_{\varepsilon,\delta}^W(\rho) = \Theta\left(\left(\frac{t}{\varepsilon}\right)^2 \ln\left(\frac{1}{\delta}\right)\right), \quad (16)$$

showcasing that also in the considered worst case, the sample complexity has the same scaling behavior as the for the Fock, spike and Gaussian states portrayed in Eqs. (10), (14) and (15) respectively.

Negativity of the Wigner function of quantum states is an indication of nonclassicality [17–19] as it portrays the fact that the Wigner function deviates from being a proper probability distribution, implies that the quantum state cannot be written as a convex combination of coherent or, more generally, Gaussian states and furthermore is equivalent to contextuality for CV measurements [20]. It has been shown that Wigner negativity is necessary for quantum advantage, as computations involving only positive Wigner functions are efficiently classically simulable [21–23]. The L^1 -norm of the Wigner function can be regarded as a quantification of Wigner negativity [18],

with it being equal to 1 if and only if the Wigner function is non-negative and otherwise being strictly greater than 1. However, in this context, the L^1 -norm of the Wigner function provides only an upper bound on the cost of classical simulation [22–25] as there are states with a large L^1 -norm that can still be simulated efficiently [26,27]. Interestingly, for the fidelity estimation problem we consider here, we can achieve lower bounds on the sample cost in terms of the Wigner L^1 -norm. The results presented here in Eqs. (7), (10), (14), (15) and (16) show an interesting connection between the complexity of learning properties of a quantum state and its nonclassicality, and we hope the techniques we introduce here, such as the lower bound techniques and the smoothed L^1 -norm, find further applications in the context of nonclassicality.

3. Fidelity estimation in the Pauli model

Applying again the results from the abstract paradigm of black box overlap estimation to the Pauli model of fidelity estimation yields upper and lower bounds on the corresponding sample complexity $N_{\varepsilon,\delta}^P(\rho)$ for all pure states ρ on an n -qubit system (see Theorems 6 and 7).

In Proposition 4 we use these to find for Haar-random pure states ρ with probability of at least $1 - 1/d$ that the corresponding sample complexity is characterized as

$$N_{\varepsilon,\delta}^P(\rho) = \tilde{\Theta}\left(\left(\frac{\|\chi_\rho\|_1}{\varepsilon}\right)^2 \ln\left(\frac{1}{\delta}\right)\right) = \tilde{\Theta}\left(\frac{d}{\varepsilon^2} \ln\left(\frac{1}{\delta}\right)\right), \quad (17)$$

with the L^1 -norm defined by $\|\chi_\rho\|_1 = 1/d \sum_{P \in \mathcal{P}_n \setminus \{1\}} |\chi_\rho(P)|$. Here, we used $\tilde{\Theta}$ to denote the fact that the upper and lower bounds in Eq. (17) match up to factors that scale logarithmically in the leading order term. In particular, the result implies that in the worst case over all pure states ρ the sample complexity blows up as $d \rightarrow \infty$, i.e., $\sup_{\rho} N_{\varepsilon,\delta}^P(\rho) \xrightarrow{d \rightarrow \infty} \infty$. The L^1 -norm of χ_ρ has appeared in the literature under the name “st-norm” in Ref. [28] and it was shown in Ref. [29] that it provides a lower bound on a magic measure obtained via an encoding into continuous variable systems.

For stabilizer states ρ on n qubits, we have $\|\chi_\rho\|_1 = (d-1)/d = \Theta(1)$ (Proposition 5). Hence, $N_{\varepsilon,\delta}^P(\rho)$ is independent of the system size and the fidelity $F(\rho, \sigma)$ can be estimated efficiently by performing Pauli measurements on a few copies of σ . This was shown in Ref. [2].

B. Related work

In Refs. [1,2], with use of the overlap formula (3), upper bounds on the sample complexity of fidelity estimation in the Pauli model have been established. In particular, in Ref. [2] the upper bound in Proposition 5 for the sample complexity for stabilizer states was

found as well as the worst case bound $\sup_{\rho \text{ pure}} N_{\varepsilon, \delta}^P(\rho) = \mathcal{O}((1/\varepsilon^2 \delta^2) + d/\varepsilon^2 \ln(1/\delta))$ [with use of the Markov argument around Eq. (10) in Ref. [2]] or $\sup_{\rho \text{ pure}} N_{\varepsilon, \delta}^P(\rho) = \mathcal{O}(1/\varepsilon^2 \delta + d/\varepsilon^4 \ln(1/\delta))$ (with use of the “truncating bad events” technique) [30]. In Remark 8, we improve these worst case bounds by establishing $\sup_{\rho \text{ pure}} N_{\varepsilon, \delta}^P(\rho) = \mathcal{O}(d/\varepsilon^2 \ln(1/\delta))$. Furthermore, we show that this worst case sample complexity is tight (up to a logarithmic factor in the dimension) for Haar-random pure state ρ with high probability (see Proposition 4).

Leone *et al.* [31] provide upper and lower bounds on the sample complexity of the specific algorithm from Ref. [2] for estimation of fidelity in the Pauli model. There, the corresponding upper and lower bounds are given in terms of the so-called 2- and 0-stabilizer Rényi entropy respectively.

In Ref. [32], an upper bound on the sample complexity in terms of the L^1 -norm is given for the problem of estimating the expectation values of observables in the Pauli model. In this work, we use a similar sampling rule to derive (with a simpler analysis) our upper bounds. Moreover, we provide nearly matching lower bounds in the case of observables being either Haar-random pure states (see Proposition 4) or stabilizer states (see Proposition 5) as well as all pure observables in the low accuracy regime (see Remark 9).

In Refs. [1,33] use of the overlap formula in terms of Wigner functions (1) for fidelity estimation of CV systems is discussed, although without the establishment of bounds on the corresponding sample complexity. Reference [33] as well as Ref. [25], in which a method for estimating the probabilities of outcomes of a quantum circuit is presented, and also Ref. [34], in which sufficient conditions for efficient classical estimation of quantum kernel functions for bosonic systems are provided, rely on a sampling rule similar to ours [see Eq. (42)]. Moreover, in Refs. [25,33] it is claimed that this sampling procedure is optimal as it has the smallest variance. We stress that this optimality proof has implicit assumptions: the estimator should be of a particular form [see Eq. (14) in Ref. [25] or Eq. (5) in Ref. [33]], the sampling distribution remains the same during different steps, and the global estimator is given by the empirical one. We know that for some CV states for which the Wigner function has an infinite L^1 -norm (see Theorem 2 in Ref. [35]), estimators satisfying these assumptions have infinite variance but nonetheless our bound (7) gives an algorithm with finite sample complexity. Here, we prove our lower bounds on the sample complexity (see Theorems 2, 4 and 7) without making these assumptions. In particular, we allow the learner to choose adaptively the sampling procedures. This allows us to prove the optimality of the sampling rule (42) (applied on the corresponding smoothing function $\tilde{f}$) for $\mathcal{S} = \mathcal{S}_{\max}$ [see Eq. (7)], for the Fock, spike and Gaussian states as well as the worst case in the

Wigner model [see Eqs. (10), (14), (15) and (16)] and for Haar-random and stabilizer states in the Pauli model [see Eq. (17) and Proposition 5].

In Ref. [36] it was shown recently that full state tomography using local Pauli measurements requires sample complexity of at least $\Omega(9.118^n/\varepsilon^2)$. On the other hand, as shown in Remark 8, fidelity estimation can be done with sample complexity of only $\mathcal{O}(2^n/\varepsilon^2)$. Therefore, the use of protocols tailored to fidelity estimation is more advantageous than the use of Pauli tomography.

Huang *et al.* [37] recently considered the task of quantum state certification, i.e., the decision problem of whether the actually prepared state σ is close to or far from the target state ρ , on n -qubit systems from Pauli string measurements. Incorporating the full n -bit string of measurement outcomes obtained by measuring each Pauli matrix for a given $P \in \mathcal{P}_n$ individually enables them to prove an $\mathcal{O}(n^2) = \mathcal{O}(\ln^2(d))$ upper bound on the corresponding sample complexity for almost all Haar-random pure states. This is in sharp contrast to the lower bound we found in Eq. (17), which is exponential in the number of qubits due to our using only the binary measurement outcome from the full Pauli string $P \in \mathcal{P}_n$.

In Refs. [38–40] the authors consider fidelity estimation of CV systems using heterodyne measurements [41] instead of performing Wigner tomography through displaced parity measurements, and establish upper bounds on the corresponding sample complexity. Quantum state tomography for continuous variable systems for different measurement setups has been extensively studied: In Ref. [42] quantum state tomography using homodyne detection [43], a measurement scheme, which differs from displaced parity measurements considered here, was studied. Furthermore, continuous variable tomography under energy constraints on the unknown state has been a topic of interest in the literature: With restriction to classical shadow tomography and using homodyne, heterodyne, photon-number-resolving [44] or photon parity measurements, Refs. [45,46] establish performance guarantees by providing upper bounds on the corresponding sample complexity provided the unknown state satisfies certain energy constraints with respect to the photon number operator. The mentioned protocols to perform direct fidelity estimation [38–40] or shadow tomography [45,46] from heterodyne and homodyne measurements all rely on such strong energy constraints of either the target or the measured state or both. In particular, in contrast to our algorithm for the Wigner model using displaced parity measurements, these results usually [47] give infinite sample complexity in the worst case over all possible measured states σ . For the particular example of target states being the Fock states, Proposition 1 shows that the sample complexity scales linearly in the energy (or photon number) in the Wigner model whereas the above-mentioned results relying on

heterodyne and homodyne measurements give only (high degree) polynomial or even exponential scaling.

From an experimentalist point of view, heterodyne and homodyne measurements are hard to perform on confined cavity fields, e.g., resonators, which are important for quantum computing with bosonic modes [48–50]. In contrast, displaced parity measurements use qubits to encode the parity and dispersive readout and nowadays provide outcomes with almost perfect accuracy. The displaced parity measurements we consider in the Wigner model are much less sensitive to noise in the detection setup since the encoding is done with the qubit as an ancillary system.

In Ref. [51], lower bounds (for general mixed states) and upper bounds (for pure and Gaussian states) on the sample complexity of full quantum state tomography using general measurements are provided. There, the authors prove their results again under the assumption that the unknown state satisfies an energy constraint. Interestingly, their lower bound implies that the scaling of the sample complexity in terms of error parameter ε worsens when the number of modes m increases. Note that in comparison, our analysis here does not rely on an energy constraint on either the target state ρ or the prepared state σ . Our results are for arbitrary states σ and involve only the (smoothed) L^1 -norm of the Wigner function of ρ , which can remain bounded for arbitrarily high energy of ρ [52]. For an extended discussion of the relation between energy constraints and the (smoothed) L^1 -norm, see Remark 3.

C. Outline of the article

The rest of the article is outlined as follows: In Sec. II we discuss some preliminary facts which are needed for the derivations in the rest of the paper.

In Sec. III we formally define the task of black box overlap estimation and show upper and lower bounds on the corresponding sample complexity in Theorems 1 and 2 and its classification for functions with a rapidly decaying tail in Corollary 1.

In Sec. IV we formally define the task of fidelity estimation in the Wigner model, discuss its connection to the abstract black box estimation and show upper and lower bounds on the corresponding sample complexity in Theorems 3 and 4. In Sec. IV A we discuss instances for which these upper and lower bounds can be shown to be matching up to constants including the Fock states in Sec. IV A 1, spike states in Sec. IV A 2, Gaussian states in Sec. IV A 3 and the worst case behavior of the sample complexity in Theorem 5.

In Sec. V we formally define the task of fidelity estimation in the Pauli model, discuss its connection to the abstract black box estimation and show upper and lower bounds on the corresponding sample complexity in Theorems 6 and 7. We then show in Sec. V A that these

upper and lower bounds are essentially matching for Haar-random pure states in Proposition 4 and for stabilizer states in Proposition 5.

In Appendix A we study properties of the smoothed L^1 -norm. In Lemma A1 we show continuity in the sense of $\lim_{\varepsilon \downarrow 0} \|f\|_1^{(\varepsilon)} = \|f\|_1$. We then provide convergence rates for this limit for functions with a rapidly decaying tail in Appendix A 1, with an explicit example being the Wigner functions of the Fock states.

In Appendix B, we investigate the noise robustness of the fidelity estimation protocol described in Theorem 3 for continuous-variable systems.

II. PRELIMINARIES

In this section we discuss some preliminary definitions and facts which are needed for the derivations of the main results of this paper.

A. Notation

For X being a set and $f, g : X \rightarrow \mathbb{R}$ being two functions we write $f \lesssim g$ if there exists $C \geq 0$ such that for all $x \in X$ we have $f(x) \leq Cg(x)$ and furthermore $f = \mathcal{O}(g)$ if $|f| \lesssim |g|$. Similarly we write $f \gtrsim g$ if there exists $c > 0$ such that $f(x) \geq cg(x)$ for all $x \in X$ and $f = \Omega(g)$ if $|f| \gtrsim |g|$. We denote by $f \sim g$ that $f \lesssim g$ and $g \gtrsim f$ and by $f = \Theta(g)$ that $f = \mathcal{O}(g)$ and $f = \Omega(g)$.

B. L^q -spaces and smoothed L^1 -norm

Let Ω be some measurable space with some measure μ . In the following we denote for $q \in [1, \infty]$ by $L^q(\Omega, \mathbb{C})$ and $L^q(\Omega, \mathbb{R})$ the Banach spaces of complex and real valued L^q -functions respectively. For brevity we often denote for the case of complex valued functions $L^q(\Omega) \equiv L^q(\Omega, \mathbb{C})$. We denote the ball of real valued L^q -functions with radius $r \geq 0$ centered at some $f \in L^q(\Omega, \mathbb{R})$ by

$$B_r^q(f) = \{g \in L^q(\Omega, \mathbb{R}) \mid \|f - g\|_q \leq r\}$$

$$\text{where } \|f\|_q = \left(\int_{\Omega} |f(\lambda)|^q d\mu(\lambda) \right)^{1/q}. \quad (18)$$

The *smoothed L^1 -norm* [53] of a function $f \in L^2(\Omega, \mathbb{R})$ for $\varepsilon \geq 0$ is defined to be

$$\|f\|_1^{(\varepsilon)} = \inf_{\substack{\tilde{f} \in L^1(\Omega, \mathbb{R}) \\ \|f - \tilde{f}\|_2 \leq \varepsilon}} \|\tilde{f}\|_1. \quad (19)$$

As seen in Lemma A1 in Appendix A the smoothed L^1 -norm satisfies the natural continuity property $\lim_{\varepsilon \downarrow 0} \|f\|_1^{(\varepsilon)} = \|f\|_1$.

Note that for all $f \in L^2(\Omega, \mathbb{R})$ and $\varepsilon > 0$ the respective smoothed L^1 -norm is always finite, i.e., $\|f\|_1^{(\varepsilon)} < \infty$, which follows since $L^1(\Omega, \mathbb{R}) \cap L^2(\Omega, \mathbb{R})$ is dense in $L^2(\Omega, \mathbb{R})$ [54].

C. Information-theoretic divergences

The Kullback-Leibler divergence between two probability distributions P and Q on $\mathcal{X}$ is defined by

$$D(P\|Q) = \int_{\mathcal{X}} dP(x) \ln \left(\frac{dP}{dQ}(x) \right)$$

whenever P is absolutely continuous with respect to Q and $D(P\|Q) = +\infty$ otherwise.

For $p, q \in [0, 1]$, we write $D(p\|q) = D(P_{\text{Ber}}(p)\|P_{\text{Ber}}(q))$, where $P_{\text{Ber}}(p)$ denotes a Bernoulli distribution with parameter p . Note that if $\delta \leq 1/2$, and $p \geq 1 - \delta$ and $q \leq \delta$ we have $D(p\|q) \geq D(1 - \delta\|\delta) \geq \ln(1/(3\delta))$.

Furthermore, the χ^2 -divergence is defined by

$$\chi^2(P\|Q) = \int_{\mathcal{X}} \left(\frac{dP}{dQ}(x) - 1 \right)^2 Q(dx) \quad (20)$$

if P is absolutely continuous with respect to Q and $+\infty$ otherwise.

These divergences satisfy the well-known bound

$$D(P\|Q) \leq \chi^2(P\|Q). \quad (21)$$

D. Wigner functions for continuous variable systems

For $m \in \mathbb{N}$ we consider an m -mode quantum harmonic oscillator on the Hilbert space $\mathcal{H} = L^2(\mathbb{R}^m)$. For a point in phase space $\alpha \in \mathbb{C}^m \cong \mathbb{R}^{2m}$ the *displacement operator* is defined as

$$D(\alpha) = e^{\sum_{k=1}^m (\alpha_k a_k^* - \alpha_k^* a_k)} = \bigotimes_{k=1}^m e^{\alpha_k a_k^* - \alpha_k^* a_k}, \quad (22)$$

where we denote by a_k and a_k^* the canonical bosonic annihilation and creation operators of the k^{th} mode (see, e.g., Chapter 5.2 in Ref. [55] or Chapter 8.3 in Ref. [56] for a definition and discussion). Note that $D(\alpha)$ is a unitary operator with $D^*(\alpha) = D(-\alpha)$ and satisfying the relation $D(\alpha)D(\beta) = e^{(\alpha \cdot \beta^* - \alpha^* \cdot \beta)/2} D(\alpha + \beta)$, where we denoted the real dot product between $\alpha, \beta \in \mathbb{C}^m$ by $\alpha \cdot \beta = \sum_{k=1}^m \alpha_k \beta_k$.

For a state ρ we can define its *characteristic function* as

$$\chi_\rho(\alpha) := \text{Tr}(\rho D(\alpha)). \quad (23)$$

The characteristic function is complex valued and satisfies $\chi_\rho^*(\alpha) = \chi_\rho(-\alpha)$ and furthermore $\chi_\rho \in L^q(\mathbb{C}^m) \cong L^q(\mathbb{R}^{2m})$ for all $q \in [2, \infty]$ [57]. Therefore, we can define the so-called *Wigner function* of ρ as the (symplectic) Fourier transform of its characteristic function,

i.e., formally [58] through the integral expression

$$\begin{aligned} W_\rho(\alpha) &:= \frac{1}{\pi^{2m}} \int_{\mathbb{C}^m} e^{\alpha \cdot \beta^* - \alpha^* \cdot \beta} \chi_\rho(\beta) d\beta \\ &= \frac{1}{\pi^{2m}} \int_{\mathbb{C}^m} e^{2i\Im(\alpha \cdot \beta^*)} \chi_\rho(\beta) d\beta, \end{aligned} \quad (24)$$

with $d\beta \equiv d\Re(\beta)d\Im(\beta)$ denoting the Lebesgue measure on $\mathbb{C}^m$. Note that the Wigner function (24) is real valued and can have positive and negative values. For $\rho = |\psi\rangle\langle\psi|$ being a pure state, the Wigner function is directly given by

$$W_\rho(x, p) = \left(\frac{2}{\pi} \right)^m \int_{\mathbb{R}^m} \psi^*(x + y) \psi(x - y) e^{i2p \cdot y} dy, \quad (25)$$

where we changed coordinates to position and momentum variables $(x, p) \in \mathbb{R}^{2m}$ through the relation $\alpha = \frac{1}{\sqrt{2}}(x + ip)$. Note the integral in Eq. (25) is well-defined as the integrand is in $L^1(\mathbb{R}^m)$ and furthermore that this definition of the Wigner function can be extended to general mixed states by linearity.

The Wigner function W_ρ for any state ρ is a *quasiprobability distribution* as it can have positive and negative values and satisfies the normalization [59]

$$\int_{\mathbb{C}^m} W_\rho(\alpha) d\alpha = \frac{1}{2^m} \int_{\mathbb{R}^{2m}} W_\rho(x, p) dx dp = 1. \quad (26)$$

An important class of states which have a non-negative Wigner function is *Gaussian states* with defining property being that their Wigner (or equivalently characteristic) function is a Gaussian function. Notably, by Hudson's theorem [60,61], a pure state has a non-negative Wigner function if and only if it is Gaussian. On the other hand, the set of general mixed states with a non-negative Wigner function has been shown to be strictly bigger than the convex hull of all Gaussian states and is itself hard to characterize [62,63].

The Hilbert-Schmidt inner product between two states ρ and σ can by Plancherel's identity be expressed as the inner product of the corresponding characteristic or Wigner functions as

$$\begin{aligned} \text{Tr}(\rho\sigma) &= \frac{1}{\pi^m} \int_{\mathbb{C}^m} \chi_\rho^*(\alpha) \chi_\sigma(\alpha) d\alpha \\ &= \pi^m \int_{\mathbb{C}^m} W_\rho(\alpha) W_\sigma(\alpha) d\alpha. \end{aligned} \quad (27)$$

Hence, in particular we know for ρ being a state that

$$\|W_\rho\|_2 = \frac{1}{\pi^{m/2}} \|\rho\|_2 \leq \frac{1}{\pi^{m/2}}, \quad (28)$$

where we denoted the Hilbert-Schmidt norm of ρ by $\|\rho\|_2$ and used the fact that it is dominated by the trace norm.

Alternatively to Eq. (24), we can write the Wigner function of ρ as the expectation value of the *displaced parity operator* [3], i.e.,

$$W_\rho(\alpha) = \left(\frac{2}{\pi}\right)^m \text{Tr}(\rho D(\alpha) \Pi D^*(\alpha)), \quad (29)$$

with Π being the *parity operator* defined through the relation $(\Pi\psi)(x) := \psi(-x)$ for all $\psi \in L^2(\mathbb{R}^m)$ or alternatively in the Fock basis

$$\Pi = \sum_{n_1, \dots, n_m \in \mathbb{N}_0} (-1)^{n_1 + \dots + n_m} |n_1, \dots, n_m\rangle \langle n_1, \dots, n_m|.$$

From that we see that the Wigner function is bounded with

$$\|W_\rho\|_\infty \leq \left(\frac{2}{\pi}\right)^m. \quad (30)$$

From Eqs. (28) and (30) and Hölder's inequality we see that $W_\rho \in L^q(\mathbb{C}^m)$ for all $q \in [2, \infty]$. However, for all $q \in [1, 2)$ there exist explicit (pure) states ρ such that $W_\rho \notin L^q(\mathbb{C}^m)$ (see Theorem 2 in Ref. [35]). The set of functions $\psi \in L^2(\mathbb{R}^m)$ whose Wigner function of the corresponding (possibly unnormalized) pure state satisfies $W_{|\psi\rangle\langle\psi|} \in L^1(\mathbb{C}^m)$ is hence a proper subset of $L^2(\mathbb{R}^m)$. It is called the “Feichtinger algebra” and possesses rich mathematical properties which are extensively studied in the literature (see Chapter 7 in Ref. [64]).

