Lower Bounds for Quantum Parameter Estimation

Michael Walter and Joseph M. Renes
Institute for Theoretical Physics, ETH Zurich, Wolfgang–Pauli–Strasse 27, 8093 Zürich, Switzerland

The laws of quantum mechanics place fundamental limits on the accuracy of measurements and therefore on the estimation of unknown parameters of a quantum system. In this work, we prove lower bounds on the size of confidence regions reported by any region estimator for a given ensemble of probe states and probability of success. Our bounds are derived from a previously unnoticed connection between the size of confidence regions and the error probabilities of a corresponding binary hypothesis test. In group-covariant scenarios, we find that there is an ultimate bound for any estimation scheme which depends only on the representation-theoretic data of the probe system, and we evaluate its asymptotics in the limit of many systems, establishing a general “Heisenberg limit” for region estimation. We apply our results to several examples, in particular to phase estimation, where our bounds allow us to recover the well-known Heisenberg and shot-noise scaling.

I. INTRODUCTION AND SUMMARY OF RESULTS

The estimation of unknown parameters in a quantum system is a basic problem in quantum mechanics [1–3] and its many applications. For example, in the Cesium atomic clock, the unknown frequency shift of a quartz oscillator is to be estimated by measuring Cesium atoms that have been subjected to a time evolution that depends on the frequency shift, while in a Mach–Zehnder interferometer one tries to estimate the phase shift between two optical paths from the output beams of photons. The standard way to reduce the uncertainty inherent with any measurement, quantum or otherwise, is to repeat the experiment a large number of times. Equivalently, one may take many probes in an i.i.d. state $\rho \otimes \mathbb{I}_N$ and subject them all to the same dynamics. The mean square error of such a scheme scales as $1/N$, as one would expect from the central limit theorem (even when one is allowed to perform entangled measurements); this is known as the standard quantum limit or shot noise limit [4]. Interestingly, it is often possible to achieve better precision by using entangled probe states [5]. Here, it can be established by using the quantum Cramér–Rao inequality [1–6] that the ultimate lower bound on the mean-square error for any unbiased estimator, known as the Heisenberg limit, scales as $1/N^2$ [4]. There exist families of probe states that achieve this scaling even when no prior information about the parameter is available [7, 8].

Recently, it has become of interest in quantum estimation theory to consider region estimators, i.e. estimators that report a confidence region rather than a single point in the parameter space [9]. This is a way of providing rigorous error bars, and it avoids conceptual problems that are inherent with point estimators. See e.g. [10–13] for results in the context of quantum state tomography [14], which is perhaps the ultimate parameter estimation problem, as the entire quantum state is unknown. There is in fact a close connection between mean-square error and confidence regions: By Chebyshev’s inequality, any unbiased point estimator with mean-square error $\Delta^2$ can be considered as a region estimator with success probability $\rho_{\text{succ}}$ by reporting an interval of radius $\delta = \Delta / \sqrt{1 - \rho_{\text{succ}}}$ around its estimate.

In this work, we are interested in the fundamental limits of any region estimation scheme. More precisely, our goal is to prove lower bounds on the maximal volume $V_{\text{max}}$ and average volume $V_{\text{avg}}$ of the confidence regions reported by an arbitrary region estimator, depending only on the success probability $\rho_{\text{succ}}$ of the estimator and the ensemble of probe states $\{p_X, \rho_B\}$, where $X$ is the parameter space and $B$ the probe system (see [11] for precise definitions). We work in the Bayesian scenario, motivated by the information-theoretic methods that we employ to prove our bounds. But any Bayesian lower bound implies directly a lower bound for the minimax, i.e. worst-case performance: If a region estimator has success probability $\rho_{\text{succ}}$ for any fixed value of the unknown parameter, then it also succeeds with probability at least $\rho_{\text{succ}}$ for an arbitrary prior, and hence our bounds apply. Furthermore, it is often natural to consider a prior distribution over the parameter (see e.g. [15], where a Bayesian variant of the Cramér–Rao bound has been used to study the steady state of atomic clocks), and doing so also allows comparison of the local and global performance of estimation schemes (see e.g. [16, 17] and below).

Our starting point is the duality between region estimators and hypothesis tests, well-known from statistics [18, §3.5]. From a given region estimator, we construct the test $E_{XB}$ on the bipartite classical–quantum system $XB$ that accepts if the state of the classical system $X$ is contained in the region predicted by the estimator from the probe system $B$. If we take as null hypothesis the cq-state $\rho_{XB}$ that corresponds to the ensemble of probe states, then the probability that the test correctly identifies this null hypothesis is precisely equal to the success probability $\rho_{\text{succ}}$ of the original region estimator. Now consider as the alternative hypothesis any “uncoupled” quantum state of the form $\mathbb{I}_X/|X| \otimes \sigma_B$. Then we can show that the type II-error, i.e. the probability of rejecting the alternative hypothesis, is never larger than $V_{\text{max}}/|X|$, the maximal relative volume reported by the region estimator from which we constructed the test. By minimizing the type II-error over all such tests $E_{XB}$, we thus obtain the following hypothesis-testing lower bound, which holds for an arbitrary region estimator (Theorem 2 in [11]):

$$\frac{V_{\text{max}}}{|X|} \geq \sup_{\sigma_B} \frac{\beta_{\rho_{\text{succ}}}(\rho_{XB}, |X| \otimes \sigma_B)}{\mathbb{I}_X}.$$

Here and in the following, $\beta_{\rho_{\text{succ}}} (\rho, \sigma)$ denotes the minimal type-II error for any hypothesis test with null hypothesis $\rho$ and alternative hypothesis $\sigma$, if we require that the null hypothesis
is correctly identified with probability $\alpha$ or larger. The bound (1) is completely independent of the inner workings of the region estimators—it depends only on the ensemble of probe states and on the desired success probability. To our knowledge, this connection between the volume of the confidence regions and the type-II error in binary hypothesis testing has not been noticed before [19].

By choosing $\sigma_B$ as the average probe state $\int dx \rho_X^p \rho_B^p$ instead of optimizing over all $\sigma_B$ in (1), we obtain a lower bound for the average volume $V_{\text{avg}}$. Similarly, by optimizing over all $\sigma_B = \rho_B^p$, we obtain a lower bound for the average volume for any fixed value of the parameter, $\sup_x V_{\text{avg}}(x)$ [12]. We also describe a general procedure for deducing lower bounds on the average volume directly from (1) and the results presented below (see discussion after Theorem 4).

The right-hand side of (1) is closely connected to the conditional hypothesis-testing entropy, which is defined as $\log \sup_{\sigma_B} \beta_\alpha(p_{XB}, 1_X \otimes \sigma_B)$ and which shares many formal properties with the conditional von Neumann entropy [20–22]. In this way the lower bound (1) acquires an intuitive information-theoretical interpretation: The maximal log-volume reported by the estimator is at least as large as the conditional hypothesis-testing entropy of the parameter $X$ conditioned on the probe system $B$. It can also be understood as the quantum-mechanical variant of a converse bound for joint source-channel coding in [23], adapted to the case of trivial encoder. Indeed, the connection between information theory and statistics has a long and fruitful history. In the context of quantum parameter estimation, lower bounds to the mean-square error in terms of the conditional von Neumann entropy have been derived for the first time in [24] by using rate-distortion theory (cf. the references in [17]).

We then focus on group-covariant scenarios, where the family of probe states is obtained from an initial state $\rho_B^0$ by the action of a compact Lie group $G$, $\rho_B^g = g \rho_B^0 g^{-1}$ [13, 25–30, 31]. Mathematically, $X$ is a homogeneous space. Here, (1) can be “untwisted” (Proposition 3); in particular, for the uniform prior we obtain that

$$V_{\text{max}} \geq \sup_{\sigma_B} \beta_{\text{succ}}(p_{XB}^p, \sigma_B),$$

(2)

where the optimization is now over $G$-invariant states $\sigma_B$. The right-hand side of (2) is a one-shot analog of the $G$-asymmetry or relative entropy of frameness [31, 32] (cf. [33]). By minimizing over all probe states $\rho_B^g$, we establish a fundamental lower bound that holds for arbitrary region estimators and probe states, depending only on the representation-theoretic data of the probe-system Hilbert space $\mathcal{H}_B$ (Theorem 12). In the case where $X = G$, it takes the particularly simple form

$$V_{\text{max}} \geq \frac{\beta_{\text{succ}}(p_X, 1_X/|X|)}{\sum_\lambda d_{\lambda} \tau_\lambda},$$

(3)

Here, $\lambda$ labels the irreducible representations of $G$ that occur in $\mathcal{H}_B$, $d_\lambda$ denotes the corresponding dimension, and $r_\lambda := \min\{d_\lambda, m_\lambda\}$, with $m_\lambda$ the multiplicity. Observe that (3) is essentially independent of the multiplicities $m_\lambda$ (small multiplicities can improve the bound, but large multiplicities do not enter). This can be rather intuitively understood: Since the group $G$ acts the same way on each copy of an irreducible representation, the multiplicities should only matter insofar as they allow for entanglement between the representation and the multiplicity space, and this entanglement is bounded by $r_\lambda \leq d_\lambda$ [28]. The nominator in (3) measures the deviation of the prior from being uniform; it is equal to $p_{\text{succ}}$ if the prior is uniform, and otherwise smaller.

Phase estimation is an important instance of covariant estimation. Here, the probe states are obtained by the evolution of an initial state under a given periodic Hamiltonian, $\rho_B^0 = e^{iH/\hbar} \rho_B^0 e^{-iH/\hbar}$, with $H \in [0, 2\pi]$ the unknown phase, and the lower bound (3) reduces to

$$\frac{V_{\text{max}}}{2\pi} \geq \frac{\beta_{\text{succ}}(p_X, 1_X/2\pi)}{J},$$

(4)

where $J$ denotes the number of distinct eigenvalues of the Hamiltonian $H$. In particular, consider a single-body Hamiltonian of the form $H_N = \sum_n H^{(n)}$ acting on $N$ particles, with each $H^{(n)}$ acting in the same way. It is easy to see that the number of eigenvalues of any such periodic Hamiltonian scales at most linearly with $N$, so that $J = O(N)$ in (4). Heisenberg scaling for the mean-square error of point estimators is an easy consequence of (3). Conversely, since the latter is well-known to be achievable [7, 8], there exist confidence regions for phase estimation whose volume scales as $1/N$. Thus our lower bound is necessarily tight (up to constants). In contrast, separable probe states fulfill a tighter lower bound of order $1/\sqrt{N}$. This is because any pure separable probe state is concentrated only on $O(1/\sqrt{N})$ many eigenvectors of the Hamiltonian.

More generally, (4) implies that the global precision in phase estimation can only be improved by either acquiring additional prior information or increasing the number of eigenvalues of the Hamiltonian. This is in stark contrast to local arguments based on the Cramér–Rao bound, which show that the local sensitivity of an estimator is determined by the gap between the minimal and maximal eigenvalue; see [16, 17] for discussions in the context of the mean-square error. This phenomenon is also visible on the level of individual probe states. For example, the oft-mentioned GHZ or NOON state $(|0\rangle^N + |1\rangle^N)/\sqrt{2}$ is supported only on two eigenspaces of the Hamiltonian $H = \sum_n \sigma_{\lambda(n)}$, and hence is only useful if prior information on the phase is available; any ostensibly scaling of the precision with $N$ can be seen purely a consequence of the prior information available about the phase [17]. This can be directly verified by evaluating (2) for the GHZ state (Figure 2).

For a different instance of this distinction between local and global performance, we also consider a “non-linear” Hamilto-
nian, where \( N \) probes are coupled via two-body interactions to an auxiliary system, \( H_{N+1} = \sum_{n=1}^{N} \sigma_z^{(n)} (\sigma_z^{(N+1)}) \). This interaction can generate entanglement, and local arguments suggest that \( 1/N \) scaling should be possible even if the initial probe state \( \rho_B^0 \) is separable. By evaluating (2), we show that this conclusion does not hold globally; instead, we find a lower bound of order \( 1/\sqrt{N} \).

We also revisit some other paradigmatic scenarios, such as phase estimation with energy-bounded probe states and state estimation of a pure state, and establish corresponding lower bounds (5).

Notation and Conventions. We denote the set of states, i.e. density operators, on a Hilbert space \( \mathcal{H}_B \) by \( \mathcal{S}_B \), and the set of observables, i.e. Hermitian operators, by \( \mathcal{O}_B \). The expectation value of an observable \( \mathcal{O}_B \) in state \( \rho_B \) will be denoted by the pairing \( \langle \rho_B, \mathcal{O}_B \rangle := \text{tr} \rho_B \mathcal{O}_B \). A POVM with outcomes in a measurable space \( Y \) is a \( \sigma \)-additive function \( M_B \) that assigns to each measurable subset \( O \subseteq Y \) a positive semidefinite operator \( M_B(O) \) with \( M_B(\emptyset) = 0 \) and \( M_B(Y) = \mathbb{1}_B \). For any state \( \rho_B \), \( \langle \rho_B, M_B(-) \rangle \) is the probability distribution of measurement outcomes. We will later generalize these definitions to bipartite classical–quantum systems with continuous classical parameter \( X \).

