Optimal Heisenberg-style bounds for the average performance of arbitrary phase estimates

Dominic W. Berry[1], Michael J. W. Hall[2], Marcin Zwierz[2] and Howard M. Wiseman[2]
[1] *Department of Physics and Astronomy, Macquarie University, Sydney, NSW 2109, Australia*
[2] *Centre for Quantum Computation and Communication Technology (Australian Research Council), Centre for Quantum Dynamics, Griffith University, Brisbane, QLD 4111, Australia*
(Dated: November 20, 2012)

The ultimate bound to the accuracy of phase estimates is often assumed to be given by the Heisenberg limit. Recent work seemed to indicate that this bound can be violated, yielding measurements with much higher accuracy than was previously expected. The Heisenberg limit can be restored as a rigorous bound to the accuracy provided one considers the accuracy averaged over the possible values of the unknown phase, as we have recently shown [Phys. Rev. A **85**, 041802(R) (2012)]. Here we present an expanded proof of this result together with a number of additional results, including the proof of a previously conjectured stronger bound in the asymptotic limit. Other measures of the accuracy are examined, as well as other restrictions on the generator of the phase shifts. We provide expanded numerical results for the minimum error and asymptotic expansions. The significance of the results claiming violation of the Heisenberg limit is assessed, followed by a detailed discussion of the limitations of the Cramér-Rao bound.

PACS numbers: 42.50.St, 03.65.Ta, 06.20.Dk

I. INTRODUCTION

Phase estimation is the basis for much precision measurement. Optical interferometers offer highly accurate measurements of length, and atomic phase measurements provide highly accurate measurements of time, as well as other physical quantities like magnetic field [1–5]. In optics, most measurements are limited by the shot-noise limit, where the accuracy scales as $1/\sqrt{\langle N\rangle}$, where N is the photon number operator. In contrast, it is normally assumed that the fundamental limit is the Heisenberg limit, where the accuracy scales as $1/\langle N\rangle$ [6–8]. This potentially provides far greater accuracy, but is extremely difficult to achieve in practice because it requires highly nonclassical states of light, as well as arbitrarily high efficiencies [9–11]. Any amount of loss will cause the scaling to revert to $1/\sqrt{\langle N\rangle}$ for large $\langle N\rangle$ [11].

Recently a number of papers suggested that the Heisenberg limit is not the fundamental limit to accuracy, and that a better scaling constant or even a higher power of $\langle N\rangle$ might be possible. In Ref. [12], Anisimov *et al.* gave a proposal for violating the Heisenberg limit by a small amount. In another work, Zhang *et al.* [13] proposed a scheme offering zero phase uncertainty with finite $\langle N\rangle$. Finally, in Ref. [14] Rivas and Luis presented a proposal for obtaining scaling as $1/\langle N\rangle^p$ for $p > 1$. A qualitatively different proposal for violating the Heisenberg limit is that based on nonlinear interferometry [15, 16]. However, that work differs in its use of terminology; it does not violate the Heisenberg limit in the sense we use here (see Sec. VII D) [7].

A common feature of proposals to violate the Heisenberg limit is that they only work for a limited range of phases. Additional phase information would be needed to confine the phase to within the region where the measurement is accurate. One can consider first using a sequence of measurements to ensure that the phase lies within a suitable region, then using the super-Heisenberg measurement. If the *overall* measurement (consisting of the sequence of individual measurements) could yield better accuracy than the Heisenberg limit, then it could be regarded as providing a true improvement. On the other hand, if the resources required to localise the phase to the required region result in an overall measurement with accuracy that is not better than the Heisenberg limit, then the accuracy of the super-Heisenberg measurements would seem to be illusory.

An analogous situation was seen in considering the reciprocal-peak-likelihood as a measure of uncertainty. In Ref. [17, 18] a technique was proposed that would apparently yield super-Heisenberg accuracy in terms of reciprocal-peak-likelihood. Later work found that, in practice, the proposal resulted in accuracy that was worse than the Heisenberg limit [19]. Another example is that of NOON states. NOON states yield phase information scaling as the Heisenberg limit, but require initial phase information with similar accuracy. In that case, it is known how to combine measurements from multiple states to obtain an overall measurement that scales at the Heisenberg limit [20, 21].

To evaluate whether the super-Heisenberg measurements would be able to yield an overall measurement violating the Heisenberg limit, we examined the case where the mean-square error is averaged over all phase shifts. We showed that the Heisenberg limit provides a rigorous lower bound to the square root of the average mean-square error (RAMSE) in such a case [22]. Therefore, no scheme that apparently beats the Heisenberg limit for a small range of phase could be used to construct an overall measurement starting from an unknown phase that beats the Heisenberg limit. An alternative approach is to determine the bound if the initial phase is restricted

to a given range. In Ref. [23] it was shown that with such a restriction the usual Heisenberg limit can be multiplied by a factor proportional to the phase range, and further results have been given in Refs. [24–27]. An alternative approach has yielded a bound on the average of the error at just two locations [28].

The specific result from Ref. [22] is

$$\delta\hat{\Phi} \geq \frac{k}{\langle G+1\rangle}, \tag{1}$$

where $\delta\hat{\Phi}$ is the RAMSE, G is the generator of the phase shifts, which is here assumed to have nonnegative integer eigenvalues, and k is a constant. These quantities are explained in Sec. II below. We have analytically proven that this inequality holds with $k = k_A := \sqrt{2\pi/e^3} \approx 0.5593$ [22]. In Sec. III we give the full proof of that result, as well as a generalised result in terms of the absolute value of G in the case where G also has negative integer eigenvalues.

Numerical calculations indicate that the inequality holds with the larger scaling constant $k = k_C \approx 1.3761$. We give the detailed numerical results in Sec. IV, indicating that this result holds both for the RAMSE and the error estimated using the Holevo variance. In Sec. V we calculate asymptotic expansions for the RAMSE, providing strong analytic support for the scaling constant k_C, and proving that $k = k_C$ is valid in the asymptotic limit $\langle G\rangle \to \infty$. We examine the scaling with the number of probe states in Sec. VI, then give a detailed discussion of the papers claiming violation of the Heisenberg limit in Sec. VII. The Cramér-Rao bound and the error propagation formula are commonly used in examining the Heisenberg limit, but have some limitations; these are discussed in Sec. VIII.

II. FIGURES OF MERIT FOR AVERAGE PHASE RESOLUTION

There are a number of different figures of merit for phase measurements. Before describing these, we first introduce some notation, largely following Ref. [22]. The random variable for the phase shift of the system is Φ, and the random variable for the estimate of that phase shift is $\hat{\Phi}$. The error in the phase estimate is $\Theta = \hat{\Phi} - \Phi$. We use capital letters for the random variables; the corresponding values and measurement outcomes are denoted by the corresponding lower case letters (ϕ, $\hat{\phi}$, and θ).

We consider a Hilbert space with a phase shift operator G. In the completely general case, the only restriction is that the eigenvalues of G must be integers. We may also consider the specific case where the eigenvalues are all nonnegative integers, in which case we denote the operator by N. This includes, for example, the case of photon number. Alternatively, if the eigenvalues include all integers, such as for angular momentum, we use the symbol J.

The phase shift is described by the unitary operator $\exp(-iG\Phi)$. That is, the probe state ρ_0 becomes $\rho_\phi := e^{-iG\phi}\rho_0 e^{iG\phi}$. The detection method used to estimate ϕ is described by a positive-operator valued measure (POVM) $\{M_{\hat{\phi}}\}$. Hence, the probability distribution is given by $p(\hat{\phi}|\phi) = \mathrm{Tr}(M_{\hat{\phi}}\rho_\phi)$. Because phase is only defined modulo 2π, we do not distinguish between ϕ and $\phi + 2\pi$, or between $\hat{\phi}$ and $\hat{\phi} + 2\pi$. This means that $p(\hat{\phi}|\phi) = p(\hat{\phi} + 2\pi k|\phi)$ for any integer k, and $p(\hat{\phi}|\phi)$ is normalised over a (arbitrary) 2π interval.

A. Root-mean-square error

The most common figure of merit for a measurement is the square root of the mean-square error (MSE). We will call this the RMSE. For a specific phase shift, ϕ, the MSE is given by

$$(\Delta_\phi^{\phi_r}\hat{\Phi})^2 := \int_{\phi_r-\pi}^{\phi_r+\pi} d\hat{\phi}\,(\hat{\phi}-\phi)^2 p(\hat{\phi}|\phi). \tag{2}$$

There is a subtlety in that for phase, values that differ by 2π are equivalent, which means that a range of 2π must be specified for the integral. However, the reference phase shift, ϕ_r, is arbitrary, and the value that is obtained for the MSE will depend on ϕ_r. Ideally ϕ should be near the centre of the range. If it is near one of the bounds of the range then the MSE will be unreasonably large.

To solve this problem, it is convenient to take the difference $\hat{\phi}-\phi$ modulo $(-\pi,\pi]$. That is, we add or subtract a multiple of 2π such that the value obtained is in the range $(-\pi,\pi]$. It is important to note that this convention can only decrease the value obtained for the MSE. In this work we are concerned with placing lower bounds on the MSE. We prove these lower bounds for the MSE with the difference defined modulo 2π. Because this MSE is no larger than that obtained without taking the difference modulo $(-\pi,\pi]$, all results hold for that case as well. Thus, it is natural to work with the minimum MSE, given by

$$\begin{aligned}(\Delta_\phi\hat{\Phi})^2 &:= \int_{\phi_r-\pi}^{\phi_r+\pi} d\hat{\phi}\ \left\{(\hat{\phi}-\phi)\mod(-\pi,\pi]\right\}^2 p(\hat{\phi}|\phi)\\ &= \int_{\phi-\pi}^{\phi+\pi} d\hat{\phi}\,(\hat{\phi}-\phi)^2 p(\hat{\phi}|\phi),\end{aligned} \tag{3}$$

where we have used the fact that $p(\hat{\phi}|\phi)$ repeats modulo 2π. It follows that

$$\Delta_\phi\hat{\Phi} \equiv \Delta_\phi^{\phi}\hat{\Phi} \leq \Delta_\phi^{\phi_r}\hat{\Phi} \tag{4}$$

for any reference phase ϕ_r.

The above $\Delta_\phi\hat{\Phi}$ is a measure of the accuracy of the phase measurement only for a specific phase shift ϕ. It is trivial to see that one can always choose a measurement such that the MSE can be zero for a specific phase

shift, ϕ_0: the trivial measurement that always yields the result $\hat{\phi} = \phi_0$. In reality, for a phase measurement the phase shift is unknown; otherwise a measurement would be unnecessary. To be useful, a measurement must give accurate results for a range of phase shifts.

A rigorous way of taking account of the range of phase is to average the figure of merit over the phase shift. For the MSE one would use

$$\int_{-\pi}^{\pi} d\phi\, p(\phi)(\Delta_\phi \hat{\Phi})^2, \tag{5}$$

where $p(\phi)$ is a probability distribution describing the prior information about the phase shift. In this work we consider the case that there is no prior information, so $p(\phi) = 1/2\pi$. Then the average MSE (AMSE) is given by

$$(\delta\hat{\Phi})^2 := \frac{1}{2\pi}\int_{-\pi}^{\pi} d\phi\, (\Delta_\phi \hat{\Phi})^2. \tag{6}$$

One then finds that

$$(\delta\hat{\Phi})^2 = \int_{-\pi}^{\pi} d\theta\, \theta^2\, \bar{p}(\theta) = \langle \Theta^2 \rangle. \tag{7}$$

Here $\bar{p}(\theta)$ is the probability density for the error in the phase estimate $\Theta = \hat{\Phi} - \Phi$, and is defined by

$$\bar{p}(\theta) := \frac{1}{2\pi}\int_{-\pi}^{\pi} d\phi\, p(\theta + \phi|\phi). \tag{8}$$

We call $\delta\hat{\Phi}$ the RAMSE, because it is averaged over ϕ before taking the square root, whereas the RMSE $\Delta_\phi\hat{\Phi}$ is for a specific ϕ.

Equation (7) holds because the mean-square error is a linear figure of merit. A general figure of merit for the accuracy of a phase estimate $\hat{\Phi}$ can be defined as a functional, F, that takes as input a probability density in $\hat{\phi}$, and outputs a scalar. In the case that F is linear, we find that

$$\int_{-\pi}^{\pi} d\phi\, p(\phi) F(p(\hat{\Phi}|\phi)) = F(\bar{p}(\Theta)). \tag{9}$$

This means that, for linear measures, the average figure of merit and the figure of merit of the average distribution are equivalent.

More generally, consider a convex figure of merit; that is, one that satisfies

$$F(tp_1(\hat{\Phi}) + (1-t)p_2(\hat{\Phi})) \leq tF(p_1(\hat{\Phi})) + (1-t)F(p_2(\hat{\Phi})), \tag{10}$$

for $t \in [0,1]$. By using Jensen's inequality, one obtains

$$\begin{aligned}\int_{-\pi}^{\pi} d\phi\, p(\phi) F(p(\hat{\Phi}|\phi)) &\geq F\left(\int_{-\pi}^{\pi} d\phi\, p(\phi) p(\hat{\Phi}|\phi)\right) \\ &= F(\bar{p}(\Theta)).\end{aligned} \tag{11}$$

What this means is that, if the figure of merit is convex, then placing a lower bound on the figure of merit for the average distribution also provides a lower bound on the average of the figure of merit. That is the approach we use in this work; we find lower bounds on the figure of merit for the average distribution, which also hold for the average of the figure of merit.

B. Holevo variance and average bias

There are alternative measures of the spread which are similar to the MSE but which are specifically defined for phase. These are typically defined in terms of the average of the exponential of the phase, $\langle e^{i\hat{\Phi}}\rangle$. In the case that the phase distribution is sharply peaked, then this quantity will be close to 1. One possibility for quantifying the uncertainty in the phase is $2(1 - |\langle e^{i\hat{\Phi}}\rangle|)$ [29]; another is $1 - |\langle e^{i\hat{\Phi}}\rangle|^2$ [30, 31].

A measure of this type with some nice properties is that proposed by Holevo [32],

$$V_{H,\phi}(\hat{\Phi}) := |\langle e^{i\hat{\Phi}}\rangle_\phi|^{-2} - 1, \tag{12}$$

which has been dubbed the Holevo variance [33]. Here the subscript ϕ indicates that the variance is determined for a specific value of the phase shift. That is,

$$V_{H,\phi}(\hat{\Phi}) := \left|\left\langle \int_{-\pi}^{\pi} d\hat{\phi}\, e^{i\hat{\phi}} p(\hat{\phi}|\phi)\right\rangle\right|^{-2} - 1. \tag{13}$$

In this case there is no ambiguity in choosing the bounds of the integral, because the argument is clearly periodic modulo 2π.

A minor problem with this definition is that it does not penalise biased estimates. However, this is easily corrected by using the modified definition

$$\mathrm{Var}_{H,\phi}\hat{\Phi} := \mathrm{Re}\langle e^{i(\hat{\Phi}-\phi)}\rangle_\phi^{-2} - 1. \tag{14}$$

If the measurement is "$U(1)$-unbiased", in the sense that

$$\phi = \arg[\langle e^{i\hat{\Phi}}\rangle_\phi], \tag{15}$$

then these two expressions for the Holevo variance are equivalent.

The Holevo variance is a convex functional of the probability distribution. From Eq. (11), this means that one can place a lower bound on the average Holevo variance by considering the Holevo variance of the average distribution. That is,

$$(\delta_H\hat{\Phi})^2 := (\mathrm{Re}\langle e^{i\Theta}\rangle)^{-2} - 1 \tag{16}$$

is a lower bound on the average value of $\mathrm{Var}_{H,\phi}(\hat{\Phi})$. In this paper we do not discuss the Holevo variance without averaging over ϕ, so we will refer to $(\delta_H\hat{\Phi})^2$ as the Holevo variance.