III. BLACK BOX OVERLAP ESTIMATION

We consider some measurable space Ω with measure μ and use the notation from Sec. II B. Let $r > 0$ be some number, which is fixed throughout the whole section. We say one has *black box access to some function* $g \in B_r^\infty(0)$ if for every $\lambda \in \Omega$ one can obtain samples from the unique [65] random variable $Y_\lambda \equiv Y_\lambda(g)$ with outcome space $\{r, -r\}$ and

$$\mathbb{E}[Y_\lambda] = g(\lambda). \quad (31)$$

In the following we consider subsets $\mathcal{S} \subseteq B_r^\infty(0)$ and assume we have black box access to the functions $g \in \mathcal{S}$.

With that terminology the task of *black box overlap estimation* is defined to be the following: Let $f \in L^2(\Omega, \mathbb{R})$ and $\mathcal{S} \subseteq L^2(\Omega, \mathbb{R}) \cap B_r^\infty(0)$. Then, provided we have full access to the function f and black box access to a function $g \in \mathcal{S}$, the goal is to estimate

$$\int_\Omega f(\lambda)g(\lambda)d\mu(\lambda) = \int_\Omega f(\lambda)\mathbb{E}[Y_\lambda]d\mu(\lambda) \quad (32)$$

using as few samples as possible of the random variable $Y_\lambda \equiv Y_\lambda(g)$ for different points $\lambda \in \Omega$. More precisely, for $N \in \mathbb{N}$, an *N-query algorithm* $\mathcal{A}$ for the task of black box overlap estimation takes points $\lambda_1, \dots, \lambda_N \in \Omega$ and for each λ_t a sample $y_{\lambda_t} \in \{r, -r\}$ from Y_{λ_t} . Here, we allow for adaptive algorithms, meaning that the points λ_t , or their distributions when chosen at random, can depend on f and the previous points $\lambda_1, \dots, \lambda_{t-1}$ and corresponding samples, i.e.,

$$\begin{aligned} \lambda_1 &\equiv \lambda_1(f) && \longrightarrow y_{\lambda_1} \text{ sample from } Y_{\lambda_1} \\ \lambda_2 &\equiv \lambda_2(f, \lambda_1, y_{\lambda_1}) && \longrightarrow y_{\lambda_2} \text{ sample from } Y_{\lambda_2} \\ &\vdots && \vdots \\ \lambda_N &\equiv \lambda_N(f, \lambda_1, y_{\lambda_1}, \dots, \lambda_{N-1}, y_{\lambda_{N-1}}) && \longrightarrow y_{\lambda_N} \text{ sample from } Y_{\lambda_N}. \end{aligned} \quad (33)$$

The algorithm then outputs a value $\phi_{\mathcal{A}}(f, \lambda_1, y_{\lambda_1}, \dots, \lambda_N, y_{\lambda_N}) \in \mathbb{R}$. We demand that $\mathcal{A}$ works for any function $g \in \mathcal{S}$, i.e., for all $g \in \mathcal{S}$ fixed its output should estimate the corresponding overlap (32). We say the algorithm $\mathcal{A}$ has accuracy $\varepsilon > 0$ with failure probability bounded by $\delta > 0$ if for all $g \in \mathcal{S}$

$$\mathbb{P}\left(\left|\phi_{\mathcal{A}}(f, \lambda_1, y_{\lambda_1}, \dots, \lambda_N, y_{\lambda_N}) - \int_\Omega f(\lambda)g(\lambda)d\mu(\lambda)\right| > \varepsilon\right) \leq \delta.$$

The optimal sample complexity of black box overlap estimation for the function f and set $\mathcal{S}$ is hence defined as

$$N_{\varepsilon, \delta}(f, \mathcal{S}) = \inf\{N \in \mathbb{N} \mid \exists N\text{-query algorithm}$$

with accuracy ε and failure probability bounded by $\delta\}$.

Theorem 1, with the proof given in Sec. III A, proposes an algorithm achieving this objective and provides an upper bound on the number of points λ needed to take samples from Y_λ . Furthermore, Theorem 2 provides corresponding lower bounds and is proved in Sec. III B. We see that in

the case of $\mathcal{S} \equiv \mathcal{S}_{\max} = B_1^2(0) \cap B_r^\infty(0)$, the optimal sample complexity is characterized by the smoothed L^1 -norm of f [see Eq. (8) for the definition] as we show for all $\varepsilon, \delta > 0$

$$\begin{aligned} \left(\frac{r\|f\|_1^{(2\varepsilon)}}{2\varepsilon} \right)^2 \ln \left(\frac{1}{3\delta} \right) &\leq N_{\varepsilon,\delta}(f, \mathcal{S}_{\max}) \\ &\leq 2 \inf_{\varepsilon' \in [0, \varepsilon]} \left(\frac{r\|f\|_1^{(\varepsilon')}}{\varepsilon - \varepsilon'} \right)^2 \ln \left(\frac{1}{\delta} \right). \end{aligned} \quad (34)$$

Theorem 1 (Upper bound for black box overlap estimation). Let $f \in L^2(\Omega, \mathbb{R})$, $r > 0$ and $\mathcal{S} \subseteq B_1^2(0) \cap B_r^\infty(0)$. Then we have

$$N_{\varepsilon,\delta}(f, \mathcal{S}) \leq 2 \inf_{\varepsilon' \in [0, \varepsilon]} \left(\frac{r\|f\|_1^{(\varepsilon')}}{\varepsilon - \varepsilon'} \right)^2 \ln \left(\frac{1}{\delta} \right). \quad (35)$$

The upper bound holds true even when we restrict ourselves to nonadaptive algorithms, i.e., where the λ_t in Eq. (33) can depend on the function f but not on the previous points $\lambda_1, \dots, \lambda_{t-1}$ and the corresponding samples $\mathcal{Y}_{\lambda_1}, \dots, \mathcal{Y}_{\lambda_{t-1}}$.

Remark 1. The algorithm proposed in the proof of Theorem 1 also works more generally if instead of having access to the two-outcome random variable Y_λ , we have access to a general bounded random variable $\tilde{Y}_\lambda$ satisfying

$$|\tilde{Y}_\lambda| \leq r \quad \text{and} \quad \mathbb{E}[\tilde{Y}_\lambda] = g(\lambda) \quad (36)$$

for all $\lambda \in \Omega$. For this slightly changed model of black box overlap estimation, we obtain the same upper bound as in Eq. (35) for the corresponding sample complexity.

Theorem 2 (Lower bound for black box overlap estimation). Let $f \in L^2(\Omega, \mathbb{R})$, $r > 0$ and $\mathcal{S} \subseteq L^2(\Omega, \mathbb{R}) \cap B_r^\infty(0)$. Then for $\varepsilon, \delta > 0$ the sample complexity of black box overlap estimation is lower bounded as

$$N_{\varepsilon,\delta}(f, \mathcal{S}) \geq \sup_{\substack{g_1, g_2 \in \mathcal{S}, \\ |f f(g_1 - g_2)| > 2\varepsilon}} \left\| \frac{(g_1 - g_2)^2}{r^2 - g_2^2} \right\|_\infty^{-1} \ln \left(\frac{1}{3\delta} \right). \quad (37)$$

Furthermore, for $\mathcal{S} \equiv \mathcal{S}_{\max} = B_1^2(0) \cap B_r^\infty(0)$ and $f \in \mathcal{S}_{\max}$ with $\|f\|_2 = 1$ this gives

$$N_{\varepsilon,\delta}(f, \mathcal{S}) \geq \left(\frac{r\|f\|_1^{(2\varepsilon)}}{2\varepsilon} \right)^2 \ln \left(\frac{1}{3\delta} \right). \quad (38)$$

For a special class of functions f , satisfying a certain decay of tail condition, we show in Corollary 1 that the

upper and lower bounds in Eq. (34) match up to constants. Here, it is important to have good control over the behavior of the smoothed L^1 -norm $\|f\|_1^{(\varepsilon)}$ for different values of ε . As we see below, the sample complexity of black box overlap estimation is in this case then characterized by the usual L^1 -norm of the function f .

For $\gamma, \kappa > 0$ and $\Omega_0 \subseteq \Omega$ a finite measure set, we say a measurable function f has a *rapidly decaying tail of order* (γ, κ) *outside* Ω_0 if for all $\delta > 0$ we have

$$\int_{\{|f| \leq \delta\} \cap \Omega_0^c} |f(\lambda)| d\mu(\lambda) \leq \kappa \delta^\gamma. \quad (39)$$

Note that a function $f \in L^2(\Omega, \mathbb{R})$ satisfying this condition can be also shown to be in $L^1(\Omega, \mathbb{R})$. Functions satisfying this condition are discussed in detail in Appendix A 1. In particular, in Lemma A2 it is shown that their smoothed L^1 -norm satisfies the lower bound

$$\|f\|_1^{(\varepsilon)} \geq \|f\|_1 - \kappa' \varepsilon^\beta - \varepsilon \sqrt{\mu(\Omega_0)} \quad (40)$$

with $\beta := \frac{\gamma}{\gamma+1} \in (0, 1]$ and $\kappa' := 2\kappa^{1/(1+\gamma)}$.

Corollary 1. Let $f \in \mathcal{S}_{\max} = B_1^2(0) \cap B_r^\infty(0)$ with $\|f\|_2 = 1$ and a rapidly decaying tail of order (γ, κ) outside Ω_0 for some $\gamma, \kappa > 0$ and finite measure set $\Omega_0 \subseteq \Omega$ such that $\kappa, \sqrt{\mu(\Omega_0)} \leq c\|f\|_1$ for some universal constant $c \geq 0$. Then we have

$$N_{\varepsilon,\delta}(f, \mathcal{S}_{\max}) = \Theta \left(\left(\frac{r\|f\|_1}{\varepsilon} \right)^2 \ln \left(\frac{1}{\delta} \right) \right) \quad (41)$$

for $\varepsilon, \delta > 0$ small enough, i.e., for all $\varepsilon, \delta \in (0, c']$ where $c' > 0$ depends on c but not on f .

Proof. We get from Theorem 1 and the definition of the smoothed L^1 -norm in Eq. (19) the upper bound

$$N_{\varepsilon,\delta}(f, \mathcal{S}_{\max}) \leq 2 \left(\frac{r\|f\|_1}{\varepsilon} \right)^2 \ln \left(\frac{1}{\delta} \right).$$

On the other hand from Lemma A2 and the fact that $\kappa, \sqrt{\mu(\Omega_0)} \leq \mathcal{O}(\|f\|_1)$, we have

$$\|f\|_1^{(2\varepsilon)} \geq \|f\|_1 (1 - C\varepsilon^\beta)$$

for some $C \geq 0$ independent of f and ε . Hence, by the lower bound (38) in Theorem 2 we find for all $\varepsilon, \delta > 0$ small enough

$$\begin{aligned} N_{\varepsilon,\delta}(f, \mathcal{S}_{\max}) &\geq \left(\frac{r\|f\|_1^{(2\varepsilon)}}{2\varepsilon} \right)^2 \ln \left(\frac{1}{3\delta} \right) \\ &\gtrsim \left(\frac{r\|f\|_1}{2\varepsilon} \right)^2 \ln \left(\frac{1}{3\delta} \right). \end{aligned}$$

■

A. Proof of the upper bound on the sample complexity $N_{\varepsilon, \delta}(f, \mathcal{S})$

In this section we prove the upper bound on the sample complexity of black box overlap estimation given in Theorem 1. The proof works by providing a specific algorithm for the estimation task: Here, the tester chooses random points $\lambda \in \Omega$ by sampling from the distribution

$$p(\lambda) = \frac{|\tilde{f}(\lambda)|}{\|\tilde{f}\|_1}, \quad (42)$$

where $\tilde{f}$ is essentially the minimizer in the optimization of the smoothed L^1 -norm in Eq. (19). For every λ chosen, a sample of the bounded random variable $Y_\lambda \|\tilde{f}\|_1 \operatorname{sgn}(f(\lambda))$ is obtained. With use of these samples, an empirical estimator is calculated which, provided the number of samples is high enough, is close to the overlap

$$\int_{\Omega} \tilde{f}(\lambda) g(\lambda) d\mu(\lambda)$$

by Hoeffding's inequality [66]. This overlap approximates the desired one in Eq. (32) since $\tilde{f}$ and the target function f are close in the L^2 -norm.

In Ref. [33] the concrete task of fidelity estimation in the Wigner model is considered. There, the authors use the sampling distribution (42), however with the smoothing function $\tilde{f}$ replaced by the (nonsmoothed) Wigner function of the target state. Note that for continuous variable systems there exist pure states whose Wigner functions have an infinite L^1 -norm (see Theorem 2 in Ref. [35]) and for which, hence, the sampling procedure in Ref. [33] is not well-defined. Similarly, in Refs. [25,34] the study authors use the sampling distribution (42) relying on (nonsmoothed) Wigner functions for the tasks of estimating the probabilities of outcomes of a quantum circuit and efficient classical estimation of quantum kernel functions respectively.

In Ref. [32] the problem of estimating expectation values of observables in the Pauli model is considered, with the sampling procedure relying on the distribution (42) with $\tilde{f}$ being replaced by the (nonsmoothed) characteristic function of the respective quantum state. However, the analysis of the resulting sample complexity for the proposed algorithm is different from ours.

In Refs. [1,2] a similar algorithm is considered for the concrete case of fidelity estimation in the Pauli model. Formulating their approach in terms in the abstract black box overlap estimation framework, the study authors consider the distribution

$$p_2(\lambda) = \frac{f^2(\lambda)}{\|f\|_2^2}, \quad (43)$$

instead of the distribution $p(\lambda)$ above, for sampling $\lambda \in \Omega$. For every λ chosen they then take samples from the

random variable $\|f\|_2^2 Y_\lambda / f(\lambda)$ to construct an empirical estimator for the overlap (32). This random variable is no longer bounded, leading to a worse upper bound on the corresponding sample complexity for many functions f .

Proof of Theorem 1. Let $\varepsilon > 0$, $\varepsilon' \in [0, \varepsilon]$ and $\tilde{f} \in L^2(\Omega, \mathbb{R}) \cap L^1(\Omega, \mathbb{R})$ such that $\|f - \tilde{f}\|_2 \leq \varepsilon'$. Note that such a function always exists since $L^2(\Omega, \mathbb{R}) \cap L^1(\Omega, \mathbb{R})$ is dense in $L^2(\Omega, \mathbb{R})$ as remarked in Sec. II B. Consider the random variable Λ with outcome space Ω and distributed according to the probability density

$$p(\lambda) = \frac{|\tilde{f}(\lambda)|}{\|\tilde{f}\|_1}.$$

Let $g \in \mathcal{S} \subseteq B_1^2(0) \cap B_r^\infty(0)$, for which we assume we have black box access, and let $Y_\lambda \equiv Y_\lambda(g)$ be the corresponding random variable with outcome space $\{r, -r\}$ satisfying Eq. (31). Consider the random variable

$$X = Y_\Lambda \|\tilde{f}\|_1 \operatorname{sgn}(\tilde{f}(\Lambda)).$$

Note that by definition we have

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[Y_\Lambda \|\tilde{f}\|_1 \operatorname{sgn}(\tilde{f}(\Lambda))] \\ &= \int_{\Omega} \frac{|\tilde{f}(\lambda)|}{\|\tilde{f}\|_1} g(\lambda) \|\tilde{f}\|_1 \operatorname{sgn}(\tilde{f}(\lambda)) d\mu(\lambda) \\ &= \int_{\Omega} \tilde{f}(\lambda) g(\lambda) d\mu(\lambda). \end{aligned}$$

For $N \in \mathbb{N}$ to be determined later we take samples $\lambda_1, \dots, \lambda_N$ of Λ and then for each λ_i a sample y_{λ_i} of Y_{λ_i} . Furthermore, we define the empirical average

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N y_{\lambda_i} \|\tilde{f}\|_1 \operatorname{sgn}(\tilde{f}(\lambda_i)).$$

Noting that by definition and the assumption (31) the random variable X is bounded as

$$|X| \leq r \|\tilde{f}\|_1,$$

we can use Hoeffding's inequality [66], which gives

$$\begin{aligned} \mathbb{P}\left(\left|\bar{X} - \int_{\Omega} \tilde{f}(\lambda) g(\lambda) d\mu(\lambda)\right| \geq \varepsilon - \varepsilon'\right) \\ \leq \exp\left(-\frac{N(\varepsilon - \varepsilon')^2}{2r^2 \|\tilde{f}\|_1^2}\right). \end{aligned} \quad (44)$$

Furthermore, note that by the Cauchy-Schwarz inequality and the assumption that $\|g\|_2 \leq 1$ we have

$$\begin{aligned} \left|\int_{\Omega} f(\lambda) g(\lambda) d\mu(\lambda) - \int_{\Omega} \tilde{f}(\lambda) g(\lambda) d\mu(\lambda)\right| \\ \leq \|f - \tilde{f}\|_2 \leq \varepsilon'. \end{aligned}$$

Hence, we see that for $\delta > 0$ a total number of

$$N = 2 \left(\frac{r \|\tilde{f}\|_1}{\varepsilon - \varepsilon'} \right)^2 \ln \left(\frac{1}{\delta} \right)$$

samples suffices to have with probability greater than $1 - \delta$

$$\left| \bar{X} - \int_{\Omega} \tilde{f}(\lambda) g(\lambda) d\mu(\lambda) \right| \leq \varepsilon - \varepsilon' \quad \text{and} \quad \left| \int_{\Omega} f(\lambda) g(\lambda) d\mu(\lambda) - \int_{\Omega} \tilde{f}(\lambda) g(\lambda) d\mu(\lambda) \right| \leq \varepsilon'$$

and thus, by the triangle inequality, give a good approximation of the desired overlap as

$$\begin{aligned} \left| \int_{\Omega} f(\lambda) g(\lambda) d\mu(\lambda) - \bar{X} \right| &\leq \left| \int_{\Omega} f(\lambda) g(\lambda) d\mu(\lambda) - \int_{\Omega} \tilde{f}(\lambda) g(\lambda) d\mu(\lambda) \right| + \left| \bar{X} - \int_{\Omega} \tilde{f}(\lambda) g(\lambda) d\mu(\lambda) \right| \\ &\leq \varepsilon' + \varepsilon - \varepsilon' = \varepsilon. \end{aligned}$$

Since $\varepsilon' \in [0, \varepsilon)$ and $\tilde{f}$ under the constraint above were arbitrary, this proves Eq. (35). ■

B. Proof of the lower bound on the sample complexity $N_{\varepsilon, \delta}(f, \mathcal{S})$

Here, we prove the lower bound on the sample complexity of black box overlap estimation given in Theorem 2. The first statement of Theorem 2, i.e. Eq. (37), follows by some information-theoretic argument in which we show the hardness of the estimation task by embedding the problem of distinguishing two hypotheses. For the second statement of Theorem 2, i.e. Eq. (38), showing the lower bound for the choice $\mathcal{S} \equiv \mathcal{S}_{\max} = B_1^2(0) \cap B_r^{\infty}(0)$, we use the established Eq. (37) together with a specific test function $g \in \mathcal{S}_{\max}$, whose existence is provided in Lemma 1. We first state and prove this lemma and then give the proof of Theorem 2.

Lemma 1. Let $f \in L^2(\Omega, \mathbb{R})$ and $\varepsilon \in (0, \|f\|_2)$. Then there exists $g \in L^2(\Omega, \mathbb{R}) \cap L^{\infty}(\Omega, \mathbb{R})$ such that the following three points hold: (1) $\int_{\Omega} f(\lambda) g(\lambda) d\mu(\lambda) \geq \varepsilon$, (2) $\|g\|_2 \leq 1$ and (3) $\|g\|_{\infty} \leq \varepsilon / \|f\|_1^{(\varepsilon)}$.

Remark 2. In the context of Secs. IV and V we consider the specific choices of $\mathcal{S}$ being either the set of Wigner functions or characteristic functions for valid quantum states on $L^2(\mathbb{R}^m)$ or $(\mathbb{C})^{\otimes n}$ respectively. There, the f considered is a Wigner function or a characteristic function of some pure state ρ . However, the function g constructed in Lemma 1 does in general not correspond to the Wigner function or characteristic function of some state σ . Hence, in the context of fidelity estimation for continuous or discrete variable systems in Secs. IV and V, we cannot use Lemma 1 to establish the corresponding lower bound (38) when σ is assumed to be a valid state.

Proof of Lemma 1. In the following we explicitly construct a function $\tilde{f} \in L^1(\Omega, \mathbb{R}) \cap L^2(\Omega, \mathbb{R})$ with

$$\|f - \tilde{f}\|_2 \leq \varepsilon \quad (45)$$

and furthermore $g \in L^2(\Omega, \mathbb{R}) \cap L^{\infty}(\Omega, \mathbb{R})$ satisfying points (1) and (2) and furthermore

$$\|g\|_{\infty} \leq \frac{\varepsilon}{\|\tilde{f}\|_1} \leq \frac{\varepsilon}{\|f\|_1^{(\varepsilon)}},$$

where the last inequality followed by the definition of the smoothed L^1 -norm in Eq. (19).