II. REGION ESTIMATORS AND THE HYPOTHESIS-TESTING LOWER BOUND

Let \( X \) be the measurable space of parameters, equipped with a measure \( \mu_X \), which we shall later use to measure the volume of confidence regions, such that \((X, \mu_X)\) is a standard measure space. Let \( \rho_X = (\rho_X^x) \in L^1(X, \mu_X) \) be the probability density of the prior distribution. For each parameter value \( x \in X \) we are given a probe state \( \rho_B^x \) on some (possibly infinite-dimensional) Hilbert space \( \mathcal{H}_B \), and we shall assume that the family \( \{\rho_B^x\} \) is \( \mu_X \)-measurable in \( x \) (see Appendix A for a precise definition).

A region estimator for this scenario consists of a POVM \( M_B \) on \( \mathcal{H}_B \) with outcomes in some measurable space \( Y \), together with a measurable subset \( \mathcal{E} \subseteq X \times Y \), the look-up table. The interpretation is that the experimenter first performs the POVM measurement and obtains some result \( y \in Y \) according to the probability measure \( (\rho, M_B(-)) \); the estimated region is then given by the set \( \mathcal{E}_y := \{ x : (x, y) \in \mathcal{E} \} \) (Figure 1). In other words, if \( x \) is the true value of the underlying parameter then the estimator succeeds if the POVM measurement outcome is an element of the "compatible set" \( \mathcal{E}_y = \{ y : (x, y) \in \mathcal{E} \} \). Thus the (average) success probability of the estimator is

\[
p_{\text{succ}} = \int_X d\mu_X(x) \rho_X^x \langle \rho_B^x, \mathcal{E}_y \rangle ,
\]

and the maximal volume of any region reported by the estimator is

\[
V_{\text{max}} = \sup_y \mu_X(\mathcal{E}_y) \in [0, \infty].
\]

We show in Appendix A that our assumptions guarantee that these quantities are in fact well-defined.

Figure 1. Definition of a region estimator. A POVM measurement \( M_B \) is performed on the probe system. The region estimate is then obtained from the look-up table \( \mathcal{E} \) as the set of all \( x \) that are compatible with the measurement outcome \( y \).

The above way of encoding a random subset using a "look-up table" avoids subtle measurability issues that may arise if one tries to define random subsets as subset-valued random variables [39]. The common case where the reported regions are \( \delta \)-balls with respect to some metric \( d_X \) is recovered by setting \( Y = X \) and \( \mathcal{E} = \{ (x, y) : d_X(x, y) \leq \delta \} \).

Lemma 1. For any region estimator, we have the lower bound

\[
V_{\text{max}} \geq \sup_{\sigma_B} \int_X d\mu_X(x) \langle \sigma_B, M_B(\mathcal{E}_y) \rangle
\]

where the supremum runs over all quantum states \( \sigma_B \) on \( \mathcal{H}_B \).

Proof. Let us denote by \( \nu_Y = \langle \sigma_B, M_B(-) \rangle \) the probability measure induced by the POVM \( M_B \) and a quantum state \( \sigma_B \). Then,

\[
\int_X d\mu_X(x) \langle \sigma_B, M_B(\mathcal{E}_x) \rangle
\]

\[
= \int_X d\mu_X(x) \int_Y d\nu_Y(y) \mathbb{1}_{\mathcal{E}}(x, y)
\]

\[
= \int_Y d\nu_Y(y) \int_X d\mu_X(x) \mathbb{1}_{\mathcal{E}}(x, y)
\]

\[
= \int_Y d\nu_Y(y) \mu_X(\mathcal{E}_y) \leq V_{\text{max}},
\]

where we have used Tonelli’s theorem to swap the order of integration in the second step [40, Theorem 2.36].

Interestingly, both (5) and (7) depend only on the ensemble of probe states \( \{\rho_X^x, \rho_B^x\} \) and on the operators \( E_B := M_B(\mathcal{E}_x) \). We observe that \( E_{XB} = (E_B) \) is an operator-valued function of \( x \) which takes values between 0 and \( \mathbb{1}_B \), and so can be interpreted as a POVM element on the classical–quantum system \( X \hat{B} \) (see below for a precise definition). Therefore, by optimizing over all such POVM elements \( E_{XB} = (E_B) \), we obtain a relation between success probability and maximal volume that holds for all region estimators for a given ensemble: We have that

\[
V_{\text{max}} \geq \inf \left\{ \int_X d\mu_X(x) (\langle \sigma_B, E_B^x \rangle : 0 \leq E_B^x \leq \mathbb{1}_B, \int_X d\mu_X(x) \rho_X^x \langle \rho_B^x, E_B^x \rangle \geq p_{\text{succ}} \right\}.
\]
for all states $\sigma_B$. The integral of the non-negative function $\langle \sigma_B, E_{XB}^p \rangle$ can be infinite if $X$ has infinite measure (e.g., when estimating a shift parameter in $\mathbb{R}$). This makes sense, since in this case an estimator might report regions of infinite volume. However, even if $X$ has infinite measure the lower-bound is still mathematically correct and can be finite.

If, on the other hand, we assume that $X$ has finite measure then the right-hand side of (9) has a nice interpretation in terms of binary hypothesis testing. Making this precise is slightly complicated by the fact that, formally, the ensemble $\{p^X_x, \rho^B_B\}$ cannot be described as a density operator if $X$ is a continuous space – indeed, the usual identification between ensembles $\{p^X_x, \rho^B_B\}$ and density operators $\sum_x p^X_x |x\rangle \langle x| \otimes \rho^B_B$ is only possible for discrete $X$.

**Classical–Quantum Hypothesis Testing.** Instead, we should model a bipartite classical–quantum system with classical part $(X, \mu_X)$ and quantum part $\mathcal{H}_B$ by the von Neumann algebra $M_{XB} = L^\infty(X, \mu_X; B(\mathcal{H}_B))$. Its elements are operator-valued functions $E_{XB} = (E_{XB}^p): X \to B(\mathcal{H}_B)$ with finite norm $\|E_{XB}\|_\infty = \sup_x \|E_{XB}^p\|_\infty$. Thus a classical–quantum observable (cq-observable) is a (weak-=-measurable) bounded function on $X$ that takes values in the space of Hermitian operators on $\mathcal{H}_B$, i.e. an element of

$$\mathcal{O}_{XB} = \{E_{XB} = (E_{XB}^p): \|E_{XB}\|_\infty < \infty, E_{XB} \in \mathcal{O}_B\},$$

Similarly, a cq-state is a $(\mu_X, \text{-measurable})$ normalized function on $X$ that takes values in the space of positive semidefinite operators on $\mathcal{H}_B$, i.e. an element of

$$\mathcal{S}_{XB} = \{\rho_{XB} = (\rho^B_B): \|\rho_{XB}\|_1 = 1, \rho^B_B \geq 0\},$$

where $\|f_y\|_1 := \int d\mu_X(x) \|f^y_x\|_1$. Any positive multiple of such a state is called an unnormalized cq-state. It will be convenient to use the notation $p^X \otimes \rho_B := (p^X_x \rho^B_B)$ for $p^X = (p^X_x) \in L^1(X, \mu_X)$ and $\rho_B$ a density operator on $\mathcal{H}_B$.

The expectation value of an observable $E_{XB} = (E_{XB}^p)$ in the state $\rho_{XB} = (\rho^B_B)$ is given by the pairing

$$\langle \rho_{XB}, E_{XB} \rangle = \int d\mu_X(x) \langle \rho^B_B, E_{XB}^p \rangle.$$

Finally, a cq-POVM with outcomes in some measurable space $Y$ is a function $M_{XB}(O)$ that assigns to each measurable set $O \subseteq Y$ a positive semidefinite element $M_{XB}(O) = (M_{XB}^p(O))$ of $\mathcal{O}_{XB}$, with $M_{XB}(\emptyset) = 0$ and $M_{XB}(Y) = \mathbb{1}_{XB}$, and which is (weakly) $\sigma$-additive [25]. For each state $\rho_{XB}$, the probability that the measurement outcome is an element of $O \subseteq Y$ is given by the formula

$$\langle \rho_{XB}, M_{XB}(O) \rangle = \int d\mu_X(x) \langle \rho^B_B, M_{XB}^p(O) \rangle.$$

These definitions arise naturally in the algebraic approach to quantum mechanics [21] [22], and they reduce to the usual definitions if $X$ is finite or discrete.

A (binary) cq-hypothesis test with null hypothesis $\rho_{XB}$ and alternative hypothesis $\sigma_{XB}$ is a measurement with two possible outcomes, $\rho_{XB}$ or $\sigma_{XB}$. It is fully determined by the POVM element $E_{XB} = M_{XB}(\{\rho_{XB}\})$ corresponding to the null hypothesis. We set

$$\beta_\alpha(\rho_{XB}, \sigma_{XB}) = \inf\{\|\sigma_{XB}, E_{XB} \in \mathcal{O}_{XB}, 0 \leq E_{XB} \leq \mathbb{1}_{XB}(\rho_{XB}, E_{XB} \geq \alpha)\}.$$  \hspace{1cm} (9)

It is the minimal type-II error if we require that the type-I error be no larger than $1 - \alpha$.

Using this notation, we can now rewrite (8) in the following way (still assuming that $|X| := \mu_X(X) < \infty$ so that $\mathbb{1}_X / |X| \otimes \sigma_B$ is a normalized cq-state):

**Theorem 2** (Hypothesis-testing lower bound). For any region estimator, we have that

$$\frac{\sup_{\sigma_B} \beta_{p_{\text{succ}}}(\rho_{XB}, \mathbb{1}_X / |X| \otimes \sigma_B)}{|X|} \geq \beta_{p_{\text{avg}}}(\rho_{XB}, \mathbb{1}_X / |X| \otimes \rho_B),$$

where $\rho_{XB} = (\rho^X_x \rho^B_B)$ is the cq-state corresponding to the ensemble of probe states $\{\rho^X_x, \rho^B_B\}$.

**Theorem 2** allows us to lower-bound the maximal volume reported by an arbitrary region estimator by analyzing binary hypothesis tests between the ensemble of probe states and arbitrary “decoupled” probe states of the form $\mathbb{1}_X / |X| \otimes \sigma_B$.

**Average vs. Worst Case.** Instead of studying the average success probability (5) of a region estimator, we may also consider its worst-case success probability, which is defined as $p_{\text{wmax}} = \inf_x \langle \rho^B_B, M_B(\mathcal{E}_y) \rangle$. Since $p_{\text{wmax}} \geq p_{\text{wcast}}$ with respect to any prior, we can maximize (10) over all priors $p_X$ to obtain a lower bound purely in terms of the given family of probe states $\{\rho^B_B\}$ and the desired worst-case success probability. As this strategy is easily implemented for the lower bounds established in the following, we will focus on the Bayesian scenario throughout the remainder of this work.

On the other hand, we might also be interested in lower-bounding the average volume reported by the estimator rather than the worst-case volume (6). For a fixed value $x$ of the unknown parameter, it is given by

$$V_{\text{avg}}(x) = \int_Y d\mu_X(x') \langle \rho^B_B, M_B(\mathcal{E}_{x'}) \rangle.$$

Since $V_{\text{avg}}(x) = \int_X d\mu_X(x') \langle \rho^B_B, M_B(\mathcal{E}_{x'}) \rangle$ by the proof of (7), we obtain the following variant of (10) by proceeding as above:

$$\frac{V_{\text{avg}}(x)}{|X|} \geq \beta_{p_{\text{avg}}}(\rho_{XB}, \mathbb{1}_X / |X| \otimes \rho_B).$$  \hspace{1cm} (11)

Thus we may bound the “worst-case average volume” $\sup_x V_{\text{avg}}(x)$ simply by optimizing (11) over all $x$. On the other hand, if our goal is to bound the average volume with respect to the prior, which is given by

$$V_{\text{avg}} = \int d\mu_X(x) \rho^X_x V_{\text{avg}}(x) = \int_Y d\mu_Y(y) \mu_X(\mathcal{E}_y),$$

with $\rho_B = \int d\mu_X(x) \rho^X_x \rho^B_B$ the average probe state, then we may do so directly by using

$$\frac{V_{\text{avg}}}{|X|} \geq \beta_{p_{\text{avg}}}(\rho_{XB}, \mathbb{1}_X / |X| \otimes \rho_B),$$  \hspace{1cm} (12)
In general, $V_{\text{max}} \geq \max_x V_{\text{avg}}(x) \geq V_{\text{avg}}$ (but see [12, Theorem 1]).

We now describe a different approach to bound average volumes, based on the intuition that the volume of the reported region is typically of the order of $V_{\text{avg}}$. To make this precise, let $Y$ be the random variable describing the POVM measurement outcome of a given region estimator (for $x$ chosen according to the prior). Then the volume of the corresponding confidence region $V := \mu_X(\mathcal{E}_Y)$ is itself a random variable and satisfies

$$P(V \geq V_{\text{avg}}/\varepsilon) \leq \frac{\mathbb{E}[V]}{V_{\text{avg}}/\varepsilon} = \varepsilon$$

for any $\varepsilon > 0$ (by Chebyshev’s inequality). By truncating those regions $\mathcal{E}_y$ that are too large, we may therefore construct a new region estimator with maximal volume $V_{\text{avg}}/\varepsilon$ for which the success probability is still at least $p_{\text{success}} - \varepsilon$. So if we apply the hypothesis-testing lower bound for the maximal volume (10) to this new estimator, we obtain that

$$V_{\text{avg}} \geq \varepsilon \sup_{\beta} p_{\text{success}} - \varepsilon(\rho_{X|B}, I_X \otimes \sigma_B)$$

for any choice of $\varepsilon \in (0, p_{\text{success}})$, e.g. for $\varepsilon = p_{\text{success}}/2$. In this way, the results of this work, in particular our scaling results, imply directly corresponding lower bounds for the average volume, since they are all obtained by evaluating (10).