In the case that the average distribution $\overline{p}(\theta)$ is $U(1)$-unbiased, in the sense that $\langle e^{i\Theta}\rangle$ is real and positive, then

$$(\delta_H\hat{\Phi})^2 = |\langle e^{i\Theta}\rangle|^{-2} - 1. \tag{17}$$

If the average distribution is biased, then it can be modified to obtain a $U(1)$-unbiased measurement. Taking $\theta_{\rm av} := \arg\langle e^{i\Theta}\rangle$, we can replace measurement operators $M_{\hat{\Phi}}$ with

$$M'_{\hat{\Phi}} = M_{\hat{\Phi}+\theta_{\rm av}}. \tag{18}$$

Then, for these new measurement operators, $p'(\hat{\phi}|\phi) = p(\hat{\phi}+\theta_{\rm av}|\phi)$, so

$$\begin{aligned}\langle e^{i\Theta}\rangle_{M'} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} d\hat{\phi}\int_{-\pi}^{\pi} d\phi\, e^{i(\hat{\phi}-\phi)}p(\hat{\phi}+\theta_{\rm av}|\phi)\\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} d\hat{\phi}\int_{-\pi}^{\pi} d\phi\, e^{i(\hat{\phi}-\theta_{\rm av}-\phi)}p(\hat{\phi}|\phi)\\ &= e^{-i\theta_{\rm av}}\langle e^{i\Theta}\rangle_M.\end{aligned} \tag{19}$$

Hence this modification of the measurement yields a $U(1)$-unbiased average measurement.

With this condition, we can bound the mean-square error by using the following inequality,

$$|\langle e^{i\Theta}\rangle| = \langle\cos\Theta\rangle \geq \cos\sqrt{\langle\Theta^2\rangle}, \tag{20}$$

where we have used the fact that $\cos\sqrt{x}$ is a convex function, along with Jensen's inequality. There are alternative ways to bound the mean-square error, but this particular inequality will be useful in Appendix C. It also has the nice property that it can be saturated, for a probability distribution that is just delta functions at $\pm\sqrt{\langle\Theta^2\rangle}$. Now consider the limit where the mean-square error $(\delta\hat{\Phi})^2 = \langle\Theta^2\rangle$ is small. Expanding as a Maclaurin series in this small parameter, we obtain

$$\begin{aligned}(\delta\hat{\Phi})^2 &\geq \left(\arccos\{[(\delta_H\hat{\Phi})^2+1]^{-1/2}\}\right)^2\\ &= (\delta_H\hat{\Phi})^2 - \frac{2}{3}(\delta_H\hat{\Phi})^4 + O((\delta_H\hat{\Phi})^6).\end{aligned} \tag{21}$$

This means that, except for higher-order terms, the Holevo variance lower bounds the AMSE from below, so asymptotically we have $(\delta_H\hat{\Phi})^2 \lesssim (\delta\hat{\Phi})^2$.

We can also use the Holevo variance to bound the AMSE from above. Using the fact that $\cos\theta \leq 1-2\theta^2/\pi^2$ on the interval $[-\pi,\pi]$, we have the inequality

$$\langle\cos\Theta\rangle \leq 1 - \frac{2\langle\Theta^2\rangle}{\pi^2}. \tag{22}$$

Using this, we have

$$\begin{aligned}(\delta\hat{\Phi})^2 &\leq \frac{\pi^2}{2}\left(1-[(\delta_H\hat{\Phi})^2+1]^{-1/2}\right)\\ &= \frac{\pi^2}{4}(\delta_H\hat{\Phi})^2 - \frac{3\pi^2}{16}(\delta_H\hat{\Phi})^4 + O((\delta_H\hat{\Phi})^6).\end{aligned} \tag{23}$$

The reason for the factor of $\pi^2/4$ is that even for small variance, the main contribution to the AMSE can be from large phase errors. The inequality (22) is saturated for a distribution that has contributions at $\pm\pi$.

Returning to the asymptotic lower bound (21), its significance is that any lower bound on the Holevo variance is also asymptotically a lower bound on the AMSE. In particular, it is known that for canonical phase measurements on a single-mode field there is the tight asymptotic lower bound on the Holevo variance $\delta_H\hat{\Phi} \gtrsim k_C/\langle N\rangle$ with $k_C := 2(-z_A/3)^{3/2} \approx 1.3761$ (where z_A is the first zero of the Airy function) [34, 35]. This is tight in the sense that, asymptotically, the Holevo variance is equal to this value with any difference being of higher order. Because the Holevo variance is asymptotically a lower bound on the usual AMSE, we must also asymptotically have the lower bound $\delta\hat{\Phi} \gtrsim k_C/\langle N\rangle$ for canonical phase measurements on a single-mode field. It will be shown in Sec. III that this is in fact a tight lower bound.

C. Entropic length

Another measure of concentration is the *entropic length* [36, 37]. This is given by

$$L(\hat{\Phi}) := e^{H(\Theta)}, \tag{24}$$

where $H(\Theta)$ is the entropy of the error probability density,

$$H(\Theta) = -\int_{-\pi}^{\pi}\bar{p}(\theta)\ln(\bar{p}(\theta))d\theta. \tag{25}$$

The entropy takes its largest positive value for a flat distribution, and takes large negative values as the distribution provides more information about the phase. The negative of the entropy provides a measure of how much information about the phase is available. The entropic length is correspondingly small for a distribution providing a lot of information about the phase.

Similar to the AMSE or the Holevo variance, the entropic length will be small for a sharply peaked distribution. However, in contrast to those measures, the entropic length will also be small if there are multiple sharp peaks, with a value roughly equal to the total width of those peaks. The entropic length satisfies several basic properties expected for a length, discussed in Ref. [36]. It can also be used to provide a lower bound to the RAMSE via the relation [37]

$$\delta\hat{\Phi} \geq (2\pi e)^{-1/2}L(\hat{\Phi}). \tag{26}$$

This is because, if one were considering a distribution on the infinite line, the entropy is maximised for fixed $\hat{\Phi}$ by a Gaussian distribution, in which case $\delta\hat{\Phi} = (2\pi e)^{-1/2}L(\hat{\Phi})$. For the case of phase, we are limited to the interval $[-\pi,\pi]$, which means that the Gaussian distribution cannot be obtained exactly. Therefore the inequality still holds, but cannot be saturated except asymptotically.

In contrast to the other measures considered here, the entropy is *not* convex. This means that one needs to

be cautious when considering the average entropy. The entropy of the average distribution does not provide a lower bound on the average of the entropies. We do not determine the lower bound on the average of entropies; this is an open problem.

III. OBTAINING UNIVERSAL BOUNDS FROM NONDEGENERATE BOUNDS

We now present the universal form of the Heisenberg limit, which was first derived in Ref. [22]. In subsection A we present the theorem showing that bounds which hold for canonical measurements on nondegenerate systems also hold for completely arbitrary measurements on general systems. In optics a single-mode field is nondegenerate, whereas the general case includes multimode interferometry. In subsection B we use this to provide our universal form of the Heisenberg limit. In Sec. IV we will present numerical results indicating that a better scaling constant is possible.

A. Mapping the general problem to a nondegenerate problem

As discussed at the start of Sec. II, the detection method may be described by a POVM $\{M_{\hat{\phi}}\}$, which gives the probability distribution via

$$p(\hat{\phi}|\phi) = \mathrm{Tr}(M_{\hat{\phi}}\rho_{\phi}). \tag{27}$$

A particularly useful form of POVM is a *covariant* POVM. Whereas for an arbitrary POVM the individual $M_{\hat{\phi}}$ can be chosen independently of each other (except for the normalisation requirement), for a covariant POVM only one measurement operator may be chosen, then all others are related via the generator of shifts. In particular,

$$\overline{M}_{\hat{\phi}} = e^{-iG\hat{\phi}}\overline{M}_0 e^{iG\hat{\phi}}. \tag{28}$$

For a covariant POVM, the probability distribution for the error in the estimate is independent of the phase shift. This may be shown via

$$\begin{aligned} p(\theta+\phi|\phi) &= \mathrm{Tr}(\overline{M}_{\theta+\phi}\rho_{\phi}) \\ &= \mathrm{Tr}(e^{-iG(\theta+\phi)}\overline{M}_0 e^{iG(\theta+\phi)}e^{-iG\phi}\rho_0 e^{iG\phi}) \\ &= \mathrm{Tr}(e^{-iG\theta}\overline{M}_0 e^{iG\theta}\rho_0). \end{aligned} \tag{29}$$

A particular form of covariant POVM is the *canonical* POVM. This can be defined as $\{e^{-iG\phi}C_0 e^{iG\phi}\}$, with [38]

$$C_0 = \frac{1}{2\pi}\sum_d \sum_{n,n'\in S; d\leq D(n), D(n')} |n,d\rangle\langle n',d|. \tag{30}$$

Here we have labelled the states with n and n' indicating the eigenvalues of G, and d the degeneracy. The function $D(n)$ gives the degeneracy for eigenvalue n. S denotes the spectrum of eigenvalues of G, which we have assumed to be the integers or a subset thereof. This definition of a canonical POVM is not unique in general, because it depends on the labelling of the degenerate states; a fact which was not noted in Refs. [22, 38], and which does not affect the results therein. However, we will only require the simpler case of no degeneracies in what follows, where for this case the POVM is uniquely given by $\{e^{-iG^{(s)}\phi}C_0^{(s)}e^{iG^{(s)}\phi}\}$, with

$$C_0^{(s)} = \frac{1}{2\pi}\sum_{n,n'\in S} |n\rangle\langle n'|. \tag{31}$$

We use $G^{(s)}$ to denote a generator with the same spectrum of eigenvalues as G, but nondegenerate.

We now show that any average phase distribution, $\bar{p}(\theta)$, can be obtained by a covariant measurement, and that the covariant measurement result can be obtained by a canonical measurement on a system without degeneracy. In Ref. [22] we obtained this result by a three-step process: first that any average phase distribution, $\bar{p}(\theta)$, can be obtained by a covariant measurement; second that the covariant measurement result can be obtained by a canonical measurement; and third that the canonical measurement result can be obtained by a canonical measurement on a system without degeneracy. Here we simplify the proof by combining the second two steps.

To express these results it is convenient to modify the notation slightly. We will use subscripts on the probability p to indicate the POVM used. In addition, we will indicate the state used in the probability. In the case of the probability for the measurement error θ for the covariant POVM, we omit ϕ, because the probability is independent of ϕ as discussed above. Therefore, we replace $p(\theta+\phi|\phi)$ with $p_{\overline{M}}(\theta|\rho_0)$.

Expressed in terms of this notation, the first result is as follows.

Lemma 1. *For any POVM $\{M_{\hat{\phi}}\}$, there exists a covariant POVM $\{\overline{M}_{\hat{\phi}}\}$ such that for all states ρ_0,*

$$p_{\overline{M}}(\theta|\rho_0) = \bar{p}_M(\theta|\rho_0). \tag{32}$$

Proof. This result is well known [32], but we provide a proof here for completeness. Given POVM $\{M_{\hat{\phi}}\}$, we define the covariant POVM via

$$\overline{M}_0 := \frac{1}{2\pi}\int_{-\pi}^{\pi} d\phi\, e^{iG\phi}M_{\phi}e^{-iG\phi}. \tag{33}$$

Then we find

$$\begin{aligned} p_{\overline{M}}(\theta|\rho_0) &= \mathrm{Tr}(e^{-iG\theta}\overline{M}_0 e^{iG\theta}\rho_0) \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} d\phi\, \mathrm{Tr}(e^{-iG\theta}e^{iG\phi}M_{\phi}e^{-iG\phi}e^{iG\theta}\rho_0) \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} d\phi\, p_M(\phi|\phi-\theta) \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} d\phi\, p_M(\phi+\theta|\phi) = \bar{p}_M(\theta|\rho_0). \end{aligned} \tag{34}$$

In the last line we have shifted the variable of integration. This shows the relation (32) required. □

The second result, which is a combination of the two steps given in Ref. [22], is as follows.

Lemma 2. *Given any covariant POVM $\{\overline{M}_{\hat{\phi}}\}$ and state ρ_0, there exists a state without degeneracies $\rho_0^{(s)}$ such that the probability distribution of $G^{(s)}$ for $\rho_0^{(s)}$ is the same as that of G for ρ_0, and*

$$p_{C^{(s)}}(\theta|\rho_0^{(s)}) = p_{\overline{M}}(\theta|\rho_0). \tag{35}$$

Proof. Choose the state without degeneracies via

$$\rho_0^{(s)} = 2\pi \sum_{n,n'\in S} |n'\rangle\langle n| \sum_{d\le D(n),d'\le D(n')} \langle n',d'|\rho_0|n,d\rangle \times \langle n,d|\overline{M}_0|n',d'\rangle. \tag{36}$$

Note that $\rho_0^{(s)}$ is indeed a density operator, and yields the same distribution for $G^{(s)}$ as ρ_0 does for G. The details for how to prove these facts are given in Appendix A.

It can be shown that the average phase distributions are also the same, via

$$\begin{aligned}
p_{C^{(s)}}(\theta|\rho_0^{(s)}) &= \mathrm{Tr}(e^{-iG^{(s)}\theta}C_0^{(s)}e^{iG^{(s)}\theta}\rho_0^{(s)}) \\
&= \frac{1}{2\pi}\mathrm{Tr}\left(e^{-iG^{(s)}\theta}\sum_{m,m'\in S}|m\rangle\langle m'|e^{iG^{(s)}\theta}2\pi\sum_{n,n'\in S}|n'\rangle\langle n|\sum_{d\le D(n),d'\le D(n')}\langle n',d'|\rho_0|n,d\rangle\times\langle n,d|\overline{M}_0|n',d'\rangle\right) \\
&= \sum_{n,n'\in S} e^{i(n'-n)\theta}\sum_{d\le D(n),d'\le D(n')}\langle n',d'|\rho_0|n,d\rangle\langle n,d|\overline{M}_0|n',d'\rangle \\
&= \sum_{n,n'\in S}\sum_{d\le D(n),d'\le D(n')}\langle n',d'|\rho_0|n,d\rangle\langle n,d|e^{-iG\theta}\overline{M}_0e^{iG\theta}|n',d'\rangle \\
&= \sum_{n'\in S}\sum_{d'\le D(n')}\langle n',d'|\rho_0 e^{-iG\theta}\overline{M}_0e^{iG\theta}|n',d'\rangle = \mathrm{Tr}(e^{-iG\theta}\overline{M}_0e^{iG\theta}\rho_0) = p_{\overline{M}}(\theta|\rho_0).
\end{aligned} \tag{37}$$

This shows the relation (35) required. □

Using these lemmas then enables us to prove our theorem that the average distribution can always be obtained by a canonical measurement on a system without degeneracies.

Theorem 1. *Any bound on the concentration of the canonical phase distribution of a nondegenerate system with shift generator $G^{(s)}$, under some constraint $\mathcal{C}$ on the distribution of $G^{(s)}$, is also a bound on the concentration of the average phase distribution $\overline{p}(\theta)$ of an arbitrary phase estimate for any shift generator G having the same eigenvalue spectrum as $G^{(s)}$, providing that the probe state satisfies the same constraint $\mathcal{C}$ with respect to the distribution of G.*

Remarks. A measure of the concentration is a functional of the probability distribution, and includes the mean-square error, the Holevo variance, and the entropic length. For measures that are convex, such as the mean-square error and Holevo variance, lower bounds on the measure for the average distribution provide lower bounds on the average of that measure (see Sec. II A). By the distribution of G, we mean the probability distribution for the eigenvalues of G. Examples of constraints on the distribution of G are a fixed mean $\langle G\rangle$, an upper bound on the eigenvalues, or a fixed mean absolute value $\langle|G|\rangle$.

Proof. Consider any state ρ_0 that satisfies the constraint $\mathcal{C}$ on the distribution of G. Given an arbitrary measurement described by a POVM $\{M_{\hat{\phi}}\}$, we obtain an average phase distribution $\bar{p}(\theta)$. Using Lemma 1, we find that there exists a covariant POVM $\{\overline{M}_{\hat{\phi}}\}$ such that the same probability distribution is obtained with the same state, ρ_0. Next, using Lemma 2, there exists a state without degeneracy, $\rho_0^{(s)}$, such that the nondegenerate canonical measurement on $\rho_0^{(s)}$ produces the same phase distribution, and the distribution of $G^{(s)}$ is the same as the distribution of G for ρ_0.