For that, note first that by the dominated convergence theorem we have $\lim_{c \rightarrow 0} \int_{\{|f| < c\}} |f(\lambda)|^2 d\mu(\lambda) = 0$ and hence we can define

$$c_{\star} = \sup \left\{ c \in [0, \infty] \mid \int_{\{|f| < c\}} |f(\lambda)|^2 d\mu(\lambda) \leq \varepsilon^2 \right\} > 0.$$

Note that since $\varepsilon < \|f\|_2$ we have $c_{\star} < \infty$. Furthermore, since for all $\lambda \in \Omega$ we have $\lim_{c \uparrow c_{\star}} |f(\lambda)|^2 \mathbf{1}_{\{|f| < c\}}(\lambda) = |f(\lambda)|^2 \mathbf{1}_{\{|f| < c_{\star}\}}(\lambda)$, we have again by the dominated convergence theorem

$$\int_{\{|f| < c_{\star}\}} |f(\lambda)|^2 d\mu(\lambda) = \lim_{c \uparrow c_{\star}} \int_{\{|f| < c\}} |f(\lambda)|^2 d\mu(\lambda) \leq \varepsilon^2. \quad (46)$$

Therefore, we can define

$$\tilde{f} = f \mathbf{1}_{\{|f| \geq c_{\star}\}} \neq 0, \quad (47)$$

which by Eq. (46) approximates f in the L^2 -norm, i.e., satisfies Eq. (45). Note that by definition we have that

$\tilde{f} \in L^1(\Omega, \mathbb{R})$ as

$$\begin{aligned} \|\tilde{f}\|_1 &= \int_{\{|f| \geq c_\star\}} |f(\lambda)| d\mu(\lambda) \\ &\leq \frac{1}{c_\star} \int_{\{|f| \geq c_\star\}} |f(\lambda)|^2 d\mu(\lambda) \leq \frac{\|f\|_2^2}{c_\star}. \end{aligned}$$

We define the function g differently depending on whether $c_\star < \varepsilon^2/\|\tilde{f}\|_1$ or $c_\star \geq \varepsilon^2/\|\tilde{f}\|_1$. In the first case $c_\star < \varepsilon^2/\|\tilde{f}\|_1$, we define

$$h = f \mathbf{1}_{\{|f| < \frac{\varepsilon^2}{\|\tilde{f}\|_1}\}}$$

and the normalized function $g = h/\|h\|_2$. By the definition of $c_\star$ and using $\varepsilon^2/\|\tilde{f}\|_1 > c_\star$, we have

$$\|h\|_2 = \sqrt{\int_{\{|f| < \frac{\varepsilon^2}{\|\tilde{f}\|_1}\}} |f(\lambda)|^2 d\mu(\lambda)} > \varepsilon \quad (48)$$

and hence

$$\|g\|_\infty = \frac{1}{\|h\|_2} \left\| |f| \mathbf{1}_{\{|f| < \frac{\varepsilon^2}{\|\tilde{f}\|_1}\}} \right\|_\infty \leq \frac{\varepsilon}{\|\tilde{f}\|_1}.$$

Furthermore, using Eq. (48) again, we see

$$\int_{\Omega} f(\lambda) g(\lambda) d\mu(\lambda) = \sqrt{\int_{\{|f| < \frac{\varepsilon^2}{\|\tilde{f}\|_1}\}} |f(\lambda)|^2 d\mu(\lambda)} \geq \varepsilon.$$

Therefore, the function g satisfies points (1)–(3).

On the other hand, in the case $c_\star \geq \varepsilon^2/\|\tilde{f}\|_1$ we define

$$g = \frac{\varepsilon}{\|\tilde{f}\|_1} \operatorname{sgn}(f) \mathbf{1}_{\{|f| \geq c_\star\}}.$$

Clearly, we have $\|g\|_\infty \leq \varepsilon/\|\tilde{f}\|_1$ and furthermore by the definition of $\tilde{f}$ in Eq. (47)

$$\int_{\Omega} f(\lambda) g(\lambda) d\mu(\lambda) = \frac{\varepsilon}{\|\tilde{f}\|_1} \int_{\{|f| \geq c_\star\}} |f(\lambda)| d\mu(\lambda) = \varepsilon.$$

Lastly, we have using $c_\star \geq \varepsilon^2/\|\tilde{f}\|_1$ that also

$$\begin{aligned} \|g\|_2^2 &= \frac{\varepsilon^2}{\|\tilde{f}\|_1^2} \mu(\{|f| \geq c_\star\}) \\ &\leq \frac{\varepsilon^2}{\|\tilde{f}\|_1^2 c_\star} \int_{\{|f| \geq c_\star\}} |f(\lambda)| d\mu(\lambda) = \frac{\varepsilon^2}{\|\tilde{f}\|_1 c_\star} \leq 1 \end{aligned}$$

and hence, also in this case, the defined function g satisfies points (1)–(3). $\blacksquare$

Proof of Theorem 2. Let $g_1, g_2 \in \mathcal{S}$ such that

$$\left| \int_{\Omega} f(\lambda) (g_1(\lambda) - g_2(\lambda)) d\mu(\lambda) \right| > 2\varepsilon. \quad (49)$$

If such functions do not exist, Eq. (37) is trivially satisfied. Consider for $i = 1, 2$ and $\lambda \in \Omega$ the random variable $Y_\lambda^{(i)}$ with outcome space $\{r, -r\}$, with $r > 0$ being the number which was fixed at the beginning of Sec. III, and distributed according to

$$\begin{aligned} \mathbb{P}(Y_\lambda^{(i)} = r) &= \frac{1 + g_i(\lambda)/r}{2}, \quad \mathbb{P}(Y_\lambda^{(i)} = -r) \\ &= \frac{1 - g_i(\lambda)/r}{2}. \end{aligned} \quad (50)$$

Note that both pairs $(g_1, (Y_\lambda^{(1)})_{\lambda \in \Omega})$ and $(g_2, (Y_\lambda^{(2)})_{\lambda \in \Omega})$ satisfy the assumption (31). We consider two situations:

- (1) $\mathcal{H}_1$: The tester has black box access to g_1 through the random variable $Y_\lambda^{(1)}$.
- (2) $\mathcal{H}_2$: The tester has black box access to g_2 through the random variable $Y_\lambda^{(2)}$.

Hence, if

$$(g, (Y_\lambda)_{\lambda \in \Omega}) \in \left\{ \left(g_1, (Y_\lambda^{(1)})_{\lambda \in \Omega} \right), \left(g_2, (Y_\lambda^{(2)})_{\lambda \in \Omega} \right) \right\},$$

and with the use of Eq. (49) a black box overlap estimation algorithm with success probability $1 - \delta$ can find which pair is present only by estimating the overlap $\int_{\Omega} f(\lambda) g(\lambda) d\mu(\lambda)$ up to additive error ε . Let N be a sufficient number of steps for this test and let $\lambda_1, \dots, \lambda_N \in \Omega$ be the points where the tester chooses to sample the random variable Y_λ . Let $y_{\lambda_1}, \dots, y_{\lambda_N} \in \{r, -r\}$ be the corresponding samples obtained. As explained around Eq. (33), we allow for the algorithm to be adaptive, i.e., at each step t the variable λ_t can depend on the previous points $\lambda_1, \dots, \lambda_{t-1}$, outcomes $y_{\lambda_1}, \dots, y_{\lambda_{t-1}}$ and the function f , i.e.,

$$\lambda_t \equiv \lambda_t(f, \lambda_1, y_{\lambda_1}, \dots, \lambda_{t-1}, y_{\lambda_{t-1}}).$$

We denote $Z_t = (\lambda_t, Y_{\lambda_t})$ and let $\mathbb{P}_{\mathcal{H}_1}^{Z_1, \dots, Z_N}$ be the distribution of $Z_1, \dots, Z_N$ under $\mathcal{H}_1$ and let $\mathbb{P}_{\mathcal{H}_2}^{Z_1, \dots, Z_N}$ be the law of $Z_1, \dots, Z_N$ under $\mathcal{H}_2$. Let $\mathcal{E}$ be the event that the tester returns the first hypothesis. We have by the data processing inequality of the D divergence

$$\begin{aligned} D\left(\mathbb{P}_{\mathcal{H}_1}^{Z_1, \dots, Z_N} \parallel \mathbb{P}_{\mathcal{H}_2}^{Z_1, \dots, Z_N}\right) &\geq D(\mathbb{P}_{\mathcal{H}_1}(\mathcal{E}) \parallel \mathbb{P}_{\mathcal{H}_2}(\mathcal{E})) \\ &\geq D(1 - \delta \parallel \delta) \\ &\geq \ln\left(\frac{1}{3\delta}\right). \end{aligned} \quad (51)$$

On the other hand, using the notation $\mathbf{Z}_{<t} = (Z_1, \dots, Z_{t-1})$ for $t \in [N]$, we have

$$\begin{aligned}
D\left(\mathbb{P}_{\mathcal{H}_1}^{Z_1, \dots, Z_N} \parallel \mathbb{P}_{\mathcal{H}_2}^{Z_1, \dots, Z_N}\right) &\stackrel{(a)}{=} \sum_{t=1}^N \mathbb{E}_{\mathbf{Z}_{<t} \sim \mathbb{P}_{\mathcal{H}_1}} \left[D\left(\mathbb{P}_{\mathcal{H}_1}^{Z_t | \mathbf{Z}_{<t}} \parallel \mathbb{P}_{\mathcal{H}_2}^{Z_t | \mathbf{Z}_{<t}}\right) \right] \\
&\stackrel{(b)}{=} \sum_{t=1}^N \mathbb{E}_{(\lambda_t, \mathbf{Z}_{<t}) \sim \mathbb{P}_{\mathcal{H}_1}} \left[D\left(\mathbb{P}_{\mathcal{H}_1}^{Y_{\lambda_t} | (\lambda_t, \mathbf{Z}_{<t})} \parallel \mathbb{P}_{\mathcal{H}_2}^{Y_{\lambda_t} | (\lambda_t, \mathbf{Z}_{<t})}\right) \right] \\
&\stackrel{(c)}{\leq} \sum_{t=1}^N \mathbb{E}_{(\lambda_t, \mathbf{Z}_{<t}) \sim \mathbb{P}_{\mathcal{H}_1}} \left[\chi^2\left(\mathbb{P}_{\mathcal{H}_1}^{Y_{\lambda_t} | (\lambda_t, \mathbf{Z}_{<t})} \parallel \mathbb{P}_{\mathcal{H}_2}^{Y_{\lambda_t} | (\lambda_t, \mathbf{Z}_{<t})}\right) \right] \\
&\stackrel{(d)}{=} \sum_{t=1}^N \mathbb{E}_{(\lambda_t, \mathbf{Z}_{<t}) \sim \mathbb{P}_{\mathcal{H}_1}} \left[\frac{1}{2} \left(\frac{(1 + g_1(\lambda_t)/r)^2}{1 + g_2(\lambda_t)/r} + \frac{(1 - g_1(\lambda_t)/r)^2}{1 - g_2(\lambda_t)/r} \right) - 1 \right] \\
&= \sum_{t=1}^N \mathbb{E}_{(\lambda_t, \mathbf{Z}_{<t}) \sim \mathbb{P}_{\mathcal{H}_1}} \left[\frac{(g_1(\lambda_t) - g_2(\lambda_t))^2}{r^2 - g_2^2(\lambda_t)} \right] \\
&\leq N \left\| \frac{(g_1 - g_2)^2}{r^2 - g_2^2} \right\|_{\infty}, \tag{52}
\end{aligned}$$

where in (a) we have used the chain rule of the D divergence; in (b) we have used the fact that the distribution of λ_t conditioned on $\mathbf{Z}_{<t}$ does not depend on $\mathcal{H}_1$ or $\mathcal{H}_2$ combined again with the chain rule; in (c) we have used the well-known inequality $D \leq \chi^2$ [see Eq. (21)]; in (d) we have used the definition of Y_{λ} [see Eq. (50)] and the definition of the χ^2 -divergence [see Eq. (20)].

Therefore, combining the inequalities (51) and (52), we obtain

$$\ln\left(\frac{1}{3\delta}\right) \leq D\left(\mathbb{P}_{\mathcal{H}_1}^{Z_1, \dots, Z_N} \parallel \mathbb{P}_{\mathcal{H}_2}^{Z_1, \dots, Z_N}\right) \leq N \left\| \frac{(g_1 - g_2)^2}{r^2 - g_2^2} \right\|_{\infty}$$

and hence

$$N \geq \left\| \frac{(g_1 - g_2)^2}{r^2 - g_2^2} \right\|_{\infty}^{-1} \ln\left(\frac{1}{3\delta}\right).$$

Since $g_1, g_2 \in \mathcal{S}$ under the constraint (49) were arbitrary, this proves Eq. (37).

To prove Eq. (38) in the case of $\mathcal{S} = \mathcal{S}_{\max} = B_1^2(0) \cap B_r^{\infty}(0)$ and $f \in \mathcal{S}_{\max}$ with $\|f\|_2 = 1$, we first use the fact that since $g_2 = 0 \in \mathcal{S}_{\max}$ we can conclude from Eq. (37) that

$$N_{\varepsilon, \delta}(f, \mathcal{S}_{\max}) \geq \sup_{\substack{g \in \mathcal{S}_{\max}, \\ |fg| \geq 2\varepsilon}} \frac{r^2}{\|g\|_{\infty}^2} \ln\left(\frac{1}{3\delta}\right).$$

Furthermore, we can, without loss of generality, restrict ourselves to $2\varepsilon < \|f\|_2 = 1$ as otherwise $\|f\|_1^{(2\varepsilon)} = 0$ and hence Eq. (38) is trivially satisfied.

Under this constraint, we want to show

$$\sup_{\substack{g \in \mathcal{S}_{\max}, \\ |fg| \geq 2\varepsilon}} \frac{1}{\|g\|_{\infty}^2} = \sup_{\substack{g \in \mathcal{S}_{\max}, \\ |fg| \geq 2\varepsilon}} \frac{1}{\|g\|_{\infty}^2}. \tag{53}$$

Note that the left-hand side of Eq. (53) is clearly less than or equal to the right-hand side, so we need to show only the opposite inequality. Let $(\varepsilon_n)_{n \in \mathbb{N}} \subseteq (0, \infty)$ be a sequence such that $\varepsilon_n > \varepsilon$ and $\lim_{n \rightarrow \infty} \varepsilon_n = \varepsilon$. Define furthermore

$$\kappa_n = \frac{2(\varepsilon_n - \varepsilon)}{1 - 2\varepsilon}, \tag{54}$$

which for $n \in \mathbb{N}$ large enough satisfies $\kappa_n \in (0, 1]$ and to which we shall restrict ourselves in the following. Let $g \in \mathcal{S}_{\max}$ such that $|\int_{\Omega} f(\lambda)g(\lambda)d\mu(\lambda)| \geq 2\varepsilon$ and define

$$g_n = (1 - \kappa_n)g + \kappa_n \operatorname{sgn}\left(\int_{\Omega} f(\lambda)g(\lambda)d\mu(\lambda)\right)f. \tag{55}$$

Note first of all that since $f, g \in \mathcal{S}_{\max} = B_1^2(0) \cap B_r^{\infty}(0)$ we also have by the triangle inequality that $g_n \in \mathcal{S}_{\max}$.

Moreover, we have

$$\left| \int_{\Omega} f(\lambda) g_n(\lambda) d\mu(\lambda) \right| = (1 - \kappa_n) \left| \int_{\Omega} f(\lambda) g(\lambda) d\mu(\lambda) \right| + \kappa_n \geq 2\varepsilon_n > 2\varepsilon.$$

Since we also have

$$\|g_n\|_{\infty} \leq (1 - \kappa_n) \|g\|_{\infty} + \kappa_n \|f\|_{\infty} \xrightarrow{n \rightarrow \infty} \|g\|_{\infty},$$

this proves Eq. (53).

To prove Eq. (38) note that we can restrict ourselves to $\delta \leq 1/3$ as otherwise Eq. (38) is trivially satisfied. For the case $2\varepsilon/\|f\|_1^{(2\varepsilon)} > r$ the statement follows by simply lower bounding the supremum in Eq. (53) by choosing the feasible function $g = f$. On the other hand, for the case $2\varepsilon/\|f\|_1^{(2\varepsilon)} \leq r$, we can use Lemma 1 and use the function g (with parameter 2ε) from there as a feasible function, which gives

$$\begin{aligned} N_{\varepsilon, \delta}(f, \mathcal{S}_{\max}) &\geq \sup_{\substack{g \in \mathcal{S}_{\max}, \\ |f|g| \geq 2\varepsilon}} \frac{r^2}{\|g\|_{\infty}^2} \ln \left(\frac{1}{3\delta} \right) \\ &\geq \left(\frac{r\|f\|_1^{(2\varepsilon)}}{2\varepsilon} \right)^2 \ln \left(\frac{1}{3\delta} \right) \end{aligned}$$

and finishes the proof. $\blacksquare$

IV. FIDELITY ESTIMATION OF CONTINUOUS VARIABLE SYSTEMS

In this section we consider for $m \in \mathbb{N}$ an m -mode harmonic oscillator on the Hilbert space $\mathcal{H} = L^2(\mathbb{R}^m)$. We study the task of estimating the fidelity between a desired pure state ρ on $L^2(\mathbb{R}^m)$, for which we have a theoretical description, and a general state σ , which is prepared experimentally, by performing measurements on σ . Since ρ is a pure state, we can write the fidelity by Eq. (27) as

$$F(\rho, \sigma) = \pi^m \int_{\mathbb{C}^m} W_{\rho}(\alpha) W_{\sigma}(\alpha) d\alpha. \quad (56)$$

Assume that we are able to perform measurements of the displaced parity operator $(2/\pi)^m D(\alpha) \Pi D(\alpha)$ with outcomes in $\{(2/\pi)^m, -(2/\pi)^m\}$ on the state σ for chosen values of $\alpha \in \mathbb{C}^m$. Here, the values of α can be chosen adaptively and hence depend on the previous values and corresponding measurement outcomes. In the following we refer to this model as the “Wigner model of fidelity estimation.”

For this model, $\varepsilon > 0$ and $\delta > 0$ being the allowed additive error and probability of failure respectively and pure state ρ fixed, which we call the “instanced-based setup,”

we denote the *sample complexity of fidelity estimation in the Wigner model* by $N_{\varepsilon, \delta}^W(\rho)$: that is, $N_{\varepsilon, \delta}^W(\rho)$ is the optimal number of measurements on different copies of the unknown state σ needed to estimate $F(\rho, \sigma)$ in the worst case over all states σ .

Due to the relation (29), the assumption that we are able to perform measurements of the displaced parity operator on σ exactly means that we have black box access to the Wigner function W_{σ} . Hence, estimation of the fidelity (56) in this setup becomes a special instance of black box overlap estimation as discussed in Sec. III. In particular we can make the identifications

$$\begin{aligned} \Omega &\mapsto \mathbb{C}^m \cong \mathbb{R}^{2m} \quad \text{and} \quad d\mu(\lambda) \mapsto \pi^m d\alpha \\ r &\mapsto (2/\pi)^m \\ f(\lambda) &\mapsto W_{\rho}(\alpha) \quad \text{to which we have full access} \\ g(\lambda) &\mapsto W_{\sigma}(\alpha) \quad \text{to which we have black box access} \\ Y_{\lambda} &\mapsto Y_{\alpha} \quad \text{measurement outcome of observable} \\ &\quad (2/\pi)^m D(\alpha) \Pi D(\alpha) \text{ on state } \sigma. \end{aligned} \quad (57)$$

Since $\mu \equiv \mu_W$ is given by the Lebesgue measure rescaled by π^m , we introduce the corresponding L^q -norms for $q \in [1, \infty)$ for notational clarity and to distinguish them from the usual L^q -norms by

$$\begin{aligned} \|W_{\rho}\|_{L^q(\mathbb{C}^m, \mu_W)} &= \left(\int_{\mathbb{C}^m} |W_{\rho}(\alpha)|^q d\mu_W(\alpha) \right)^{1/q} \\ &= \pi^{m/q} \|W_{\rho}\|_q. \end{aligned} \quad (58)$$

Note in particular, with that convention, we have from Eq. (28) for all states ρ

$$\|W_{\rho}\|_{L^2(\mathbb{C}^m, \mu_W)} \leq 1,$$

with equality if the state is pure.

For the subset $\mathcal{S}$ used in Sec. III, we are interested here in the particular choice $\mathcal{S}_W$ consisting of all Wigner functions W_{σ} for valid states σ . Using Eqs. (28) and (30), we see that $\mathcal{S}_W \subseteq B_1^2(0) \cap B_{(2/\pi)^m}^{\infty}(0)$, again with balls with respect to the rescaled Lebesgue measure $d\mu(\alpha) = \pi^m d\alpha$.

With the notation above, the sample complexity of fidelity estimation can be written as

$$N_{\varepsilon, \delta}^W(\rho) \equiv N_{\varepsilon, \delta}(W_{\rho}, \mathcal{S}_W), \quad (59)$$

with $N_{\varepsilon, \delta}(W_{\rho}, \mathcal{S}_W)$ being defined in Sec. III.

Formulating the achievability result of black box estimation, Theorem 1, for this special case leads to following result:

Theorem 3 (Upper bound on fidelity estimation for CV systems). Let ρ be a pure state on $L^2(\mathbb{R}^m)$ for which

we have a theoretical description. Then, for any state σ , we can estimate the fidelity $F(\rho, \sigma)$ with precision $\varepsilon > 0$ and failure probability at most $\delta > 0$ by measuring the displaced parity operator on number

$$N_{\varepsilon, \delta}^W(\rho) \leq 2^{2m+1} \inf_{\varepsilon' \in [0, \varepsilon)} \left(\frac{\|W_\rho\|_1^{(\varepsilon')}}{\varepsilon - \varepsilon'} \right)^2 \ln \left(\frac{1}{\delta} \right) \quad (60)$$

copies of the state σ . Note that by the argument around Eq. (19), this in particular shows $N_{\varepsilon, \delta}^W(\rho) < \infty$. Furthermore, the upper bound holds true even when we restrict ourselves to nonadaptive algorithms.

Remark 3 (Relation to energy constraints). For the tasks of shadow and full state tomography of CV systems, the proposed algorithms in the literature usually require that the measured state has finite energy with respect to the bosonic number operator [45, 46, 51]. This assumption, which is often referred to as an energy constraint, is used to localize the state within a bounded region in the overall infinite phase space.