Relationship to Information Theory. As described in the introduction, the right-hand side of (10) is closely related to the conditional hypothesis-testing entropy as defined in the literature [20–22]. Moreover, (10) directly implies a lower bound in terms of the conditional von Neumann entropy, defined as $H(X|B)_{\rho_{X|B}} := -D(\rho_{X|B} || I_X \otimes \sigma_B)$ in terms of the relative entropy $D$ [43,44], that is closely related to Fano’s inequality.

Proposition 3. For any region estimator, we have the lower bound

$$h(p_{\text{success}}) + p_{\text{success}} \log V_{\text{avg}} + (1 - p_{\text{success}}) \log |X| \geq H(X|B)_{\rho_{X|B}},$$

where $\rho_{X|B}$ is the cq-state corresponding to the ensemble of probe states.

Proof. We start with the data-processing inequality for the relative entropy which, when applied to the optimal test for $\beta_{\alpha} = \beta_{\alpha}(\rho_{X|B}, I_X \otimes \sigma_B)$, shows that

$$-H(X|B) = D(\rho_{X|B} || I_X \otimes \rho_B) \geq \log(\alpha/\beta_{\alpha}) - \log|X|,$$

where $d(p||q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}$ denotes the binary relative entropy function (see [42] in Appendix B). Using the lower bound $d(p||q) \geq -h(p) + p \log 1/q$ [43,156], with $h(p) = -p \log p - (1-p) \log (1-p)$ the binary entropy function, we find that

$$-H(X|B) \geq -h(p_{\text{success}}) + p_{\text{success}} \log \frac{|X|}{V_{\text{avg}}} - \log|X|,$$

For any region estimator, the hypothesis-testing lower bound (12) now gives that

$$-H(X|B) \geq -h(p_{\text{success}}) + p_{\text{success}} \log \frac{|X|}{V_{\text{avg}}} - \log|X|,$$

which is equivalent to the advertised lower bound.

There are several ways in which the parameter estimation problem in the form considered here can be related to well-known tasks in quantum information theory. For one, any region estimator can be understood as a list decoder [46] for joint source-channel coding, for the source described by the prior $\rho_X$ and the cq-channel $x \mapsto \rho_{X|B}$ induced by the family of probe states. Indeed, (10) can be understood as a quantum version of the list-decoding lower bound [23, Theorem 4], adapted to the case of a trivial encoder. This observation was one of the starting points of this work.

Parameter estimation in general can also be related to source compression with side information. Here, $\rho_{Y|X}$ is the side information, and the goal in parameter estimation is to decompress given only the side information. In fact, one can also establish the following, perhaps unexpected link to lossless source compression. Suppose that the parameter space is finite (i.e., we are concerned with multiple hypothesis testing), and that we have a region estimator that reports $\delta$-balls with respect to some metric on $X$. Then we can turn $X$ into a graph by connecting any two points that have distance less than $\delta$ with respect to the metric. Such a graph can be colored with $V_{\text{max}}$ colors, and the knowledge of the color can be used to pick out the correct point from any $\delta$-ball reported by the estimator with probability $p_{\text{success}}$. This defines a lossless compression scheme for the source $p_X$ and side information $(p_{Y|X})$ which uses $m = \log_2 V_{\text{max}}$ bits and which works with error probability at most $1 - p_{\text{success}}$. It is amusing to note that our hypothesis-testing lower bound (10) applied to this construction agrees precisely with the converse bound for source compression established in [21, Theorem 9].

Mean-Square Error. The bound (10) can also be used to derive statements about the mean-square error of point estimators. Recall that a point estimator is a simply a POVM $M_B$ on the probe system $H_B$ with outcomes in the parameter space $X$, which we assume is a metric measure space with metric $d_X$ (i.e., the measure $\mu_X$ is defined on the corresponding Borel $\sigma$-algebra). Let $X$ and $\hat{X}$ denote random variables describing the prior and the estimate, respectively. That is, $X$ is distributed according to $p_X$, and $\hat{X}$ is the random variable with distribution $\rho_{\hat{X}B}$. Then conditioned on $X = x$, the mean square error of the estimator is defined as

$$\Delta^2 := \mathbb{E}[d_X(X, \hat{X})^2].$$

Now observe that we can also consider $M_B$ as a region estimator by interpreting its estimate as the center of a $\delta$-ball for some fixed radius $\delta > 0$. Mathematically, this corresponds to the choice $Y = X$ and $E = \{x, \hat{x} : d_X(x, \hat{x}) < \delta\}$. Clearly, the average volume reported by this estimator can be upper-bounded by the maximal volume of a $\delta$-ball in $X$,

$$V_{\text{avg}} \leq V_{\text{max}} \leq b_X(\delta) := \sup_X \mu_X(\{x : d_X(x, \hat{x}) < \delta\}).$$

On the other hand, Chebyshev’s inequality shows that

$$p_{\text{success}} = 1 - \mathbb{P}(d_X(X, \hat{X}) \geq \delta) \geq 1 - \frac{\Delta^2}{\delta^2}.$$
apply this to recover the well-known Heisenberg and shot-noise scaling bounds for phase estimation \([XY]\). Conversely, any achievability result for the mean-square error can in this way be lifted to region estimation. As a corollary of Proposition 3 we can also prove the following entropic lower bound for the mean-square error:

**Corollary 4.** For any point estimator with mean-square error \(\Delta^2\), we have the lower bound

\[
b_X(\sqrt{2}\Delta) \geq \frac{1}{4|X|} e^{2H(X|B)},
\]

where \(b_X(\delta)\) denotes the maximal volume of a \(\delta\)-ball as defined in [13].

**Proof.** Using Proposition 3 (14), and (13), we get that

\[
H(X|B) \leq \log 2 + p_{\text{succ}} \log V_{\text{avg}} + (1 - p_{\text{succ}}) \log |X|
\]

\[
\leq \log 2 + \left(1 - \frac{\Delta^2}{\delta^2}\right) \log V_{\text{avg}} + \frac{\Delta^2}{\delta^2} \log |X|
\]

\[
\leq \log 2 + \left(1 - \frac{\Delta^2}{\delta^2}\right) b_X(\delta) + \frac{\Delta^2}{\delta^2} \log |X|
\]

as long as we choose \(\delta \geq \Delta\), since a convex combination of the two quantities \(\log V_{\text{avg}}\) and \(\log |X|\) can only increase if weight is shifted away from the lesser (\(\log V_{\text{avg}}\)). In particular, for \(\delta = \sqrt{2}\Delta\) we find that

\[
H(X|B) \leq \log 2 + \frac{1}{2} \log b_X(\delta) + \frac{1}{2} \log |X|,
\]

which implies the asserted bound.

For example, for phase estimation, where \(X = U(1) = \{0, 2\pi\}/\sim\) and \(b_X(\delta) \leq 2\delta\), we find that

\[
\Delta \geq \frac{1}{16\sqrt{2\pi}} e^{2H(X|B)}.
\]

For comparison, the well-known entropic lower bound for an arbitrary, not necessarily periodic, “shift parameter” in \(X = \mathbb{R}\) that follows from rate-distortion theory [17] [47] reads

\[
\Delta \geq \frac{1}{2\pi e} e^{2H(X)} \geq \frac{1}{2\pi e} e^{2H(X|B)}.
\]

**Properties of \(\beta_\alpha\).** In the remainder of this section we collect some useful properties of \(\beta_\alpha\) that will later facilitate the computation of the hypothesis-testing lower bound.

Observe that the right-hand side of (9) is a linear cone program of locally convex topological vector spaces in duality (see, e.g., [48] §IV.6). The dual program is given by

\[
\beta_\alpha^*(\rho_X B, \sigma_{XB}) = \sup\{\alpha \mu - \langle \tau_{XB}, 1_{XB} \rangle : \mu \geq 0, \tau_{XB} \geq 0, \tau_{XB} \geq \mu\rho_X - \sigma_{XB}\}.
\]

(15)

It can be established that the primal value, i.e., the infimum in (9) is always attained. Moreover, there is zero duality gap: we have \(\beta_\alpha^* = \beta_\alpha\) for all \(\alpha \in [0, 1]\). We note that the quantity \(\beta_\alpha\) is monotonously increasing and continuous on \(\alpha \in [0, 1]\) (see Appendix B where we establish these claims for hypothesis testing in general von Neumann algebras).

In the case where \(X\) is a singleton space, (9) and (15) reduce to the usual primal and dual formulation for hypothesis testing between two quantum states \(\rho_B\) and \(\sigma_B\).

\[
\beta_\alpha(\rho_B, \sigma_B) = \inf\{\{\rho_B, E_B\} : 0 \leq E_B \leq 1_B, \langle \rho_B, E_B \rangle \geq \alpha\} = \sup\{\alpha \mu - \langle \tau, 1_B \rangle : \mu \geq 0, \tau \geq 0, \tau \geq \mu\rho_B - \sigma_B\}.
\]

(16)

It is easy to see that

\[
\beta_\alpha(p_X \otimes \rho_B, p_X \otimes \sigma_B) = \beta_\alpha(\rho_B, \sigma_B).
\]

(17)

for all probability densities \(p_X\) and quantum states \(\rho_B, \sigma_B\).

Moreover, we have the following data-processing inequality: For any CPTP map \(\Lambda\),

\[
\beta_\alpha(\Lambda(\rho), \Lambda(\sigma)) \geq \beta_\alpha(\rho, \sigma),
\]

(18)

which follows easily from the fact that the dual channel \(\Lambda^*\) is unital and positive, and therefore maps feasible tests onto feasible tests.

**Lemma 5.** Given any \(\rho, \rho'\) for which \(\frac{1}{2} \|\rho - \rho'\|_1 \leq \delta\) for some \(\delta \geq 0\), then for any \(\alpha\)

\[
\beta_{\alpha+\delta}(\rho, \sigma) \geq \beta_\alpha(\rho', \sigma).
\]

**Proof.** The claim follows by observing that any feasible test \(E\) for \(\beta_{\alpha+\delta}(\rho, \sigma)\) is also feasible for \(\beta_\alpha(\rho', \sigma)\). Indeed, by the properties of the trace distance,

\[
\delta \geq \frac{1}{2} \|\rho - \rho'\|_1 = \max_{0 \leq P \leq 1} \langle \rho - \rho', P\rangle \geq \langle \rho, E\rangle - \langle \rho', E\rangle \geq (\alpha + \delta) - \langle \rho', E\rangle,
\]

so \(E\) is indeed feasible for \(\beta_\alpha(\rho', \sigma)\).

**Lemma 6.** Let \(\rho\) be a quantum state, \((P_j)_{j=1}^J\) a projective measurement with \(J\) outcomes, and \(\sigma\) the corresponding post-measurement state. Then, \(\beta_\alpha(\rho, \sigma) \geq \alpha/J\).

**Proof.** Let \(Z\) denote the operator that acts by multiplication with \(\omega_j^*\) on the support of \(P_j\), with \(\omega\) a primitive \(J\)-th root of unity. Then we can write \(\sigma\) as the group average

\[
\sigma = \frac{1}{J} \sum_{j=1}^J Z_j \rho(Z)^j.
\]

Clearly, \(\frac{1}{J} \rho \leq \sigma\), so that \(\mu = 1/J, \tau = 0\) are feasible for the dual formulation of \(\beta_\alpha(\rho, \sigma)\). We conclude that \(\beta_\alpha(\rho, \sigma) \geq \alpha/J\) by (weak) duality.
III. COVARIANT ESTIMATION

In the following, let the parameter space $X$ be a smooth manifold equipped with the transitive action of a compact, connected Lie group $G$ and $G$-invariant metric $\mu_X$. Thus $\mu_X$ is induced by the unique Haar measure $\mu_G$ on $G$ with $|G| = |X|$. We moreover assume that the Hilbert space $H_B$ is a unitary $G$-representation. We will be interested in studying probe states $|\tilde{\beta}\rangle_B$ that form a covariant family, i.e. for which $B_{\tilde{\beta}} = gB_{\tilde{\beta}} g^{-1}$.

Let $x_0 \in X$ be an arbitrary base point. Then we have that $X = G \cdot x_0 \cong G/K$, where $K$ is the stabilizer of $x_0$. It follows that the probe state $|\tilde{\beta}\rangle_B$ has an additional symmetry, $k|\tilde{\beta}\rangle_B k^{-1} = |\tilde{\beta}\rangle_B \quad (\forall k \in K)$ \hfill (19)

The stabilizer subgroup $K \subseteq G$ is uniquely determined up to conjugation.

Proposition 7. For any region estimator for a covariant family $|\tilde{\beta}\rangle_B$ and prior $p_X$, we have the lower bound \hfill \hfill (20)

$$V_{\text{max}} \geq \sup_{\tilde{\sigma}_B \in \mathcal{O}} \beta_{\text{pacc}}(p_X \otimes |\tilde{\beta}\rangle_B \otimes \tilde{\sigma}_B) \quad (\forall \tilde{\sigma}_B \in \mathcal{O}) 
$$

where the supremum runs over all $G$-invariant states $\tilde{\sigma}_B$.