Therefore, the distribution of $G^{(s)}$ for $\rho_0^{(s)}$ still satisfies the same constraint $\mathcal{C}$. Furthermore, because the probability distribution for the canonical measurement $p_{C^{(s)}}(\theta|\rho_0^{(s)})$ is equal to the average phase distribution $\bar{p}_M(\theta|\rho_0)$, any measure of the concentration of the probability distribution is unchanged. Because any value of the concentration that can be obtained for the arbitrary measurement under constraint $\mathcal{C}$ can also be obtained for the concentration of the canonical phase distribution under the same constraint, the arbitrary measurement must satisfy the same bound as the canonical measurement. □

B. Analytic bounds via an entropic uncertainty relation

It is possible to obtain a number of bounds by using entropic uncertainty relations. The entropic uncertainty relation for canonical phase measurements and a nondegenerate shift generator G is given by [39, 40]

$$H(\Theta) + H(G) \geq \ln 2\pi. \tag{38}$$

This can then be used to obtain bounds on the RAMSE [22]. In particular, combining Eqs. (24), (26) and (38) yields

$$\delta\hat{\Phi} \geq (2\pi e)^{-1/2} e^{H(\Theta)} \geq (2\pi/e)^{1/2} e^{-H(G)}. \tag{39}$$

We first specialise to the case where the eigenvalue spectrum S includes all nonnegative integers, so we denote the generator by N. The entropy for fixed mean number is maximised for the thermal (negative exponential) distribution. By a straightforward calculation, one can show that this results in the inequality

$$H(N) \leq \ln\langle N+1\rangle + \langle N\rangle \ln(1 + 1/\langle N\rangle). \tag{40}$$

Because $x\ln(1+1/x) < 1$, this yields (for finite expectation values)

$$H(N) < \ln\langle N+1\rangle + 1. \tag{41}$$

Substitution into Eq. (39) then gives

$$\delta\hat{\Phi} > \frac{k_A}{\langle N+1\rangle}, \tag{42}$$

where $k_A = \sqrt{2\pi/e^3} \approx 0.5593$ (defined in the Introduction). Using Theorem 1, this result holds for the RAMSE for all possible phase measurements, and for any shift generator with nonnegative integer eigenvalues. Recall that, because the MSE is a linear measure, the RMSE of the average distribution is equivalent to the RAMSE [see Eqs. (7) and (9)].

We can also use this result to infer the result in the more general case where there is some lower bound g on the eigenvalues of G. Then we can take $G = g\mathbb{1} + N$, so $\langle N\rangle = \langle G - g\rangle$. Then one obtains

$$\delta\hat{\Phi} > \frac{k_A}{\langle G - g + 1\rangle}. \tag{43}$$

Note that, for this result, it is not necessary for the spectrum S to include all integers above g. This is because, in minimising $\delta\hat{\Phi}$ for given $\langle G - g\rangle$, removing some integers restricts the possible states, and therefore can only increase the AMSE.

An alternative restriction that one may wish to consider is, instead of a fixed mean, a fixed mean of the absolute value, $\langle|G|\rangle$. This is of particular interest in the case of angular momentum, where $G = J$. Then fixed $\langle|J|\rangle$ corresponds to a mean absolute value of the angular momentum. The maximum entropy for fixed $\langle|G-g|\rangle$, where g is any real number, can be obtained by finding a critical point of the variational quantity

$$\Lambda = -\sum_{n\in S} p_n \ln p_n - \alpha \sum_{n\in S} p_n - \beta \sum_{n\in S} |n-g| p_n, \tag{44}$$

where α and β are variational parameters. As shown in Appendix B, this yields

$$H(G) < \ln(2\langle|G-g|\rangle + 1) + 1. \tag{45}$$

Substitution in Eq. (39) then gives

$$\delta\hat{\Phi} > \frac{k_A}{\langle 2|G-g|+1\rangle}. \tag{46}$$

Once again we note that this result holds both when S includes all integers, so $G = J$, and when S does not include all integers. In the latter cases, the maximum entropy distribution can not be obtained exactly, but it still provides a bound.

For a given state, one can adjust the value of g in order to maximise this lower bound. The optimal value is the median; that is, the value such that there is equal probability for eigenvalues above and below g.

Another restriction on the distribution that can be considered is a finite range of eigenvalues. For example, with number we have a minimum eigenvalue of 0, and can place an upper bound of $n_{\max}$ on the eigenvalues. Then the entropy is bound as

$$H(G) \leq \ln(n_{\max} + 1), \tag{47}$$

because the maximum entropy is for the flat distribution. Then, combining with (39) gives

$$\delta\hat{\Phi} > \frac{\sqrt{2\pi/e}}{n_{\max}+1}. \tag{48}$$

In the specific case of the Holevo variance, there is a well-known result for canonical measurements [33, 41],

$$\delta_H\hat{\Phi} \geq \tan\left(\frac{\pi}{n_{\max}+2}\right). \tag{49}$$

This result is achievable for arbitrary $n_{\max}$. Using our Theorem, this result also holds for the average distribution for arbitrary measurements. Furthermore, because the Holevo variance is a convex functional of the probability distribution, this bound holds for the root-mean value of the Holevo variance (averaging over phase shifts).

IV. OPTIMAL BOUNDS VIA NUMERICAL CALCULATIONS

The bound in Eq. (42) has a scaling constant of $k_A = \sqrt{2\pi/e^3} \approx 0.5593$. In contrast, based on the asymptotic result for Holevo variance [34, 35], we expect $\delta\hat{\Phi} \gtrsim k_C/\langle N\rangle$ with $k_C = 2(-z_A/3)^{3/2} \approx 1.3761$ for

large $\langle N\rangle$, where z_A is defined in Sec. II B. This indicates that the scaling constant of the bound in Eq. (42) is not optimal, and suggests the conjecture [22]

$$\delta\hat{\Phi} > \frac{k_C}{\langle N+1\rangle}. \tag{50}$$

In order to test this conjecture, we solved the variational problem to find the minimum value of the RAMSE or Holevo variance as a function of $\langle N\rangle$. The results supporting this conjecture are given in this section. In Sec. V the conjecture is proved analytically for the special case of the asymptotic limit $\langle N\rangle \to \infty$.

A. Holevo variance

The case of the Holevo variance is simplest, because the problem is to maximise $|\langle e^{i\Theta}\rangle|$. Given a state

$$|\psi\rangle = \sum_{n=0}^{\infty} \psi_n |n\rangle, \tag{51}$$

we have

$$|\langle e^{i\Theta}\rangle| = \left|\sum_{n=0}^{\infty} \psi_{n+1}\psi_n^*\right|. \tag{52}$$

Note that we can upper bound this expression via

$$\left|\sum_{n=0}^{\infty} \psi_{n+1}\psi_n^*\right| \leq \sum_{n=0}^{\infty} |\psi_{n+1}\psi_n|. \tag{53}$$

The normalisation and $\langle N\rangle$ are unaffected by replacing the coefficients ψ_n with their absolute values. Therefore, for maximisation of $|\langle e^{i\Theta}\rangle|$, we can always take ψ_n to be real and nonnegative.

From the above, the variational problem is thus to find a critical point of

$$\Lambda = \sum_{n=0}^{\infty} \left(\psi_n\psi_{n+1} - \alpha\psi_n^2 - \beta n\psi_n^2\right), \tag{54}$$

where α and β correspond to normalisation and mean photon number constraints. The variational condition $\partial\Lambda/\partial\psi_n = 0$ leads directly to the eigenvalue equation

$$\psi_{n-1} + \psi_{n+1} = 2(\alpha+\beta n)\psi_n \tag{55}$$

for $n \geq 1$, and $\psi_1 = 2(\alpha+\beta n)\psi_0$. To avoid the need to specify a different equation for $n = 0$, we can simply define $\psi_{-1} := 0$.

B. Root-mean-square error

The problem for minimising the RAMSE is somewhat more difficult, because we do not have a simple expression like Eq. (52). However, any well-behaved function (i.e., satisfying the Dirichlet conditions) can be expanded in a Fourier series on the interval $[-\pi, \pi]$ as

$$f(\theta) = \sum_{m=-\infty}^{\infty} z_m e^{im\theta}. \tag{56}$$

For $m \geq 0$, the expectation values of the exponentials are given by

$$\langle e^{im\Theta}\rangle = \sum_{n=0}^{\infty} \psi_{n+m}\psi_n^*. \tag{57}$$

For $m < 0$, the expectation values are just the complex conjugate of those for positive m.

Unlike the case of the Holevo variance, it is not obvious at first sight that we can take the state coefficients to be real. However, if $f(\theta)$ is real, and symmetric about $\theta = 0$, then $z_{-m} = z_m^* = z_m$. Therefore the expectation value of $f(\Theta)$ is given by

$$\langle f(\Theta)\rangle = \sum_{m,n=0}^{\infty} \psi_m^* Z_{mn}\psi_n \tag{58}$$

where Z is the real symmetric matrix with coefficients $Z_{mn} := z_{|m-n|}$.

The variational problem is then to find a critical point of

$$\begin{aligned}\Lambda &= \langle f(\Theta)\rangle - \alpha - \beta\langle N\rangle \\ &= \sum_{m,n=0}^{\infty} \psi_m^* \left[Z_{mn} - (\alpha+\beta n)\delta_{mn}\right]\psi_n.\end{aligned} \tag{59}$$

The variational condition leads to

$$\sum_{n=0}^{\infty} Z_{mn}\psi_n = (\alpha+\beta m)\psi_m. \tag{60}$$

This equation is solved as an eigenvalue equation with α as the eigenvalue. Because the corresponding matrix $Z - \beta N$ is real and symmetric in the number state basis, the eigenvectors are real in this basis (up to a global phase factor). This means that the state coefficients can indeed be taken to be real.

In the specific case of $f(\theta) = \theta^2$, the Fourier series is

$$\theta^2 = \frac{\pi^2}{3} + 4\sum_{m=1}^{\infty} \frac{(-1)^m}{m^2}\cos(m\theta). \tag{61}$$

We then obtain the eigenvalue equation

$$\left(\frac{\pi^2}{3} - \beta m\right)\psi_m + \sum_{\substack{n=-m \\ n\neq 0}}^{\infty} \frac{2(-1)^n}{n^2}\psi_n = \alpha\,\psi_m. \tag{62}$$

Numerical solution of this eigenvalue equation is difficult, because there are an infinite number of Fourier coefficients. The problem can be truncated at some maximum number, but solution still requires finding the eigenvalues of a full matrix. In contrast, the problem for the Holevo variance is sparse, and can therefore be solved much more efficiently.

C. Bounding the AMSE

As we are interested in testing a lower bound on the AMSE, we can alternatively use an expression with a finite number of Fourier coefficients, but that forms a lower bound on θ^2. One alternative is to use $f_1(\theta) := 2(1-\cos\theta)$, which is the same optimisation problem as for the Holevo variance. To show $f_1(\theta) \leq \theta^2$ on $[-\pi, \pi]$, we can use a Taylor expansion to third order with the Lagrange form of the remainder

$$f_1(\theta) = \theta^2 + \frac{f_1^{(3)}(\xi)}{3!}\theta^3 = \theta^2 - \frac{1}{3}\theta^3 \sin\xi, \qquad (63)$$

where $\xi \in [0, \theta]$. Because $\sin\xi$ has the same sign as θ^3, the remainder term is negative, and $f_1(\theta) \leq \theta^2$.

The drawback to this alternative is that it yields results that do not satisfy the conjectured lower bound. We therefore use a higher-order approximation given by

$$f_2(\theta) := \frac{5}{2} - \frac{8}{3}\cos\theta + \frac{1}{6}\cos 2\theta. \qquad (64)$$

Again expanding in a Taylor series,

$$f_2(\theta) = \theta^2 + \frac{f_2^{(3)}(\xi)}{3!}\theta^3 = \theta^2 - \frac{8}{3}\theta^3(1-\cos\xi)\sin\xi \leq \theta^2. \qquad (65)$$

We can also obtain an upper bound using (see Appendix C)

$$f_3(\theta) := (\pi^2/4 - 1)[2(1-\cos\theta) - (1-\cos 2\theta)/2] + 2(1-\cos\theta). \qquad (66)$$

In the following we will use $(\delta_m\hat{\Phi})^2 := \langle f_m(\Theta)\rangle$ for $m \in \{1, 2, 3\}$.

D. Numerical results

The minimal Holevo variance, as well as the minimal values of $\langle\Theta^2\rangle$ and $\delta_2\hat{\Phi}$, have been determined by numerically solving the eigenvalue equations. In each case, a number cutoff was used that was about 10 times the value of $\langle N\rangle$, or 100 for small $\langle N\rangle$. At this point the magnitude of the state coefficients had fallen to less than $1/10^6$ of the maximum value, and increasing the cutoff beyond this did not alter the results by more than 1 part in 10^6. For the results for $\langle\Theta^2\rangle$, the maximum $\langle N\rangle$ was about 5000, due to the difficulty in finding eigenvalues of a full matrix. In contrast, for the Holevo variance and for $\delta_2\hat{\Phi}$, the maximum $\langle N\rangle$ was over 10^6.

The results for the Holevo variance are given in Fig. 1. In this figure the square root of the Holevo variance is plotted multiplied by $\langle N+1\rangle$. Therefore, if $k_C/\langle N+1\rangle$ provides a lower bound to the RAMSE, the curve should be above k_C (also shown in the figures). It is clear from the figure that the numerical results indicate that

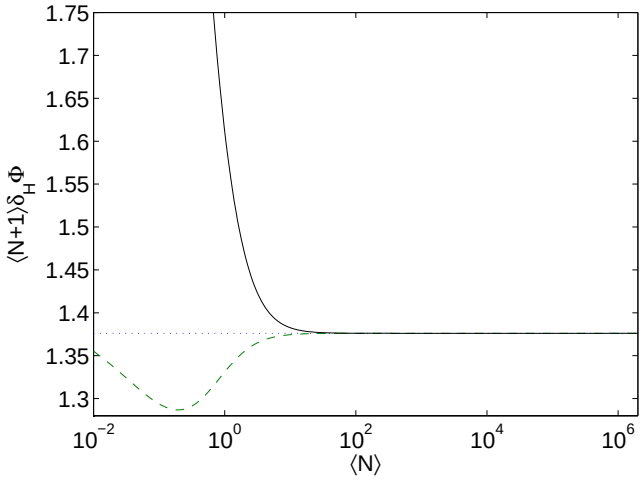

FIG. 1: Minimum possible value of $\langle N+1\rangle\delta_H\hat{\Phi}$, plotted as a function of $\langle N\rangle$ (solid curve). The case where $\delta_1\Phi$ is used instead of $\delta_H\hat{\Phi}$ is shown as the dashed curve (green). The asymptotic value of $k_C \approx 1.3761$ is shown as the horizontal dotted line (blue).

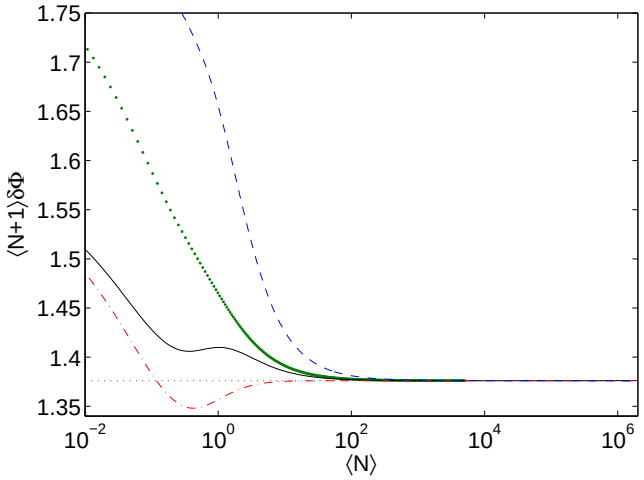

FIG. 2: Minimum possible value of $\langle N+1\rangle\delta\hat{\Phi}$, plotted as a function of $\langle N\rangle$. The dotted curve (green) shows the values obtained for $\delta\hat{\Phi}$ (i.e., the RAMSE). The solid curve (black) uses $\delta_2\hat{\Phi}$ instead of $\delta\hat{\Phi}$. The dash-dotted curve (red) is the lower bound using $\arccos[1-(\delta_1\hat{\Phi})^2/2]$. The dashed curve (blue) is the upper bound using $\delta_3\hat{\Phi}$ calculated for the state that minimises $\delta_1\hat{\Phi}$. The asymptotic value of $k_C \approx 1.3761$ is again shown as the horizontal dotted line (blue).