In contrast, for the task of fidelity estimation considered here, localizing the measured state σ within phase space in a similar manner is not necessary. Intuitively, this is because, with the use of Eq. (56), the relevant part of phase space is determined by the regions where the (absolute value of the) Wigner function of the target state ρ is large. Hence, the algorithm proposed in the proof of Theorem 1, which is at the heart of the achievability result of Theorem 3, does not require bounded energy of the target state ρ or the prepared state σ and the obtained upper bound on the sample complexity is given in terms of the smoothed L^1 -norm of W_ρ instead. From this we can already conclude that energy is not the right quantity to determine hardness for the task of fidelity estimation. To see this, note that the smoothed L^1 -norm and hence also the sample complexity can remain bounded even for arbitrary high energies of the states as the following example shows: Consider for some state ρ the family $\rho_\alpha = D(\alpha)\rho D^*(\alpha)$ for arbitrary parameters $\alpha \in \mathbb{C}^m$. The resulting Wigner functions of ρ_α are simply translations of the original Wigner function of ρ , i.e., $W_{\rho_\alpha}(\beta) = W_\rho(\alpha + \beta)$. Hence, by translation invariance of the smoothed L^1 -norm, we have by Theorem 3 that the sample complexity of fidelity estimation stays uniformly bounded as

$$\begin{aligned} \sup_\alpha N_{\varepsilon, \delta}^W(\rho_\alpha) &\lesssim \sup_\alpha \inf_{\varepsilon' \in [0, \varepsilon)} \left(\frac{\|W_{\rho_\alpha}\|_1^{(\varepsilon')}}{\varepsilon - \varepsilon'} \right)^2 \ln \left(\frac{1}{\delta} \right) \\ &= \inf_{\varepsilon' \in [0, \varepsilon)} \left(\frac{\|W_\rho\|_1^{(\varepsilon')}}{\varepsilon - \varepsilon'} \right)^2 \ln \left(\frac{1}{\delta} \right) < \infty \end{aligned}$$

while on the other hand the energy blows up for large displacements as

$$\text{Tr}(N\rho_\alpha) \sim \text{Tr}(N\rho) + |\alpha|^2 \xrightarrow{|\alpha| \rightarrow \infty} \infty.$$

In contrast, understanding whether finite energy implies a finite L^1 -norm remains, to the best of our knowledge, an open problem. Additionally, we leave as an interesting open problem for future research the question of whether imposing additional energy constraints on the target state ρ can further improve the sample complexity of fidelity estimation.

Notably, the upper bound on the sample complexity of fidelity estimation of Theorem 3 holds true in the worst case over all prepared states σ . For certain target states ρ , however, it is plausible that the sample complexity can be improved when there is an additional energy constraint on σ . For example, performing full state tomography using the algorithm in Ref. [51] for prepared states σ with small energy and then calculating the fidelity $F(\rho, \sigma)$ using the knowledge of the target state ρ can require fewer measurements than performing direct fidelity estimation using the algorithms of Theorem 1 and 3 if the smoothed L^1 -norm of the Wigner function of ρ is large.

Remark 4 (Robustness regarding noise). In experiments, the quantum system corresponding to the prepared state σ is usually stored in a stationary bosonic mode, e.g., a resonator, to perform the displaced parity measurements. The most prominent noise present in this setup is photon loss. In Appendix B, we conduct a robustness analysis of the fidelity estimation protocol of Theorem 3. In particular, as argued there, without any assumption on the energy of the prepared state σ , the quantum system can be heavily distorted under photon loss [see Eq. (B1)], leading to inaccuracy of the fidelity estimation protocol.

However, under the natural assumption that preparing σ consumes only a finite amount of energy $E \geq 0$ with respect to the bosonic number, i.e., precisely that $\text{Tr}(N\sigma) \leq E$, and for target states ρ satisfying a similar energy constraint [see Eq. (B4)], we can bound the additional error due to photon loss when estimating the fidelity [see Eq. (B3)], showcasing the robustness of the protocol of Theorem 3.

Proof of Theorem 3. Here, with the conventions from Eq. (57), the smoothed L^1 -norm in Theorem 1 is given by $\|W_\rho\|_{L^1(\mathbb{C}^m, \mu_W)}^{(\varepsilon')} = \pi^m \|W_\rho\|_1^{(\varepsilon')}$. Hence, we can immediately read off from Eq. (35) that

$$\begin{aligned} N_{\varepsilon, \delta}^W(\rho) &\leq 2 \left(\frac{2}{\pi} \right)^{2m} \inf_{\varepsilon' \in [0, \varepsilon)} \left(\frac{\|W_\rho\|_{L^1(\mathbb{C}^m, \mu)}^{(\varepsilon')}}{\varepsilon - \varepsilon'} \right)^2 \ln \left(\frac{1}{\delta} \right) \\ &= 2^{2m+1} \inf_{\varepsilon' \in [0, \varepsilon)} \left(\frac{\|W_\rho\|_1^{(\varepsilon')}}{\varepsilon - \varepsilon'} \right)^2 \ln \left(\frac{1}{\delta} \right). \quad \blacksquare \end{aligned}$$

Furthermore, from the lower bound on the sample complexity of black box estimation in Theorem 2, we find the corresponding lower bound on $N_{\varepsilon,\delta}^W(\rho)$:

Theorem 4 (Lower bound on fidelity estimation for CV systems). Let ρ be a pure state on $L^2(\mathbb{R}^m)$ for which we have a theoretical description. Then for $0 < \varepsilon < 1/2$ and $0 < \delta \leq 1/3$ we have that the sample complexity of fidelity estimation in the Wigner model is lower bounded as

$$\begin{aligned} N_{\varepsilon,\delta}^W(\rho) &\geq \sup_{\substack{\sigma_1, \sigma_2 \text{ state,} \\ |F(\rho, \sigma_1) - F(\rho, \sigma_2)| > 2\varepsilon}} \left\| \frac{(W_{\sigma_1} - W_{\sigma_2})^2}{r^2 - W_{\sigma_2}^2} \right\|_{\infty}^{-1} \ln \left(\frac{1}{3\delta} \right) \\ &\geq \left(\frac{2}{\pi} \right)^{2m} \sup_{\substack{\sigma \text{ state,} \\ F(\rho, \sigma) \geq 2\varepsilon}} \frac{1}{\|W_{\sigma}\|_{\infty}^2} \ln \left(\frac{1}{3\delta} \right). \end{aligned} \quad (61)$$

Remark 5. Note that the restriction $0 < \varepsilon < 1/2$ in Theorem 4 is sensible as for allowed additive error $\varepsilon \geq 1/2$ one can always output $1/2$ to estimate $F(\rho, \sigma)$ with demanded precision without performing any measurements on σ . In other words, for $\varepsilon \geq 1/2$ we have the trivial result

$$N_{\varepsilon,\delta}^W(\rho) = 0.$$

Proof of Theorem 4. Using Eq. (56) and the identifications (57), we immediately get from Eq. (37) in Theorem 2 that

$$N_{\varepsilon,\delta}^W(\rho) \geq \sup_{\substack{\sigma_1, \sigma_2 \text{ state,} \\ |F(\rho, \sigma_1) - F(\rho, \sigma_2)| > 2\varepsilon}} \left\| \frac{(W_{\sigma_1} - W_{\sigma_2})^2}{r^2 - W_{\sigma_2}^2} \right\|_{\infty}^{-1} \ln \left(\frac{1}{3\delta} \right). \quad (62)$$

We now focus on proving the second inequality in Eq. (61). For that, consider that σ_2 in Eq. (62) is the sequence $(\rho_n)_{n \in \mathbb{N}}$ from Lemma 2. Since ρ is pure, Eq. (91) yields $\lim_{n \rightarrow \infty} F(\rho, \rho_n) = 0$. Combining this with the fact that $\lim_{n \rightarrow \infty} \|W_{\rho_n}\|_{\infty} = 0$ [see Eq. (90)], we obtain for every state σ that

$$\left\| \frac{(W_{\sigma} - W_{\rho_n})^2}{r^2 - W_{\rho_n}^2} \right\|_{\infty} \leq \frac{\|W_{\sigma} - W_{\rho_n}\|_{\infty}^2}{r^2 - \|W_{\rho_n}\|_{\infty}^2} \xrightarrow{n \rightarrow \infty} \left(\frac{\|W_{\sigma}\|_{\infty}}{r} \right)^2$$

and hence

$$\begin{aligned} N_{\varepsilon,\delta}^W(\rho) &\geq \left(\frac{2}{\pi} \right)^{2m} \sup_{\substack{\sigma \text{ state,} \\ F(\rho, \sigma) \geq 2\varepsilon}} \frac{1}{\|W_{\sigma}\|_{\infty}^2} \ln \left(\frac{1}{3\delta} \right) \\ &= \left(\frac{2}{\pi} \right)^{2m} \sup_{\substack{\sigma \text{ state,} \\ F(\rho, \sigma) \geq 2\varepsilon}} \frac{1}{\|W_{\sigma}\|_{\infty}^2} \ln \left(\frac{1}{3\delta} \right). \end{aligned} \quad (63)$$

Here, the last equality follows due to $\varepsilon < 1/2$ and the same argument as for Eq. (53). Note that the corresponding function g_n defined in Eq. (55) in the context here is the convex

combination of two Wigner functions and hence a valid Wigner function itself. ■

A. Matching upper and lower bounds for $N_{\varepsilon,\delta}^W(\rho)$ for examples of states

In this section we discuss two examples of families of states for which the upper and lower bounds on $N_{\varepsilon,\delta}^W(\rho)$ provided in Theorems 3 and 4 match. In particular we give a full characterization of the sample complexity of fidelity estimation in the Wigner model for (1) the Fock states $(|n\rangle\langle n|)_{n \in \mathbb{N}}$ corresponding to the eigenstates of the (1-mode) harmonic oscillator, (2) the *spike states*, which we define in Sec. IV A 2 and (3) Gaussian states, in terms of the L^1 -norms of the corresponding Wigner functions.

In the case of the Fock states the key insight is that the corresponding Wigner functions can explicitly be expressed in terms of Laguerre polynomials [see Eq. (71)]. Combining this with the results on the asymptotic behavior of the Laguerre polynomials in Refs. [16,67–69], we show in Proposition 1 that the upper bound in Theorem 3 and the first lower bound in Theorem 4 match, giving

$$\begin{aligned} N_{\varepsilon,\delta}^W(|n\rangle\langle n|) &= \Theta \left(\left(\frac{\|W_{|n\rangle\langle n|}\|_1}{\varepsilon} \right)^2 \ln \left(\frac{1}{\delta} \right) \right) \\ &= \Theta \left(\frac{n}{\varepsilon^2} \ln \left(\frac{1}{\delta} \right) \right). \end{aligned}$$

The spike states on the other hand are a sequence of pure states $(\rho_n)_{n \in \mathbb{N}}$ on $L^2(\mathbb{R}^m)$ constructed in such a way that their Wigner functions uniformly vanish as n increases, i.e.,

$$\|W_{\rho_n}\|_{\infty} \xrightarrow{n \rightarrow \infty} 0. \quad (64)$$

Furthermore, we show in Lemma 2 that the L^1 -norms of their Wigner functions blow up and exactly scale reciprocally to the L^{∞} -norms, i.e.,

$$\|W_{\rho_n}\|_1 \sim \|W_{\rho_n}\|_{\infty}^{-1}. \quad (65)$$

Hence, we can use the upper bound in Theorem 3 and the second lower bound in Theorem 4, showing that

$$N_{\varepsilon,\delta}^W(\rho_n) = \Theta \left(\left(\frac{\|W_{\rho_n}\|_1}{\varepsilon} \right)^2 \ln \left(\frac{1}{\delta} \right) \right). \quad (66)$$

We believe the existence of such a sequence of pure states satisfying Eqs. (64) and (65) to be potentially of independent interest.

Lastly, for Gaussian states ρ , the L^1 -norm of the corresponding Wigner function is by the normalization (26)

equal to 1, i.e., $\|W_\rho\|_1 = 1$. Combining this with the lower bound provided in Theorem 4, we see in Proposition 3 that

$$N_{\varepsilon,\delta}^W(\rho) = \Theta\left(\frac{1}{\varepsilon^2} \ln\left(\frac{1}{\delta}\right)\right) \quad (67)$$

for all pure Gaussian states ρ .

As a consequence of these results we find the following worst case characterization of the sample complexity of fidelity estimation in the Wigner model:

Theorem 5 (Sample complexity of fidelity estimation in Wigner model in the worst case). Let $0 < \varepsilon < 1/2$ and $0 < \delta < 1/4$. Then for all $t \geq 1$ we have

$$\sup_{\substack{\rho \text{ pure state,} \\ \|W_\rho\|_1 \leq t}} N_{\varepsilon,\delta}^W(\rho) = \Theta\left(\left(\frac{t}{\varepsilon}\right)^2 \ln\left(\frac{1}{\delta}\right)\right), \quad (68)$$

where the constants in $\Theta(\cdot)$ can be m -dependent. In particular, this shows that in the worst case over all pure states ρ the sample complexity of fidelity estimation in the Wigner model is infinite, i.e.,

$$\sup_{\rho \text{ pure state}} N_{\varepsilon,\delta}^W(\rho) = \infty. \quad (69)$$

In the following sections we formally state and prove the above-mentioned results for the Fock states, spike states and Gaussian states and then provide the proof of Theorem 5.

1. Sample complexity for Fock states

Proposition 1 (Sample complexity for Fock states). Let $0 < \varepsilon, \delta < 1/4$ and $(|n\rangle\langle n|)_{n \in \mathbb{N}}$ be the 1-mode Fock states. Then we have the following characterization of the corresponding sample complexity of fidelity estimation in the Wigner model:

$$\begin{aligned} N_{\varepsilon,\delta}^W(|n\rangle\langle n|) &= \Theta\left(\left(\frac{\|W_{|n\rangle\langle n|}\|_1}{\varepsilon}\right)^2 \ln\left(\frac{1}{\delta}\right)\right) \\ &= \Theta\left(\frac{n}{\varepsilon^2} \ln\left(\frac{1}{\delta}\right)\right). \end{aligned} \quad (70)$$

Remark 6. As seen in Proposition A1 in Appendix A 1, the smoothed L^1 -norm of the Wigner functions of the Fock states satisfies

$$\|W_{|n\rangle\langle n|}\|_1^{(\varepsilon)} \geq \|W_{|n\rangle\langle n|}\|_1 (1 - C\varepsilon^\beta)$$

for all $\varepsilon > 0$, $\beta \in (0, 1/2)$ and some $C > 0$ possibly dependent on β but independent of n and ε . As discussed

at the beginning of Appendix A 1, this gives

$$\inf_{\varepsilon' \in [0, \varepsilon]} \frac{\|W_{|n\rangle\langle n|}\|_1^{(\varepsilon')}}{\varepsilon - \varepsilon'} = \Theta\left(\frac{\|W_{|n\rangle\langle n|}\|_1}{\varepsilon}\right).$$

Hence, for the Fock states, the upper bound on the sample complexity in terms of the smoothed L^1 -function in Theorem 3 does essentially not become weaker when it is expressed directly through the nonsmoothed L^1 -norm. This further justifies our finding in Proposition 1 of a characterization of the sample complexity of fidelity estimation for the Fock states in terms of the L^1 -norm of the corresponding Wigner functions.

For the proof of Proposition 1, we use the fact that the corresponding Wigner functions can be expressed in terms of the Laguerre polynomials L_n as [see Eq. (36) in Ref. [70] and Eq. (7.3) in Ref. [71]]

$$W_{|n\rangle\langle n|}(\alpha) = \left(\frac{2}{\pi}\right) (-1)^n e^{-2|\alpha|^2} L_n(4|\alpha|^2). \quad (71)$$

Using the results from Ref. [16] on the L^p -norms of the Laguerre polynomials, we can deduce the following scaling of the L^1 -norms of the Wigner function of the Fock states:

$$\|W_{|n\rangle\langle n|}\|_1 \sim \sqrt{n}. \quad (72)$$

This can be combined with Theorem 3, yielding the desired upper bound on $N_{\varepsilon,\delta}^W(|n\rangle\langle n|)$. On the other hand, using the multiple asymptotic expressions of the Laguerre polynomials from Refs. [67–69] for different regions in phase space, we show explicitly below that

$$\left\| \frac{(W_{|n+2\rangle\langle n+2|} - W_{|n\rangle\langle n|})^2}{(2/\pi)^2 - W_{|n\rangle\langle n|}^2} \right\|_\infty \lesssim \frac{1}{n}.$$

Application of this to the first lower bound in Theorem 4 for specific choices of σ_1 and σ_2 provides the corresponding lower bound of $N_{\varepsilon,\delta}^W(|n\rangle\langle n|)$ in Proposition 1.

Proof of Proposition 1. We first start by proving the corresponding upper bound in Eq. (70). We use the upper bound in Theorem 3, which gives

$$N_{\varepsilon,\delta}^W(|n\rangle\langle n|) \lesssim \left(\frac{\|W_{|n\rangle\langle n|}\|_1}{\varepsilon}\right)^2 \ln\left(\frac{1}{\delta}\right).$$

As noted in Eq. (71), the Wigner function of the Fock states $|n\rangle\langle n|$ can be written explicitly as [see Eq. (36) in Ref. [70]]

and Eq. (7.3) in Ref. [71]]

$$W_{|n\rangle\langle n|}(\alpha/2) = r(-1)^n e^{-|\alpha|^2/2} L_n(|\alpha|^2), \quad (73)$$

with L_n denoting the *Laguerre polynomial* of n^{th} order and $r = 2/\pi$. From that we see that the scaling of the corresponding L^1 -norms is given as

$$\begin{aligned} \|W_{|n\rangle\langle n|}\|_1 &\sim \int_0^\infty e^{-s^2/2} |L_n(s^2)| s ds \\ &= \int_0^\infty e^{-s^2/2} |L_n(s)| ds \sim \sqrt{n}, \end{aligned} \quad (74)$$

where we used Lemma 1 in Ref. [16] for the last step. Hence, we have shown

$$N_{\varepsilon,\delta}^W(|n\rangle\langle n|) \lesssim \frac{n}{\varepsilon^2} \ln\left(\frac{1}{\delta}\right). \quad (75)$$

We finish the proof by providing the corresponding lower bound in Eq. (70): Since $\varepsilon < 1/4$ we can pick states $\sigma_1^{(n)} = (1 - 3\varepsilon)|n\rangle\langle n| + 3\varepsilon|n+2\rangle\langle n+2|$ and $\sigma_2^{(n)} = |n\rangle\langle n|$, which satisfy

$$\left|F(|n\rangle\langle n|, \sigma_1^{(n)}) - F(|n\rangle\langle n|, \sigma_2^{(n)})\right| = 3\varepsilon > 2\varepsilon$$

Hence, we see from the first lower bound in Theorem 4 that

$$\begin{aligned} N_{\varepsilon,\delta}^W(|n\rangle\langle n|) &\gtrsim \left\| \frac{(W_{\sigma_1^{(n)}} - W_{\sigma_2^{(n)}})^2}{r^2 - W_{\sigma_2^{(n)}}^2} \right\|_\infty^{-1} \ln\left(\frac{1}{3\delta}\right) \\ &= \frac{1}{9\varepsilon^2} \left\| \frac{(W_{|n+2\rangle\langle n+2|} - W_{|n\rangle\langle n|})^2}{r^2 - W_{|n\rangle\langle n|}^2} \right\|_\infty^{-1} \ln\left(\frac{1}{3\delta}\right). \end{aligned} \quad (76)$$

In the following, we estimate the quotient appearing in the infinity norm in the second line of Eq. (76) to find the desired lower bound on the sample complexity. We do this by using the explicit form of the Wigner functions in Eq. (73) and provide bounds in three different regions of phase space, i.e., for phase space points α satisfying (1) $|\alpha|^2 \leq 1/\tilde{n}$, (2) $|\alpha|^2 \in [\tilde{n}^{-1}, 1]$ or (3) $|\alpha|^2 > 1$. Here, we introduced the shorthand notation $\tilde{n} = n + 1/2$. For the first two regions we use the *formula of Hilb's type* (see Theorem 8.22.4 in Ref. [69]), which for $|\alpha|^2 \in [0, 1]$ gives

$$\begin{aligned} e^{-|\alpha|^2/2} L_n(|\alpha|^2) &= J_0(2\sqrt{\tilde{n}}|\alpha|) \\ &+ \begin{cases} \mathcal{O}(|\alpha|^4 \ln(|\alpha|^{-2}n^{-1})) & \text{for } |\alpha|^2 \in [0, \tilde{n}^{-1}], \\ \mathcal{O}(n^{-3/4}) & \text{for } |\alpha|^2 \in [\tilde{n}^{-1}, 1]. \end{cases} \end{aligned} \quad (77)$$