Proof. To see this, let $E_{XB} = \langle E_{XB} \rangle_B$ be a feasible test for $\beta_{\text{pacc}}(p_X \otimes |\tilde{\beta}\rangle_B \otimes \tilde{\sigma}_B)$. We define the “untwisted” test $\tilde{E}_{XB} = \langle \tilde{E}_{XB} \rangle_B$ by $\tilde{E}_{XB} := g^{-1} E_{XB} g$ where $g = g(x)$ is chosen such that $g x_0 = x$ [49]. Then, $\langle \tilde{E}_{XB} \rangle_B = \int d\mu_X(x) p_X(x) \langle \tilde{E}_{XB} \rangle_B = \int d\mu_X(x) \langle p_X, \tilde{E}_{XB} \rangle_B = \langle p_X \otimes |\tilde{\beta}\rangle_B, \tilde{E}_{XB} \rangle_B$. On the other hand, by using the $G$-invariance of $\tilde{\sigma}_B$, we find \hfill \hfill (21)

$$\langle b_{\text{pacc}}(p_X \otimes |\tilde{\beta}\rangle_B \otimes \tilde{\sigma}_B) \geq \beta_{\text{pacc}}(p_X \otimes |\tilde{\beta}\rangle_B \otimes \tilde{\sigma}_B), \quad \forall \tilde{\sigma}_B \in \mathcal{O} 
$$

and hence the claim follows from (10).

Corollary 8. For any region estimator for a covariant family $|\tilde{\beta}\rangle_B$ and invariant (hence uniform) prior, we have \hfill \hfill (22)

$$V_{\text{max}} \geq \sup_{\tilde{\sigma}_B \in \mathcal{O}} \beta_{\text{pacc}}(p_X \otimes |\tilde{\beta}\rangle_B \otimes \tilde{\sigma}_B) \quad (\forall \tilde{\sigma}_B \in \mathcal{O}) 
$$

where $\tilde{\sigma}_B, x_0$ is an arbitrary point in $X$, and $E_{\tilde{\beta}}$ denotes the $G$-average of the operator $E_B$.$\square$

Proof. Since the action is transitive, any invariant prior is equal to the uniform prior, so that $p_X = \mathbb{1}_X/|X|$. Thus the first lower bound follows from combining (17) and (20).

We now compute the supremum: By Fan’s minimax theorem \hfill \hfill (23)

$$\sup_{\tilde{\sigma}_B \in \mathcal{O}} \beta_{\text{pacc}}(p_X \otimes |\tilde{\beta}\rangle_B \otimes \tilde{\sigma}_B) 
$$

Example 9. In general, the optimal $\tilde{\sigma}_B$ in (21) is not given by the $G$-average of the probe state. Consider e.g. the action of $U(1) \subseteq SU(2)$ on $\mathbb{C}^2$ generated by the Pauli $\sigma_z$-operator, with $\tilde{\beta}_0^\mathcal{B}$ the pure state $\sqrt{0.9}|0\rangle + \sqrt{0.1}|1\rangle$. For $p_{\text{pacc}} = 1$, the optimal test $E_{\tilde{\beta}}$ is given precisely by $\tilde{\beta}_0^\mathcal{B}$, so that $E_{\tilde{\beta}} = \langle \tilde{\beta}_0^\mathcal{B} \rangle_B = 0.9|0\rangle \langle 0| + 0.1|1\rangle \langle 1|$. Clearly, the optimal $\tilde{\sigma}_B$ is not the $G$-average of $\tilde{\beta}_0^\mathcal{B}$ but rather the pure state $|0\rangle \langle 0|$. By continuity the same conclusion holds for $p_{\text{pacc}} \approx 1$.

We now give a state-independent lower bound for the volume of any confidence region for a given confidence level. Such a lower bound is an ultimate limit to the precision of any region estimator for an arbitrary covariant ensemble on $H_B$. From now on we will assume that $H_B$ is finite-dimensional.

Before we state the result, let us consider the isotypical decomposition of $H_B$ with respect to the group $G$, i.e. the decomposition \hfill \hfill (24)

$$H_B \cong \bigoplus_{\lambda} V_{G,\lambda} \otimes \mathbb{C}^{m_\lambda}, \quad (\forall \lambda) 
$$

where $m_\lambda$ denotes the multiplicity of an irreducible representation $V_{G,\lambda}$ of $G$. Fix a base point $x_0 \in X$ and denote by $K \subseteq G$ the stabilizer of $x_0$. By decomposing each irreducible $G$-representation $V_{G,\lambda}$ into irreducible $K$-representations $V_{K,\mu}$, we obtain (23) that \hfill \hfill (25)

$$H_B \cong \bigoplus_{\lambda} \left( \bigoplus_{\mu} V_{K,\mu} \otimes \mathbb{C}^{m_\mu} \right) \otimes \mathbb{C}^{m_\lambda}, \quad (\forall \lambda) 
$$

where $m_\mu$ denotes the multiplicity of $V_{K,\mu}$ in $V_{G,\lambda}$.

Corollary 10. For any region estimator for a covariant family and uniform prior, we have \hfill \hfill (26)

$$V_{\text{max}} \geq \inf_{p_{\text{pacc}}} \sup_{\tilde{\sigma}_B \in \mathcal{O}} \beta_{\text{pacc}}(p_X \otimes |\tilde{\beta}\rangle_B \otimes \tilde{\sigma}_B) = \frac{p_{\text{pacc}}}{\max \lambda \sum_\lambda \lambda d_{\lambda} m_\lambda}, \quad (\forall \lambda) 
$$

where $d_{\lambda}$ denotes the dimension of the irreducible representation $V_{G,\lambda}$. The result follows from combining (17) and (20).
where the infimum runs over all $K$-invariant states $\rho_B^{\sigma}$ and the supremum over all $G$-invariant states $\tilde{\sigma}_B$ on $\mathcal{H}_B$. Moreover, $d_\lambda$ and $d_\mu$ denote the dimension of $V_{G,\lambda}$ and $V_{K,\mu}$, respectively, and $r_\mu := \min\{m_\mu, m_\lambda\}$, where we use the same notation as in the decompositions (23) and (24).

In particular, if $X = G$ is a group then we have the lower bound

$$\frac{V_{\text{max}}}{|X|} \geq \frac{p_{\text{succ}}}{\sum_{\lambda} d_\lambda r_\lambda},$$

(26)

where $r_\lambda := \min\{d_\lambda, m_\lambda\}$.

**Proof.** Recall from [19] that the probe state $\rho_B^{\sigma}$ is necessarily $K$-invariant. Thus we get a state-independent lower bound by optimizing (22) over all $K$-invariant probe states $\rho_B^{\sigma}$.

$$\frac{V_{\text{max}}}{|X|} \geq \inf_{\rho_B^{\sigma}} \sup_{p_{\text{succ}}} \beta_{p_{\text{succ}}} (\rho_B^{\sigma}, \tilde{\sigma}_B)$$

$$= \inf_{\rho_B^{\sigma}} \sup_{\rho_B^{\sigma}} \left\{ \|E_B^G\|_\infty : 0 \leq E_B \leq 1, \|\rho_B^{\sigma} \otimes E_B\| \geq p_{\text{succ}} \right\}$$

$$= \inf_{\rho_B^{\sigma}} \sup_{\rho_B^{\sigma}} \left\{ \|E_B^G\|_\infty : 0 \leq E_B \leq 1, \max_{\rho_B^{\sigma}} \|\rho_B^{\sigma} \otimes E_B\| \geq p_{\text{succ}} \right\}$$

$$= \inf_{\rho_B^{\sigma}} \sup_{\rho_B^{\sigma}} \left\{ \|E_B^G\|_\infty : 0 \leq E_B \leq 1, \|E_B^G\|_\infty \geq p_{\text{succ}} \right\}$$

$$= \inf_{p_{\text{succ}}} \inf_{\rho_B^{\sigma}} \left\{ \|E_B^G\|_\infty : 0 \leq E_B = E_B^G, \|E_B^G\|_\infty = 1 \right\}.$$

By Schur’s lemma, each such $E_B$ can be written in the form $E_B = \bigoplus_{\lambda} \mathbb{1}_{V_{G,\lambda}} \otimes E_{\lambda}$ with respect to the isotypical decomposition (24). We may in fact assume that

$$E_B = \mathbb{1}_{V_{K,\mu}} \otimes |\psi_\mu\rangle \langle \psi_\mu| = d_\mu \frac{\rho_{\text{red}, \mu}}{d_\mu} \otimes |\psi_\mu\rangle \langle \psi_\mu|.$$

Indeed, restricting to a single summand and replacing $E_{\mu}$ by the rank-one projector onto a maximal eigenvector will never increase $\|E_B^G\|_\infty$. Now decompose $|\psi_\mu\rangle = \sum_{\lambda} \sqrt{\rho_\lambda} |\psi_\lambda\rangle$ according to the direct sum $\bigoplus_{\lambda} \mathbb{C}^{m_\lambda} \otimes \mathbb{C}^{m_\lambda}$, and denote by $\rho_\lambda$ the reduced density matrix of $|\psi_\lambda\rangle$ on $\mathbb{C}^{m_\lambda}$. It follows from another application of Schur’s lemma that

$$E_B^G = d_\mu \sum_{\lambda} p_{\lambda} \frac{\rho_{\text{red}, \lambda}}{d_\mu} \otimes \rho_\lambda.$$

By the Schmidt decomposition, the rank of $\rho_\lambda$ is at most $r_\mu$, while its trace is one. Therefore,

$$\|E_B^G\|_\infty = d_\mu \max_{\lambda} p_{\lambda} \frac{1}{d_\lambda} \|\rho_\lambda\|_\infty$$

$$\geq \max_{\lambda} p_{\lambda} \frac{1}{d_\lambda} \frac{1}{r_\mu} \geq \frac{1}{\sum_{\lambda} d_\lambda r_\lambda / d_\mu}.$$

By minimizing over $\mu$ we arrive at the advertised lower bound. \qed

By examining the final inequalities in the above proof, it is easy to extract the form of probe states $\rho_B^{\sigma}$ for which the state-dependent lower bound (21) attains the universal lower bound (25). E.g., in the case where $X = G$ we can choose $\rho_B = |\psi\rangle \langle \psi| \sum_{\lambda} z_\lambda |\psi_\lambda\rangle \langle \psi_\lambda|$, $|z_\lambda|^2 = \frac{d_\lambda r_\lambda}{\sum_{\lambda} d_\lambda r_\lambda}$ and $|\psi_\lambda\rangle = \frac{1}{\sqrt{\sum_{\lambda} |m_\lambda|^2}} \sum_{m_\lambda} |m_\lambda\rangle \otimes |m_\lambda\rangle$ (cf. [22]), where it was shown that such states also achieve “maximal $G$-asymmetry” as defined in that work, and (28), where it was shown that states of the general form $\sum_{\lambda} z_\lambda |\psi_\lambda\rangle \langle \psi_\lambda|$ are optimal for group element estimation with respect to a wide class of risk functions.

Although we have so far established the lower bounds (21) and (25) for uniform priors only, it is easy to generalize these to general priors by adapting a chain rule proved in [22]:

**Lemma 11.** Let $p_X \in L^1(X, \mu_X)$ be a probability density, $\rho_B$ and $\sigma_B$ be quantum states on $\mathcal{H}_B$, and $\alpha \geq 0, \alpha' > 0$ such that $\alpha + \sqrt{2\alpha'} \leq 1$. Then we have

$$\beta_{\alpha + \sqrt{2\alpha'}} (p_X \otimes \rho_B, \frac{f_X}{|X|} \otimes \sigma_B) \geq \beta_{\alpha} (p_X, \frac{f_X}{|X|}) \cdot \frac{1}{\alpha'} \beta_{\alpha'} (\rho_B, \sigma_B).$$

**Proof.** We closely follow the proof of [22 Prop. 5.1]. Let $\mu_X, \tau_X$ be feasible for the dual formulation of $\beta_\alpha (p_X, \frac{f_X}{|X|})$, and let $\mu_B, \tau_B$ be feasible for the dual formulation of $\beta_{\alpha'} (\rho_B, \sigma_B)$. Then,

$$\mu_B \mu_X \otimes p_X \otimes \rho_B \leq \mu_B \left( \frac{f_X}{|X|} + \tau_X \right) \otimes \rho_B$$

$$\leq \frac{f_X}{|X|} \otimes (\mu_B \rho_B) + \mu_B \tau_X \otimes \rho_B$$

$$\leq \frac{f_X}{|X|} \otimes (\sigma_B + \tau_B) + \mu_B \tau_X \otimes \rho_B$$

Define $T_B := \sigma_B^{1/2} (\sigma_B + \tau_B)^{-1/2}$, where the inverse is taken on $\text{supp} \sigma_B \subseteq \text{supp} \sigma_B + \tau_B$ [31]. Conjugating the above operator inequality with $T_B$, we find that

$$\mu_B \mu_X \otimes T_B \rho_B T_B^\dagger \leq \frac{f_X}{|X|} \otimes \sigma_B + \mu_B \tau_X \otimes T_B \rho_B T_B^\dagger.$$

Thus $\mu = \mu_B \mu_X, \tau_X B = \mu_B \tau_X \otimes T_B \rho_B T_B^\dagger$ are feasible for the dual formulation of the hypothesis test between $p_X \otimes T_B \rho_B T_B^\dagger$ and $\frac{f_X}{|X|} \otimes \sigma_B$, so that

$$\beta_{\alpha} (p_X \otimes T_B \rho_B T_B^\dagger, \frac{f_X}{|X|} \otimes \sigma_B)$$

$$\geq \mu_B \mu_X \alpha - \langle \mu_B \tau_X \otimes T_B \rho_B T_B^\dagger, \frac{f_X}{|X|} \rangle$$

$$= \mu_B \left( \mu_X \alpha - \langle \tau_X \otimes \rho_B T_B^\dagger T_B, \frac{f_X}{|X|} \rangle \right)$$

$$\geq \mu_B \left( \mu_X \alpha - \langle \tau_X, \frac{f_X}{|X|} \rangle \right)$$

$$\geq \frac{1}{\alpha'} (\mu_B \alpha' - \langle \tau_X, \frac{f_X}{|X|} \rangle) \left( \mu_X \alpha - \langle \tau_X, \frac{f_X}{|X|} \rangle \right)$$

for any $\alpha' \in [0, 1]$. Here we have used that $T_B$ is a contraction, i.e. that $T_B^\dagger T_B \leq 1$. By optimizing over all dual feasible points, we find that

$$\beta_{\alpha} (p_X \otimes T_B \rho_B T_B^\dagger, \frac{f_X}{|X|} \otimes \sigma_B) \geq \alpha \beta_{\alpha} (p_X, \frac{f_X}{|X|}) \cdot \frac{1}{\alpha'} \beta_{\alpha'} (\rho_B, \sigma_B).$$

(27)
On the other hand, we have that
\[ \frac{1}{2} \left\| p_X \otimes T_B \rho_B T_B^\dagger - p_X \otimes \rho_B \right\|_1 = \frac{1}{2} \left\| \rho_B - T_B \rho_B T_B^\dagger \right\|_1 \leq \sqrt{2\alpha}. \]
where the inequality is established just like in the proof of [22] Proposition 5.1. Thus the claim follows from [27] and Lemma 5.