$k_C/\langle N+1\rangle$ provides a strict lower bound to $\delta_H\hat{\Phi}$. In this figure $\delta_1\hat{\Phi}$ is also shown, and $\delta_1\hat{\Phi} < k_C/\langle N+1\rangle$ in the range shown.

The results calculated for $\delta\hat{\Phi}$ are shown in Fig. 2. It can be seen that these results are also above the line for k_C, indicating that $\delta\hat{\Phi} > k_C/\langle N+1\rangle$. One would like to provide more easily calculated lower bounds on $\delta\hat{\Phi}$ to test this inequality more thoroughly. It is clear that $\delta_1\hat{\Phi}$ is not useful for this purpose, because the curve in Fig. 1

is below k_C. It is also possible to obtain a tighter lower bound on $\delta\hat{\Phi}$ using $\delta_1\hat{\Phi}$ (see Appendix C), but this curve is still not above k_C for all $\langle N\rangle$.

A better lower bound to $\delta\hat{\Phi}$ is $\delta_2\hat{\Phi}$, which is also shown in Fig. 2, and is above the k_C line. This quantity can be calculated more rapidly and reliably than $\delta\hat{\Phi}$, and results are given up to $\langle N\rangle \approx 2\times 10^6$. This provides further numerical evidence that $\delta\hat{\Phi} > k_C/\langle N+1\rangle$. Results were also calculated for $\langle N\rangle$ down to about 10^{-6}. These are not shown in the figures, but the curves that are above k_C do not cross below k_C.

E. Angular momentum calculations

We have also calculated the corresponding results with a fixed value of $\langle|J|\rangle$. The variational problem is exactly the same as before, except now we sum over positive and negative values of j (as opposed to n), and replace n with $|j|$. That is, the variational problem is to find a critical point of

$$\Lambda = \langle f(\Theta)\rangle - \alpha - \beta\langle|J|\rangle. \tag{67}$$

As before, for a real function f symmetric about zero we can assume that the state coefficients are real, so the variational condition yields

$$\sum_{m=-\infty}^{\infty} a_m\psi_{j+m} = (\alpha+\beta j)\psi_j. \tag{68}$$

In the case of $f(\theta)=\theta^2$, we obtain the eigenvalue equation

$$\frac{\pi^2}{3}\psi_j + \sum_{\substack{m=-\infty\\ m\neq 0}}^{\infty}\frac{2(-1)^m}{m^2}\psi_{j+m} = (\alpha+\beta|j|)\psi_j. \tag{69}$$

The eigenvalue equation for the case of $f_1(\theta)$ is

$$\psi_{j-1}+\psi_{j+1} = 2(\alpha+\beta|j|)\psi_j. \tag{70}$$

We will not consider $f_2(\theta)$ for this problem.

The results for $\delta\Phi$, $\delta_H\Phi$, and $\delta_1\Phi$ were all determined numerically, and the results are shown in Fig. 3. It will be shown in the next section that the asymptotic optimal value for $\delta_1\hat{\Phi}$ is

$$\delta_1\hat{\Phi} \sim \frac{k'_C}{\langle 2|J|+1\rangle}, \tag{71}$$

with $k'_C = 4(-z'_A/3)^{3/2} \approx 0.7916$, where z'_A is the first zero of the derivative of the Airy function. We have therefore plotted the results for $\delta\hat{\Phi}$ multiplied by $\langle 2|J|+1\rangle$ in Fig. 3. It can be seen in this figure that all the results are above k'_C, supporting the conjecture that there is strict inequality with the scaling constant k'_C.

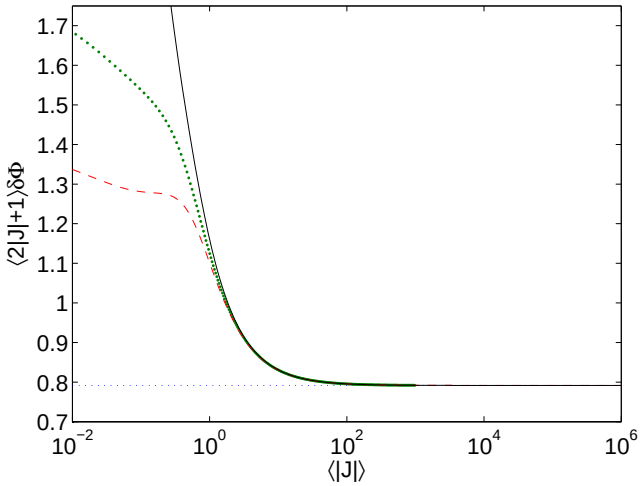

FIG. 3: Minimum possible value of $\langle 2|J|+1\rangle\delta\hat{\Phi}$ is plotted as a function of $\langle N\rangle$ as the dotted curve (green). The minimum value of $\langle 2|J|+1\rangle\delta_H\hat{\Phi}$ is shown as the solid curve (black), and $\langle 2|J|+1\rangle\delta_1\hat{\Phi}$ is shown as the dashed curve (red). The asymptotic value of $k'_C \approx 0.7916$ is shown as the horizontal dotted line (blue).

V. ASYMPTOTIC EXPANSIONS

A. Holevo variance

In the specific case of the Holevo variance, it is possible to obtain analytic results in terms of Bessel functions to provide further support to the conjecture that $\delta_H\hat{\Phi} > k_C/\langle N+1\rangle$. The recurrence relation (55) has a known solution in terms of Bessel functions [35]. Bessel functions of the first kind satisfy the recurrence relation $J_{k-1}(z)+J_{k+1}(z) = (2k/z)J_k(z)$. Therefore the solution is of the form

$$\psi_n(x,z) = AJ_{x+n+1}(z), \tag{72}$$

with $x := \alpha/\beta - 1$, $z := 1/\beta$. Bessel functions of the second kind can be ignored, because they diverge for large values of the order. The condition that $\psi_{-1}=0$ implies the restriction

$$J_x(z) = 0 \tag{73}$$

on the parameter z, thus confining its allowed values to the (countable) set of zeroes of J_x.

To obtain the smallest Holevo variance for a given mean photon number, we wish to take the solution for the *largest* value of α. This corresponds to the largest solution of Eq. (73) in terms of x for given z. Conversely, for given x we want the first positive zero of J_x. The normalisation constraint yields

$$A^{-2} = \sum_{n=0}^{\infty}[J_{x+n+1}(z)]^2 = \sum_{k=1}^{\infty}[J_{x+k}(z)]^2, \tag{74}$$

and hence one has

$$\langle N\rangle = A^2\sum_{n=0}^{\infty} n\,[J_{x+n+1}(z)]^2 = A^2\sum_{k=1}^{\infty}(k-1)[J_{x+k}(z)]^2$$
$$= \frac{\sum_{k=1}^{\infty} k\,[J_{x+k}(z)]^2}{\sum_{k=1}^{\infty}[J_{x+k}(z)]^2} - 1, \tag{75}$$

$$\langle e^{i\Theta}\rangle = A^2\sum_{n=0}^{\infty} J_{x+n+1}(z)\,J_{x+n+2}(z)$$
$$= \frac{\sum_{k=1}^{\infty} J_{x+k}(z)\,J_{x+k+1}(z)}{\sum_{k=1}^{\infty}[J_{x+k}(z)]^2}. \tag{76}$$

Using Eq. (55), we have

$$\langle e^{i\Theta}\rangle = (\alpha+\beta\langle N\rangle) = (x+\langle N\rangle+1)/z. \tag{77}$$

Up until this point, these results for the Bessel functions are the same as those of Ref. [35]. Reference [35] then uses an approximation in terms of Airy functions. We have determined more accurate results using formulae for sums of Bessel functions (see Appendix D). We find that

$$|\langle e^{i\Theta}\rangle|^{-2} - 1 = \sum_{k=1}^{5}\frac{b_{2k}}{\langle N+1\rangle^{2k}} + O\left(\frac{1}{\langle N+1\rangle^{12}}\right), \tag{78}$$

where $b_2 = k_C^2$, and b_2 to b_{10} are all positive and close to 2. The fact that each b_j that has been calculated is positive strongly supports the conjecture that the Holevo variance is strictly lower bounded by the first term.

B. Upper bounding the optimal mean-square error

It would be desirable to obtain a similar approximation for the exact RAMSE $\delta\Phi$. However, the eigenvalue equation does not have any solution in terms of elementary functions that we have been able to find. Even the lower bounding quantity $\delta_2\Phi$ yields an eigenvalue equation that does not appear to have an analytic solution. However, we can place an upper bound on the optimal value of $\delta\Phi$, using

$$\theta^2 \le f_3(\theta), \tag{79}$$

for $\theta\in[-\pi,\pi]$. We can calculate $\delta_3\hat{\Phi}$, except for the state that minimises $\delta_1\hat{\Phi}$. This value is shown in Fig. 2 for comparison with $\delta\hat{\Phi}$.

Using the properties of Bessel functions, this leads to the result that the optimal value of $\delta\hat{\Phi}$ satisfies (see Appendix D)

$$(\delta\hat{\Phi})^2 \le \frac{k_C^2}{\langle N+1\rangle^2} + O\left(\frac{1}{\langle N+1\rangle^3}\right). \tag{80}$$

This means that, asymptotically, the optimal value of $\delta\hat{\Phi}$ cannot be larger than $k_C/\langle N+1\rangle$. Because $\delta\hat{\Phi}$ cannot be smaller than $\delta_H\hat{\Phi}$ except for higher-order terms [see Eq. (21)], this means the optimal $\delta\hat{\Phi}$ must be asymptotically equal to $k_C/\langle N+1\rangle$ [i.e., k_C is the largest value of k for which Eq. (1) can be true].

C. Angular momentum calculations

Next we consider the problem with fixed $\langle|J|\rangle$. Recall that the variational problem yields an eigenvalue problem given in Eq. (70). This is solved by taking [42]

$$\psi_j(x,z) = A_1\,J_{x+j}(z), \tag{81}$$

for $j\ge 0$, and

$$\psi_j(x,z) = A_2\,J_{x-j}(z), \tag{82}$$

for $j\le 0$. In this case we take $x := \alpha/\beta$, $z := 1/\beta$. We again may ignore Bessel functions of the second kind, because they diverge. The restriction that the solutions coincide for $n=0$ means that $A_1 = A_2 = A$, and

$$\psi_j(x,z) = A\,J_{x+|j|}(z), \tag{83}$$

for all j. The condition that the recurrence relation holds for $j=0$ means that

$$J_{x+1}(z) = \frac{x}{z}J_x(z) = \frac{1}{2}[J_{x-1}(z)+J_{x+1}(z)]. \tag{84}$$

This implies

$$J_{x-1}(z) - J_{x+1}(z) = 0. \tag{85}$$

Then, using $[J_{x-1}(z)-J_{x+1}(z)]/2 = J_x'(z)$, this means we must have $J_x'(z)=0$.

Performing series expansions similar to that for the first case, gives (see Appendix E)

$$2(1-|\langle e^{i\Theta}\rangle|) = \sum_{k=2}^{9}\frac{d_k}{\langle 2|J|+1\rangle^k} + O\left(\frac{1}{\langle 2|J|+1\rangle^{10}}\right), \tag{86}$$

where $d_2 = {k'_C}^2$, and coefficients up to d_5 are positive, but d_6 is negative. This strongly supports the numerical results that the strict inequality

$$\delta_1\hat{\Phi} \ge \frac{k'_C}{\langle 2|J|+1\rangle}, \tag{87}$$

holds. In turn, because $\delta\hat{\Phi}\ge\delta_1\hat{\Phi}$, this also supports the conjecture that the inequality holds for $\delta\hat{\Phi}$. In addition, $\delta_H\hat{\Phi}\ge\delta_1\hat{\Phi}$, so this supports the conjecture that the inequality holds for $\delta_H\hat{\Phi}$.

Similarly to the case for fixed $\langle N\rangle$, one can use Eq. (79) to find a series expansion for an upper bound on the optimal value of $\delta\hat{\Phi}$, giving

$$(\delta\hat{\Phi})^2 \le \frac{{k'_C}^2}{\langle 2|J|+1\rangle^2} + O\left(\frac{1}{\langle 2|J|+1\rangle^3}\right). \tag{88}$$

This means that we have upper and lower bounds on the optimal $\delta\hat{\Phi}$, showing that it is asymptotically equal to $k'_C/\langle 2|J|+1\rangle$.

VI. SCALING WITH NUMBER OF PROBE STATES

Another question is what the scaling of the lower bound is if there are m identical probe states. Normally it is expected that the MSE will scale like $1/\sqrt{m}$ if there are m copies of the state. This is because, for estimates formed by the average of the individual estimates, the standard error scales as $1/\sqrt{m}$. Similarly, the Cramér-Rao bound for m identical probe states yields the following bound for estimates that are unbiased (in the standard statistical sense, not what we have called $U(1)$-unbiased in Sec. II B) [32, 43]:

$$\Delta_\phi \hat{\Phi} \geq \frac{1}{2\sqrt{m}\Delta N}. \tag{89}$$

We will call this the Helstrom-Holevo bound.

Because of these results one might expect that one could derive a lower bound to the uncertainty in terms of $\langle N\rangle$ of the form $k/(\sqrt{m}\langle N+1\rangle)$. On the other hand, directly using the above methods yields a lower bound of

$$\delta\hat{\Phi} \geq \frac{k}{\langle mN+1\rangle}, \tag{90}$$

because the overall average number is $m\langle N\rangle$. Recall that we have proven this inequality for $k = k_A$, and have extremely strong numerical evidence for the inequality for $k = k_C$.

We can prove that there is no lower bound scaling as $1/(\sqrt{m}\langle N+1\rangle)$ in the following way. Let $m \in \mathbb{N}$, $\langle N\rangle$, and $\delta > 0$ be given. We use μ for the required value of $\langle N\rangle$, to avoid confusion with intermediate states we use in this discussion with different values of $\langle N\rangle$.

Let $|\chi_{n-1}\rangle$ be the state with the minimum phase uncertainty for mean number $\langle N\rangle = n-1$, and let $|\chi'_{n-1}\rangle$ be the corresponding state with the same amplitudes, but shifted up by one. This means that there is no vacuum component, the phase uncertainty is unchanged, and $\langle N\rangle = n$. We are considering small δ and large m, so we expect that $\mu^\delta \leq m$. In that case, we take $n = (m\mu)^{1/(1+\delta)}$, and consider m copies of

$$|\psi\rangle = \sqrt{1-\mu/n}|0\rangle + \sqrt{\mu/n}|\chi'_{n-1}\rangle. \tag{91}$$

For this state, $\langle N\rangle = \mu$. Now consider a phase measurement that first distinguishes between $|0\rangle$ and $|\chi'_{n-1}\rangle$ on all copies of the state. If the $|\chi'_{n-1}\rangle$ result is found, then a canonical phase measurement is performed.