Here, the *Bessel function* of order 0 was denoted by

$$J_0(y) = \sum_{k=0}^{\infty} \frac{(-1)^k (y/2)^{2k}}{(k!)^2} = 1 - \left(\frac{y}{2}\right)^2 + R(y), \quad (78)$$

with the last equation holding by Taylor's theorem, including up to third order for small y , say $y \in [0, 3]$, with $R(y)$ being the corresponding remainder. The remainder satisfies

$$|R(y)| \leq \frac{y^4}{4!}, \quad (79)$$

which follows from the fact [72] that $|J_0^{(4)}(y)| \leq 1$ together with the standard estimate for the remainder in Taylor's theorem (see Theorem 7.7 in Ref. [73]). Let us focus on the first phase space region close to the origin, i.e., $|\alpha|^2 \leq 1/\tilde{n}$, in which case we get from the series expression of the Bessel function

$$\begin{aligned} &\left|J_0(2\sqrt{\tilde{n}}|\alpha|) - J_0(2\sqrt{\tilde{n}}|\alpha|)\right| \\ &= \left| \sum_{k=0}^{\infty} \frac{(-1)^k ((\tilde{n}+2)^k - \tilde{n}^k) |\alpha|^{2k}}{(k!)^2} \right| \\ &\lesssim \sum_{k=1}^{\infty} \frac{\tilde{n}^{k-1} |\alpha|^{2k}}{k!} \\ &\leq |\alpha|^2 e^{\tilde{n}|\alpha|^2} \lesssim |\alpha|^2, \end{aligned}$$

where we used for the first inequality that $|(\tilde{n}+2)^k - \tilde{n}^k| = |\sum_{l=0}^{k-1} \tilde{n}^l 2^{k-l} \binom{k}{l}| \leq \tilde{n}^{k-1} 3^k$ together with the fact that $3^k/k! \leq C$ for some $C > 0$. From this and Eq. (77) we see

$$\begin{aligned} &r^{-2} (W_{|n+2\rangle\langle n+2|}(\alpha/2) - W_{|n\rangle\langle n|}(\alpha/2))^2 \\ &= \left(e^{-|\alpha|^2/2} L_{n+2}(|\alpha|^2) - e^{-|\alpha|^2/2} L_n(|\alpha|^2)\right)^2 \\ &\lesssim (|\alpha|^2 + \mathcal{O}(|\alpha|^4 \ln(|\alpha|^{-2}n^{-1})))^2 \lesssim |\alpha|^4, \end{aligned}$$

where we used in the last inequality that $|\alpha|^2 \ln(|\alpha|^{-2}) \leq C$ and $|\alpha|^2 \ln(n) \leq C$ for some $C > 0$. Furthermore, again for $|\alpha|^2 \leq 1/\tilde{n}$, we have

$$\begin{aligned} &1 - r^{-2} W_{|n\rangle\langle n|}^2(\alpha/2) \\ &= 1 - e^{-|\alpha|^2} L_n^2(|\alpha|^2) \\ &= 1 - (J_0(2\sqrt{\tilde{n}}|\alpha|) + \mathcal{O}(|\alpha|^4 \ln(|\alpha|^{-2}n^{-1})))^2 \\ &= 1 - (1 - n|\alpha|^2 + R(\sqrt{\tilde{n}}|\alpha|) \\ &\quad + \mathcal{O}(|\alpha|^4 \ln(|\alpha|^{-2}n^{-1})))^2 \\ &\geq 1 - (1 - cn|\alpha|^2)^2 \geq cn|\alpha|^2, \end{aligned}$$

where the first inequality holds for n large enough and some $0 < c < 1$ and we used Eq. (79). From this we see

$$\begin{aligned} & \sup_{0 \leq |\alpha|^2 \leq 1/\tilde{n}} \frac{(W_{|n+2\rangle\langle n+2|}(\alpha/2) - W_{|n\rangle\langle n|}(\alpha/2))^2}{r^2 - W_{|n\rangle\langle n|}^2(\alpha/2)} \\ & \lesssim \sup_{0 \leq |\alpha|^2 \leq \frac{1}{\tilde{n}}} \frac{|\alpha|^2}{n} \lesssim \frac{1}{n^2}. \end{aligned} \quad (80)$$

For the other two regions in phase space we use the results highlighted in the table on page 699 of Ref. [67], which gives

$$\sup_{\alpha \in \mathbb{C}} \left(e^{-|\alpha|^2/2} (L_{n+2}(|\alpha|^2) - L_n(|\alpha|^2)) \right) = \mathcal{O}(n^{-1/2}). \quad (81)$$

Let us focus first on the second region in phase space, i.e., $|\alpha|^2 \in [\tilde{n}^{-1}, 1]$: For those phase space points we use the fact that $|J_0(y)| \leq 0.8/y^{1/3}$ [68] for all $y \in \mathbb{R}$, which gives by Eqs. (73) and (77) that

$$\begin{aligned} & 1 - r^{-2} W_{|n\rangle\langle n|}^2(\alpha/2) \\ & = 1 - \left(J_0(2\sqrt{\tilde{n}}|\alpha|) + \mathcal{O}(n^{-3/4}) \right)^2 \geq c \end{aligned} \quad (82)$$

for n large enough and some $c > 0$. Dividing Eq. (81) by Eq. (82) we find

$$\begin{aligned} & \sup_{|\alpha|^2 \in [\tilde{n}^{-1}, 1]} \frac{(W_{|n+2\rangle\langle n+2|}(\alpha/2) - W_{|n\rangle\langle n|}(\alpha/2))^2}{r^2 - W_{|n\rangle\langle n|}^2(\alpha/2)} \\ & \leq \sup_{|\alpha|^2 \in [\tilde{n}^{-1}, 1]} \frac{\left(e^{-|\alpha|^2/2} (L_{n+2}(|\alpha|^2) - L_n(|\alpha|^2)) \right)^2}{c} \lesssim \frac{1}{n}. \end{aligned} \quad (83)$$

For the last phase space region, we again use the results highlighted in the table on page 699 of Ref. [67], which for $|\alpha| > 1$ gives

$$e^{-|\alpha|^2/2} L_n(|\alpha|^2) = \mathcal{O}(n^{-1/4})$$

and hence using again Eq. (73), we find

$$1 - r^{-2} W_{|n\rangle\langle n|}^2(\alpha/2) \geq c$$

for n large enough and some $c > 0$. Hence, combining the last inequality with Eq. (81), we see

$$\sup_{|\alpha| > 1} \frac{(W_{|n+2\rangle\langle n+2|}(\alpha/2) - W_{|n\rangle\langle n|}(\alpha/2))^2}{r^2 - W_{|n\rangle\langle n|}^2(\alpha/2)} \lesssim \frac{1}{n}. \quad (84)$$

Therefore, combining Eqs. (80), (83) and (84), we see

$$\left\| \frac{(W_{|n+2\rangle\langle n+2|} - W_{|n\rangle\langle n|})^2}{r^2 - W_{|n\rangle\langle n|}^2} \right\|_{\infty} \lesssim \frac{1}{n}$$

and therefore from Eq. (76)

$$N_{\varepsilon, \delta}^W(|n\rangle\langle n|) \gtrsim \frac{n}{\varepsilon^2} \ln \left(\frac{1}{3\delta} \right),$$

which together with the established Eq. (75) finishes the proof. $\blacksquare$

2. Sample complexity for spike states

In this section we define the *spike states* and analyze the scaling behavior of the L^1 - and L^∞ -norms of their Wigner functions. In particular for every number of modes $m \in \mathbb{N}$ the spike states are a sequence of pure states $(\rho_n)_{n \in \mathbb{N}} = (|\psi_n\rangle\langle\psi_n|)_{n \in \mathbb{N}}$ on $L^2(\mathbb{R}^m)$ whose Wigner functions satisfy

$$\|W_{\rho_n}\|_{\infty} \xrightarrow{n \rightarrow \infty} 0. \quad (85)$$

Here, the idea of the construction is to pick wave functions ψ_n with increasing uncertainty in both position and momentum, which then leads to a further and further spread out Wigner function. We pick such wave functions by super-positioning “spikes” in position space, which each have a high momentum uncertainty. In the particular construction of Lemma 2 these “spikes” are simply given by Gaussian functions. Super-positioning sufficiently many at increasing distance also leads to high position uncertainty and to a pointwise uniformly vanishing Wigner function.

More precisely, the spike states are defined as follows: Let $\phi \in L^2(\mathbb{R}^m)$ be the normalized (within $L^2(\mathbb{R}^m)$, i.e., $\|\phi\|_2 = 1$) Gaussian function centered at the origin, i.e., for $x \in \mathbb{R}^m$

$$\phi(x) = \left(\frac{2}{\pi} \right)^{m/4} e^{-x^2}. \quad (86)$$

Consider for $n \in \mathbb{N}$ the wave function defined for $x \in \mathbb{R}^m$ by

$$\psi_n(x) = \frac{1}{\sqrt{c_n}} \sum_{k=1}^n \phi(x - \mu_k), \quad (87)$$

with

$$\mu_k = n3^k e_1, \quad (88)$$

where $e_1 = (1, 0, \dots, 0) \in \mathbb{R}^m$ and where we drop the n -dependence for brevity. Furthermore, c_n is the normalization constant such that $\|\psi_n\|_2 = 1$, which in particular satisfies

$$\begin{aligned} c_n &= \sum_{k,l=1}^n \int_{\mathbb{R}^m} \phi(x - \mu_k) \phi(x - \mu_l) dx \\ &\geq \sum_{k=1}^n \int_{\mathbb{R}^m} |\phi(x - \mu_k)|^2 dx = n. \end{aligned} \quad (89)$$

The n^{th} spike state is defined to be the corresponding pure state $\rho_n = |\psi_n\rangle\langle\psi_n|$.

The particular spreading of the spikes parameterized by the μ_k in Eq. (88) is chosen in such a way that for every fixed point in phase space essentially at most only two spikes contribute to the Wigner function in Eq. (25); see the proof of Lemma 2 for more details.

Notably, we provide control of the speed of convergence for the limit Eq. (85) and also on the blow up of the corresponding L^1 -norms. In particular, we see that the sequence of pure states satisfies

$$\|W_{\rho_n}\|_{\infty}^{-1} \sim \|W_{\rho_n}\|_1$$

(see Remark 7). This is then used as key insight for proving the worst case characterization (68) of the sample complexity in Theorem 4.

In the following lemma we provide the precise scaling behavior of the L^1 - and L^{∞} -norms of the spike states:

Lemma 2 (Asymptotic behavior of spike states). The pure states $(\rho_n)_{n \in \mathbb{N}} = (|\psi_n\rangle\langle\psi_n|)_{n \in \mathbb{N}}$ on $L^2(\mathbb{R}^m)$ defined by Eq. (87) have Wigner functions satisfying $W_{\rho_n} \in L^1(\mathbb{C}^m)$ and

$$\|W_{\rho_n}\|_{\infty} \leq 4 \left(\frac{2}{\pi}\right)^m \frac{1}{n} \quad \text{and} \quad \|W_{\rho_n}\|_1 \leq n. \quad (90)$$

Moreover, the corresponding sequence $(\psi_n)_{n \in \mathbb{N}}$ converges weakly to 0 in $L^2(\mathbb{R}^m)$, i.e., for all $\varphi \in L^2(\mathbb{R}^m)$ we have

$$\lim_{n \rightarrow \infty} \langle \varphi, \psi_n \rangle = 0. \quad (91)$$

Remark 7. Note that the n -dependence of the bounds in Eq. (90) is tight. This can be seen by use of the fact that that since ρ_n is pure we have $\|W_{\rho_n}\|_2^2 = \pi^{-m}$ by Eq. (28) and hence

$$\|W_{\rho_n}\|_{\infty}^{-1} \leq \pi^m \|W_{\rho_n}\|_1.$$

Use of this and the established bounds in Eq. (90) shows

$$\|W_{\rho_n}\|_{\infty} \sim \frac{1}{n} \quad \text{and} \quad \|W_{\rho_n}\|_1 \sim n. \quad (92)$$

Before giving the proof of Lemma 2, we show how we can use it to obtain a characterization of the sample complexity $N_{\varepsilon,\delta}^W(\rho_n)$ from the upper bound and the lower bound in Theorems 3 and 4.

Proposition 2 (Sample complexity for spike states). Let $0 < \varepsilon < 1/2$, $0 < \delta < 1/4$ and $(\rho_n)_{n \in \mathbb{N}} = (|\psi_n\rangle\langle\psi_n|)_{n \in \mathbb{N}}$ be the spike states defined in Eq. (87). Then we have the following characterization of the corresponding sample complexity of fidelity estimation in the Wigner model:

$$N_{\varepsilon,\delta}^W(\rho_n) = \Theta \left(\left(\frac{\|W_{\rho_n}\|_1}{\varepsilon} \right)^2 \ln \left(\frac{1}{\delta} \right) \right),$$

where the constants in $\Theta(\cdot)$ can be m -dependent.

Proof. By the upper bound provided in Theorem 3 we immediately see

$$\begin{aligned} N_{\varepsilon,\delta}^W(\rho_n) &\leq 2^{2m+1} \left(\frac{\|W_{\rho_n}\|_1}{\varepsilon} \right)^2 \ln \left(\frac{1}{\delta} \right) \\ &\leq 2^{2m+1} \left(\frac{n}{\varepsilon} \right)^2 \ln \left(\frac{1}{\delta} \right), \end{aligned}$$

where for the second inequality we have used Lemma 2.

To prove the corresponding lower bound we take $n' = \lceil n/\varepsilon \rceil$ and state $\sigma_n = 2\varepsilon\rho_n + (1 - 2\varepsilon)\rho_{n'}$, which satisfies $F(\rho_n, \sigma_n) \geq 2\varepsilon$. Furthermore, we have by Lemma 2 that

$$\begin{aligned} \|W_{\sigma_n}\|_{\infty} &\leq 2\varepsilon\|W_{\rho_n}\|_{\infty} + (1 - 2\varepsilon)\|W_{\rho_{n'}}\|_{\infty} \\ &\leq 16 \left(\frac{2}{\pi} \right)^m \frac{\varepsilon}{n}. \end{aligned}$$

Therefore, using the second lower bound in Theorem 4, we see that

$$\begin{aligned} N_{\varepsilon,\delta}^W(\rho_n) &\geq \left(\frac{2}{\pi} \right)^{2m} \frac{1}{\|W_{\sigma_n}\|_{\infty}^2} \ln \left(\frac{1}{3\delta} \right) \\ &\geq \frac{1}{16} \left(\frac{n}{\varepsilon} \right)^2 \ln \left(\frac{1}{3\delta} \right), \end{aligned}$$

which finishes the proof. ■

Proof of Lemma 2. We prove the result with phase space being parameterized with position and momentum variables $(x, p) \in \mathbb{R}^{2m}$, which corresponds to the change of coordinates given by $\alpha = \frac{1}{\sqrt{2}}(x + ip)$.

Using Eq. (25) and the explicit form of the wave functions in Eq. (87), we can calculate the Wigner function of ρ_n at phase space point $(x, p) \in \mathbb{R}^{2m}$ by

$$\begin{aligned}
W_{\rho_n}(x, p) &= \frac{1}{c_n} \left(\frac{2}{\pi} \right)^{3m/2} \sum_{k,l=1}^n \int_{\mathbb{R}^m} e^{-(x-\mu_k+y)^2-(x-\mu_l-y)^2} e^{i2p \cdot y} dy \\
&= \frac{1}{c_n} \left(\frac{2}{\pi} \right)^{3m/2} \sum_{k,l=1}^n e^{-(x-\mu_k)^2-(x-\mu_l)^2+\frac{(\mu_k-\mu_l)^2}{2}} \int_{\mathbb{R}^m} e^{-2(y-(\mu_k-\mu_l)/2)^2} e^{i2p \cdot y} dy \\
&= \frac{1}{c_n} \left(\frac{2}{\pi} \right)^m \sum_{k,l=1}^n e^{-2(x-(\mu_k+\mu_l)/2)^2} e^{ip \cdot (\mu_k-\mu_l)-p^2/2},
\end{aligned} \tag{93}$$

where for the last line we used the standard relation for Fourier transforming a Gaussian function (see, e.g., Lemma 7.3 in Ref. [56]).

Note that because of the specific choice of μ_k in Eq. (88), we have for $x \in \mathbb{R}^m$ fixed that there exists at most one $(k_x, l_x) \in \mathbb{N}^2$ with $k_x \geq l_x$ such that for the euclidean norm we have

$$\left\| x - \frac{\mu_{k_x} + \mu_{l_x}}{2} \right\| \leq \frac{n}{2}, \tag{94}$$

which can be seen by the following argument: Assume for contradiction that there exists $(k, l) \neq (i, j)$ with $k \geq l$ and $i \geq j$ and such that $\|x - (\mu_k + \mu_l)/2\| \leq n/2$

and $\|x - (\mu_i + \mu_j)/2\| \leq n/2$. This hence gives

$$\|\mu_k + \mu_l - \mu_i - \mu_j\| = n |3^k + 3^l - 3^i - 3^j| \leq 2n. \tag{95}$$

Without loss of generality we can assume that $i > k$ as for $k = i$ we immediately also have $l = j$ from Eq. (95). With use of $k \geq l$ this gives

$$3^i + 3^j \geq 3^i \geq 3^{k+1} \geq 3^k + 3^l + 3,$$

which contradicts Eq. (95) and hence proves the desired claim.

Using for $(x, p) \in \mathbb{R}^{2m}$ the uniqueness of $(k_x, l_x) \in \mathbb{N}^2$ with $k_x \geq l_x$ and satisfying Eq. (94) and plugging this into Eq. (93), we obtain together with the use of Eq. (89)

$$\begin{aligned}
|W_{\rho_n}(x, p)| &\leq \frac{1}{n} \left(\frac{2}{\pi} \right)^m \sum_{k,l=1}^n e^{-2(x-(\mu_k+\mu_l)/2)^2} \\
&\leq \frac{2}{n} \left(\frac{2}{\pi} \right)^m \left(e^{-2(x-(\mu_{k_x}+\mu_{l_x})/2)^2} + \sum_{\substack{k,l=1 \\ k \geq l, (k,l) \neq (k_x, l_x)}}^n e^{-2(x-(\mu_k+\mu_l)/2)^2} \right) \\
&\leq \frac{2}{n} \left(\frac{2}{\pi} \right)^m (1 + n^2 e^{-n^2/2}) \leq \frac{4}{n} \left(\frac{2}{\pi} \right)^m,
\end{aligned}$$

where we used $t^2 e^{-t^2/2} \leq 1$ for all $t \in \mathbb{R}$ and which hence shows the L^∞ -bound in Eq. (90).

On the other hand, using again the explicit form of the Wigner function of ρ_n in Eq. (93), we see

$$\|W_{\rho_n}\|_1 = \int_{\mathbb{C}^m} |W_{\rho}(\alpha)| d\alpha = 2^{-m} \int_{\mathbb{R}^{2m}} |W_{\rho}(x, p)| dx dp$$

$$\leq \frac{1}{n\pi^m} \sum_{k,l=1}^n \int_{\mathbb{R}^{2m}} e^{-2(x-(\mu_k+\mu_l)/2)^2} e^{-p^2/2} dx dp$$

$$= \frac{n}{\pi^m} \int_{\mathbb{R}^{2m}} e^{-x^2} e^{-p^2} dx dp = n.$$

Lastly we want to show the weak convergence of the sequence $(\psi_n)_{n \in \mathbb{N}}$ to 0, i.e., Eq. (91). For that let $\varphi \in L^2(\mathbb{R})$ and $\varepsilon > 0$ and pick $R \geq 0$ such that

$$\int_{\|x\| > R} |\varphi(x)|^2 dx < \varepsilon,$$

which is possible due to the dominated convergence theorem. Furthermore, note that we have for the Gaussian function ϕ defined in Eq. (86), $k \in [n]$ and $n3^k \geq R$

$$\begin{aligned} \int_{\|x\| \leq R} |\phi(x - \mu_k)|^2 dx &\leq \sqrt{\frac{2}{\pi}} \int_{-R}^R e^{-2(x_1 - n3^k)^2} dx_1 \\ &\leq \sqrt{\frac{2}{\pi}} \int_{-R+n3^k}^{\infty} e^{-2x_1^2} dx_1 \leq e^{-2(n3^k - R)^2}, \end{aligned}$$

where the first inequality follows by the definition of μ_k in Eq. (88) and third inequality follows by standard Gaussian tail bounds. Hence, using the Cauchy-Schwarz inequality and the fact that ψ_n is normalized in L^2 , we see for n large enough

$$\begin{aligned} |\langle \varphi, \psi_n \rangle| &\leq \sqrt{\int_{\|x\| > R} |\varphi(x)|^2 dx} + \frac{\|\varphi\|_2}{\sqrt{c_n}} \\ &\quad \times \sum_{k=1}^n \sqrt{\int_{\|x\| \leq R} |\phi(x - \mu_k)|^2 dx} \\ &\leq \sqrt{\varepsilon} + \frac{\|\varphi\|_2}{\sqrt{c_n}} \sum_{k=1}^n e^{-(n3^k - R)^2} \xrightarrow{n \rightarrow \infty} \sqrt{\varepsilon}. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, this proves Eq. (91). $\blacksquare$

3. Sample complexity for Gaussian states

Proposition 3. Let ρ be a pure Gaussian state on $L^2(\mathbb{R}^m)$ and $0 < \varepsilon, \delta < 1/4$. Then

$$N_{\varepsilon, \delta}^W(\rho) = \Theta\left(\frac{1}{\varepsilon^2} \ln\left(\frac{1}{\delta}\right)\right), \quad (96)$$

where the constants in $\Theta(\cdot)$ can be m -dependent.

Proof. For ρ being a Gaussian state, we have that $W_\rho \geq 0$ and hence by the normalization (26)

$$\|W_\rho\|_1 = 1. \quad (97)$$

Therefore, we see from Theorem 3 the upper bound

$$N_{\varepsilon, \delta}^W(\rho) = \mathcal{O}\left(\frac{1}{\varepsilon^2} \ln\left(\frac{1}{\delta}\right)\right).$$

For the corresponding lower bound we use the second inequality in Eq. (61) in Theorem 4 for the sequence

of feasible states given by $\sigma_n = 2\varepsilon\rho + (1 - 2\varepsilon)\rho_n$, where $(\rho_n)_{n \in \mathbb{N}}$ is the sequence of spike states defined in Sec. IV A 2, which satisfies $\lim_{n \rightarrow \infty} \|W_{\rho_n}\|_\infty = 0$ and hence $\lim_{n \rightarrow \infty} \|W_{\sigma_n}\|_\infty = 2\varepsilon\|W_\rho\|_\infty \leq 2\varepsilon$. $\blacksquare$

4. Proof of Theorem 5

In this section we give the proof of Theorem 5, which essentially follows the same argument as the proof of Proposition 2.