**Theorem 12.** For any region estimator for a covariant family and prior \( p_X \), we have the lower bound
\[
\frac{V_{\text{max}}}{|X|} \geq \beta_{p_{\text{cov}}} \left( \frac{\|X\|}{|X|} \right) \frac{1}{\max_\mu} \sum_\lambda d_\lambda r_\mu^\lambda / d_\mu. \tag{28}
\]
In particular, if \( X = G \) is a group then we have the lower bound
\[
\frac{V_{\text{max}}}{|X|} \geq \beta_{p_{\text{cov}}} \left( \frac{\|X\|}{|X|} \right) \frac{1}{\sum_\lambda d_\lambda r_\mu^\lambda}. \tag{29}
\]
Here, \( d_\lambda, d_\mu, r_\mu^\lambda \) and \( r_\lambda \) are defined as in the statement of Corollary 10.

**Proof.** For all \( \alpha' > 0 \) small that
\[
\frac{V_{\text{max}}}{|X|} \geq \sup_{\sigma_B} \beta_{p_{\text{cov}}} \left( p_X \otimes p_B^\sigma \right) \left( \frac{\|X\|}{|X|} \otimes \sigma_B \right) \geq \beta_{p_{\text{cov}}} - \sqrt{2\alpha'} \left( p_X \otimes \sigma_B \right) \geq \beta_{p_{\text{cov}}} - \sqrt{2\alpha'} \left( p_X \otimes \sigma_B \right) \geq \beta_{p_{\text{cov}}} - \sqrt{2\alpha'} \left( p_X \otimes \sigma_B \right).
\]
where we have used [20], Lemma 11 and the identity in [25] (in this order). Now let \( \alpha' \to 0 \) and use continuity of \( \beta_\alpha \).

Observe that (28) and (29) reduce to (25) and (26) in the case of a uniform prior.

**IV. ASYMPTOTICS**

We will now analyze the scaling of the lower bound (28) when the probe system is a tensor power \( H_B^N = H_B^\otimes N \) of a fixed, finite-dimensional representation \( H_B \). Physically, this corresponds to the case where the probe states are generated by symmetric single-body operators, e.g. single-body Hamiltonians in the important case of \( U(1) \)-phase estimation. We shall only treat the case where \( X = G \) (but see [51]).

**Lemma 13.** Let \( T \subseteq G \) be a maximal torus (i.e. a maximal compact, connected, abelian subgroup). Then:

1. The number of isotypical components in \( H_B^\otimes N \) is \( O(N^{\dim T}) \).
2. Each irreducible representation that occurs in \( H_B^\otimes N \) has dimension \( O(N^{(\dim G - \dim T)/2}) \).

**Proof for \( G = T = U(1) \):** For the first claim, we need to show that the Hamiltonian generating the \( U(1) \)-action on \( H_B \) has at most linearly many eigenvalues. After a choice of basis we may assume that \( H_B = \mathbb{C}^d \), and that the \( U(1) = \{ 0, 2\pi \}/\sim \)-action is generated by a diagonal Hamiltonian \( H = \text{diag} h \) with integer entries \( h \in \mathbb{Z}^d \). Thus the action on \( H_B^\otimes N \) is generated by the one-body Hamiltonian \( H_N = H \otimes I + \ldots + I \otimes H \). Clearly, \( H_N \) is diagonal in the computational product basis, and the eigenvalue of a basis vector \( |\vec{x} \rangle \) is equal to the inner product \( \langle \vec{x}, h \rangle \), where \( \vec{x} \in \mathbb{Z}^d \) is the type of \( \vec{x} \), specifying the number of occurrences of the symbols \( 1, \ldots, d \) in a string \( \vec{x} \).

But \( \langle \vec{x}, h \rangle \) is an integer such that
\[
|\langle \vec{x}, h \rangle| \leq \|| \vec{x} || | h ||_\infty = N \| h \|_\infty.
\]

It follows that there are at most \( 2N \| h \|_\infty + 1 \) eigenvalues.

For the second claim, recall that the irreducible representations of abelian groups are one-dimensional.

**Proof for general \( G \):** We will use some basic notions of the theory of compact Lie groups [52]–[54]. Without loss of generality we may assume that \( G \) is semisimple, since we can always treat the connected part of the center via the above proof for \( U(1) \). Let us denote by \( t^\dagger \) the dual of the Lie algebra \( t \), equipped with the inner product \( \langle \cdot, \cdot \rangle \) induced by the Killing form. We can choose a finite set of positive roots \( R_+ = \{ \alpha \} \subseteq t^\dagger \). They span a proper cone; the dual cone with respect to the Killing form is called the positive Weyl chamber and denoted by \( t_+^\dagger \). The Weyl vector \( \rho = \frac{1}{2} \sum_\alpha \alpha \) is an element in the interior of both cones. We can use it to define a partial order \( t^\dagger_+ : \xi \geq \xi' \) if and only if \( \langle \xi, \rho \rangle \geq \langle \xi', \rho \rangle \). In particular, \( \xi \geq 0 \) for all \( \xi \in t_+^\dagger \). There is a lattice \( \Lambda^+ \subseteq t^\dagger \), called the weight lattice, which corresponds to the generators of one-parameter subgroups \( U(1) \subseteq T \). The intersection \( \Lambda_+^\dagger = \Lambda^+ \cap t_+^\dagger \) is called the set of dominant weights. The fundamental theorem of the representation theory of compact, connected Lie groups states that the irreducible representations of \( G \) are labeled by an element \( \lambda \in \Lambda_+^\dagger \), called the highest weight. In the familiar case where \( G = SU(2) \), \( \Lambda_+^\dagger \) can be identified with the set of non-negative half-integers, and \( \lambda \) is the spin \( j \) of the representation.

Let us now consider the tensor product of two irreducible representations, and decompose it into irreducible representations,
\[
V_{G,\lambda} \otimes V_{G,\mu} = \bigoplus_\nu V_{G,\nu} \otimes C^{\lambda,\mu}.\tag{28}
\]
Then it is well-known that \( \lambda + \mu \) is the highest weight in this decomposition. That is, \( \lambda + \mu \geq \nu \) for all \( \nu \) with respect to the order defined above. This generalizes the fact that the sum of the two spins is the largest term in the Clebsch–Gordan series for \( SU(2) \).

Now let \( H_B = \bigoplus_{j=1}^J V_{G,\lambda_j} \otimes C^{m_j} \) be the isotypical decomposition of \( H_B \). Then,
\[
H_B^\otimes N = \bigoplus_{i_1, \ldots, i_N} V_{G,\lambda_{j_1}} \otimes \cdots \otimes V_{G,\lambda_{j_N}} \otimes C^{m_{j_1} \cdots m_{j_N}}.
\]
which scales at most as \(N^{\dim T}\), since \(\dim T = \dim \mathfrak{t}^*\) is the dimension of the weight lattice. This establishes the first claim.

For the second claim, recall that the Weyl dimension formula asserts that the dimension of an irreducible representation \(V_{G,\lambda}\) is given by the polynomial

\[
p(\lambda) = \prod_{\alpha \in R^+} (\alpha, \lambda + \rho)/(\alpha, \rho).
\]  

(30)

The degree of \(p(\lambda)\) is equal to \(|R_+|\), the number of positive roots, so that

\[
p(\nu) \leq N^{\dim T} \sup_{\nu \in \Lambda_N^+} \{p(\nu) : \nu \in \Lambda_N^+, \nu \leq \lambda\}
\]

for all \(\nu \in \Lambda_N^+\). The second claim follows from this, since \(|R_+| = (\dim G - \dim T)/2\). \(\square\)

**Theorem 14 (Heisenberg limit).** For any region estimator for a covariant family on \(\mathcal{H}_{B}^{\otimes N}\) and prior \(p_G\), we have that

\[
\frac{V_{\max}}{|G|} \geq C \beta_{\max}(p_G; \frac{|G|}{|G|}) N^{-\dim G}
\]

(31)

for \(N\) large, where the constant \(C > 0\) only depend on the representation \(\mathcal{H}_{B}\).

**Proof.** By Lemma 13 we can estimate the right-hand side quantity in (29) by

\[
\sum_\lambda d_{\Lambda_N^+} \leq \sum_\lambda d_{\Lambda_N^+} \leq O(N^{\dim T} N^{\dim G - \dim T}) = O(N^{\dim G}).
\]

Thus the assertion is a consequence of the bound (29). \(\square\)

**Theorem 14** is the analog of Heisenberg scaling for region estimators. It provides an ultimate lower bound for any region estimator and family of probe states that is covariant for the tensor product action on \(\mathcal{H}_{B}^{\otimes N}\). In fact, by interpreting point estimators as region estimators and using (13) and (14), we obtain Heisenberg scaling for the mean-square error as a direct consequence of our bound, generalizing results in the literature for \(U(1)\) [4] and \(SU(d)\) [28,34] (see [8] for worked examples).

**Separable States.** We will now show that for a covariant family of separable probe states and abelian \(G\), any region estimator satisfies a stronger lower bound than the one just established:

**Proposition 15.** Let \(G = T\) be abelian. Let \(\rho_{B_N}\) be a (fully) separable state on \(\mathcal{H}_{B}^{\otimes N}\). Then there exists a constant \(D > 0\), only depending on the representation \(\mathcal{H}_{B}\), such that

\[
\sup_{\tilde{\sigma}_{B_N}} (\rho_{B_N}, \tilde{\sigma}_{B_N}) \geq D \alpha^{\dim T + 1} N^{-\dim T/2}
\]

for all \(\alpha > 0\), where the supremum runs over all \(T\)-invariant states \(\tilde{\sigma}_{B_N}\).

**Proof.** We first consider the case where \(\rho_{B_N}\) is a pure product state \(|\psi_{B_N}\rangle = |\psi_{B_1}\rangle \otimes \cdots \otimes |\psi_{B_N}\rangle\). Recall that any compact abelian Lie group is a torus \(T = U(1)^k\), where \(k = \dim T\). Let us choose generators \(H_1, \ldots, H_k\) of the action of each \(U(1)\)-factor of \(\mathcal{H}_{B}\). The generators commute and can therefore be jointly diagonalized, i.e., there exists a decomposition \(\mathcal{H}_{B} = \bigoplus_{\omega} \mathcal{H}_{\omega}\) into joint eigenspaces, with \(\omega \in \mathbb{Z}^k\) encoding the (integral) eigenvalues. The vector \(\omega\) is commonly called a weight. Let us decompose each tensor factor of \(|\psi_{B_N}\rangle\) accordingly, \(|\psi_{B_N}\rangle = \sum_\omega \sqrt{p_{N,\omega}} |n, \omega\rangle\), where each \(|n, \omega\rangle\) \(\in \mathcal{H}_{\omega}\). Thus,

\[
|\psi_{B_N}\rangle = \sum_\omega \sqrt{p_{N,\omega}} |\omega_1, \ldots, \omega_N\rangle
\]

where \(p_{\omega_1, \ldots, \omega_N} := |\omega_1, \ldots, \omega_N\rangle \neq 0\). Consider the probability distribution of \(N\) independent random variables \(X_1, \ldots, X_N\), each with values in the finite set \(|\omega\rangle\) of possible weights.