The probability of getting the $|\chi'_{n-1}\rangle$ result is μ/n. For m repetitions, the probability of projecting every single copy onto the state $|0\rangle$ is $(1-\mu/n)^m \leq \exp(-m\mu/n) = \exp(-(m\mu)^{\delta/(1+\delta)})$. This probability scales exponentially in $m\mu$ and may be ignored for asymptotically large $m\mu$. The phase uncertainty is therefore (up to an exponentially small correction) no more than that for $|\chi'_{n-1}\rangle$, which is

$$\begin{aligned}\delta\hat{\Phi} &= k_C/n + O(1/n^2)\\ &= k_C/(m\mu)^{1/(1+\delta)} + O(1/(m\mu)^{2/(1+\delta)}).\end{aligned} \tag{92}$$

For $\mu^\delta > m$, we can just take $n = \mu$, and $|\psi\rangle = |\chi'_{n-1}\rangle$. In this case we have $1/(m\mu)^{1/(1+\delta)} \geq 1/\mu$. Therefore, considering just the uncertainty for a single copy of the state gives

$$\begin{aligned}\delta\hat{\Phi} &= k_C/\mu + O(1/\mu^2)\\ &< k_C/(m\mu)^{1/(1+\delta)} + O(1/(m\mu)^{2/(1+\delta)}).\end{aligned} \tag{93}$$

This provides an upper bound to the uncertainty for m copies of the state.

Therefore, we find that, for any $\delta > 0$, $m \in \mathbb{N}$ and $\mu = \langle N\rangle$, we can find a state such that the uncertainty is no greater than $k_C/(m\mu)^{1/(1+\delta)}$ to leading order. Because we can choose any $\delta > 0$, this means that, for fixed $\langle N\rangle$, the lower bound to the scaling must be arbitrarily close to $1/m$.

This result is counterintuitive, because for a state that does not depend on m, the uncertainty can be expected to scale as $1/\sqrt{m}$, similarly to the Helstrom-Holevo bound (89). However, the Helstom-Holevo bound, in terms of m and ΔN, holds even for states that depend on m. Similarly, a bound in terms of m and $\langle N\rangle$ must hold for states that are chosen based on m. We have shown that the potential dependence of states upon m means that it is not possible to obtain a universal bound that scales as $1/\sqrt{m}$ for given $\langle N\rangle$.

VII. PAPERS CLAIMING VIOLATION OF HEISENBERG LIMIT

In the following, we present some recent measurement schemes claiming violation of the Heisenberg limit. We summarise the techniques used in these schemes, and explain why they appear to violate the Heisenberg limit. We argue that the accuracy of these super-Heisenberg measurements should be considered illusory, primarily because they only work for a very restricted range of phase.

A. Anisimov et al.

Anisimov *et al.* [12] describe a noncovariant phase estimation method having a minimum RMSE

$$\Delta^0_\phi \hat{\Phi} = \frac{1}{[\langle N\rangle(\langle N\rangle+2)]^{1/2}}. \tag{94}$$

This quantity is for a particular phase shift, as opposed to the average over the phase shift, $\delta\hat{\Phi}$. Also, the RMSE is here using a reference phase of 0, rather than the reference phase of ϕ that we use (see Sec. II A). This result violates an alternative definition of the Heisenberg limit, given by Anisimov *et al.* as [12]

$$\Delta^0_\phi \hat{\Phi} \geq 1/\langle N\rangle. \tag{95}$$

First, it should be noted that this does not give a different power of $\langle N\rangle$, and does not change the scaling constant.

It only violates this form of the Heisenberg limit by an amount which is significant for small $\langle N\rangle$, and is of higher order for large $\langle N\rangle$. In later work [44], they have modified their claim to that of achieving the Heisenberg limit.

In fact, it is easy to see that the above form (95) of the Heisenberg limit cannot be a strict limit for small $\langle N\rangle$. For any $\langle N\rangle$ less than $1/\pi$ it must be violated, because the maximum RMSE possible is π^2. For the same reason, any bound of the form $k/\langle N\rangle$ cannot hold for all $\langle N\rangle$. It is for this reason that we have used $\langle N+1\rangle$ (or $\langle G+1\rangle$ more generally).

Note from Eq. (94) that the minimum RMSE satisfies

$$\Delta^0_\phi\hat{\Phi} > \frac{1}{[\langle N\rangle(\langle N\rangle+2)+1]^{1/2}} = \frac{1}{\langle N+1\rangle}. \tag{96}$$

Hence, the RAMSE, $\delta\hat{\Phi}$, trivially satisfies our analytical lower bound (42). However, the minimum value is below our conjectured best possible bound (50), by a factor of $k_C \approx 1.3761$ in the asymptotic limit. This does not contradict our conjectured bound, because the conjectured bound is for $\delta\hat{\Phi}$, whereas the above value is for a specific value of the phase shift. This is an important aspect of our result. It is possible to obtain smaller errors for specific values of the phase shift [23–27], but not when the average is taken over the phase. That is, the noncovariance of the scheme in Ref. [12] does allow beating our conjectured bound, in a small range of phase shifts about $\phi = 0$, but only at the expense of worse phase resolution over the remainder of possible phase shift values.

B. Zhang et al.

Zhang *et al.* [13] propose a superposition state with arbitrarily high phase sensitivity but finite average photon number. They consider a two-mode Mach-Zehnder interferometer (MZI) system, with a probe state of the form

$$|\psi\rangle := \sum_{n\geq 1} c_n|\psi_n\rangle,$$
$$|\psi_n\rangle := \frac{1}{\sqrt{2}}\left[|n,0\rangle + |0,n\rangle\right]. \tag{97}$$

That is, $|\psi\rangle$ is a superposition of (mutually orthogonal) NOON states.

They use the quantum Cramér-Rao bound (QCRB) to derive the ultimate limit to the uncertainty of phase measurement as

$$\Delta^0_\phi\hat{\Phi} \geq \frac{1}{\sqrt{\langle N^2\rangle}}, \tag{98}$$

in contrast to Eq. (95). Also $N = N_a + N_b$ is the total photon number operator for the two modes a and b, rather than just the number operator N_a for the mode passing through the phase shift. They call Eq. (98) the "proper" Heisenberg limit, and Eq. (95) the "generally accepted form" of the Heisenberg limit. This result is similar to the result given in a number of other works [45, 46]. By choosing $c_n \propto n^{-3/2}$, they obtain $\langle N\rangle < \infty$ and $\langle N^2\rangle = \infty$, which gives $\Delta^0_\phi\hat{\Phi} \geq 0$. They further claim in Sec. V of Ref. [13] that this lower bound is achievable (i.e., that the uncertainty can be zero for finite $\langle N\rangle$).

An interesting feature of their result is that the Fisher information can be infinite for finite $\langle N\rangle$. Therefore, it should not be expected that the QCRB can give a nontrivial lower bound on the uncertainty for fixed $\langle N\rangle$. Furthermore, the Fisher information is infinite for all ϕ.

However, there are some problems with the result presented. First, they give no proof that the lower bound provided by the QCRB is achievable. In many cases Fisher's theorem [47] allows the QCRB to be achieved asymptotically (i.e., with a scaling constant of $1/\sqrt{m}$ for m probe states). However, Fisher's theorem is not universally applicable, because it requires a unique maximally likely estimate [48]. In contrast, here the measurements will yield multiple maximally likely estimates.

Second, the form of the QCRB given is for unbiased measurements, but it is unclear how to perform an unbiased measurement here. For biased measurements this lower bound does not hold. In fact, the obvious measurement technique is biased, and will only yield zero error for $\phi = 0$ and π, similar to the example in Sec. VIII C. This can be achieved with a very simple choice of state.

However, measurements that yield zero error only for isolated values of ϕ will not be useful. Further, based on the results presented here, the *average* performance of any two-mode MZI estimate must satisfy

$$\delta\hat{\Phi} \geq \frac{k_A}{\langle N_a+1\rangle}, \tag{99}$$

as a consequence of Eq. (42).

C. Rivas and Luis

Rivas and Luis [14] consider a linear phase estimation procedure that employs as the probe state the coherent superposition

$$|\psi\rangle = \mu|0\rangle + \nu|\xi\rangle \tag{100}$$

of the vacuum $|0\rangle$ and a squeezed state $|\xi\rangle$. The authors consider the case with $\nu \ll 1$, $\mu \simeq 1$ and also assume that the phase shift is known to be small: $\phi \ll 1$. The fixed mean photon number of the probe state is then given by

$$\langle N\rangle = \nu^2\bar{n}_\xi\,, \tag{101}$$

where $\bar{n}_\xi$ is the (average) number of photons in the squeezed state. Using conventional error propagation arguments, they find for this state

$$(\Delta^0_\phi\hat{\Phi})^2 \geq \frac{\nu^2}{4m\langle N\rangle^2}, \tag{102}$$

where m is the number of repetitions of the measurement. The lower bound here is arbitrarily below the usual Heisenberg limit by a factor $O(\nu^2)$.

We note that similar results can be obtained in a simplified scenario by employing the probe state

$$|\psi\rangle = \mu|0\rangle + \nu|\bar{n}_\xi\rangle\,, \tag{103}$$

where $|\bar{n}_\xi\rangle$ denotes a number state with $\bar{n}_\xi$ photons. The interference fringes obtained from this state are high frequency, but low visibility. A calculation using the error propagation formula based on the observable $X = |0\rangle\langle\bar{n}_\xi| + |\bar{n}_\xi\rangle\langle 0|$ yields

$$(\Delta^0_\phi\hat{\Phi})^2 = \frac{(\delta X)^2}{|d\langle X\rangle/d\phi|^2} \approx \frac{\nu^2}{4\langle N\rangle^2}\,. \tag{104}$$

Taking $\nu \propto \langle N\rangle^{1-p}$ gives

$$\Delta^0_\phi\hat{\Phi} \propto \frac{1}{\langle N\rangle^p}, \tag{105}$$

which, in principle, gives an accuracy that scales arbitrarily well with $\langle N\rangle$ (for large p).

The problem with this scheme is that the high accuracy predicted by the error propagation formula is given by high frequency fringes with low visibility. It would take a great deal of additional phase information to resolve the ambiguity in the fringes, as well as many repetitions of the measurement to obtain a reasonable estimate of the observable X so that the error propagation formula would become accurate.

The scheme presented in Ref. [14] is a little more complicated (including an analysis of the efficiency), but similar considerations apply. A quadrature measurement is considered for a fixed phase, which means that the analysis essentially gives an estimate of the uncertainty for a given value of the phase. As we have noted above, it is possible to obtain higher accuracy for a particular value of the phase shift. For example, it is trivial to design a measurement that gives zero error for a single value of the phase shift. The bound (42) must hold when averaging over the phase shift.

D. Nonlinear interferometry

A qualitatively different type of proposal for beating the Heisenberg limit is that based on nonlinear interferometry [15, 16]. The basis of these proposals is that the generator of the phase shifts is nonlinear in the number operator. For example, $G = N^q$ for some $q > 1$. It is then found that the phase uncertainty can scale as $1/\langle N\rangle^q$. Subtleties involved in achieving such scalings are discussed in Ref. [27].

These proposals do not contradict the results presented here; they are just using the terminology differently [7]. In Refs. [15, 16], the Heisenberg limit is given as $1/\langle N\rangle$, where N is the number of particles. In contrast, here we give the Heisenberg limit in terms of the generator of the phase shifts. That is, the bound is

$$\delta\hat{\Phi} \geq \frac{k}{\langle G+1\rangle} = \frac{k}{\langle N^q+1\rangle}, \tag{106}$$

which typically scales as $k/\langle N\rangle^q$. Therefore the results do not violate the Heisenberg limit (106) given here. In Refs. [15, 16], they call this limit the "quantum limit", rather than the Heisenberg limit.

VIII. LIMITATIONS OF THE CRAMÉR-RAO BOUND

The Cramér-Rao bound for the RMSE $\Delta^0_\phi\hat{\Phi}$ is often used as motivation for the Heisenberg limit, but it has limitations which mean that it does not provide a rigorous basis for the Heisenberg limit. There are a number of different variations of the way the Cramér-Rao bound is used. First, the classical Cramér-Rao bound (CRB), $1/\sqrt{mF_C(\phi)}$, is in terms of the classical Fisher information $F_C(\phi)$ of a specific probability distribution, so in quantum mechanics it is calculated for a given state and measurement.

Second, the *quantum* Cramér-Rao bound (QCRB) replaces $F_C(\phi)$ by the quantum Fisher information, $F_Q(\phi)$ (corresponding to the classical Fisher information optimised over all quantum measurements), but is still calculated for a given state [49]. Third, the Helstrom-Holevo bound (HHB), as in Eq. (89), is optimised over both the quantum measurement and the quantum state, with the optimisation being for a given ΔN. Because these bounds use successively more optimisation, one has the ordering $CRB \geq QCRB \geq HHB$. In particular, for any estimate that is *unbiased* for phase shift ϕ, one has

$$\Delta^0_\phi\hat{\Phi} \geq \frac{1}{\sqrt{mF_C(\phi)}} \geq \frac{1}{\sqrt{mF_Q(\phi)}} \geq \frac{1}{2\sqrt{m}\Delta N}. \tag{107}$$

The most obvious limitation in using the HHB is that it is a limit in terms of ΔN, whereas the Heisenberg limit is in terms of $\langle N\rangle$. This means that, for states with large uncertainty in N as compared to the mean value, the HHB does not imply the Heisenberg limit. This is taken advantage of in Refs. [13, 14].

A fixed value of ΔN is just a choice of constraint. One could also consider optimisation for fixed $\langle N\rangle$, as a method to obtain the Heisenberg limit. However, it is easily seen that there is no upper bound on the Fisher information for a given $\langle N\rangle$. In the example of Zhang *et al.*, they find a state with infinite Fisher information for finite $\langle N\rangle$ (see Sec. VII B). Note also that if there were such a bound, then the CRB would imply a $1/\sqrt{m}$ scaling for fixed $\langle N\rangle$, whereas we have found that such scaling is impossible (see Sec. VI). The difficulty of using the CRB was also noted in Ref. [28].

A. Bias in phase estimation

Another major factor that needs to be taken into account when considering the CRB and related bounds is that of bias. Note, for example, that the value of ΔN in Eq. (107) can be arbitrarily small, whereas the RMSE cannot be larger than π. It follows that any phase estimate must be biased for sufficiently small ΔN. In fact, one can show that covariant phase measurements cannot be unbiased for every phase shift value, in the sense needed for the QCRB and HHB.

In particular, when considering the RMSE with reference phase ϕ_r, $\Delta_\phi^{\phi_r}\hat{\Phi}$, one needs to define the bias function

$$b_{\phi_r}(\phi) := \langle\hat{\Phi}\rangle_\phi^{\phi_r} - \phi, \tag{108}$$

with

$$\langle\hat{\Phi}\rangle_\phi^{\phi_r} := \int_{\phi_r-\pi}^{\phi_r+\pi} d\hat{\phi}\,\hat{\phi}\,p(\hat{\phi}|\phi). \tag{109}$$

Then the CRB with bias is [50]

$$(\Delta_\phi^{\phi_r}\hat{\Phi})^2 \geq \frac{[1+b'_{\phi_r}(\phi)]^2}{mF_C(\phi)} + b_{\phi_r}(\phi)^2. \tag{110}$$

The QCRB and HHB in Eq. (107) similarly generalise (see also [51]).

If one is to use the form of the CRB *without* bias, then one needs $b_{\phi_r}(\phi) = 0$ and $b'_{\phi_r}(\phi) = 0$. This is highly problematic if one is to consider the full range of values of ϕ with a fixed reference phase ϕ_r. This is because $\langle\hat{\Phi}\rangle_\phi^{\phi_r}$ would need to change discontinuously at $\phi = \phi_r + \pi$. But, for finite $\langle N\rangle$, it is easily shown that $\langle\hat{\Phi}\rangle_\phi^{\phi_r}$ is a continuous function of ϕ (see Appendix F). Therefore it is not possible for the phase to be globally unbiased unless $\langle N\rangle$ is infinite. Moreover, there must be a region of size scaling as $1/\langle N\rangle$ where the measurement is biased [this follows from Eq. (F14)].