Proof. Let $t \geq 1$ and use Theorem 3 for the corresponding upper bound as

$$\begin{aligned} \sup_{\substack{\rho \text{ pure state,} \\ \|W_\rho\|_1 \leq t}} N_{\varepsilon, \delta}(W_\rho, \mathcal{S}_W) &\leq 2^{2m+1} \sup_{\substack{\rho \text{ pure state,} \\ \|W_\rho\|_1 \leq t}} \left(\frac{\|W_\rho\|_1}{\varepsilon}\right)^2 \ln\left(\frac{1}{\delta}\right) \\ &\leq 2^{2m+1} \left(\frac{t}{\varepsilon}\right)^2 \ln\left(\frac{1}{\delta}\right). \end{aligned}$$

For the corresponding lower bound in Eq. (68) we take the second lower bound in Theorem 4 for the specific choice $\rho \equiv \rho_n$, where $n = \lfloor t \rfloor$ and $(\rho_n)_{n \in \mathbb{N}}$ are the spike states defined and discussed in Sec. IV A 2. Note that ρ_n is feasible in Eq. (68) as by Eq. (90) we have $\|W_{\rho_n}\|_1 \leq n \leq t$. Furthermore, let $\sigma_n = 2\varepsilon\rho_n + (1 - 2\varepsilon)\rho_{n'}$, where $n' = \lceil t/\varepsilon \rceil$. Note that for that choice we have $F(\rho_n, \sigma_n) \geq 2\varepsilon$ and furthermore by Eq. (90) and the fact that $1/n \leq 2/t$ since $t \geq 1$ that

$$\|W_{\sigma_n}\|_\infty \leq 2\varepsilon\|W_{\rho_n}\|_\infty + (1 - 2\varepsilon)\|W_{\rho_{n'}}\|_\infty \leq C \left(\frac{2}{\pi}\right)^m \frac{\varepsilon}{t}$$

for some $C \geq 0$ independent of t, ε and m . Hence, σ_n is a feasible state in Eq. (61) and we obtain

$$\begin{aligned} \sup_{\substack{\rho \text{ pure state,} \\ \|W_\rho\|_1 \leq t}} N_{\varepsilon, \delta}(W_\rho, \mathcal{S}_W) &\geq N_{\varepsilon, \delta}(W_{\rho_n}, \mathcal{S}_W) \\ &\geq \left(\frac{2}{\pi}\right)^{2m} \frac{1}{\|W_{\sigma_n}\|_\infty^2} \ln\left(\frac{1}{3\delta}\right) \\ &\geq C' \left(\frac{t}{\varepsilon}\right)^2 \ln\left(\frac{1}{3\delta}\right) \end{aligned}$$

for some $C' \geq 0$. $\blacksquare$

V. FIDELITY ESTIMATION FOR DISCRETE VARIABLE SYSTEMS

We consider for $n \in \mathbb{N}$ an n -qubit quantum system on the Hilbert space $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$ with dimension $d = 2^n$. The corresponding set of Pauli strings is denoted by

$\mathcal{P}_n = \{\mathbb{1}, X, Y, Z\}^{\otimes n}$. Note that $\mathcal{P}_n$ forms an orthogonal basis on the space of operators on $(\mathbb{C}^2)^{\otimes n}$ and we have for all $P, Q \in \mathcal{P}_n$

$$\frac{1}{d} \text{Tr}(PQ) = \delta_{PQ}. \quad (98)$$

For ρ a state on $(\mathbb{C}^2)^{\otimes n}$ we can define its *characteristic function* as

$$\begin{aligned} \chi_\rho : \mathcal{P}_n &\rightarrow [-1, 1] \\ P &\mapsto \text{Tr}(P\rho). \end{aligned}$$

In the following we consider the task of estimating the fidelity between a desired n -qubit pure state ρ , for which we have a theoretical description, and a general state σ , which we prepared experimentally, by performing measurements on σ . Since ρ is a pure state, we can write the fidelity by using Eq. (98) as

$$\begin{aligned} F(\rho, \sigma) &= \frac{1}{d} \sum_{P \in \mathcal{P}_n} \chi_\rho(P) \chi_\sigma(P) = \frac{1}{d} \\ &+ \sum_{P \in \mathcal{P}_n \setminus \{\mathbb{1}\}} \chi_\rho(P) \chi_\sigma(P). \end{aligned} \quad (99)$$

Assume that we are able to measure any observable $P \in \mathcal{P}_n$ with outcome in $\{1, -1\}$ on the state σ . Here, the observables P can be chosen adaptively and hence depend on the previous ones and their corresponding measurement outcomes. Since $\chi_\sigma(\mathbb{1}) = 1$ we can restrict ourselves to Pauli measurements with $P \in \mathcal{P}_n \setminus \{\mathbb{1}\}$ as for $P = \mathbb{1}$ a tester does not gain any information. In the following we refer to this model as the “Pauli-parity model of fidelity estimation.”

For this model, $\varepsilon > 0$ and $\delta > 0$ are the allowed additive error and the probability of failure respectively and pure state ρ is fixed, which we call the “instanced-based setup.” We denote the *sample complexity of fidelity estimation in the Pauli-parity model* by $N_{\varepsilon, \delta}^P(\rho)$: that is, $N_{\varepsilon, \delta}^P(\rho)$ is the optimal number of measurements on different copies of the unknown state σ needed to estimate $F(\rho, \sigma)$ in the worst case over all states σ .

Due to the definition of the characteristic function χ_σ , the assumption that we are able to perform measurements of the observables $P \in \mathcal{P}_n$ on σ exactly means that we have black box access to the χ_σ . Hence, estimation of the fidelity (99) in this setup becomes a special instance of black box overlap estimation as discussed in Sec. III. In particular we can make the identifications

$$\begin{aligned} \Omega &\mapsto \mathcal{P}_n \setminus \{\mathbb{1}\} \quad \text{and} \quad d\mu(\lambda) \mapsto \frac{1}{d} d\mu_{\text{count}}(P) \text{ with } \mu_{\text{count}} \text{ being the counting measure on } \mathcal{P}_n \setminus \{\mathbb{1}\} \\ r &\mapsto 1 \\ f(\lambda) &\mapsto \chi_\rho(P) \quad \text{to which we have full access} \\ g(\lambda) &\mapsto \chi_\sigma(P) \quad \text{to which we have black box access} \\ Y_\lambda &\mapsto Y_P \quad \text{measurement outcome of observable } P \in \mathcal{P}_n \setminus \{\mathbb{1}\} \text{ on state } \sigma. \end{aligned} \quad (100)$$

In particular this leads to the following convention of L^q -norms for $q \in [1, \infty)$

$$\|\chi_\sigma\|_q := \left(\frac{1}{d} \sum_{P \in \mathcal{P}_n \setminus \{\mathbb{1}\}} |\chi_\sigma(P)|^q \right)^{1/q} \quad (101)$$

and

$$\|\chi_\sigma\|_\infty := \max_{P \in \mathcal{P}_n \setminus \{\mathbb{1}\}} |\chi_\sigma(P)|. \quad (102)$$

With these definitions we have $\|\chi_\sigma\|_2 = \sqrt{\text{Tr}(\sigma^2) - 1/d} \leq 1$ and hence by the Cauchy-Schwarz inequality

$$\|\chi_\rho\|_1 = \frac{1}{d} \sum_{P \in \mathcal{P}_n \setminus \{\mathbb{1}\}} |\chi_\rho(P)| \leq \sqrt{d} \|\chi_\rho\|_2 \leq \sqrt{d}. \quad (103)$$

For the subset $\mathcal{S}$ used in Sec. III, we are interested here in the particular choice $\mathcal{S}_P$ consisting of all characteristic functions χ_σ for valid states σ . We see that $\mathcal{S}_P \subseteq B_1^2(0) \cap B_1^\infty(0)$.

With the notation above, the sample complexity of fidelity estimation can be written as

$$N_{\varepsilon, \delta}^P(\rho) \equiv N_{\varepsilon, \delta}(\chi_\rho, \mathcal{S}_P), \quad (104)$$

with $N_{\varepsilon, \delta}(\chi_\rho, \mathcal{S}_P)$ being defined in Sec. III.

Formulation of the achievability result (Theorem 1) for this special case leads to the following theorem:

Theorem 6 (Upper bound on fidelity estimation with Pauli measurements). Let ρ be a pure state on $(\mathbb{C}^2)^{\otimes n}$ for which we have a theoretical description. Then, for any

state σ , we can estimate the fidelity $F(\rho, \sigma)$ with precision $\varepsilon > 0$ and failure probability at most $\delta > 0$ by performing Pauli measurements on number

$$N_{\varepsilon, \delta}^P(\rho) \leq 2 \inf_{\varepsilon' \in [0, \varepsilon]} \left(\frac{\|\chi_\rho\|_1^{(\varepsilon')}}{\varepsilon - \varepsilon'} \right)^2 \ln \left(\frac{1}{\delta} \right) \quad (105)$$

copies of the state σ . Furthermore, the upper bound holds true even when we restrict ourselves to nonadaptive algorithms.

Remark 8 (Worst case bound). Using Eq. (103), we see that for all pure states ρ the sample complexity in the Pauli model satisfies

$$N_{\varepsilon, \delta}^P(\rho) = \mathcal{O} \left(\frac{d}{\varepsilon^2} \ln \left(\frac{1}{\delta} \right) \right). \quad (106)$$

As we see below, this worst case bound is essentially tight as Theorems 7 and 4 show a lower bound on the sample complexity for Haar-random states which matches with high probability Eq. (106) up to factors logarithmic in the dimension d . Furthermore, we believe that this tightness result can be strengthened by not relying on a randomized argument but, possibly inspired by the construction of the spike states in Sec. IV A 2, finding explicit examples of states for which the lower bound in Theorem 7 can be shown to match Eq. (106) up to constants which are independent of the system size. We, however, leave this as an open problem.

Furthermore, from the lower bound on the sample complexity of black box estimation in Theorem 2, we find the corresponding lower bound on $N_{\varepsilon, \delta}^P(\rho)$:

Theorem 7 (Lower bound on fidelity estimation with Pauli measurements). Let ρ be a pure state on $(\mathbb{C}^2)^{\otimes n}$ for which we have a theoretical description. Then for $0 < \varepsilon < 1/2$ and $0 < \delta \leq 1/3$ we have that the sample complexity of fidelity estimation in the Pauli model is lower bounded as

$$\begin{aligned} N_{\varepsilon, \delta}^P(\rho) &\geq \sup_{\substack{\sigma_1, \sigma_2 \text{ state,} \\ |F(\rho, \sigma_1) - F(\rho, \sigma_2)| > 2\varepsilon}} \left\| \frac{(\chi_{\sigma_1} - \chi_{\sigma_2})^2}{1 - \chi_{\sigma_2}^2} \right\|_\infty^{-1} \ln \left(\frac{1}{3\delta} \right) \\ &\geq \sup_{\substack{\sigma \text{ state,} \\ |F(\rho, \sigma) - 1/d| > 2\varepsilon}} \frac{1}{\|\chi_\sigma\|_\infty^2} \ln \left(\frac{1}{3\delta} \right). \end{aligned} \quad (107)$$

Proof. The first inequality immediately follows from Theorem 2. For the second we use the first with the choice $\sigma_2 = \mathbb{1}/d$, which satisfies $\chi_{\sigma_2} = 0$ on $\mathcal{P}_n \setminus \{\mathbb{1}\}$ and $F(\rho, \sigma_2) = 1/d$. ■

Remark 9 (Matching upper and lower bound for small ε). Inspired from the construction in Lemma 1 consider for some pure state ρ and $\varepsilon > 0$ the operator

$$\omega = \frac{1}{d} \left(\mathbb{1} + \frac{3\varepsilon}{\|\chi_\rho\|_1} \sum_{P \in \mathcal{P}_n \setminus \{\mathbb{1}\}} \text{sgn}(\chi_\rho(P)) P \right),$$

which has characteristic function $\chi_\omega(P) = \text{Tr}(P\omega) = 3\varepsilon/\|\chi_\rho\|_1 \text{sgn}(\chi_\rho(P))$ for all $P \in \mathcal{P}_n \setminus \{\mathbb{1}\}$. This operator ω has normalized trace but is in general not positive semi-definite and hence not a state (see Remark 2) unless $\varepsilon > 0$ is small enough. In particular, a naive operator norm bound shows that ω is in fact positive semi-definite if $0 < \varepsilon \leq 1/3d(d+1)$.

Whenever ω is a state we can combine Eq. (107) with the argument in Lemma 1 to show the instanced-based lower bound on the sample complexity of fidelity estimation in the Pauli model

$$N_{\varepsilon, \delta}^P(\rho) \geq \left(\frac{\|\chi_\rho\|_1}{3\varepsilon} \right)^2 \ln \left(\frac{1}{3\delta} \right). \quad (108)$$

Note that

$$\begin{aligned} F(\rho, \omega) &= \frac{1}{d} \sum_{P \in \mathcal{P}_n} \chi_\rho(P) \chi_\omega(P) \\ &= \frac{1}{d} \left(1 + \sum_{P \in \mathcal{P}_n \setminus \{\mathbb{1}\}} \frac{3\varepsilon |\chi_\rho(P)|}{\|\chi_\rho\|_1} \right) = \frac{1}{d} + 3\varepsilon, \end{aligned}$$

from which we see that ω is a feasible state in the second line of Eq. (107). Furthermore, since by definition $\|\chi_\omega\|_\infty \leq 3\varepsilon/\|\chi_\rho\|_1$ we can conclude the validity of Eq. (108).

A. Matching upper and lower bounds for $N_{\varepsilon, \delta}^P(\rho)$ for examples of states

In this section we consider two examples of families of states ρ for which the upper and lower bounds on $N_{\varepsilon, \delta}^P(\rho)$ provided in Theorems 6 and 7 can be shown to be essentially matching. In particular we give a full characterization of the sample complexity of fidelity estimation in the Pauli model for (1) Haar-random states and (2) stabilizer states in terms of the L^1 -norms of the corresponding characteristic function.

For a Haar-random pure state $\rho = |\Psi\rangle\langle\Psi|$ we find in Proposition 4 with probability of at least $1 - 1/d$ that

$$N_{\varepsilon, \delta}^P(\rho) = \tilde{\Theta} \left(\left(\frac{\|\chi_\rho\|_1}{\varepsilon} \right)^2 \ln \left(\frac{1}{\delta} \right) \right) = \tilde{\Theta} \left(\frac{d}{\varepsilon^2} \ln \left(\frac{1}{\delta} \right) \right). \quad (109)$$

Here, we used $\tilde{\Theta}$ to denote the fact that the upper and lower bounds in Eq. (109) match up to factors that scale logarithmically in the leading order term.

On the other hand, for stabilizer states ρ on n qubits, we have that the L^1 -norm satisfies $\|\chi_\rho\|_1 \leq 1$ and is hence independent of the system size. Using this, we can characterize in Proposition 5 the corresponding sample complexity of fidelity estimation in the Pauli model as

$$N_{\varepsilon, \delta}^P(\rho) = \Theta\left(\frac{1}{\varepsilon^2} \ln\left(\frac{1}{\delta}\right)\right). \quad (110)$$

The upper bound in Eq. (110) can also be found in Ref. [2].

1. Sample complexity for Haar-random states

The following proposition shows that the upper and lower bounds of Theorems 6 and 7 on the sample complexity $N_{\varepsilon, \delta}^P(\rho)$ match (up to factors logarithmic in the leading order) for Haar-random pure states with high probability:

Proposition 4. Let $\rho = |\Psi\rangle\langle\Psi|$ be a Haar-random pure state on $(\mathbb{C}^2)^{\otimes n}$, $0 < \varepsilon, \delta \leq 1/5$. Then with probability of at least $1 - 1/d$, with $d = 2^n$, we have

$$N_{\varepsilon, \delta}^P(\rho) = \tilde{\Theta}\left(\left(\frac{\|\chi_\rho\|_1}{\varepsilon}\right)^2 \ln\left(\frac{1}{\delta}\right)\right) = \tilde{\Theta}\left(\frac{d}{\varepsilon^2} \ln\left(\frac{1}{\delta}\right)\right). \quad (111)$$

Here, we used $\tilde{\Theta}$ to denote the fact that the upper and lower bounds in Eq. (111) match up to factors that scale logarithmically in the leading order term.

Proof. For the upper bound we use Theorem 6, which for all pure states ρ gives

$$N_{\varepsilon, \delta}^P(\rho) \leq 2\left(\frac{\|\chi_\rho\|_1}{\varepsilon}\right)^2 \ln\left(\frac{1}{\delta}\right).$$

To prove the lower bound, we show in the following that with high probability a Haar-random pure state $\rho = |\Psi\rangle\langle\Psi|$ satisfies

$$\|\chi_\rho\|_\infty \lesssim \sqrt{\frac{\ln d}{d}}. \quad (112)$$

Let $U \in \mathbb{U}(d)$ be a Haar unitary such that $|\Psi\rangle = U^\dagger|0\rangle$. For a fixed $P \in \mathcal{P}_n \setminus \{\mathbb{1}\}$ we consider the random variable

$$f(U) = |\langle\Psi|P|\Psi\rangle| = |\langle 0|UPU^\dagger|0\rangle|.$$

The function $f : \mathbb{U}(d) \rightarrow \mathbb{R}$ is 2-Lipschitz with respect to the Frobenius norm denoted $\|\cdot\|_F$. Let $U, V \in \mathbb{U}(d)$ be two

unitary matrices. We have

$$\begin{aligned} |f(U) - f(V)| &= ||\langle 0|UPU^\dagger|0\rangle| - |\langle 0|VPV^\dagger|0\rangle|| \\ &\stackrel{(a)}{\leq} |\langle 0|(U - V)PU^\dagger|0\rangle| \\ &\quad + |\langle 0|VP(U - V)^\dagger|0\rangle| \\ &\stackrel{(b)}{\leq} \|P(U - V)^\dagger|0\rangle\|_2 \|U^\dagger|0\rangle\|_2 \\ &\quad + \|V^\dagger|0\rangle\|_2 \|P(U - V)^\dagger|0\rangle\|_2 \\ &\stackrel{(c)}{\leq} 2\|P(U - V)^\dagger\|_F = 2\|U - V\|_F, \end{aligned}$$

where (a) follows from the triangle inequality, (b) follows from the Cauchy-Schwarz inequality and (c) uses the fact that $\|A|x\rangle\|_2 \leq \|A\|_F$ for a pure $|x\rangle$ as well as the Frobenius norm is unitarily invariant. So by the concentration inequality of Lipschitz functions of Haar unitaries [74], for all $s > 0$

$$\mathbb{P}(|f(U) - \mathbb{E}(f)| \geq s) \leq \exp\left(-\frac{ds^2}{48}\right).$$

Moreover we have by the Cauchy-Schwarz inequality

$$\begin{aligned} \mathbb{E}(f) &\leq \sqrt{\mathbb{E}(f^2)} = \sqrt{\mathbb{E}(\langle\Psi|P|\Psi\rangle^2)} \\ &= \sqrt{\frac{\text{Tr}(PP^\dagger) + |\text{Tr}(P)|^2}{d(d+1)}} = \sqrt{\frac{1}{d+1}}. \end{aligned}$$

So by the union bound

$$\begin{aligned} \mathbb{P}\left(\exists P \in \mathcal{P}_n \setminus \{\mathbb{1}\} : |\langle\Psi|P|\Psi\rangle| > 2\sqrt{\frac{48 \ln d^3}{d}}\right) \\ \leq \sum_{P \in \mathcal{P}_n \setminus \{\mathbb{1}\}} \mathbb{P}\left(|\langle\Psi|P|\Psi\rangle| > 2\sqrt{\frac{48 \ln d^3}{d}}\right) \\ \leq d^2 \mathbb{P}\left(f(U) - \mathbb{E}(f) > 2\sqrt{\frac{48 \ln d^3}{d}} - \sqrt{\frac{1}{d+1}}\right) \\ \leq d^2 \mathbb{P}\left(f(U) - \mathbb{E}(f) > \sqrt{\frac{48 \ln d^3}{d}}\right) \\ \leq d^2 \exp\left(-\frac{d \cdot \frac{48 \ln d^3}{d}}{48}\right) = \frac{1}{d}. \end{aligned}$$

Hence, when sampling from the Haar measure, with probability of at least $1 - 1/d$, we find a pure state $|\Psi\rangle$ such that for all $P \in \mathcal{P}_n \setminus \{\mathbb{1}\}$

$$|\langle\Psi|P|\Psi\rangle| \leq 2\sqrt{\frac{48 \ln d^3}{d}}$$

and therefore the corresponding pure state $\rho = |\Psi\rangle\langle\Psi|$ satisfies Eq. (112). Consider the state $\sigma = 5\varepsilon\rho + (1 -$

$5\varepsilon\mathbb{1}/d$, which satisfies $F(\rho, \sigma) - 1/d = 5\varepsilon(1 - 1/d) > 2\varepsilon$ and has characteristic function $\chi_\sigma = 5\varepsilon\chi_\rho$ on $\mathcal{P}_n \setminus \{\mathbb{1}\}$. Hence, this state is feasible in the optimization in the second lower bound provided in Theorem 7, which with the use of Eq. (103) gives

$$\begin{aligned} N_{\varepsilon, \delta}^P(\rho) &\gtrsim \frac{1}{\varepsilon^2 \|\chi_\rho\|_\infty^2} \ln \left(\frac{1}{3\delta} \right) \gtrsim \frac{d}{\varepsilon^2 \ln d} \ln \left(\frac{1}{3\delta} \right) \\ &\geq (\ln d)^{-1} \left(\frac{\|\chi_\rho\|_1}{\varepsilon} \right)^2 \ln \left(\frac{1}{3\delta} \right). \end{aligned} \quad (113)$$

■

2. Sample complexity for stabilizer states

We say a pure state $\rho = |\psi\rangle\langle\psi|$ on $(\mathbb{C}^2)^{\otimes n}$ is a stabilizer state if there are 2^n Pauli strings $P \in \mathcal{P}_n$ satisfying $P|\psi\rangle = |\psi\rangle$.