Note that each \(|\omega_1, \ldots, \omega_N\rangle\) is a joint eigenvector of the \(T\)-action on \(\mathcal{H}_{B}^{\otimes N}\), which is generated by the single-body Hamiltonians \(H_{N,j} = H_j \otimes \mathbb{1} + \ldots + \mathbb{1} \otimes H_j\). Clearly, the weight of \(|\omega_1, \ldots, \omega_N\rangle\) is given by the sum of the individual weights, \(\sum_n \omega_n\). Let us thus consider the set of all eigenvectors whose weight is less than \(\varepsilon\) away from the mean \(\bar{m} := \mathbb{E}[X_1 + \cdots + X_N] = \sum_n \omega_n\). Let \(\varepsilon\) be chosen appropriately. Then we can lower-bound the overlap between \(|\psi_{B_N}\rangle\) and its normalized truncation \(|\psi_{B_N}'\rangle\) by

\[
|\langle \psi_{B_N}', \psi_{B_N}' \rangle| \geq \sum_{|n, \omega\rangle \in \Omega_\varepsilon} p_{\omega_1, \ldots, \omega_N} = 1 - \sum_{|n, \omega\rangle \notin \Omega_\varepsilon} p_{\omega_1, \ldots, \omega_N}
\]

(32)

\[
= 1 - \mathbb{P}(\|\sum_{n=1}^N X_n - \mathbb{E}[X_1 + \cdots + X_N]\|_2 \geq \varepsilon)
\]

(33)

\[
\geq 1 - \mathbb{E}[\|\sum_{n=1}^N X_n - \mathbb{E}[X_1 + \cdots + X_N]\|_2^2]\varepsilon^2
\]

(34)

\[
\geq 1 - \sum_{n=1}^N \mathbb{E}[\|X_n\|_2^2]\varepsilon^2
\]

(35)
where we have used Chebyshev’s inequality and independence of the \( \{X_n\} \). Now, each random variable \( X_n \) takes values in the finite set \( \{\omega\} \), so that \( \mathbb{E}[|X_n|^2] \leq \max_{\omega} |\omega|^2 =: M^2 \). It follows that

\[
|\langle \psi_{BN}, \psi_{\bar{B}N} \rangle| \leq 1 - \frac{NM^2}{\varepsilon^2}.
\]

By the bounds relating the fidelity and trace distance applied to \( \rho_{BN} = |\psi\rangle \langle \psi|_{BN} \) and \( \rho_{\bar{B}N} = |\psi\rangle \langle \psi|_{\bar{B}N} \),

\[
\frac{1}{2} \|\rho_{BN} - \rho_{\bar{B}N}\|_1 \leq \sqrt{1 - |\langle \psi_{BN}, \psi_{\bar{B}N} \rangle|^2} \leq \sqrt{2\frac{NM^2}{\varepsilon^2}} =: f(\varepsilon),
\]

so that Lemma 5 gives

\[
\beta_a(\rho_{BN}, \sigma_{BN}) \geq \beta_a - f(\varepsilon)(\rho_{\bar{B}N}, \sigma_{BN}),
\]

for all \( \sigma_{BN} \).

On the other hand, note that we can certainly write \( |\psi_{BN}\rangle \) as a sum of \((2e+1)^k\) joint eigenvectors of the \( T \)-action on \( \mathcal{H}_{BN} \), as the remaining weights \( \sum_n \tilde{\omega}_n \) are integral vectors in \( \mathbb{Z}^k \) that are all within \( \pm \varepsilon \) of the mean \( \tilde{m} \). Since the effect of averaging under the group action amounts to a projective measurement in this eigenbasis (dephasing!), Lemma 6 implies that

\[
\beta_a(\rho_{BN}, \tilde{\rho}(BN)) \geq \frac{\alpha - f(\varepsilon)}{2(2e+1)^k},
\]

where \( \tilde{\rho} \) denotes the \( T \)-average of \( \rho' \).

Finally, choose \( \varepsilon = \sqrt{8NM}/\alpha \) so that \( f(\varepsilon) = \alpha/2 \), and conclude that

\[
\beta_a(\rho_{BN}, \tilde{\rho}(BN)) \geq \frac{\alpha}{2(2e+1)^k} \geq \frac{\alpha^{k+1}}{2(2\sqrt{8NM})^k} \geq \frac{D\alpha^{k+1}}{\sqrt{N^{k/2}}},
\]

for some constant \( D \geq 0 \) independent of \( \alpha \).

If \( \rho_{BN} \) is an arbitrary separable state then we can still write it as a convex combination \( \rho_{BN} = \sum_ip_ip_{BN}^{(i)} \) of pure product states. Thus it follows from (22) and the above that

\[
\sup_{\bar{\rho}_{BN}} \beta_a(\rho_{BN}, \bar{\sigma}_{BN}) \geq \min_i \sup_{\sigma_{BN}} \beta_a(\rho_{BN}^{(i)}, \sigma_{BN}) \geq \frac{D\alpha^{k+1}}{\sqrt{N^{k/2}}},
\]

since any \( E_B \) that is feasible for \( \rho_{BN} \) in (22) is also feasible for at least one of the \( \rho_{BN}^{(i)} \).

**Corollary 16** *(Shot-noise limit).* Let \( G = T \) be abelian. For any region estimator for a covariant family of separable probe states on \( \mathcal{H}_{BN} \) and uniform prior, we have a lower bound

\[
\frac{V_{\text{max}}}{|G|} \geq Dp_{\text{acc}}^{\dim T+1}N^{-\dim T/2}, \tag{32}
\]

where the constant \( D \geq 0 \) depends only on the representation \( \mathcal{H}_{BN} \). More generally, for an arbitrary prior \( p_G \) and all \( \alpha < p_{\text{acc}}/2 \), we have that

\[
\frac{V_{\text{max}}}{|G|} \geq D\alpha^{\dim T}p_{\text{acc}}^{\sqrt{\alpha}}(p_G, |G|)N^{-\dim T/2}. \tag{33}
\]

Just as (31) could be interpreted as a Heisenberg limit, the lower bound (33) can be understood as a shot-noise bound in each component of the unknown parameter \( g \in G \). It would be interesting to generalize (33) to non-abelian groups \( G \).

**Stein’s lemma.** In view of the generalized Stein’s lemma proved in (35), one might expect that any lower bound on \( \sup_{\bar{\rho}_{BN}}(\rho_{BN}, \bar{\sigma}_{BN}) \) should decay exponentially in \( N \) rather than polynomially with \( N \). However, while this Stein’s lemma is indeed applicable to the situation at hand (with alternative hypothesis the set of \( G \)-invariant states), it can be shown that the corresponding error exponent, which has been called the \( G \)-asymmetry or regularized relative entropy of frameness (31), (32), is always equal to zero – thus there is no contradiction. This has been proved in (32) (Corollary 11), but also follows from Theorem 14 (which furthermore gives more precise information than the subexponential decay predicted by Stein’s lemma).

**V. EXAMPLES**

**A. Phase Estimation**

In this section we apply our bounds to various scenarios for the estimation of a phase parameter in \( U(1) = [0, 2\pi]/\sim \). We shall always equip \( U(1) \) with the usual metric and Lebesgue measure of total volume \( 2\pi \).

**Single-Body Hamiltonian.** We first consider a probe system \( \mathcal{H}_{BN} = \mathcal{H}_{BN}^0 \) and a family of probe states that is obtained by the evolution of an initial state \( \rho_B^0 \) under the action of a single-body Hamiltonian \( H_B = \sum_n H_n^{(n)} \), where \( H_n^{(n)} = \mathbb{1}_{\otimes N-n} \otimes H \otimes \mathbb{1}_{\otimes N-n} \) and \( H \) is a periodic Hamiltonian on \( \mathcal{H}_B \). Thus,

\[
\rho_B^\theta = e^{i\theta H_B} \rho_B^0 e^{-i\theta H_B} = (e^{i\theta H} \otimes N) \rho_B^0 (e^{-i\theta H} \otimes N).
\]

for \( \theta \in [0, 2\pi] \). Mathematically, \( H \) generates a representation of \( U(1) \) on the single-body Hilbert space \( \mathcal{H}_B \), and the representation generated by \( H_B \) on \( \mathcal{H}_{BN} = \mathcal{H}^0_{BN} \) is precisely its \( N \)-fold tensor product. Thus, Theorem 14 applied to \( X = G = U(1) \) gives the following lower bound that holds for any region estimator and prior \( p_U(1) \):

\[
\frac{V_{\text{max}}}{2\pi} \geq C\frac{\beta_{\text{acc}}(p_U(1), |U(1)|/2\pi)}{N}. \tag{34}
\]

The constant \( C \) depends only on the generator \( H \); see the first proof of Lemma 13 for an explicit bound. In particular, for the uniform prior \( p_U(1) = |U(1)|/2\pi \) we find that

\[
\frac{V_{\text{max}}}{2\pi} \geq C\frac{p_{\text{acc}}}{N}. \tag{34}
\]

If we restrict to separable probe states then by (32) this can be strengthened to give

\[
\frac{V_{\text{max}}}{2\pi} \geq D\frac{p_{\text{acc}}^2}{\sqrt{N}}, \tag{35}
\]

where again the constant \( D \) depends only on the generator \( H \).

We can also evaluate the state-dependent lower bound (21) numerically for individual probe states \( \rho_B^0 \), since it is given by
lower bound to $V_{\text{max}}/2$

Thus we have recovered Heisenberg scaling. In contrast, the tensor powers of the plus state $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ (dash-dotted line) cannot overcome the shot-noise limit \cite{Mooij}. For the GHZ-state $\frac{1}{\sqrt{2}}(|0\rangle^N + |1\rangle^N)$ (stars), though entangled, we get a constant lower bound. Indeed, it is well-known that the GHZ-state achieves Heisenberg scaling only if additional prior information about the phase is available \cite{Giovannetti} §III.C].

a semi-definite program (note that the supremum over $\delta_B$ can be incorporated into the dual program). See Figure 2 for some illustrative numerical results.

We now illustrate how to recover the well-known Heisenberg and shot-noise scaling for the mean-square error $\Delta^2$ of a point estimator. As explained in \cite{Giovannetti} we can always consider a point estimator as a region estimator by interpreting its estimate as the center of a $\delta$-ball, for any choice of $\delta > 0$. If we do so, then \cite{Giovannetti, Giovannetti2} and \cite{Mooij} combine to

$$2\delta \geq \Delta^2/\Delta^2/\beta^2 \cdot C \cdot 1 - \Delta^2/\beta^2 \frac{N}{\Delta^2}.$$  

By optimizing over $\delta > 0$, we arrive at the following lower bound, which holds for an arbitrary point estimator and uniform prior:

$$\Delta^2 \geq 4C^2\pi^2 \frac{1}{27} \frac{N^2}{\Delta^2}.$$ 

Thus we have recovered Heisenberg scaling for the mean-square error, which is normally established by evaluating the Cramér-Rao bound and averaging over the prior information. Conversely, since we know that Heisenberg scaling is asymptotically achievable for the mean-square error \cite{Giovannetti, Giovannetti2}, it follows from \cite{Mooij} that there exist probe states and region estimators for which $V_{\text{max}}$ scales as $1/N$ for any fixed value of $\rho_{\text{acc}}$.

Similarly, \cite{Giovannetti, Giovannetti2} and \cite{Mooij} combine to

$$2\delta \geq D \frac{(1 - \Delta^2/\Delta^2)^2}{\sqrt{N}} \geq D \frac{1 - 2\Delta^2/\Delta^2}{\sqrt{N}}.$$  

which gives the following lower bound that holds for any point estimator, uniform prior and separable probe states,

$$\Delta^2 \geq \frac{2D^2\pi^2}{27} \frac{1}{N}. $$

It exhibits the expected shot-noise scaling behavior. The same procedure can be used to see that the large-$N$ scaling of the curves in Figure 2 is consistent with the literature.

The above discussions can be readily generalized to the estimation of multiple phases (where $T = U(1)^k$).

**Energy-Bounded Probe States.** We now consider phase estimation where the probe system is the single-mode bosonic Fock space $\mathcal{H}_B = L^2(\mathbb{R}) \cong \text{Sym}(\mathbb{C})$. Let $n = a^\dagger a$ denote the bosonic number operator and $(|n\rangle)_{n=0}^\infty$ the occupation number basis. We will establish a fundamental lower for all covariant families of probe states generated by the Hamiltonian $H = a^\dagger a$ with bounded mean energy $\langle \rho_B^0, H \rangle = \langle \rho_B^0, H \rangle \leq E$ \cite{Giovannetti}.

For this, let $P_B^{(N)}$ denote the projector onto the finite-dimensional subspace $\mathcal{H}_B^{(N)}$ spanned by $(|n\rangle)_{n<N}$. Then,

$$\langle \rho_B^0, P_B^{(N)} \rangle = 1 - \sum_{n=N}^\infty \langle n | \rho_B^0 | n \rangle \geq 1 - \frac{E}{N},$$

and hence the gentle measurement lemma implies that the state $\rho_B' := P_N \rho_B^0 P_N / \langle \rho_B^0, P_N \rangle$ has trace-norm distance at most

$$\frac{1}{2} || \rho_B - \rho_B'|| \leq \sqrt{E/N}. $$

Observe that $\rho_B'$ is supported on the subspace $\mathcal{H}_B^{(N)}$, which is $N$-dimensional. It follows from Lemma 3 and Lemma 4 that

$$\beta_\alpha(\rho_B', \tilde{\sigma}_B) \geq \beta_\alpha - \sqrt{E/N} (\rho_B', \tilde{\sigma}_B) \geq \frac{\alpha - \sqrt{E/N}}{N},$$

where $\tilde{\sigma}_B$ denotes the $U(1)$-average of $\rho_B'$. Set $N = \lfloor 4E/\alpha^2 \rfloor$, so that $\sqrt{E/N} \leq \alpha/2$ and

$$\alpha - \sqrt{E/N} \geq \frac{\alpha^3}{8E + 2\alpha^2}.$$ 

In view of \cite{Giovannetti} we thus arrive at the following lower bound, which holds for all region estimators, uniform prior, and probe states whose mean energy is bounded by $E$:

$$V_{\text{max}} \geq \frac{2\delta}{2\pi} \geq \frac{4C^2\pi^2}{27} \frac{1}{\Delta^2} \frac{N}{\Delta^2} = \frac{p_{\text{acc}}^2}{8E + 2p_{\text{acc}}^2}.$$ 

which scales as $1/E$ for large $E$. The “Heisenberg scaling” for the mean-square error that was established in \cite{Giovannetti} (11) is an easy consequence of this bound (cf. \cite{Giovannetti}).
Nonlinear Interactions. The Heisenberg scaling limit can be exceeded in local estimation when the Hamiltonian contains interaction terms. For a Hamiltonian with identical k-body interactions between all subsets of k probe systems, the Heisenberg limit for the mean-square error becomes $N^{-2k+1/2}$ [56-62]. For instance, for $H_N$ the single-body Hamiltonian from above, a generator of the form $H_N^2$ gives a $1/N^2$ scaling. Additionally, the use of unentangled probe states still gives a scaling of $N^{-k+1/2}$, the possibility of the Hamiltonian itself generating entanglement closes the gap between the two types of probes. However, as mentioned in the introduction, such “super-Heisenberg” scalings are not to be found in global estimation. Hall and Wiseman offer a resolution of this apparent paradox by analyzing iterative schemes in which local estimation is repeatedly employed to perform global estimation [17].