On the other hand, one can consider applying the CRB to $\Delta_\phi\hat{\Phi}$ in Eq. (3); that is, to the RMSE modulo $(-\pi,\pi]$. Because $\Delta_\phi\hat{\Phi} \leq \Delta_\phi^0\hat{\Phi}$ (see Sec. II A), using the CRB to bound $\Delta_\phi\hat{\Phi}$ also yields a bound on $\Delta_\phi^0\hat{\Phi}$. Also, because $\Delta_\phi\hat{\Phi} \equiv \Delta_\phi^\phi\hat{\Phi}$, the conditions for the measurement to be unbiased become $b_\phi(\phi) = 0$ and $b'_\phi(\phi) = 0$. It is important to note that $b'_\phi(\phi)$ is *not* the same as $\frac{d}{d\phi}b_\phi(\phi)$. In fact, the restriction $\frac{d}{d\phi}b_\phi(\phi) = 0$ implies

$$\begin{aligned} 0 &= \frac{d}{d\phi_r}b_{\phi_r}(\phi)\bigg|_{\phi_r=\phi} + b'_\phi(\phi) \\ &= 2\pi p(\phi+\pi|\phi) + b'_\phi(\phi). \end{aligned} \tag{111}$$

That is, if $b_\phi(\phi) = 0$, then $\frac{d}{d\phi}b_\phi(\phi)$ will automatically be zero, but $b'_\phi(\phi)$ will only be zero if $p(\phi+\pi|\phi) = 0$. In fact, for $b_\phi(\phi) = 0$, the condition $b'_\phi(\phi) = 0$ is equivalent to $p(\phi+\pi|\phi) = 0$.

The conditions for the measurement to be unbiased (when applying the CRB to the RMSE modulo $(-\pi,\pi]$) can therefore be given as $b_\phi(\phi) = 0$ and $p(\phi+\pi|\phi) = 0$. The condition $b_\phi(\phi) = 0$ can be satisfied relatively easily, because it will be satisfied whenever the probability distribution for the error in the phase estimate is symmetric, so $p(\phi+\theta|\phi) = p(\phi-\theta|\phi)$. However, it is not possible to satisfy $p(\phi+\pi|\phi) = 0$ for all ϕ when $\langle N\rangle$ is small. This is also the parameter regime where the HHB without bias must break down, because it would predict an impossibly large uncertainty.

Hence, the bias of a given estimate is crucial in any application of the Cramér-Rao bound to the RMSE. This is in strong contrast to the Heisenberg-type bounds for the RAMSE derived in this paper, which are independent of the bias function.

B. Asymptotic achievability

It is often stated that the CRB (and QCRB and HHB) is asymptotically achievable in the limit of many probe states, without any further qualification. However, for example, it is important to note from Eq. (110) that, in the asymptotic limit $m \to \infty$, the RMSE does not approach zero for a biased estimate — it is always bounded below by $|b_{\phi_r}(\phi)|$.

Furthermore, Fisher's theorem that Eq. (110) is itself asymptotically achievable, as $m \to \infty$, does not hold in all cases of physical interest [48]. In particular, this theorem assumes that there is a unique maximally likely estimate [47]. However, this is not the case for many states considered in quantum phase estimation, including the NOON states as per Eq. (97), which are the states that minimize the QCRB. The reason is of course that there is nothing to distinguish phase shifts modulo $2\pi/n$, regardless of the number of samples, unless the phase shift is in fact already known to this accuracy. That is, there are n maximally likely estimates, so Fisher's theorem does not apply.

In contrast, the above qualifications do not apply to the Heisenberg-type bounds for the RAMSE derived in this paper, which are independent of the bias of the estimate, and which are asymptotically achievable in the sense described in Sec. VI.

C. Example

There are obvious phase estimates that are not unbiased, where the RMSE obtained is qualitatively different from what would be expected from the Cramér-Rao bound without correcting for bias. Consider a simple measurement with a single photon in a MZI, in the state

$$\rho_\phi = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) + \frac{v}{2}(e^{i\phi}|0\rangle\langle 1| + e^{-i\phi}|1\rangle\langle 0|), \tag{112}$$

with visibility $v < 1$. The photon-counting measurement at the output of the interferometer gives probabilities of measurement results

$$p(\pm|\phi) = (1 \pm v\cos\phi)/2. \tag{113}$$

For the + measurement result, the optimal (least-square-error) estimate is $\hat{\phi} = 0$, and for the − measurement result the optimal estimate is $\hat{\phi} = \pi$. With these estimates, the RMSE is given by

$$\Delta_\phi\hat{\Phi} = \sqrt{\phi^2(1 + v\cos\phi)/2 + (\pi - |\phi|)^2(1 - v\cos\phi)/2}. \tag{114}$$

The absolute value of ϕ is taken above to take account of the fact that the difference should be determined modulo 2π. For $v = 1$, the error is zero at $\phi = 0$ and $\phi = \pi$.

In contrast, using the inverse square-root of the Fisher information (as for the CRB without correcting for bias) would give the lower bound

$$\Delta_\phi\hat{\Phi} \geq \frac{\sqrt{1 - v^2\cos^2\phi}}{v|\sin\phi|}. \tag{115}$$

In the limit $v \to 1$, the uncorrected CRB gives a result exactly equal to 1. This is already greater than the actual RMSE for some ϕ. For imperfect visibility, the contrast is even stronger. The uncorrected bound *diverges* at $\phi = 0$ and π, even though the actual measurement error is a *minimum* there. This result is illustrated in Fig. 4. It is therefore clear that the CRB can give completely misleading results if it is not corrected for bias. On the other hand, correcting the CRB for bias, via Eq. (110), yields a bound exactly equal to the RMSE (114).

The QCRB in Eq. (107), which assumes zero bias, gives $1/v$, and is also violated near $\phi = 0$ and $\phi = \pm\pi$ by the biased measurements considered here. Similarly, the HHB in Eq. (107) yields a lower bound of 1 (the same as the QCRB for $v = 1$), which is also violated by the biased measurements considered here. Thus we can see that caution needs to be employed in using Eq. (107), because it requires unbiased measurements. Such measurements are impossible in some cases, and even reasonable measurements can give highly biased estimates, resulting in a violation of the QCRB and HHB in Eq. (107).

It is also interesting to compare these results to the error propagation formula, which is often used to estimate the measurement error. The error propagation formula leads to the estimate of the error (using measurement operator $X = |0\rangle\langle 1| + |1\rangle\langle 0|$),

$$\begin{aligned}\Delta_\phi\hat{\Phi} &\approx \frac{\sqrt{\langle X^2\rangle_\phi - \langle X\rangle_\phi^2}}{\left|\frac{d}{d\phi}\langle X\rangle_\phi\right|} \\ &= \frac{\sqrt{1 - v^2\cos^2\phi}}{v|\sin\phi|}.\end{aligned} \tag{116}$$

Thus the error propagation formula gives an estimate of the uncertainty that is identical to the uncorrected Cramér-Rao bound.

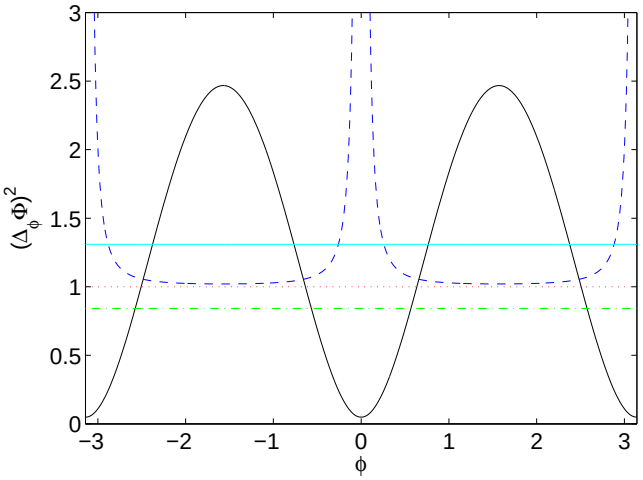

FIG. 4: MSE for phase measurements with a single photon using an interferometer with visibility $v = 0.99$ and photon counting at the outputs. The actual MSE, as well as the Cramér-Rao bound with the correction for bias, is given by the solid curve (black). The uncorrected Cramér-Rao bound, as well as the estimate given by the error propagation formula, is given as the dashed curve (dark blue). The horizontal dash-dotted line (green) is the conjectured bound on the AMSE, $k_C^2/\langle N+1\rangle^2$. The horizontal solid line (light blue) is the actual AMSE for these measurements (obtained by averaging the MSE over ϕ), and the horizontal dotted line (red) is the Helstrom-Holevo bound, which in this case is only slightly smaller than the quantum Cramér-Rao bound of $1/v$.

IX. CONCLUSIONS

We have rigorously proven that the square root of the average mean-square error (RAMSE) of phase measurements is lower bounded by the Heisenberg limit $k/\langle G+1\rangle$ [Eq. (1)]. The inequality with $k = k_A \approx 0.56$ holds in the case where the generator of the phase shifts has nonnegative integer eigenvalues. We obtain a very similar result in the case where G also has negative integer eigenvalues. The result is as in Eq. (46), where the absolute value of G is used, and the scaling constant is again k_A.

These results mean that the accuracy of super-Heisenberg measurement schemes is essentially illusory. They may work for a small range of phases, but if one considers the additional resources needed to locate an unknown phase to within the required range, the overall measurement will not violate the Heisenberg limit.

A new feature of our form of the Heisenberg limit is that it holds for all $\langle G\rangle$, not just in the asymptotic limit of large $\langle G\rangle$. We achieve this by adding 1 to the denominator. This modification is necessary, because otherwise the inequality would indicate that the error must approach infinity in the limit $\langle G\rangle \to 0$. This is impossible because phase has a bounded range.

As well as the analytical result stated above, we have very powerful evidence for a stronger bound with k_A replaced by $k_C \approx 1.38$. We have provided extensive numer-

ical evidence that the inequality holds with this larger scaling constant, both in the case of the RAMSE and for the square root of the Holevo variance. In the case where G has negative eigenvalues, the numerical results indicate that Eq. (46) holds with the scaling constant $k'_C \approx 0.79$, which is again larger than k_A.

These stronger lower bounds are also supported by asymptotic expansions of the exact solution for minimal Holevo variance, both for generators with nonnegative eigenvalues and generators without this restriction. A similar result for the RAMSE has also been obtained, via an asymptotic expansion of a lower bound for this quantity, for the case of a generator that is *not* restricted to nonnegative eigenvalues. The case where the eigenvalues of G are restricted to nonnegative eigenvalues is a possible area for future study. The asymptotic expansions also enable us to show that these stronger lower bounds are asymptotically achievable. That is, the minimum RAMSE is equal to the lower bounds to leading order.

We showed how various schemes that have been proposed to break the Heisenberg limit do not break our bound on the RAMSE. The primary reason for this is that they can only violate the Heisenberg limit scaling if the phase shift is already known, not when averaging over the phase shift. Another factor is that they typically consider the Cramér-Rao bound, which is problematic for the Heisenberg limit. It cannot provide a nontrivial lower bound for fixed mean photon number, as is required for the Heisenberg limit. In addition, it requires knowledge of the bias. Our alternative approach circumvents these limitations.

Our bound also differs from the Cramér-Rao bound in that it scales as $1/m$ in the number of copies of the state. The Cramér-Rao bound scales as $1/\sqrt{m}$, but we have found that such scaling is impossible for a fixed $\langle N\rangle$. This indicates that it is fundamentally impossible to obtain the Heisenberg limit from the Cramér-Rao bound.

Acknowledgments

DWB is funded by an ARC Future Fellowship (FT100100761). MJWH, MZ, and HMW are supported by the ARC Centre of Excellence CE110001027.

Appendix A: Details for the proof of Lemma 2

There are two results used in the proof of Lemma 2, which are proven here. First, we show that for this operator the distribution of $G^{(s)}$ is the same as the distribution of G for ρ_0. The normalisation condition for the covariant measurement gives

$$\mathbb{1} = \int d\theta\, e^{-iG\theta}\overline{M}_0 e^{iG\theta}, \tag{A1}$$

so

$$\begin{aligned}\delta_{n,n'}\delta_{d,d'} &= \int d\theta\, \langle n,d|e^{-iG\theta}\overline{M}_0 e^{iG\theta}|n',d'\rangle \\ &= \int d\theta\, e^{i(n'-n)\theta}\langle n,d|\overline{M}_0|n',d'\rangle \\ &= 2\pi\delta_{n,n'}\langle n,d|\overline{M}_0|n',d'\rangle. \end{aligned}\tag{A2}$$

This means that $\langle n,d|\overline{M}_0|n,d'\rangle = \delta_{d,d'}/2\pi$. Then, evaluating the distribution for $G^{(s)}$ gives

$$\begin{aligned}\langle n|\rho_0^{(s)}|n\rangle &= 2\pi \sum_{d,d'\le D(n)} \langle n,d'|\rho_0|n,d\rangle\, \langle n,d|\overline{M}_0|n,d'\rangle \\ &= 2\pi \sum_{d,d'\le D(n)} \langle n,d|\rho_0|n,d'\rangle\, (2\pi)^{-1}\delta_{d,d'} \\ &= \mathrm{Tr}(\rho_0 P_n), \end{aligned}\tag{A3}$$

where $P_n := \sum_d |n,d\rangle\langle n,d|$ denotes the projection onto eigenvalue n of G. The expression in the last line is the distribution of G for ρ_0.

Second, we must show that $\rho_0^{(s)}$ is positive and has trace one, and is therefore a valid density operator. Note one can always write the positive operators ρ_0 and $\overline{M}_0$ as sums of (not necessarily normalised or orthogonal) kets:

$$\rho_0 = \sum_\lambda |\lambda\rangle\langle\lambda|, \quad \overline{M}_0 = \sum_\mu |\mu\rangle\langle\mu|. \tag{A4}$$

Hence, for any state $|\psi\rangle = \sum_n \psi_n|n\rangle$,

$$\begin{aligned}\langle\psi|\rho_0^{(s)}|\psi\rangle &= 2\pi \sum_{n,n'\in S, d\le D(n), d'\le D(n')} \psi_n\psi^*_{n'} \\ &\quad\times \langle n',d'|\rho_0|n,d\rangle\langle n,d|\overline{M}_0|n',d'\rangle \\ &= 2\pi\sum_{\lambda,\mu}|X_{\lambda,\mu}|^2 \ge 0, \end{aligned}\tag{A5}$$

where

$$X_{\lambda,\mu} := \sum_{n\in S, d\le D(n)} \psi_n\langle\lambda|n,d\rangle\langle n,d|\mu\rangle. \tag{A6}$$

Hence $\rho_0^{(s)} \ge 0$, as required. Summing Eq. (A3) over n yields $\mathrm{Tr}(\rho_0^{(s)}) = 1$, so $\rho_0^{(s)}$ is a valid density operator.

Appendix B: Details for Eq. (45)

Here we give the details of the derivation of Eq. (45). First, variation of Eq. (44) gives the optimising distribution

$$p_n = e^{-(\alpha+1)}e^{-\beta|n-g|}, \tag{B1}$$

which is the double exponential (Laplace) distribution found in [23]. In Ref. [23] it was assumed that g was an integer, but here we consider the more general case that g is an arbitrary real number.