Proposition 5. Let $\rho = |\psi\rangle\langle\psi|$ be a stabilizer state on $(\mathbb{C}^2)^{\otimes n}$. Then for all $0 < \varepsilon, \delta < 1/5$ we have that

$$N_{\varepsilon, \delta}^P(\rho) = \Theta \left(\frac{1}{\varepsilon^2} \ln \left(\frac{1}{\delta} \right) \right). \quad (114)$$

Proof. It is well-known that all the 2^n Pauli strings $P \in \mathcal{P}_n$ which satisfy $P|\psi\rangle = |\psi\rangle$ commute, and furthermore for all remaining $Q \in \mathcal{P}_n$ we have $\chi_\rho(Q) = \langle\psi, Q\psi\rangle = 0$. Hence, we have that the L^1 -norm of the characteristic function of ρ satisfies

$$\|\chi_\rho\|_1 = \frac{1}{d} \sum_{P \in \mathcal{P}_n \setminus \{\mathbb{1}\}} |\chi_\rho(P)| = \frac{d-1}{d} \leq 1. \quad (115)$$

From Theorem 6 we hence see

$$N_{\varepsilon, \delta}^P(\rho) = \mathcal{O} \left(\frac{1}{\varepsilon^2} \ln \left(\frac{1}{\delta} \right) \right). \quad (116)$$

For the remaining lower bound we use the second line in Eq. (107) in Theorem 7 for the state $\sigma = 5\varepsilon\rho + (1 - 5\varepsilon)\mathbb{1}/d$, which is feasible since $F(\rho, \sigma) - 1/d = 5\varepsilon(1 - 1/d) > 2\varepsilon$ and satisfies $\|\chi_\sigma\|_\infty = \max_{P \in \mathcal{P}_n \setminus \{\mathbb{1}\}} |\chi_\sigma(P)| \leq 5\varepsilon$. ■

ACKNOWLEDGMENTS

The authors thank Ulysse Chabaud, Giulia Ferrini, Niklas Galke, Naixu Guo, Benjamin Huard, Hector Hutin, Ludovico Lami, Lauritz van Luijk and Corentin Moumard for helpful discussions. O.F. and R.S. acknowledge funding from the European Research Council (AlgoQIP, Grant Agreement No. 851716), from the QuantERA II program of the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 101017733 (VERIQ-TAS) and a government grant managed by the

Agence Nationale de la Recherche under Plan France 2030 (reference ANR-22-PETQ-0007). A.O. acknowledges funding by the European Research Council (Grant Agreement No. 948139).

DATA AVAILABILITY

No data were created or analyzed in this study.

APPENDIX A: SOME FACTS ABOUT THE SMOOTHED L^1 -NORM

For Ω a measure space with measure μ , we defined in Sec. II B the smoothed L^1 -norm of a function $f \in L^2(\Omega)$ as

$$\|f\|_1^{(\varepsilon)} = \inf_{\substack{\tilde{f} \in L^1(\Omega) \\ \|f - \tilde{f}\|_2 \leq \varepsilon}} \|\tilde{f}\|_1. \quad (A1)$$

Differently from the scope of the rest of this paper, in which we restricted ourselves to real valued functions, we allow in this section for real or complex valued functions. Hence, for simplicity, we write as $L^q(\Omega)$ the corresponding L^q -spaces without denoting the field over which the functions considered are taking values.

The following lemma establishes that in the limit $\varepsilon \rightarrow 0$ the smoothed L^1 -norm converges to the usual (non-smoothed) L^1 -norm:

Lemma A1. For all $f \in L^2(\Omega)$ we have

$$\lim_{\varepsilon \downarrow 0} \|f\|_1^{(\varepsilon)} = \|f\|_1, \quad (A2)$$

where the right-hand side is understood as infinity if $f \notin L^1(\Omega)$.

Proof. We use the relation for a measurable function f

$$\int_{\Omega} |f(\lambda)| d\mu(\lambda) = \sup_{\substack{\Omega_0 \subseteq \Omega \\ \mu(\Omega_0) < \infty}} \int_{\Omega_0} |f(\lambda)| d\mu(\lambda), \quad (A3)$$

which follows directly from the definition of the Lebesgue integral.

First we consider the case $f \in L^1(\Omega)$. Let $\delta > 0$. Using Eq. (A3), we can pick a set of finite measure $\Omega_0 \subseteq \Omega$ such that

$$\int_{\Omega_0^c} |f(\lambda)| d\mu(\lambda) \leq \frac{\delta}{2}.$$

Let $\varepsilon \in (0, \delta/2\sqrt{\mu(\Omega_0)})$ and $\tilde{f} \in L^1(\Omega) \cap L^2(\Omega)$ be such that $\|f - \tilde{f}\|_2 \leq \varepsilon$. Then

$$\begin{aligned} \|f\|_1 &\leq \int_{\Omega_0} |f(\lambda) - \tilde{f}(\lambda)| d\mu(\lambda) + \int_{\Omega_0} |\tilde{f}(\lambda)| d\mu(\lambda) \\ &\quad + \int_{\Omega_0^c} |f(\lambda)| d\mu(\lambda) \\ &\leq \int_{\Omega_0} |f(\lambda) - \tilde{f}(\lambda)| d\mu(\lambda) + \|\tilde{f}\|_1 + \frac{\delta}{2}. \end{aligned}$$

For the first term we use the Cauchy-Schwarz inequality, which gives

$$\int_{\Omega_0} |f(\lambda) - \tilde{f}(\lambda)| d\mu(\lambda) \leq \|f - \tilde{f}\|_2 \sqrt{\mu(\Omega_0)} \leq \frac{\delta}{2}.$$

Using the fact that $\tilde{f}$ was arbitrary under the constraints above, we hence obtain

$$\|f\|_1^{(\varepsilon)} \geq \|f\|_1 - \delta.$$

Noting that the opposite inequality $\|f\|_1^{(\varepsilon)} \leq \|f\|_1$ is trivially true and using the fact that $\delta > 0$ was arbitrary, we obtain $\lim_{\varepsilon \downarrow 0} \|f\|_1^{(\varepsilon)} = \|f\|_1$.

To finish the proof we consider the case $f \notin L^1(\Omega)$. We have to show $\liminf_{\varepsilon \downarrow 0} \|f\|_1^{(\varepsilon)} = \infty$. Assume for contradiction that there exists $C > 0$ such that for all $\varepsilon > 0$ we have the uniform bound

$$\|f\|_1^{(\varepsilon)} \leq C.$$

Let $\Omega_0 \subseteq \Omega$ be a set of finite measure. Note by arguments similar to those above, it holds true that $\|f \mathbf{1}_{\Omega_0}\|_1^{(\varepsilon)} \leq \|f\|_1^{(\varepsilon)} \leq C$ for all $\varepsilon > 0$. Hence, using the established Eq. (A2) for the function $f \mathbf{1}_{\Omega_0} \in L^1(\Omega)$, we see

$$\int_{\Omega_0} |f| d\mu(\lambda) = \lim_{\varepsilon \downarrow 0} \|f \mathbf{1}_{\Omega_0}\|_1^{(\varepsilon)} \leq C.$$

But use of the fact that Ω_0 was an arbitrary set of finite measure together with Eq. (A3) shows that $f \in L^1(\Omega)$, which gives a contradiction and finishes the proof. ■

1. Lower bound on the smoothed L^1 -norm for functions with a rapidly decaying tail

The smoothed L^1 -norm of a function f , defined in Eq. (19), satisfies the trivial upper bound $\|f\|_1^{(\varepsilon)} \leq \|f\|_1$ for all $\varepsilon > 0$. The gap here between the smoothed L^1 -norm and the nonsmoothed L^1 -norm can in general be arbitrarily large even for L^1 -functions for which the right-hand side is finite. In particular, although by Lemma A1 we know that $\lim_{\varepsilon \downarrow 0} \|f\|_1^{(\varepsilon)} = \|f\|_1$, in general we have no control of the speed of convergence when we are taking this limit.

In this section, we, however, restrict ourselves to functions with a *rapidly decaying tail* [as made precise in Eq. (A6)] and for which we can provide good control in the relevant lower bound (see Lemma A2). Here, for instances of interest, we obtain the lower bound

$$\|f\|_1^{(\varepsilon)} \geq \|f\|_1 (1 - C\varepsilon^\beta) \quad (\text{A4})$$

for some $C \geq 0$ and $\beta \in (0, 1]$ independent of ε and the specific function f . In particular Eq. (A4) holds true for

f being the Wigner functions $W_{|n\rangle\langle n|}$ of the 1-mode Fock states as shown in Proposition A1.

For functions satisfying Eq. (A4), we find that [75]

$$\inf_{\varepsilon' \in [0, \varepsilon)} \frac{\|f\|_1^{(\varepsilon')}}{\varepsilon - \varepsilon'} = \Theta\left(\frac{\|f\|_1}{\varepsilon}\right) \quad (\text{A5})$$

for $\varepsilon > 0$ small enough. As seen in the previous sections, the sample complexity of black box overlap estimation and fidelity estimation in the Wigner model and the Pauli model can be upper bounded in terms of the smoothed L^1 -norm, with the quantity of interest being the left-hand side of Eq. (A5) (see Theorems 1, 3 and 6). Therefore, by Eq. (A5) we see that for functions satisfying Eq. (A4), we do not weaken this upper bound when choosing $\varepsilon' = 0$ and providing it in terms of the (nonsmoothed) L^1 -norm. This, hence, further justifies that the characterization of the sample complexity of the Fock states found in Proposition 1 is in terms of the usual L^1 -norms of the Wigner functions $W_{|n\rangle\langle n|}$.

In the remainder of this section we formally define the notion of functions with a rapidly decaying tail and prove the mentioned lower bound on their smoothed L^1 -norm in Lemma A2. After that we discuss as an example the relevant class of *eventually exponentially decaying* functions on $\mathbb{R}^m$, which includes the Wigner functions of Fock states.

Let Ω be some measurable space with some measure μ . For $\gamma, \kappa > 0$ and $\Omega_0 \subseteq \Omega$ a finite measure set, we say a measurable function f has a *rapidly decaying tail of order* (γ, κ) *outside* Ω_0 if for all $\delta > 0$ we have

$$\int_{\{|f| \leq \delta\} \cap \Omega_0^c} |f(\lambda)| d\mu(\lambda) \leq \kappa \delta^\gamma. \quad (\text{A6})$$

For such functions we have good control of the respective smoothed L^1 -norm as proven in the following lemma. In particular the proof follows lines similar to those for Lemma A1, however providing additionally a convergence rate as $\varepsilon \rightarrow 0$ by use of the rapidly decaying tail assumption.

Lemma A2 (Lower bound on the smoothed L^1 -norm). Let $\varepsilon, \gamma, \kappa > 0$ and $\Omega_0 \subseteq \Omega$ be a set of finite measure. Furthermore, let $f \in L^2(\Omega)$ with $\|f\|_2 \leq 1$ have a rapidly decaying tail of order (γ, κ) outside Ω_0 . Then

$$\|f\|_1^{(\varepsilon)} \geq \|f\|_1 - \kappa' \varepsilon^\beta - \varepsilon \sqrt{\mu(\Omega_0)}, \quad (\text{A7})$$

with $\beta := \gamma/(\gamma + 1) \in (0, 1]$ and $\kappa' := 2\kappa^{1/(1+\gamma)}$.

Remark A1. Note that since f in Lemma A2 is supposed to be in $L^2(\Omega)$ and have a rapidly decaying tail outside a finite measure set, it is immediately also in $L^1(\Omega)$. Hence, the right-hand side of Eq. (A7) is finite.

Proof. Let $\tilde{f} \in L^1(\Omega) \cap L^2(\Omega)$ with $\|f - \tilde{f}\|_2 \leq \varepsilon$ and $\delta > 0$ to be determined later. We have

$$\begin{aligned} \|f\|_1 &= \int_{\{|f| > \delta\}} |f(\lambda)| d\mu(\lambda) + \int_{\{|f| \leq \delta\} \cap \Omega_0} |f(\lambda)| d\mu(\lambda) \\ &\quad + \int_{\{|f| \leq \delta\} \cap \Omega_0^c} |f(\lambda)| d\mu(\lambda) \\ &\leq \int_{\{|f| > \delta\}} |f(\lambda) - \tilde{f}(\lambda)| d\mu(\lambda) \\ &\quad + \int_{\Omega_0} |f(\lambda) - \tilde{f}(\lambda)| d\mu(\lambda) + \|\tilde{f}\|_1 + \kappa \delta^\gamma. \quad (\text{A8}) \end{aligned}$$

For the first term we use with the Cauchy-Schwarz inequality together with the fact that $\|f\|_2 \leq 1$, which gives

$$\begin{aligned} &\int_{\{|f| > \delta\}} |f(\lambda) - \tilde{f}(\lambda)| d\mu(\lambda) \\ &\leq \|f - \tilde{f}\|_2 \sqrt{\mu(\{|f| > \delta\})} \leq \frac{\varepsilon}{\delta}. \end{aligned}$$

Similarly, we see for the second term in Eq. (A8) that

$$\int_{\Omega_0} |f(\lambda) - \tilde{f}(\lambda)| d\mu(\lambda) \leq \varepsilon \sqrt{\mu(\Omega_0)}.$$

Plugging this into Eq. (A8) and using the fact that $\tilde{f}$ is an arbitrary feasible function in the optimization corresponding to the smoothed L^1 -norm in Eq. (19), we see

$$\|f\|_1^{(\varepsilon)} \geq \|f\|_1 - \varepsilon \sqrt{\mu(\Omega_0)} - \frac{\varepsilon}{\delta} - \kappa \delta^\gamma.$$

Optimization over $\delta > 0$ and the choosing of $\delta = (\varepsilon/\kappa)^{1/(\gamma+1)}$ finishes the proof. ■

For $m \in \mathbb{N}$ and the special case $\Omega = \mathbb{R}^m$ equipped with the Lebesgue measure, a class of particular interest which satisfies the condition (A6) is *eventually exponentially decaying functions*, i.e., $f \in L^1(\mathbb{R}^m)$, which satisfy for some $t, C > 0$ and $\Omega_0 \subset \mathbb{R}^m$ of finite measure

$$|f(\alpha)| \leq C e^{-t|\alpha|} \quad (\text{A9})$$

for all $\alpha \in \mathbb{R}^m \setminus \Omega_0$. Such functions have a rapidly decaying tail of order $(\gamma, \kappa_{C,\gamma,t,m})$ outside Ω_0 for every $\gamma \in (0, 1)$ fixed and $\kappa_{C,\gamma,t,m}$ being linear in C and $\kappa_{C,\gamma,t,m}/C$ depending on γ, t, m but not f . This essentially follows due to the standard inequality [76]

$$\begin{aligned} \int_{\{e^{-t|\alpha|} \leq \delta\}} e^{-t|\alpha|} d\alpha &\lesssim \int_{\ln((1/\delta)^{1/t})}^{\infty} e^{-tr} r^{m-1} dr \\ &\lesssim \delta (\ln(1/\delta) + 1), \quad (\text{A10}) \end{aligned}$$

where constants independent of δ are hidden in the $\lesssim$ -notation.

Note that the Wigner functions of the (1-mode) Fock states $W_{|n\rangle\langle n|}$, which are discussed in Sec. IV A 1 and given in terms of Laguerre polynomials (71), are eventually exponentially decaying in the sense of Eq. (A9). This can be seen directly from the asymptotic expressions of the Laguerre polynomials provided in Ref. [67] (consider there the table on page 699 for convenience). From that we can conclude the validity of the mentioned lower bound on their respective smoothed L^1 -norm in the following proposition:

Proposition A1. The Wigner functions of the 1-mode Fock states $(|n\rangle\langle n|)_{n \in \mathbb{N}}$ satisfy

$$\|W_{|n\rangle\langle n|}\|_1^{(\varepsilon)} \geq \|W_{|n\rangle\langle n|}\|_1 (1 - C\varepsilon^\beta) \sim \sqrt{n}(1 - C\varepsilon^\beta) \quad (\text{A11})$$

for all $\varepsilon > 0, \beta \in (0, 1/2)$ and some $C > 0$ possibly dependent on β but independent of n and ε .

Proof. We use the fact that the Wigner function of the Fock states $|n\rangle\langle n|$ can be written explicitly [see Eq. (36) in Ref. [70] and Eq. (7.3) in [71]] in terms of the Laguerre polynomials as

$$W_{|n\rangle\langle n|}(\alpha) = (-1)^n \left(\frac{2}{\pi}\right) e^{-2|\alpha|^2} L_n(4|\alpha|^2).$$

Using the asymptotic expressions for the Laguerre polynomials in Ref. [67], we see that

$$W_{|n\rangle\langle n|}(\alpha) \leq c e^{-t|\alpha|^2}$$

for all $|\alpha| \geq s\sqrt{n}$ for some $s, c, t > 0$ independent of n . Therefore, by the discussion around Eq. (A9) we see that $W_{|n\rangle\langle n|}$ has a rapidly decaying tail of order $(\gamma, \kappa_{c,\gamma,t})$ outside $\Omega_0 = B_{s\sqrt{n}}(0) \subset \mathbb{C}$ for all $\gamma \in (0, 1)$. Here, $\kappa_{c,\gamma,t} > 0$ depends linearly on c and is independent of n . Noting that the volume of Ω_0 is of order n , we see from Lemma A2 that

$$\|W_{|n\rangle\langle n|}\|_1 - \|W_{|n\rangle\langle n|}\|_1^{(\varepsilon)} \lesssim \varepsilon^\beta + \sqrt{n}\varepsilon \lesssim \sqrt{n}\varepsilon^\beta,$$

where $\beta = \gamma/(\gamma + 1) \in (0, 1/2)$ and we are hiding all constants which are independent of n and ε in the $\lesssim$ -notation. Use of the fact that $\|W_{|n\rangle\langle n|}\|_1 \sim \sqrt{n}$ as noted in Eq. (74) as a consequence of Lemma 1 in Ref. [16] finishes the proof. ■

APPENDIX B: ROBUSTNESS OF FIDELITY ESTIMATION UNDER PHOTON LOSS NOISE

In the following, we discuss the robustness of the ideal protocol proposed in Theorem 3 in the case of photon loss

noise. Considering the 1-mode case for simplicity, photon loss is usually described as the quantum dynamical semigroup $\Lambda_t \equiv e^{t\mathcal{L}_{\text{loss}}}$ corresponding to the Lindbladian

$$\mathcal{L}_{\text{loss}}(\sigma) = \kappa(a\sigma a^* - \frac{1}{2}(N\sigma + \sigma N)),$$

with $\kappa \geq 0$ being the photon loss rate; see Sec. 5.2 in Ref. [77] for a general discussion of the photon loss semigroup.

Even for arbitrarily short times $t > 0$, it is known that the photon loss semigroup can distort the quantum system maximally with respect to trace distance when general input states σ are considered, i.e., more precisely

$$\sup_{\sigma} \frac{1}{2} \|\Lambda_t(\sigma) - \sigma\|_1 = 1 \quad (\text{B1})$$

for all $t > 0$ (see Proposition 1 in Ref. [78]). Hence, in the worst case over all prepared states σ , we do not expect that the protocol proposed for the achievability result (Theorem 3) is robust with respect to photon loss.

However, actual experiments can use only a finite amount of energy to prepare the quantum system, which is why it is reasonable to consider only *energy-constrained states* satisfying $\text{Tr}(N\sigma) \leq E$ for some fixed energy threshold $E \geq 0$. For such states, the photon loss semigroup satisfies the continuity bound (see Theorem 1 and Example 5 in Ref. [79])

$$\frac{1}{2} \|\Lambda_t(\sigma) - \sigma\|_1 \leq 3\sqrt{\kappa Et}. \quad (\text{B2})$$

This bound can be used to obtain a robustness analysis of Theorem 3: First, let us consider for simplicity the case of the target state ρ satisfying $W_{\rho} \in L^1(\mathbb{C})$. Following the protocol in Sec. III A a tester samples points in phase space according to

$$\alpha \sim p(\alpha) = \frac{|W_{\rho}(\alpha)|}{\|W_{\rho}\|_1}$$

and for obtained sample $\alpha \in \mathbb{C}$ desires to measure the prepared state σ with respect to the displaced parity operator $(2/\pi)D(\alpha)\Pi D^*(\alpha)$. In actual experiments, following the standard approach from Ref. [4], this is obtained by storing the state σ at first in a cavity and then displacing it, resulting in the state $\sigma_{\alpha} = (2/\pi)D(\alpha)\sigma D^*(\alpha)$. The cavity is then coupled to an auxiliary qubit for a fixed interaction time $t_p > 0$, after which the qubit is measured, which effectively results in the measuring of the parity operator $\Pi = (-1)^N$ for the state σ_{α} . However, assuming noise in terms of photon loss, the state which is actually measured after this interaction time is $\Lambda_{t_p}(\sigma_{\alpha})$. We denote the outcome of this measurement by $Y_{\alpha} \in \{-2/\pi, 2/\pi\}$, which in particular satisfies $\mathbb{E}[Y_{\alpha}] = W_{\Lambda_{t_p}(\sigma_{\alpha})}(0)$ for $\alpha \in \mathbb{C}$ fixed.

The tester then considers the random variable

$$X = \text{sgn}(W_{\rho}(\alpha))\|W_{\rho}\|_1 Y_{\alpha}.$$

In total the tester executes this whole process N times to draw multiple samples from X and build an empirical average.

Note that on expectation we have

$$\mathbb{E}[X] = \pi \int_{\mathbb{C}} W_{\rho}(\alpha) W_{\Lambda_{t_p}(\sigma_{\alpha})}(0) d\alpha.$$

From that and Eq. (56) we see

$$\begin{aligned} |F(\rho, \sigma) - \mathbb{E}[X]| &\leq \pi \int_{\mathbb{C}} |W_{\rho}(\alpha)| \left| W_{\sigma_{\alpha}}(0) - W_{\Lambda_{t_p}(\sigma_{\alpha})}(0) \right| d\alpha \\ &\leq 2 \int_{\mathbb{C}} |W_{\rho}(\alpha)| \|\sigma_{\alpha} - \Lambda_{t_p}(\sigma_{\alpha})\|_1 d\alpha \\ &\lesssim \sqrt{\kappa t_p} \int_{\mathbb{C}} |W_{\rho}(\alpha)| \sqrt{E + |\alpha|^2} d\alpha, \end{aligned} \quad (\text{B3})$$

where the second inequality follows by Eq. (29) and for the last inequality we used the continuity bound for the photon loss semigroup (B2) together with the fact that the displaced state has energy $\text{Tr}(N\sigma_{\alpha}) \sim \text{Tr}(N\sigma) + |\alpha|^2$.

Hence, the protocol can be made robust given that $\int_{\mathbb{C}} |W_{\rho}(\alpha)| \sqrt{E + |\alpha|^2} d\alpha < \infty$, which can be seen as some kind of energy constraint on the target state ρ [80]. In particular, for fixed desired accuracy $\varepsilon > 0$ to estimate the fidelity, the tester obtains an estimator of the fidelity in the worst case over all prepared states σ with average energy bounded by $E \geq 0$ and target states ρ satisfying

$$\eta \int_{\mathbb{C}} |W_{\rho}(\alpha)| \sqrt{E + |\alpha|^2} d\alpha < \varepsilon, \quad (\text{B4})$$

where we defined $\eta = \sqrt{\kappa t_p}$, which can be seen as a parameter modeling the overall strength of the noise.