A different nonlinear approach was proposed in [53-56], whereby the Hamiltonian couples all $N$ probe systems to an auxiliary system, i.e. the total Hamiltonian takes the form $H_N \otimes H'$. Here the goal was to show that separable probe states themselves offer 1/N scaling and that the scheme is more resilient to phase noise. Indeed, in [37] it is claimed that in this scenario a 1/N scaling is even possible with (classical) coupled harmonic oscillators. However, this conclusion does not hold for global estimation. Specifically, we now show that separable probe states have the same performance under a Hamiltonian of the form $H_N^{n+1} = J_{n,N} \otimes \sigma_z$, with $J_{n,N} = \sum_{k=1}^n \sigma_k^{(n)}$, as they do under $J_{n,N}$ itself. The reasoning is much the same as in the proof of Proposition 15. In particular, we need only consider pure product probe states, now of the form $|\psi_{B_N+1}⟩ = |\psi_{B_N}⟩ \otimes |\psi_{B_N+1}⟩$, where $|\psi_{B_N}⟩ = |\psi_{B_1}⟩ \otimes \cdots \otimes |\psi_{B_N}⟩$. The $|\psi_{B_N}⟩$ may again be truncated to a state $|\psi_\beta⟩$ with support on only $O(\sqrt{N})$ eigenvalues with weight within $\pm O(\sqrt{N})$ of the mean; note that now the weight is just a scalar as there is only one generator. The rest of the proof proceeds exactly as before, though now $n = 1$, as the advantage of the group average on the truncated probe state $|\psi_\beta'⟩_{B_N+1} = |\psi_{B_N}⟩|\psi_\beta⟩ \otimes |\psi_{B_N+1}⟩$ is to make a projective measurement in the eigenbasis of the generator $H_{n+1} = J_{n,N} \otimes \sigma_z$. While the latter has $N+1$ distinct eigenvalues, $|\psi_\beta⟩$ has support on no more than $O(\sqrt{N})$ of them. Thus we may conclude, as in Corollary 16 that

$$\frac{V_{\max}}{2\pi} \geq \frac{1}{O(\sqrt{N})}$$

for any prior distribution.

B. State Estimation

We now consider the problem of estimating a density matrix $\rho$ with known eigenvalues $r_1, \ldots, r_d$ from N i.i.d copies of the state $|\psi⟩$ (this is the “opposite” of the spectrum estimation problem a la [63]). Thus we consider the parameter space

$$X = \{ \rho = \rho^1 \in C^{d \times d} : \text{spec } \rho = \{ r_1, \ldots, r_d \} \},$$

and to each $\rho \in X$, we associate the probe state $\rho_{B_N} = \rho^{\otimes N}$ on $H_{B_N} = (C^d)^{\otimes N}$. Observe that $(\rho^{\otimes N})_{\rho \in X}$ is a covariant family with respect to the actions of the non-abelian group $G = SU(d)$ on $X$ (by conjugation) and on $H_{B_N}$ (diagonally).

We shall lower-bound [21] by evaluating the dual formulation (16) of $\beta_\alpha(\rho^{\otimes N}, \bar{\sigma})$, with $\bar{\sigma}$ the SU(d)-average of $\rho^{\otimes N}$. For this, recall that $(C^d)^{\otimes N}$ is also a representation of the symmetric group $S_N$, which acts by permuting the tensor factors. Since the action of $S_N$ commutes with the one of $SU(d)$, Schur’s lemma implies that the multiplicity spaces of the irreducible SU(d)-representations are representations of $S_N$ (and vice versa). Schur–Weyl duality asserts that these representations are in fact irreducible and pairwise distinct (e.g. [65]). In other words, we have a decomposition of the form

$$(C^d)^{\otimes N} = \bigoplus_\lambda V_{SU(d),\lambda} \otimes [\lambda],$$

where $V_{SU(d),\lambda}$ and $[\lambda]$ are distinct irreducible representations of $SU(d)$ and $S_N$. Since $\rho^{\otimes N}$ is permutation-invariant, it follows by Schur’s lemma that it is necessarily of the form

$$\rho^{\otimes N} = \bigoplus_\lambda p_\lambda \rho^\lambda \otimes \frac{[\lambda]}{\dim[\lambda]}.$$

Thus the $G$-average of $\rho^{\otimes N}$ is given by

$$\bar{\sigma} = \bigoplus_\lambda p_\lambda \frac{V_{SU(d),\lambda} \otimes [\lambda]}{\dim[\lambda]}.$$

Let $d_N = \max_\lambda \dim V_{SU(d),\lambda}$, where we take the maximum over all $\lambda$ that occur in the above decomposition. Then $\rho^{\otimes N}/d_N \leq \bar{\sigma}$, i.e. $\mu = 1/d_N$, $\tau = 0$ are feasible for the dual formulation (16) of $\beta_\alpha(\rho^{\otimes N}, \bar{\sigma})$, and so $\beta_\alpha(\rho^{\otimes N}, \bar{\sigma}) \geq \alpha_1/d_N$. But recall that we have shown in Lemma 13 that $d_N \leq O(N(d-1)(d-2)/2) = O(N^{d(d-1)/2})$. Thus we can conclude from [21] that we have the following lower bound, which holds for an arbitrary region estimator and uniform prior:

$$\frac{V_{\max}}{|X|} \geq C \frac{P_{\text{succ}}}{\sqrt{N^{d(d-1)}}} = C \frac{P_{\text{succ}}}{\sqrt{N^{d(d-1)}}},$$

where the latter identity holds if we assume that the eigenvalues $r_j$ are all distinct. Since the probe states $\rho^{\otimes N}$ are all separable, this is the expected shot noise scaling, namely $1/\sqrt{N}$ in each dimension.

It is interesting to compare this bound with the state-independent “Heisenberg” bound (25), where one allows an arbitrary covariant family of probe states on $(C^d)^{\otimes N}$. In the non-degenerate case where all the $r_j$ are distinct, we have $X \cong G/T$, where $T$ is the subgroup of diagonal matrices. Thus $d_\mu = 1$, and the multiplicities $r_{\mu}$ are the so-called Kostka numbers, which are of the order $O(N^{d(d-1)(d-2)/2})$. This can be seen e.g. by upper-bounding the Kostka numbers by the Kostant partition function and using [66, Lemma 1.5]. On the other hand, we recall from Lemma 13 that there are $O(N^{d(d-1)})$ isotypic components of dimension $d_\lambda = O(N^{d(d-1)/2})$ each. Consequently, the bound (25) reads

$$\frac{V_{\max}}{|X|} \geq D \frac{P_{\text{succ}}}{N^{d(d-1)+d(d-1)/2+2(d-2)/2}},$$
Analytical lower bounds for estimating the pure state of a qubit, $|\psi\rangle \in \mathbb{C}^2$, for uniform prior and varying $N$ ($n = 0.9$): Given $N$ i.i.d. copies of the state, $|\psi\rangle^\otimes N$, we cannot overcome $1/N$-scaling [36] (dashed line). In contrast, the state-independent lower bound [37], where we allow for an arbitrary covariant family of probe states on $(\mathbb{C}^2)^{\otimes N}$, scales as $1/N^2$ for large $N$ (solid line). This is consistent with shot-noise resp. Heisenberg scaling for the two-dimensional Bloch sphere.

$$V_{\text{max}} \geq C \frac{p_{\text{acc}}}{N^{d-1}} = C \frac{p_{\text{acc}}}{\sqrt{N^{d-1} \dim X}}.$$

Using the same argument as in [8VA], this implies that the mean-square error of any point estimator scales at best as $1/N$, in agreement with the central limit theorem and [67] Theorem 4) (where the minimal mean-square error was even computed to third order).

To evaluate the state-independent lower bound [37], where we allow for an arbitrary covariant family on $\mathcal{H}_B$, we use that $X \cong SU(d)/U(d-1)$. It follows by inspection of the corresponding branching rule that $\sum A dx m^2_A 2 \rho_A$ is of the order $O(X^{2(d-1)})$ [Lemma 20 in §C], so that

$$V_{\text{max}} \geq D \frac{p_{\text{acc}}}{\sqrt{N^{d-1} \dim X}}.$$

See Figure 3 for an illustration of both bounds in the case of qubits ($d = 2$ and hence $\dim X = 2$).

Thus in both cases we recover the expected shot-noise resp. Heisenberg scaling in each dimension of the parameter space, suggesting that it might be possible to generalize the covariant scaling results [Theorem 14] and [Corollary 16] to general homogeneous spaces $X$.

Acknowledgements. We thank Mario Berta, Robin Blume-Kohout, Matthias Christandl, Martin Fraas, David Gross, Volkher Scholz, and Mankei Tsang for valuable discussions and feedback. This work is supported by the European Research Council (grant 258932), the German Science Foundation (grant CH 843/2–1), the Swiss National Science Foundation (grants 200020_135048, P0002P_128455, and 20CH21_138799), and the National Center of Competence in Research ‘Quantum Science and Technology’.

Appendix A: Measurability

In this section we recall some definitions and facts on the measurability of vector-valued functions that are necessary to rigorously define cq-systems with infinite classical part (see e.g. [68] for the general theory).

Let $(X, \mu_X)$ be a standard measure space and $\mathcal{H}_B$ a (not necessarily finite-dimensional) Hilbert space. Let $B_1(\mathcal{H}_B)$ denote the space of trace-class operators, i.e. operators $A$ for which $\|A\|_1 := \text{tr}\sqrt{A^*A} < \infty$, and $B(\mathcal{H}_B)$ denote the space of bounded operators on $\mathcal{H}_B$ for which the operator norm $\|A\|_\infty < \infty$. Both spaces are Banach spaces, and the latter is the dual of the former.

A function $\rho_{XB} = (\rho^x_B) : X \rightarrow B_1(\mathcal{H}_B)$ is called simple if it is of the form $\rho^x_B = \sum_j A_j \otimes \rho_{O_j} (x)$ for disjoint measurable subsets $O_j \subseteq X$ and operators $A_j \in B_1(\mathcal{H}_B)$. It is called $\mu_X$-measurable if there exists a sequence of simple functions that converge $\mu_X$-almost everywhere to $\rho_{XB}$. In particular, this implies that the norm function $x \mapsto \|\rho^x_B\|_1$ is measurable in the usual sense.

A function $E_{XB} = (E^x_B) : X \rightarrow B(\mathcal{H}_B)$ is called weak*-measurable if the function $x \mapsto \langle \rho^x_B, E^x_B \rangle$ is measurable for each $\rho_B \in B_1(\mathcal{H})$.

We now show that the quantities $V_{\text{max}}$ and $p_{\text{acc}}$ are well-defined by [36] and [69].

Lemma 17. Let $M_B$ be a POVM with outcomes in some measurable space $Y$, $\mathcal{E} \subseteq X \times Y$ a measurable subset, and $\rho_{XB} = (\rho^x_B) = \mu_X$-measurable function which takes values in the set of density operators on $\mathcal{H}_B$.

1. The sets $E_x = \{ y : (x, y) \in \mathcal{E} \}$ and $E_y = \{ x : (x, y) \in \mathcal{E} \}$ are measurable. In particular, $V_{\text{max}}$ is well-defined.

2. The function $x \mapsto \text{tr}(\rho^x_B E_x^x_B)$ is weak*-measurable, i.e. $\langle \rho^x_B, E_x^x_B (\mathcal{E}_x) \rangle$ is measurable for each $\rho_B \in B_1(\mathcal{H}_B)$.

3. The function $x \mapsto \langle \rho^x_B, E^x_B (\mathcal{E}_x) \rangle$ is measurable. Therefore, the success probability $p_{\text{acc}}$ is well-defined.

Proof. 1. This is a standard fact that is routinely established on the route to proving Tonelli’s theorems, see e.g. [40] Proposition 2.34, a)].
2. For each $\rho_B$, $\langle \rho_B, E_B(-) \rangle$ is an ordinary probability measure on $Y$, so that the measurability of $\langle \rho_B, E_B(z) \rangle$ is again standard [49, Theorem 2.36].