We require $\beta > 0$ in order for the distribution to be normalisable. Then normalisation gives the restriction

$$\frac{e^{-\beta r} + e^{-\beta(1-r)}}{1 - e^{-\beta}} = e^{\alpha+1}, \tag{B2}$$

where $r = \lceil g \rceil - g$. We then find that the mean value is

$$\langle |G - g| \rangle = \frac{(e^{\beta r} + e^{-\beta r})(1 - r) + (e^{\beta(1-r)} + e^{-\beta(1-r)})r}{(1 - e^{-\beta})(e^{\beta r} + e^{\beta(1-r)})}, \tag{B3}$$

and

$$\begin{aligned} H(G) &= (\alpha + 1) + \beta \langle |G - g| \rangle \\ &= \ln\left(\frac{e^{-\beta r} + e^{-\beta(1-r)}}{1 - e^{-\beta}}\right) + \beta \langle |G - g| \rangle. \end{aligned} \tag{B4}$$

Without loss of generality we take $r \in [0, 1/2]$. The problem is symmetric about $r = 1/2$, so these results also apply to $r > 1/2$. Then we obtain

$$\begin{aligned} &2\langle |G - g| \rangle + 1 - \exp\left[\beta r + \ln\left(\frac{e^{-\beta r} + e^{-\beta(1-r)}}{1 - e^{-\beta}}\right)\right] \\ &= \frac{(e^{\beta} - e^{2\beta r})[1 + e^{2\beta r} + 2r(e^{\beta} - 1)]}{(e^{\beta} - 1)(e^{\beta} + e^{2\beta r})} \geq 0. \end{aligned} \tag{B5}$$

This then yields

$$\beta r + \ln\left(\frac{e^{-\beta r} + e^{-\beta(1-r)}}{1 - e^{-\beta}}\right) \leq \ln(2\langle |G - g| \rangle + 1). \tag{B6}$$

It can be shown that

$$\begin{aligned} &e^{\beta} - (1 + \beta) - e^{2\beta r}[e^{-\beta}(1 + 2\beta r) + \beta - 1 - 2\beta r] \\ &= (1 + \beta r - \beta \langle |G - g| \rangle)(1 - e^{-\beta})(e^{\beta} + e^{2\beta r}). \end{aligned} \tag{B7}$$

Next, for $\beta > 0$ and $r \geq 0$ we have

$$\begin{aligned} 0 &< \beta(1 - 2r)^2 + 4r \\ &= 1 + \beta(1 - 2r) - [1 + 2(\beta - 2)r - 4\beta r^2] \\ &\leq e^{\beta(1-2r)} - [1 + 2(\beta - 2)r - 4\beta r^2]. \end{aligned} \tag{B8}$$

This means that

$$\begin{aligned} &\frac{d}{d\beta}\left\{e^{\beta} - (1 + \beta) - e^{2\beta r}[e^{-\beta}(1 + 2\beta r) + \beta - 1 - 2\beta r]\right\} \\ &= (1 - e^{-\beta})e^{2\beta r}\left\{e^{\beta(1-2r)} - [1 + 2(\beta - 2)r - 4\beta r^2]\right\} \\ &> 0 \end{aligned} \tag{B9}$$

for $\beta > 0$. This implies that

$$e^{\beta} - (1 + \beta) - e^{2\beta r}[e^{-\beta}(1 + 2\beta r) + \beta - 1 - 2\beta r] > 0, \tag{B10}$$

for $\beta > 0$, because the left-hand side is zero for $\beta = 0$, and has positive slope for $\beta > 0$. Using Eq. (B7), this gives

$$-\beta r + \beta \langle |G - g| \rangle < 1. \tag{B11}$$

Now adding Eqs. (B6) and (B11) yields

$$\ln\left(\frac{e^{-\beta r} + e^{-\beta(1-r)}}{1 - e^{-\beta}}\right) + \beta \langle |G - g| \rangle < \ln(2\langle |G - g| \rangle + 1) + 1, \tag{B12}$$

and substitution into Eq. (B4) gives Eq. (45) as required.

Appendix C: Inequality proofs

To prove $\theta^2 \leq f_3(\theta)$, consider the function

$$\Delta(\theta) = \sqrt{(\pi^2/2)(1 - \cos\theta) - (\pi^2/4 - 1)(1 - \cos 2\theta)/2} - \theta. \tag{C1}$$

Taking the derivative with respect to θ and solving to find the turning points of $\Delta(\theta)$ yields only two in the range $[0, \pi]$. One is at $\theta = 0$, and the other is at $\theta \approx 2.23$. As $\Delta(\theta) > 0$ for $\theta \approx 2.23$, and $\Delta(\theta) = 0$ for $\theta = 0$ or π, we have $\Delta(\theta) \geq 0$ for $\theta \in [0, \pi]$. This proves Eq. (79) for $\theta \in [0, \pi]$, and the result for $\theta \in [-\pi, 0]$ follows because Eq. (79) is symmetric.

Next, to prove a lower bound on $\delta\Phi$, we use Eq. (20), which was

$$\langle \cos\Theta \rangle \geq \cos\sqrt{\langle \Theta^2 \rangle}. \tag{C2}$$

Using this, we have

$$\delta\hat{\Phi} = \sqrt{\langle \Theta^2 \rangle} \geq \arccos\langle \cos\Theta \rangle = \arccos[1 - (\delta_1\hat{\Phi})^2/2]. \tag{C3}$$

Now note that, if we have a state that minimises $\delta\hat{\Phi}$, then it can not give a value of $\arccos(1 - (\delta_1\hat{\Phi})^2/2)$ smaller than that for the minimum value of $\delta_1\hat{\Phi}$. This means that we can lower bound $\delta\hat{\Phi}$ by the minimum value of $\arccos[1 - (\delta_1\hat{\Phi})^2/2]$. This is a tighter lower bound on $\delta\hat{\Phi}$ than $\delta_1\hat{\Phi}$, because $\arccos(1 - x^2/2) = x + x^3/24 + O(x^5)$. Unfortunately $\arccos[1 - (\delta_1\hat{\Phi})^2/2]$ can still be below $k_C/\langle N + 1 \rangle$.

Appendix D: Asymptotic behaviour for Holevo variance

Here we derive the asymptotic results for the variance given in Sec. V. Using Eq. (12) of Ref. [52] (with $q = p + 1$), one has

$$\begin{aligned} \sum_{k=1}^{\infty} J_{x+k}(z)\, J_{x+k+1}(z) &= \frac{z}{2}\left[J_{x+1}^2(z) - J_x(z)\, J_{x+2}(z)\right] \\ &= \frac{z}{2}[J_{x+1}(z)]^2, \end{aligned} \tag{D1}$$

where the second equality follows from Eq. (73).

Second, from Eq. (32) of Ref. [52], one has

$$\begin{aligned} \sum_{k=1}^{\infty}[J_{x+k}(z)]^2 &= \frac{z}{2}\left[J_{x+1}(z)\frac{\partial J_x(z)}{\partial x} - J_x(z)\frac{\partial J_{x+1}(z)}{\partial x}\right] \\ &= \frac{z}{2}J_{x+1}(z)\frac{\partial J_x(z)}{\partial x}, \end{aligned} \tag{D2}$$

where the second equality similarly follows via Eq. (73). We therefore have

$$\langle e^{i\Theta} \rangle = \frac{J_{x+1}(z)}{[\partial J_x(z)/\partial x]}. \tag{D3}$$

Using Eq. (77) then yields

$$\langle N+1\rangle = \frac{zJ_{x+1}(z)}{[\partial J_x(z)/\partial x]} - x. \tag{D4}$$

There are a number of asymptotic results that we can use. From [53, 54]

$$z = x + \gamma x^{1/3} + \frac{3\gamma^2}{10x^{1/3}} + \frac{5-\gamma^3}{350x} - \frac{479\gamma^4+20\gamma}{63000x^{5/3}} + \frac{20231\gamma^5 - 27550\gamma^2}{8085000x^{7/3}} + O(x^{-3}), \tag{D5}$$

where $\gamma = |z_A|/2^{1/3}$, and z_A is the first zero of the Airy function. We can invert this relation to give

$$x = z - \gamma z^{1/3} + \frac{\gamma^2}{30z^{1/3}} - \frac{5-\gamma^3}{350z} + \frac{281\gamma^4 - 5220\gamma}{567000z^{5/3}} + \frac{73769\gamma^5 - 3312450\gamma^2}{654885000z^{7/3}} + O(z^{-3}). \tag{D6}$$

Now from Eq. (9.3.23) of [55] we have, for $z = x + yx^{1/3}$,

$$J_x(x+yx^{1/3}) \sim \frac{2^{1/3}}{x^{1/3}}\mathrm{Ai}(-2^{1/3}y)\left(1 + \sum_{k=1}^{\infty}\frac{f_k(y)}{x^{2k/3}}\right) + \frac{2^{2/3}}{x}\mathrm{Ai}'(-2^{1/3}y)\sum_{k=0}^{\infty}\frac{g_k(y)}{x^{2k/3}}. \tag{D7}$$

The functions are given by

$$f_1(y) = -\frac{1}{5}y, \tag{D8}$$

$$f_2(y) = -\frac{9}{100}y^5 + \frac{3}{35}y^2, \tag{D9}$$

$$f_3(y) = \frac{957}{7000}y^6 - \frac{173}{3150}y^3 - \frac{1}{225}, \tag{D10}$$

$$f_4(y) = \frac{27}{20000}y^{10} - \frac{23573}{147000}y^7 + \frac{5903}{138600}y^4 + \frac{947}{346500}y, \tag{D11}$$

$$g_0(y) = \frac{3}{10}y^2, \tag{D12}$$

$$g_1(y) = -\frac{17}{70}y^3 + \frac{1}{70}, \tag{D13}$$

$$g_2(y) = -\frac{9}{1000}y^7 + \frac{611}{3150}y^4 - \frac{37}{3150}y, \tag{D14}$$

$$g_3(y) = \frac{549}{28000}y^8 - \frac{110767}{693000}y^5 + \frac{79}{12375}y^2. \tag{D15}$$

Using Eq. (D5) we have

$$y = (z-x)/x^{1/3} = \gamma + \frac{3\gamma^2}{10x^{2/3}} + \frac{5-\gamma^3}{350x^{4/3}} - \frac{479\gamma^4+20\gamma}{63000x^2} + \frac{20231\gamma^5 - 27550\gamma^2}{8085000x^{8/3}} + O(x^{-3}). \tag{D16}$$

We can then substitute Eq. (D6), which gives

$$y - \gamma = \frac{3\gamma^2}{10z^{2/3}} + \frac{5+69\gamma^3}{350z^{4/3}} + \frac{9361\gamma^4+1180\gamma}{63000z^2} + \frac{8691349\gamma^5 + 1484550\gamma^2}{72765000z^{8/3}} + O(z^{-3}). \tag{D17}$$

Now to take the derivative with respect to the order, we can use

$$\frac{\partial}{\partial x}J_x(z) = \frac{d}{dx}J_x(x+yx^{1/3}) + \frac{dy}{dx}\frac{d}{dy}J_x(x+yx^{1/3}) = \frac{d}{dx}J_x(x+yx^{1/3}) - \left(\frac{z}{3x^{4/3}} + \frac{2}{3x^{1/3}}\right)\frac{d}{dy}J_x(x+yx^{1/3}). \tag{D18}$$

In the resulting expression it is possible to expand in a series for y about γ, then expand in a series about z.

It is possible to determine a series in z for

$$\frac{zJ_{x+1}(z)}{[\partial J_x(z)/\partial x]} - x. \tag{D19}$$

We can then invert this series, finding a series in $\langle N+1\rangle$ for z. Similarly, it is possible to find a series in z for

$$\frac{J_{x+1}(z)}{[\partial J_x(z)/\partial x]}. \tag{D20}$$

Then we can express

$$\left(\frac{J_{x+1}(z)}{[\partial J_x(z)/\partial x]}\right)^{-2} - 1, \tag{D21}$$

as a series in z. Substituting the series for z in $\langle N+1\rangle$, the overall result is as in Eq. (78), with

$$b_2 = \frac{4|z_A|^3}{27} = k_C^2 \approx 1.8936, \tag{D22}$$

$$b_4 = \frac{16|z_A|^6}{1215} \approx 2.1514, \tag{D23}$$

$$b_6 = \frac{16|z_A|^6(27+40|z_A|^3)}{688905} \approx 2.0424, \tag{D24}$$

$$b_8 = \frac{256|z_A|^9(3+|z_A|^3)}{4428675} \approx 1.9050, \tag{D25}$$

$$b_{10} = \frac{64|z_A|^9(2673+9252|z_A|^3+1120|z_A|^6)}{21483502425} \approx 1.8906. \tag{D26}$$

Next, to place an upper bound on the RAMSE for this state, we use Eq. (79). This equation gives

$$\langle\Theta^2\rangle \le (\pi^2/2)(1-\langle\cos\Theta\rangle) - (\pi^2/4-1)(1-\langle\cos 2\Theta\rangle)/2. \tag{D27}$$

To find the expectation value $\langle e^{i2\Theta}\rangle$, we use

$$\langle e^{i2\Theta}\rangle = A^2\sum_{n=0}^{\infty} J_{x+n+1}(z)\,J_{x+n+3}(z) = \frac{\sum_{k=1}^{\infty} J_{x+k}(z)\,J_{x+k+2}(z)}{\sum_{k=1}^{\infty}[J_{x+k}(z)]^2}. \tag{D28}$$

Again using Eq. (12) of Ref. [52], but now with $q = p+2$, we have

$$\begin{aligned}&\sum_{k=1}^{\infty} J_{x+k}(z)\, J_{x+k+2}(z)\\ &= \frac{z}{4}\left[J_{x+1}(z)J_{x+2}(z) - J_x(z)\, J_{x+3}(z)\right]\\ &= \frac{z}{4}J_{x+1}(z)J_{x+2}(z).\end{aligned} \tag{D29}$$

That then gives

$$\langle e^{i2\Theta}\rangle = \frac{J_{x+2}(z)}{2[\partial J_x(z)/\partial x]}. \tag{D30}$$

Using this expression, and expanding the Bessel function solution in a series as above, we obtain

$$\begin{aligned}\langle\Theta^2\rangle &\le (\pi^2/2)(1-\langle\cos\Theta\rangle) - (\pi^2/4-1)(1-\langle\cos 2\Theta\rangle)/2\\ &= \frac{4|z_A|^3}{27\langle N+1\rangle^2} + \frac{(\pi^2-4)|z_A|^3}{54\langle N+1\rangle^3} + O\left(\frac{1}{\langle N+1\rangle^4}\right).\end{aligned} \tag{D31}$$

This results in the inequality given in Eq. (80).