A bound on the corresponding sample complexity in the presence of photon loss noise can be obtained similarly as discussed in Sec. III A as we outline in the following: As discussed there, we can avoid the assumption $W_{\rho} \in L^1$ and in the first step of the protocol sample instead with respect to $\tilde{p}(\alpha) = \tilde{W}(\alpha)/\|\tilde{W}\|_1$ for some $\tilde{W} \in L^1(\mathbb{C})$ satisfying $\|W_{\rho} - \tilde{W}\|_2 < \varepsilon$. The tester then uses $\tilde{X} = \text{sgn}(\tilde{W}(\alpha))\|\tilde{W}\|_1 Y_{\alpha}$ as an estimator which satisfies $|\tilde{X}| \leq \|\tilde{W}\|_1$. Similarly as in the expression for the smoothed L^1 -norm, we can then perform optimization over all such functions $\tilde{W}$ satisfying additionally the energy constraint (B4) with W_{ρ} replaced by $\tilde{W}$. More precisely, we

find using Hoeffding's inequality that

$$N = \mathcal{O} \left(\inf_{\substack{\tilde{W} \in L^1(\mathbb{C}), \varepsilon_1, \varepsilon_2 > 0 \\ \text{s.t. } \varepsilon_1 + \varepsilon_2 < \varepsilon \\ \|W_\rho - \tilde{W}\|_2 \leq \varepsilon_1, \\ \eta \int_{\mathbb{C}} |\tilde{W}(\alpha)| \sqrt{E + |\alpha|^2} d\alpha \leq \varepsilon_2}} \left(\frac{\|\tilde{W}\|_1}{\varepsilon - \varepsilon_1 - \varepsilon_2} \right)^2 \ln \left(\frac{1}{\delta} \right) \right) \quad (\text{B5})$$

copies of the state σ are sufficient to estimate the fidelity $F(\rho, \sigma)$ using displaced parity measurements in the worst case over all states satisfying $\text{Tr}(N\sigma) \leq E$.

-
- [1] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Practical characterization of quantum devices without tomography, *Phys. Rev. Lett.* **107**, 210404 (2011).
 - [2] S. T. Flammia and Y.-K. Liu, Direct fidelity estimation from few Pauli measurements, *Phys. Rev. Lett.* **106**, 230501 (2011).
 - [3] Antoine Royer, Wigner function as the expectation value of a parity operator, *Phys. Rev. A* **15**, 449 (1977).
 - [4] L. G. Lutterbach and L. Davidovich, Method for direct measurement of the Wigner function in cavity QED and ion traps, *Phys. Rev. Lett.* **78**, 2547 (1997).
 - [5] G. Nogues, A. Rauschenbeutel, S. Osnaghi, P. Bertet, M. Brune, J. M. Raimond, S. Haroche, L. G. Lutterbach, and L. Davidovich, Measurement of a negative value for the Wigner function of radiation, *Phys. Rev. A* **62**, 054101 (2000).
 - [6] P. Bertet, A. Auffeves, P. Maioli, S. Osnaghi, T. Meunier, M. Brune, J. M. Raimond, and S. Haroche, Direct measurement of the Wigner function of a one-photon Fock state in a cavity, *Phys. Rev. Lett.* **89**, 200402 (2002).
 - [7] Brian Vlastakis, Gerhard Kirchmair, Zaki Leghtas, Simon E. Nigg, Luigi Frunzio, S. M. Girvin, Mazhar Mirrahimi, M. H. Devoret, and R. J. Schoelkopf, Deterministically encoding quantum information using 100-photon Schrödinger cat states, *Science* **342**, 607 (2013).
 - [8] Antoine Marquet, Simon Dupouy, Ulysse Réglade, Antoine Essig, Joachim Cohen, Emanuele Abertinale, Audrey Bienfait, Théau Peronnin, Sébastien Jezouin, Raphaël Lescanne, and Benjamin Huard, Harnessing two-photon dissipation for enhanced quantum measurement and control, *Phys. Rev. Applied* **22**, 034053 (2024).
 - [9] A. Marquet, A. Essig, J. Cohen, N. Cottet, A. Murani, E. Albertinale, S. Dupouy, A. Bienfait, T. Peronnin, S. Jezouin, R. Lescanne, and B. Huard, Autoparametric resonance extending the bit-flip time of a cat qubit up to 0.3 s, *Phys. Rev. X* **14**, 021019 (2024).
 - [10] Harald Putterman, Kyungjoo Noh, Connor T. Hann, Gregory S. MacCabe, Shahriar Aghaeimeibodi, Rishi N. Patel, Menyoun Lee, William M. Jones, Hesam Moradinejad, Roberto Rodriguez, *et al.*, Hardware-efficient quantum error correction using concatenated bosonic qubits, *Nature* **638**, 927–934 (2025).
 - [11] U. Réglade, A. Bocquet, R. Gautier, J. Cohen, A. Marquet, E. Albertinale, N. Pankratova, M. Hallén, F. Rautschke, L.-A. Sellem, P. Rouchon, A. Sarlette, M. Mirrahimi, P. Campagne-Ibarcq, R. Lescanne, S. Jezouin, and Z. Leghtas, Quantum control of a cat qubit with bit-flip times exceeding ten seconds, *Nature* **629**, 778 (2024).
 - [12] Chen Wang, Yvonne Y. Gao, Philip Reinhold, R. W. Heeres, Nissim Ofek, Kevin Chou, Christopher Axline, Matthew Reagor, Jacob Blumoff, K. M. Sliwa, L. Frunzio, S. M. Girvin, Liang Jiang, M. Mirrahimi, M. H. Devoret, and R. J. Schoelkopf, A Schrödinger cat living in two boxes, *Science* **352**, 1087 (2016).
 - [13] More precisely, we choose in Sec. V the set $\Omega_P = \mathcal{P}_n \setminus \{\mathbb{1}\}$ instead of $\mathcal{P}_n$ as on the identity string the characteristic function of any state σ trivially satisfies $\chi_\sigma(\mathbb{1}) = 1$ and hence no information is gained when we measure the observable $\mathbb{1}$ in σ .
 - [14] Gregory Valiant and Paul Valiant, An automatic inequality prover and instance optimal identity testing, *SIAM J. Comput.* **46**, 429 (2017).
 - [15] Sitan Chen, Jerry Li, and Ryan O'Donnell, Toward instance-optimal state certification with incoherent measurements, in *Conference on Learning Theory* (PMLR, London, 2022), p. 2541.
 - [16] Clemens Markett, Mean Cesàro summability of Laguerre expansions and norm estimates with shifted parameter, *Anal. Math.* **8**, 19 (1982).
 - [17] Francesco Albarelli, Marco G. Genoni, Matteo G. A. Paris, and Alessandro Ferraro, Resource theory of quantum non-Gaussianity and Wigner negativity, *Phys. Rev. A* **98**, 052350 (2018).
 - [18] Anatole Kenfack and Karol Życzkowski, Negativity of the Wigner function as an indicator of non-classicality, *J. Opt. B: Quantum Semiclassical Opt.* **6**, 396 (2004).
 - [19] K. C. Tan, S. Choi, and H. Jeong, Negativity of quasiprobability distributions as a measure of nonclassicality, *Phys. Rev. Lett.* **124**, 110404 (2020).
 - [20] R. I. Booth, U. Chabaud, and P.-E. Emeriau, Contextuality and Wigner negativity are equivalent for continuous-variable quantum measurements, *Phys. Rev. Lett.* **129**, 230401 (2022).
 - [21] S. D. Bartlett, B. C. Sanders, S. L. Braunstein, and K. Nemoto, Efficient classical simulation of continuous variable quantum information processes, *Phys. Rev. Lett.* **88**, 097904 (2002).
 - [22] A. Mari and J. Eisert, Positive Wigner functions render classical simulation of quantum computation efficient, *Phys. Rev. Lett.* **109**, 230503 (2012).
 - [23] Victor Veitch, Nathan Wiebe, Christopher Ferrie, and Joseph Emerson, Efficient simulation scheme for a class of quantum optics experiments with non-negative Wigner representation, *New J. Phys.* **15**, 013037 (2013).
 - [24] Oliver Hahn, Giulia Ferrini, and Ryuji Takagi, Bridging magic and non-Gaussian resources via Gottesman-Kitaev-Preskill encoding, *PRX Quantum* **6**, 010330 (2025).
 - [25] H. Pashayan, J. J. Wallman, and S. D. Bartlett, Estimating outcome probabilities of quantum circuits using quasiprobabilities, *Phys. Rev. Lett.* **115**, 070501 (2015).
 - [26] C. Calcluth, A. Ferraro, and G. Ferrini, Vacuum provides quantum advantage to otherwise simulatable architectures, *Phys. Rev. A* **107**, 062414 (2023).

- [27] L. García-Álvarez, C. Calcluth, A. Ferraro, and G. Ferrini, Efficient simulatability of continuous-variable circuits with large Wigner negativity, *Phys. Rev. Res.* **2**, 043322 (2020).
- [28] E. T. Campbell, Catalysis and activation of magic states in fault-tolerant architectures, *Phys. Rev. A—At., Mol. Opt. Phys.* **83**, 032317 (2011).
- [29] O. Hahn, A. Ferraro, L. Hultquist, G. Ferrini, and L. García-Álvarez, Quantifying qubit magic resource with Gottesman-Kitaev-Preskill encoding, *Phys. Rev. Lett.* **128**, 210502 (2022).
- [30] In the main text of Ref. [2] the authors claim to have proven $\sup_{\rho \text{ pure}} N_{\varepsilon, \delta}^P(\rho) = \mathcal{O}((1/\varepsilon^2 \delta) + d/\varepsilon^2 \ln(1/\delta))$. However, considering their proof in the appendix (in particular the “Truncating bad events” section), the error parameter β introduced there needs to scale as ε in order to estimate the fidelity with the desired accuracy leading to a $1/\varepsilon^4$ scaling of their actual upper bound.
- [31] L. Leone, S. F. E. Oliviero, and A. Hama, Nonstabilizer-ness determining the hardness of direct fidelity estimation, *Phys. Rev. A* **107**, 022429 (2023).
- [32] Naixu Guo, Feng Pan, and Patrick Rebentrost, Estimating properties of a quantum state by importance-sampled operator shadows, *ArXiv:quant-ph/2305.09374*.
- [33] V. V. Sivak, A. Eickbusch, H. Liu, B. Royer, I. Tsioutsios, and M. H. Devoret, Model-free quantum control with reinforcement learning, *Phys. Rev. X* **12**, 011059 (2022).
- [34] Ulysse Chabaud, Roohollah Ghobadi, Salman Beigi, and Saleh Rahimi-Keshari, Phase-space negativity as a computational resource for quantum kernel methods, *Quantum* **8**, 1519 (2024).
- [35] Barry Simon, The Weyl transform and L^p functions on phase space, *Proc. Am. Math. Soc.* **116**, 1045 (1992).
- [36] Jayadev Acharya, Abhilash Dharmavarapu, Yuhang Liu, and Nengkun Yu, Pauli Measurements Are Not Optimal for Single-Copy Tomography, in *Proceedings of the 57th Annual ACM Symposium on Theory of Computing (Association for Computing Machinery, New York, NY, USA, 2025)*, pp. 718–729.
- [37] Hsin-Yuan Huang, John Preskill, and Mehdi Soleimanifar, Certifying almost all quantum states with few single-qubit measurements, *Nat. Phys.* (2025).
- [38] Ulysse Chabaud, Tom Douce, Frédéric Grosshans, Elham Kashefi, and Damian Markham, in *15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020)*, Leibniz International Proceedings in Informatics (LIPIcs), edited by Steven T. Flammia (Schloss Dagstuhl—Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2020), Vol. 158, p. 3:1.
- [39] Ulysse Chabaud, Frédéric Grosshans, Elham Kashefi, and Damian Markham, Efficient verification of boson sampling, *Quantum* **5**, 578 (2021).
- [40] Varun Upreti and Ulysse Chabaud, An efficient quantum state verification framework and its application to bosonic systems, *ArXiv:quant-ph/2411.04688*.
- [41] Heterodyne measurements are done by performing the POVM $((1/\pi^m)|\alpha\rangle\langle\alpha|)_{\alpha \in \mathbb{C}^m}$ defined through the coherent states.
- [42] A. I. Lvovsky and M. G. Raymer, Continuous-variable optical quantum-state tomography, *Rev. Mod. Phys.* **81**, 299 (2009).
- [43] Homodyne detection is done by measuring the observable $X_\theta = \cos(\theta)X + \sin(\theta)P$, where X and P are the standard position and momentum operators and θ is an angle, which can be freely chosen and can be seen as rotating phase space.
- [44] Photon-number-resolving measurements are done by performing the measurement corresponding to the positive operator valued measure (POVM) $((1/\pi^m)D(\alpha)|n\rangle\langle n|D(\alpha)^*)_{\alpha \in \mathbb{C}^m}$ for different photon numbers $n \in \mathbb{N}$.
- [45] Simon Becker, Nilanjana Datta, Ludovico Lami, and Cambyse Rouze, Classical shadow tomography for continuous variables quantum systems, *IEEE Trans. Inf. Theory* **70**, 3427 (2024).
- [46] Srilekha Gandhari, Victor V. Albert, Thomas Gerrits, Jacob M. Taylor, and Michael J. Gullans, Precision bounds on continuous-variable state tomography using classical shadows, *ArXiv:quant-ph/2211.05149*.
- [47] As an exemption Refs. [39,40] do provide a finite sample complexity but only under the restrictive assumption that the target state is a finite superposition of Fock states.
- [48] Atharv Joshi, Kyungjoo Noh, and Yvonne Y. Gao, Quantum information processing with bosonic qubits in circuit QED, *Quantum Sci. Technol.* **6**, 033001 (2021).
- [49] Nissim Ofek, Andrei Petrenko, Reinier Heeres, Philip Reinhold, Zaki Leghtas, Brian Vlastakis, Yehan Liu, Luigi Frunzio, S. M. Girvin, L. Jiang, Mazyar Mirrahimi, M. H. Devoret, and R. J. Schoelkopf, Extending the lifetime of a quantum bit with error correction in superconducting circuits, *Nature* **53**, 441 (2016).
- [50] V. V. Sivak, A. Eickbusch, B. Royer, S. Singh, I. Tsioutsios, S. Ganjam, A. Miano, B. L. Brock, A. Z. Ding, L. Frunzio, S. M. Girvin, R. J. Schoelkopf, and M. H. Devoret, Real-time quantum error correction beyond break-even, *Nature* **616**, 50 (2023).
- [51] Francesco Anna Mele, Antonio Anna Mele, Lennart Bittel, Jens Eisert, Vittorio Giovannetti, Ludovico Lami, Lorenzo Leone, and Salvatore F. E. Oliviero, Learning quantum states of continuous variable systems, *ArXiv:quant-ph/2405.01431*.
- [52] This can for example be seen by considering the displaced state $\rho_\alpha = D(\alpha)\rho D(\alpha)^*$ which satisfies $\|\mathcal{W}_{\rho_\alpha}\|_1 = \|\mathcal{W}_\rho\|_1$ for all $\alpha \in \mathbb{C}^m$ but $\text{Tr}(N\rho_\alpha) \xrightarrow{|\alpha| \rightarrow \infty} \infty$.
- [53] Note that the smoothed L^1 -norm is not a norm itself as it does not satisfy the triangle inequality or faithfulness due to the infimum used.
- [54] Density can be seen by defining for $c > 0$ the function $\tilde{f}_c = f \mathbf{1}_{\{|f| \geq c\}}$ and noting that by the dominated convergence theorem it approximates f in the L^2 -norm as $\lim_{c \rightarrow 0} \int_\Omega |\tilde{f}_c(\lambda) - f(\lambda)|^2 d\mu(\lambda) = \lim_{c \rightarrow 0} \int_{\{|f| < c\}} |f(\lambda)|^2 d\mu(\lambda) = 0$. Furthermore, note that for all $c > 0$ we have $\tilde{f}_c \in L^1(\Omega, \mathbb{R})$ as $\int_{\{|f| \geq c\}} |f(\lambda)| d\mu(\lambda) \leq \|f\|_2^2/c$.
- [55] Ola Bratteli and Derek W. Robinson, *Operator Algebras and Quantum Statistical Mechanics II* (Springer, Berlin, Heidelberg, 1997).
- [56] Gerald Teschl, *Mathematical Methods in Quantum Mechanics: With Applications to Schrödinger Operators*, 2nd ed. Graduate Studies in Mathematics (157, American Mathematical Society, 2014), Vol.

- [57] R. Werner, Quantum harmonic analysis on phase space, *J. Math. Phys.* **25**, 1404 (1984).
- [58] The integral expression (24) is well-defined for χ_ρ being a Schwartz function or more generally in $L^1(\mathbb{C}^m) \cap L^2(\mathbb{C}^m)$. For general functions in $L^2(\mathbb{C}^m)$ the Fourier transform is then extended by density in the usual way; see, e.g., Sec 7.1 in Ref. [56].
- [59] To be precise, this property holds rigorously for all states ρ such that $W_\rho \in L^1(\mathbb{C}^m)$ (see Proposition 18 in Ref. [64]), which is a proper subset of the set of all states.
- [60] R. L. Hudson, When is the Wigner quasi-probability density non-negative? *Rep. Math. Phys.* **6**, 249 (1974).
- [61] Francisco Soto and Pierre Claverie, When is the Wigner function of multidimensional systems nonnegative? *J. Math. Phys.* **24**, 97 (1983).
- [62] Radim Filip and Ladislav Mišta, Detecting quantum states with a positive Wigner function beyond mixtures of Gaussian states, *Phys. Rev. Lett.* **106**, 200401 (2011).
- [63] A. Mandilara, E. Karpov, and N. J. Cerf, Extending Hudson's theorem to mixed quantum states, *Phys. Rev. A* **79**, 062302 (2009).
- [64] M. A. De Gosson, *The Wigner Transform*, Advanced Textbooks In Mathematics (World Scientific Publishing Company, 2017).
- [65] Note that the condition on Y_λ uniquely determines its distribution to be $\mathbb{P}(Y_\lambda = r) = (1 + g(\lambda)/r)/2$ and $\mathbb{P}(Y_\lambda = -r) = (1 - g(\lambda)/r)/2$.
- [66] Wassily Hoeffding, Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.* **58**, 13 (1963).
- [67] Richard Askey and Stephen Wainger, Mean convergence of expansions in Laguerre and Hermite series, *Am. J. Math.* **87**, 695 (1965).
- [68] L. J. Landau, Bessel functions: Monotonicity and bounds, *J. London Math. Soc.* **61**, 197 (2000).
- [69] G. Szegő, *Orthogonal Polynomials* (Colloquium Publications, American Mathematical Society, Rhode island, 1975), <https://books.google.fr/books?id=ZOhmnsXlcY0C>.
- [70] U. Leonhardt and H. Paul, Measuring the quantum state of light, *Prog. Quantum Electron.* **19**, 89 (1995).
- [71] Stefan Scheel, Quantum optics. Available at https://www.lkv.uni-rostock.de/storages/uni-rostock/Alle_MNF/Physik_Qms/Lehre_Scheel/quantenoptik/Quantenoptik-Vorlesung4.pdf.
- [72] This can be seen by denoting the Bessel function of order $n \in \mathbb{N}$ by J_n and using $J'_0 = -J_1$ and the recurrence relation $J'_n = \frac{1}{2}(J_{n-1} - J_{n+1})$ [see Eq. (9.1.27) in Ref. [81]] combined with the fact that $|J_n(y)| \leq 1$ for all $y \in \mathbb{R}$ [see Eq. (9.1.60) in Ref. [81]].
- [73] T. M. Apostol, *Calculus* (John Wiley & Sons, New York, 1991), Vol. 1, <https://books.google.fr/books?id=o2D4DwAAQBAJ>.
- [74] Elizabeth Meckes and Mark Meckes, Spectral measures of powers of random matrices, *Electron. Commun. Probab.* **18**, 1 (2013).
- [75] This can be seen by $\inf_{\varepsilon' \in [0, \varepsilon]} \|f\|_1^{(\varepsilon')} / (\varepsilon - \varepsilon') \geq (\|f\|_1 / \varepsilon) - C\varepsilon^{-(1-\beta)} = \Omega(\|f\|_1 / \varepsilon)$ together with $\|f\|_1^{(\varepsilon)} \leq \|f\|_1$.
- [76] More precisely, to show that an exponentially decaying function f satisfies Eq. (A6) for all $a \in (0, 1)$, we note that $\int_{\{|f| \leq \delta < e^{-t|\alpha|}\}} |f(\alpha)| d\alpha \leq \int_0^{\ln((1/\delta)^{1/t})} \delta d\alpha = \delta/t \ln(1/\delta)$.
- [77] Robert Salzmann, Quantitative quantum Zeno and strong damping limits in strong topology, *ArXiv:quant-ph/2409.06469*.
- [78] Andreas Winter, Energy-constrained diamond norm with applications to the uniform continuity of continuous variable channel capacities, *ArXiv:quant-ph/1712.10267*.
- [79] Simon Becker and Nilanjana Datta, Convergence rates for quantum evolution and entropic continuity bounds in infinite dimensions, *Commun. Math. Phys.* **374**, 823 (2019).
- [80] Note that finiteness of this integral is not immediately translatable into finiteness of the average photon number $\text{Tr}(N\rho) = \int W_\rho(\alpha) |\alpha|^2 d\alpha$ because of the absolute value around the Wigner function $W_\rho(\alpha)$. Regardless, finiteness of $\int_{\mathbb{C}} |W_\rho(\alpha)| \sqrt{E + |\alpha|^2}$ corresponds to fast decay of $W_\rho(\alpha)$ for large values of $|\alpha|$, which we, therefore, interpret as some form of energy constraint.
- [81] Milton Abramowitz and Irene A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1964), 9th Dover Printing, 10th GPO Printing ed.