3. This follows from the $\mu_X$-measurability of $\rho_{XB}$ and point 2 [68, Lemma 7.15].

**Appendix B: Zero Duality Gap for Binary Hypothesis Testing**

We consider binary hypothesis testing between two (normal) states $\rho$ and $\sigma$ in the positive cone $M^+_+ \subset$ of the predual $M_+\subset$ of an arbitrary von Neumann algebra $M$ [68]. In particular, all results in this section are applicable to binary hypothesis testing on classical–quantum systems as defined in §II (see below). We define the norm of a state to be $|\rho| := \langle \rho, 1_M \rangle$. It will be convenient to require that $|\rho|$ (but not $\sigma$) be normalized.

A binary hypothesis test between $\rho$ and $\sigma$ is a $\{\rho, \sigma\}$-valued measurement, or, equivalently, an observable $f \in F$ with $0 \leq f \leq 1_M$ (the POVM element corresponding to outcome $\rho$). The type I error is $1 - \alpha = \langle \rho, 1_M - f \rangle$, while the type II error is $\beta = \langle \sigma, f \rangle$. The hypothesis testing region is defined as

$$\mathcal{R}(\rho, \sigma) = \{(f, \rho), (\sigma, f) : 0 \leq f \leq 1_M\}.$$ 

Just as in the classical case [18, pp. 62] we can establish the following properties:

**Lemma 18.** The hypothesis testing region $\mathcal{R}(\rho, \sigma)$ is a compact, convex subset of the square $[0,1] \times [0,|\sigma|]$. It contains the diagonal and it is closed under the symmetry $(\rho, \sigma) \mapsto (1 - \alpha, |\sigma| - \beta)$.

**Proof.** All assertions except for compactness are immediate. To see compactness, observe that the set of POVM elements

$$\{f \in M : 0 \leq f \leq 1_M\} = B(X) \cap M^+$$

is a weak-$*$-closed subset of the unit ball, and therefore also weak-$*$-compact by the Banach–Alaoglu theorem [79, Theorem 4.3]. Since $(\langle \rho, \sigma \rangle, (\sigma, -))$ is weak-$*$-continuous (by definition), the assertion follows.

As a consequence, the infimum in the lower boundary function

$$\beta_\alpha(\rho, \sigma) = \inf\{\beta : (\alpha, \beta) \in \mathcal{R}(\rho, \sigma)\}$$

is attained, and $\beta_\alpha$ is a continuous, convex function of $\alpha \in [0,1]$ [71, Theorem 10.2].

Note that $\beta_\alpha(\rho, \sigma)$ can be understood as a linear cone program in $M$. We define the (predual) program by

$$\beta_\alpha^*(\rho, \sigma) := \sup\{\alpha \mu - \langle \tau, 1_M \rangle : \mu \in \mathbb{R}, \tau \in M_+^*, \tau \geq 0, \tau \geq \mu \rho - \sigma\}$$

and

$$\beta_\alpha^*(\rho, \sigma) := \sup\{\mu \tau - \langle \sigma, 1_M \rangle : \mu \geq 0, \tau \in M_+, \tau \geq \mu \rho - \sigma\}.$$ 

Again, it holds as in the classical case [72] that there is zero duality gap. We adapt an argument sketched in [73]:

**Proposition 19.** We have $\beta_\alpha(\rho, \sigma) = \beta_\alpha^*(\rho, \sigma)$.

**Proof.** Let us denote by $f^\bot$ the conjugate, or Legendre–Fenchel, transform, of a function $f$. Since $\beta_\alpha$ is a closed proper convex function, it is equal to its double Legendre transform [71, Corollary 12.2.1]:

$$\beta_\alpha(\rho, \sigma) = \beta_\alpha^{LL}(\rho, \sigma) = \sup\{\alpha \mu - \beta_\alpha^L(\rho, \sigma) : \mu \in \mathbb{R}\}.$$ 

We compute:

$$\beta_\alpha^L(\rho, \sigma) = \sup\{\alpha \mu - \beta_\alpha(\rho, \sigma) : \alpha \in \mathbb{R}\}$$

and

$$\beta_\alpha^{LL}(\rho, \sigma) = \sup\{\alpha \mu - \beta_\alpha(\rho, \sigma) : \alpha \in \mathbb{R}\}.$$ 

The supremum is attained when $f$ is the projection onto the positive part $\mu \rho - \sigma$ of $\mu \rho$ (which exists, since $M$ is a von Neumann algebra). Thus the Legendre transform $\beta_\alpha^*\rho$ is given by

$$\beta_\alpha^*(\rho, \sigma) = \langle \mu \rho - \sigma, 1_M \rangle.$$ 

The claim follows by plugging this into the right-hand side of [B1] and comparing it with the definition of $\beta_\alpha^*(\rho, \sigma)$. }

Let $D(\rho, \sigma)$ denote the relative entropy between two states $\rho$ and $\sigma$ in $M$. [43, 85]. Then we have the following relation between the relative entropy and the hypothesis-testing region,

$$\begin{aligned}
D(\rho|\sigma) \\
\geq D((\alpha, 1 - \alpha)) \|\beta, |\sigma| - \beta\|) \\
= d(\alpha \|\beta\|/\|\sigma\|) - \log |\sigma| \\
\forall (\alpha, \beta) \in \mathcal{R}(\rho, \sigma)
\end{aligned}$$

where $d(p|q) := p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}$ denotes the binary relative entropy function. It is an immediate consequence of the data-processing inequality [43, Corollary 5.12, (iii)], applied to the channels that correspond to tests (cf. [45, (154)]). In particular, [B2] holds for $\beta = \beta_\alpha(\rho, \sigma)$ (use the optimal test).

Now consider cq-hypothesis testing as discussed in §III. We define $M_{XB} = L^\infty(X, \mu, B(\mathcal{H}_B))$, the space of all weak-$*$-measurable functions for which $\|E_{XB}\|_\infty < \infty$, identifying any two functions that are $\mu$-almost everywhere equal. Then $M_{XB}$ is a von Neumann algebra. Its predual is given by $M_{XB^*}^\perp = L^1(X, \mu, B_1(\mathcal{H}_B))$, the space of all $\mu$-measurable functions for which $\|\rho_{XB}\|_1 := \int_X d\mu_x(\rho_{XB}) < \infty$ [68, Theorem 4.2.17]. Thus the set of unnormalized cq-states as defined in §III agrees precisely with the positive cone $(M_{XB})_+^\perp$ of the predual, and it is easily seen that the definitions of $\beta_\alpha$ and $\beta_\alpha^*$ for general von Neumann algebras reduce to the definitions given in §III for cq-systems.

**Appendix C: A Technical Lemma**

**Lemma 20.** Let $\mathcal{H}_{B\otimes N} = (\mathcal{C}^d)^\otimes N$, $G = U(d)$ acting diagonally by $U^\otimes N$, and $K = U(d - 1) \subseteq G$ embedded as the
upper-left block. Then we have that
\[
\max_{\mu} \sum_{\lambda} d_\lambda r_{\lambda\mu}^\nu / d_\mu = O(N^{2(d-1)}),
\]
where \(d_\lambda, d_\mu\) and \(r_{\lambda\mu}^\nu\) are defined as in Corollary 10.

Proof: We can label the irreducible representations of \(U(d)\) by Young diagrams, i.e. non-increasing sequences \(\lambda = (\lambda_1, \ldots, \lambda_d)\) of non-negative integers. By the Weyl dimension formula \([6]\), the dimension of the corresponding irreducible representation \(V_{U(d), \lambda}\) is then given by
\[
d_\lambda = \prod_{1 \leq i < j \leq d} \frac{\lambda_i - \lambda_j + j - i}{j - i}.	ag{C1}
\]

From Schur–Weyl duality, it is well-known that the irreducible representations \(V_{U(d), \lambda}\) that appear in \(H_{B_N}\) are precisely those that correspond to Young diagrams with \(|\lambda| := \sum \lambda_j = N\). The restriction of such an irreducible representation to the subgroup \(K = U(d-1)\) is given by the following branching rule \([74]\),
\[
V_{U(d), \lambda}|_{U(d-1)} = \bigoplus_{\mu} V_{U(d-1), \mu},
\]
where the direct sum runs over all Young diagrams \(\mu = (\mu_1, \ldots, \mu_{d-1})\) for which
\[
\lambda_j \geq \mu_j \geq \lambda_{j+1}.	ag{C2}
\]

Note that all (non-zero) multiplicities \(m^\lambda_\mu\) are equal to one. Since \(r_{\lambda\mu}^\nu \leq m_\lambda^\mu\), and using the dimension formula \(C1\), it follows that
\[
d_\lambda r_{\lambda\mu}^\nu / d_\mu = d_\lambda / d_\mu = \frac{1}{d!} \prod_{1 \leq i < j \leq d} (\lambda_i - \lambda_j + j - i) \prod_{1 \leq i < j \leq d-1} (\mu_i - \mu_j + j - i).
\]

By the branching rule \(C2\), we have that \(\mu_i - \mu_j \geq \lambda_{i+1} - \lambda_j\), and so \(\mu_i - \mu_j + j - i \geq \lambda_{i+1} - \lambda_j + j - (i + 1)\), which is non-trivial if \(i + 1 < j\). We can thus upper-bound the above expression by
\[
1 \prod_{1 \leq i < j \leq d-1} (\lambda_i - \lambda_j + j - i) \prod_{1 \leq i < j \leq d} (\lambda_i - \lambda_j + d - i) \prod_{1 \leq i < j \leq d-1} (\mu_i - \mu_{i+1} + 1) \leq O(N^{d-2} N^{d-1})
\]

since each \(\lambda_i - \lambda_j \leq \lambda_j \leq N\). On the other hand, observe that the branching rule \(C2\) also implies that \(\mu_i \geq \lambda_{i+1} \geq \mu_{i+1}\) for all \(\lambda\) that are compatible with a given \(\mu\). Since the sum of all \(\lambda_j\) has to be equal to \(N\), we find that the number of compatible \(\lambda\) can be upper-bounded by
\[
N \prod_{1 \leq i < d-1} (\mu_i - \mu_{i+1} + 1).
\]

We conclude that
\[
\sum_{\lambda} d_\lambda r_{\lambda\mu}^\nu / d_\mu \leq O(N^{2(d-1)}).
\]

[1] C. W. Helstrom, “Estimation of a Displacement Parameter of a Quantum System,” Int. J. Theo. Phys. 11, 357 (1974).
[2] A. S. Holevo, “Estimation of Shift Parameters of a Quantum State,” Rep. Math. Phys. 13, 379 (1978).
[3] R. Werner, “Screen observables in relativistic and nonrelativistic quantum mechanics,” J. Math. Phys. 27, 793 (1985).
[4] V. Giovannetti, S. Lloyd, and L. Maccone, “Quantum Metrology,” Phys. Rev. Lett. 96, 010401 (2006).
[5] V. Giovannetti, S. Lloyd, and L. Maccone, “Quantum-Enhanced Measurements: Beating the Standard Quantum Limit,” Science 306, 1330–1336 (2004).
[6] S. L. Braunstein and C. M. Caves, “Statistical Distance and the Geometry of Quantum States,” Phys. Rev. Lett. 72, 3439 (1994).
[7] B. C. Sanders and G. J. Milburn, “Optimal Quantum Measurements for Phase Estimation,” Phys. Rev. Lett. 75, 2944 (1995).
[8] D. W. Berry and H. M. Wiseman, “Optimal States and Almost Optimal Adaptive Measurements for Quantum Interferometry,” Phys. Rev. Lett. 85, 5098 (2000).
[9] J. Neyman, “On the Problem of Confidence Intervals,” Ann. Math. Stat. 6, 111 (1935).
[10] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, “Quantum State Tomography via Compressed Sensing,” Phys. Rev. Lett. 105, 150401 (2010).
[11] R. Blume-Kohout, “Optimal, reliable estimation of quantum states,” New J. Phys. 12, 043034 (2010).
[12] R. Blume-Kohout, “Robust error bars for quantum tomography,” (2012), arXiv:1202.5270.
[13] M. Christandl and R. Renner, “Reliable Quantum State Tomography,” Phys. Rev. Lett. 109, 120403 (2012).
[14] M. Paris and Réhák, eds., Quantum State Estimation, Lecture Notes in Physics, Vol. 649 (Springer, 2004).
[15] M. Fraas, “An analysis of the stationary operation of atomic clocks,” (2013), arXiv:1303.6083.
[16] G. A. Durkin and J. P. Dowling, “Local and Global Distinguishability in Quantum Interferometry,” Phys. Rev. Lett. 99, 070801 (2007).
[17] M. J. W. Hall and H. M. Wiseman, “Does Nonlinear Metrology Offer Improved Resolution? Answers from Quantum Information Theory,” Phys. Rev. X 2, 041006 (2012).
[18] E. L. Lehmann and J. P. Romano, Testing Statistical Hypotheses (Springer, 2005).
[19] Interestingly, the family of Ziv–Zakai inequalities \([75,76]\) give lower bounds on the mean square error of point estimators in terms of Bayesian binary hypothesis testing, with alternative hypotheses the shifted probe states \(\rho^{\mu+}\) for varying choices of \(\tau\).
[73] Y. Polyanskiy, “Saddle Point in the Minimax Converse for Channel Coding,” IEEE Trans. Inf. Theory 59, 2576–2595 (2013).
[74] H. Weyl, The classical groups (Princeton University Press, 1946).
[75] J. Ziv and M. Zakai, “Some Lower Bounds on Signal Parameter Estimation,” IEEE Trans. Inf. Theory 15, 386 (1969).
[76] M. Tsang, “Ziv-Zakai Error Bounds for Quantum Parameter Estimation,” Phys. Rev. Lett. 108, 230401 (2012).