Expanding in a series also gives

$$\begin{aligned}(\arccos\langle\cos\Theta\rangle)^2 = \frac{4|z_A|^3}{27\langle N+1\rangle^2} - \frac{16|z_A|^6}{10935\langle N+1\rangle^4}\\ + O\left(\frac{1}{\langle N+1\rangle^6}\right).\end{aligned} \tag{D32}$$

Using Eq. (C3), we therefore have the upper and lower bounds on $(\delta\hat{\Phi})^2$,

$$\begin{aligned}\frac{k_C^2}{\langle N+1\rangle^2} - O\left(\frac{1}{\langle N+1\rangle^4}\right) \le (\delta\hat{\Phi})^2\\ \le \frac{k_C^2}{\langle N+1\rangle^2} + O\left(\frac{1}{\langle N+1\rangle^3}\right).\end{aligned} \tag{D33}$$

Appendix E: Asymptotic behaviour for variance with fixed $\langle|J|\rangle$

The normalization constraint yields, using Eq. (32) of Ref. [52],

$$\begin{aligned}A^{-2} &= [J_x(z)]^2 + 2\sum_{j=1}^{\infty}[J_{x+j}(z)]^2\\ &= [J_x(z)]^2 + z\left[J_{x+1}(z)\frac{\partial J_x(z)}{\partial x} - J_x(z)\frac{\partial J_{x+1}(z)}{\partial x}\right].\end{aligned} \tag{E1}$$

We also have

$$\begin{aligned}\langle|J|\rangle &= 2A^2\sum_{j=0}^{\infty} j\,[J_{x+j}(z)]^2,\\ \langle e^{i\Theta}\rangle &= A^2\left[2J_x(z)\, J_{x+1}(z) + 2\sum_{j=1}^{\infty} J_{x+j}(z)\, J_{x+j+1}(z)\right].\end{aligned} \tag{E2}$$

Using Eq. (12) of [52] we have

$$\begin{aligned}\langle e^{i\Theta}\rangle = A^2\left[2J_x(z)\, J_{x+1}(z) + zJ_{x+1}^2(z)\right.\\ \left. -zJ_x(z)\, J_{x+2}(z)\right].\end{aligned} \tag{E3}$$

In this case we have

$$\langle e^{i\Theta}\rangle = (\alpha + \beta\langle|J|\rangle) = (x + \langle|J|\rangle)/z, \tag{E4}$$

so

$$\langle|J|\rangle = z\langle e^{i\Theta}\rangle - x. \tag{E5}$$

The first zero of the derivative of the Bessel function is given by [53, 54]

$$\begin{aligned}z = x + \gamma' x^{1/3} + \left(\frac{3\gamma'^2}{10} - \frac{1}{10\gamma'}\right)\frac{1}{x^{1/3}}\\ - \left(\frac{\gamma'^3}{350} + \frac{1}{25} + \frac{1}{200\gamma'^3}\right)\frac{1}{x}\\ - \frac{958\gamma'^9 - 2036\gamma'^6 - 84\gamma'^3 + 63}{126000\gamma'^5 x^{5/3}} + O(x^{-7/3}).\end{aligned} \tag{E6}$$

We have corrected an error in Ref. [53] where "840" was given instead of "84". Here $\gamma' = |z'_A|/2^{1/3}$, where z'_A is the first zero of the derivative of the Airy function. Inverting this series, and performing series expansions similar to that for the first case, gives the series in Eq. (86) with

$$d_2 = \frac{16|z'_A|^3}{27} = {k'_C}^2 \approx 0.6266, \tag{E7}$$

$$d_3 = \frac{32|z'_A|^3}{27} \approx 1.2533, \tag{E8}$$

$$d_4 = \frac{16|z'_A|^3(111 - 4|z'_A|^3)}{1215} \approx 1.4868, \tag{E9}$$

$$d_5 = \frac{64|z'_A|^3(21 - 4|z'_A|^3)}{1215} \approx 0.9341, \tag{E10}$$

$$d_6 = \frac{16|z'_A|^3(-63 - 40488|z'_A|^3 + 160|z'_A|^6)}{1148175} \approx -0.6292. \tag{E11}$$

To determine an upper bound, we need to determine

$$\begin{aligned}\langle e^{i2\Theta}\rangle &= A^2\left[J_{x+1}^2(z) + 2J_x(z)\, J_{x+2}(z)\right.\\ &\qquad \left. +2\sum_{j=1}^{\infty} J_{x+j}(z)\, J_{x+j+2}(z)\right]\\ &= A^2\left\{J_{x+1}^2(z) + 2J_x(z)\, J_{x+2}(z)\right.\\ &\qquad \left. +\frac{z}{2}[J_{x+1}(z)J_{x+2}(z) - J_x(z)\, J_{x+3}(z)]\right\}.\end{aligned} \tag{E12}$$

Expanding in a series then gives

$$\begin{aligned}\langle\Theta^2\rangle \le \frac{16|z'_A|^3}{27\langle 2|J|+1\rangle^2} + \frac{32|z'_A|^3}{27\langle 2|J|+1\rangle^3}\\ + O\left(\frac{1}{\langle 2|J|+1\rangle^4}\right).\end{aligned} \tag{E13}$$

Therefore, the MSE is upper and lower bounded as

$$\frac{{k'_C}^2}{\langle 2|J|+1\rangle^2} + O\left(\frac{1}{\langle 2|J|+1\rangle^3}\right) \leq (\delta\hat{\Phi})^2 \leq \frac{{k'_C}^2}{\langle 2|J|+1\rangle^2} + O\left(\frac{1}{\langle 2|J|+1\rangle^3}\right). \tag{E14}$$

Appendix F: Continuity of the expected phase estimate

Here we show that $\langle\hat{\Phi}\rangle_\phi^{\phi_r}$ is a continuous function of ϕ. Defining

$$X_{\phi_r} := \int_{\phi_r-\pi}^{\phi_r+\pi} d\hat{\phi}\,\hat{\phi}\,M_{\hat{\phi}}, \tag{F1}$$

we have

$$\langle\hat{\Phi}\rangle_\phi^{\phi_r} = \mathrm{Tr}(X_{\phi_r}\rho_\phi). \tag{F2}$$

The expectation value of the phase estimate at $\phi+\epsilon$ is

$$\langle\hat{\Phi}\rangle_{\phi+\epsilon}^{\phi_r} = \mathrm{Tr}(X_{\phi_r}e^{-iG\epsilon}\rho_\phi e^{iG\epsilon}). \tag{F3}$$

The difference is

$$|\langle\hat{\Phi}\rangle_{\phi+\epsilon}^{\phi_r} - \langle\hat{\Phi}\rangle_\phi^{\phi_r}| = |\mathrm{Tr}[X_{\phi_r}(e^{-iG\epsilon}\rho_\phi e^{iG\epsilon} - \rho_\phi)]|. \tag{F4}$$

Take $|\xi_j\rangle$ to be the eigenbasis of X_{ϕ_r}. Then

$$\begin{aligned}|\langle\xi_j|X_{\phi_r}|\xi_j\rangle| &\leq \int_{\phi_r-\pi}^{\phi_r+\pi} d\hat{\phi}\,|\hat{\phi}|\,\langle\xi_j|M_{\hat{\phi}}|\xi_j\rangle \\ &\leq 2\pi\int_{\phi_r-\pi}^{\phi_r+\pi} d\hat{\phi}\,\langle\xi_j|M_{\hat{\phi}}|\xi_j\rangle \\ &= 2\pi\langle\xi_j|\mathbb{1}|\xi_j\rangle = 2\pi.\end{aligned} \tag{F5}$$

Using this, we find

$$\begin{aligned}&|\mathrm{Tr}[X_{\phi_r}(e^{-iG\epsilon}\rho_\phi e^{iG\epsilon} - \rho_\phi)]| \\ &= \left|\sum_j \langle\xi_j|X_{\phi_r}|\xi_j\rangle\langle\xi_j|(e^{-iG\epsilon}\rho_\phi e^{iG\epsilon} - \rho_\phi)|\xi_j\rangle\right| \\ &\leq 2\pi\sum_j \left|\langle\xi_j|(e^{-iG\epsilon}\rho_\phi e^{iG\epsilon} - \rho_\phi)|\xi_j\rangle\right|.\end{aligned} \tag{F6}$$

Take $|\zeta_j\rangle$ to be the eigenbasis of $e^{-iG\epsilon}\rho_\phi e^{iG\epsilon} - \rho_\phi$. Then

$$\begin{aligned}&\sum_j \left|\langle\xi_j|(e^{-iG\epsilon}\rho_\phi e^{iG\epsilon} - \rho_\phi)|\xi_j\rangle\right| \\ &= \sum_j \left|\sum_k |\langle\xi_j|\zeta_k\rangle|^2\langle\zeta_k|(e^{-iG\epsilon}\rho_\phi e^{iG\epsilon} - \rho_\phi)|\zeta_k\rangle\right| \\ &\leq \sum_k \left|\langle\zeta_k|(e^{-iG\epsilon}\rho_\phi e^{iG\epsilon} - \rho_\phi)|\zeta_k\rangle\right| \\ &= \|e^{-iG\epsilon}\rho_\phi e^{iG\epsilon} - \rho_\phi\|_1.\end{aligned} \tag{F7}$$

Hence

$$|\mathrm{Tr}[X_{\phi_r}(e^{-iG\epsilon}\rho e^{iG\epsilon} - \rho_\phi)]| \leq 2\pi\|e^{-iG\epsilon}\rho_\phi e^{iG\epsilon} - \rho_\phi\|_1. \tag{F8}$$

Take the state to be given by

$$\rho_\phi = \sum_j p_j|\psi_j\rangle\langle\psi_j|. \tag{F9}$$

For $|\psi_j\rangle$,

$$\begin{aligned}\langle\psi_j|e^{-iG\epsilon}|\psi_j\rangle &= \sum_k |\psi_{jk}|^2 e^{-ik\epsilon} \\ &= 1 - \sum_k |\psi_{jk}|^2(1 - e^{-ik\epsilon}).\end{aligned} \tag{F10}$$

Evaluating the distance from 1 gives

$$\begin{aligned}\left|\sum_k |\psi_{jk}|^2(1 - e^{ik\epsilon})\right| &\leq \sum_k |\psi_{jk}|^2|1 - e^{-ik\epsilon}| \\ &\leq \sum_k |\psi_{jk}|^2|k\epsilon| \\ &= \langle\psi_j||G||\psi_j\rangle|\epsilon|,\end{aligned} \tag{F11}$$

so

$$\begin{aligned}D(|\psi_j\rangle, e^{-iG\epsilon}|\psi_j\rangle) &= 2\sqrt{1 - |\langle\psi_j|e^{-iG\epsilon}|\psi_j\rangle|^2} \\ &\leq 2\sqrt{1 - |1 - \langle\psi_j||G||\psi_j\rangle|\epsilon||^2} \\ &\leq 2\sqrt{2\langle\psi_j||G||\psi_j\rangle|\epsilon|}.\end{aligned} \tag{F12}$$

By the convexity of trace distance

$$\begin{aligned}\|e^{-iG\epsilon}\rho_\phi e^{iG\epsilon} - \rho_\phi\|_1 &\leq \sum_j p_j D(|\psi_j\rangle, e^{-iG\epsilon}|\psi_j\rangle) \\ &\leq \sum_i p_i 2\sqrt{2\langle\psi_j||G||\psi_j\rangle|\epsilon|} \\ &\leq 2\sqrt{2\langle|G|\rangle|\epsilon|},\end{aligned} \tag{F13}$$

where $\langle|G|\rangle = \mathrm{Tr}(|G|\rho)$. Hence

$$|\langle\hat{\Phi}\rangle_{\phi+\epsilon}^{\phi_r} - \langle\hat{\Phi}\rangle_\phi^{\phi_r}| \leq 4\pi\sqrt{2\langle|G|\rangle|\epsilon|}. \tag{F14}$$

Thus the expectation value of the phase estimate must be a continuous function of ϕ unless $\langle|G|\rangle$ is infinite.

[1] C. M. Caves, Phys. Rev. D **23**, 1693 (1981).
[2] V. Giovannetti, S. Lloyd, and L. Maccone, Science **306**, 1330 (2004).
[3] R. Slavik *et al.*, Nature Photonics **4**, 690 (2010).
[4] B. Yurke, S. L. McCall, and J. R. Klauder, Phys. Rev. A **33**, 4033 (1986).
[5] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control* (Cambridge University Press, Cambridge, 2010).
[6] M. J. Holland and K. Burnett, Phys. Rev. Lett. **71**, 1355 (1993); Z. Y. Ou, Phys. Rev. A **55**, 2598 (1997); D. Braun and J. Martin, Nature Communnications **2**, 223 (2011).
[7] M. Zwierz, C. A. Pérez-Delgado, and P. Kok, Phys. Rev. Lett. **105** 180402 (2010); **107**, 059904(E) (2011).
[8] V. Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. Lett. **96**, 010401 (2006).
[9] U. Dorner, R. Demkowicz-Dobrzański, B. J. Smith, J. S. Lundeen, W. Wasilewski, K. Banaszek, and I. A. Walmsley, Phys. Rev. Lett. **102**, 040403 (2009).
[10] M. Kacprowicz, R. Demkowicz-Dobrzański, W. Wasilewski, K. Banaszek, and I. A. Walmsley, Nature Photonics **4**, 357 (2010).
[11] B. M. Escher, R. L. de Matos Filho, and L. Davidovich, Nature Physics **7**, 406 (2011).
[12] P. M. Anisimov, G. M. Raterman, A. Chiruvelli, W. N. Plick, S. D. Huver, H. Lee, and J. P. Dowling, Phys. Rev. Lett. **104**, 103602 (2010).
[13] Y. R. Zhang, G. R. Jin, J. P. Cao, W. M. Liu, and H. Fan, e-print arXiv:1105.2990v4.
[14] Á. Rivas and A. Luis, New J. Phys. **14**, 093052 (2012).
[15] S. Boixo, S. T. Flammia, C. M. Caves, and J. M. Geremia, Phys. Rev. Lett. **98**, 090401 (2007); S. M. Roy and S. L. Braunstein, Phys. Rev. Lett. **100**, 220501 (2008).
[16] M. Napolitano, M. Koschorreck, B. Dubost, N. Behbood, R. J. Sewell, and M. W. Mitchell, Nature **471**, 486 (2011).
[17] J. H. Shapiro, S. R. Shepard, and N. C. Wong, Phys. Rev. Lett. **62**, 2377 (1989).
[18] J. H. Shapiro and S. R. Shepard, Phys. Rev. A **43**, 3795 (1991).
[19] S. L. Braunstein, A. S. Lane, and C. M. Caves, Phys. Rev. Lett. **69**, 2153 (1992).
[20] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde, Nature **450**, 393 (2007).
[21] D. W. Berry, B. L. Higgins, S. D. Bartlett, M. W. Mitchell, G. J. Pryde, and H. M. Wiseman, Phys. Rev. A. **80**, 052114 (2009).
[22] M. J. W. Hall, D. W. Berry, M. Zwierz, and H. M. Wiseman, Phys. Rev. A **85**, 041802(R) (2012).
[23] M. J. W. Hall and H. M. Wiseman, New J. Phys. **14** 033040 (2012).
[24] M. Tsang, Phys. Rev. Lett. **108**, 230401 (2012).
[25] V. Giovannetti and L. Maccone, Phys. Rev. Lett. **108**, 210404 (2012).
[26] R. Nair, e-print arXiv:1204.3761v1.
[27] M. J. W. Hall and H. M. Wiseman, Phys. Rev. X **2**, 041006 (2012).
[28] V. Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. Lett. **108**, 260405 (2012).
[29] M. J. Collett, Phys. Scr. **T48**, 124 (1993).
[30] T. Opatrný, J. Phys. A **28**, 6961 (1995).
[31] G. W. Forbes and M. A. Alonso, Am. J. Phys. **69**, 340 (2001), ibid **69**, 1091 (2001).
[32] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
[33] H. M. Wiseman and R. B. Killip, Phys. Rev. A **56**, 944 (1997).
[34] D. W. Berry, PhD thesis (University of Queensland, Australia, 2002), Ch. 2 (also available as e-print arXiv:quant-ph/0202136v1).
[35] A. Bandilla, H. Paul, and H. H. Ritze, Quantum Opt. **3**, 267 (1991).
[36] M. J. W. Hall, Phys. Rev. A **59**, 2602 (1999).
[37] M. J. W. Hall, Phys. Rev. A **62**, 012107 (2000).
[38] M. J. W. Hall, J. Phys. A: Math. Gen. **41**, 255301 (2008).
[39] M. J. W. Hall, J. Mod. Opt. **40**, 809 (1993).
[40] I. Białynicki-Birula and J. Mycielski, Commun. Math. Phys. **44**, 129 (1975).
[41] A. Luis and J. Peřina, Phys. Rev. A **54**, 4564 (1996).
[42] G. M. D'Ariano, C. Macchiavello, and M. F. Sacchi, Phys. Lett. A **248**, 103 (1998); G. M. D'Ariano, M. G. A. Paris, and P. Perinotti, J. Opt. B: Quantum Semiclass. Opt. **3**, 337 (2001).
[43] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
[44] K. P. Seshadreesan, P. M. Anisimov, H. Lee, and J. P. Dowling, New J. Phys. **13**, 083026 (2011).
[45] H. F. Hofmann, Phys. Rev. A **79**, 033822 (2009).
[46] P. Hyllus, L. Pezze, and A. Smerzi, Phys. Rev. Lett. **105**, 120501 (2010).
[47] R. A. Fisher, Proc. Camb. Phil. Soc. **22**, 700 (1925).
[48] G. A. Durkin and J. P. Dowling, Phys. Rev. Lett. **99**, 070801 (2007).
[49] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994).
[50] Y. C. Eldar, IEEE Trans. Signal Process. **54**, 2943 (2006).
[51] C. W. Helstrom, Phys. Lett. A **25**, 101 (1967).
[52] E. Hansen, Am. Mathemat. Monthly **73**(2), 143-150 (1966).
[53] Á. Elbert, J. Comp. App. Math. **133**, 65 (2001).
[54] F. W. J. Olver, Proc. Cambridge Philos. Soc. **47**, 699 (1951).
[55] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (U.S. Government Printing Office, Washington, D.C., 1972).