

# Does Nonlinear Metrology Offer Improved Resolution? Answers from Quantum Information Theory

Michael J. W. Hall and Howard M. Wiseman

*Centre for Quantum Dynamics, Griffith University, Brisbane, Queensland 4111, Australia*

(Received 10 May 2012; published 25 October 2012)

A number of authors have suggested that nonlinear interactions can enhance resolution of phase shifts beyond the usual Heisenberg scaling of  $1/n$ , where  $n$  is a measure of resources such as the number of subsystems of the probe state or the mean photon number of the probe state. These suggestions are based on calculations of “local precision” for particular nonlinear schemes. However, we show that there is no simple connection between the local precision and the average estimation error for these schemes, leading to a scaling puzzle. This puzzle is partially resolved by a careful analysis of iterative implementations of the suggested nonlinear schemes. However, it is shown that the suggested nonlinear schemes are still limited to an exponential scaling in  $\sqrt{n}$ . (This scaling may be compared to the exponential scaling in  $n$  which is achievable if multiple passes are allowed, even for linear schemes.) The question of whether nonlinear schemes may have a scaling advantage in the presence of loss is left open. Our results are based on a new bound for average estimation error that depends on (i) an entropic measure of the degree to which the probe state can encode a reference phase value, called the  $G$  asymmetry, and (ii) any prior information about the phase shift. This bound is asymptotically stronger than bounds based on the variance of the phase-shift generator. The  $G$  asymmetry is also shown to directly bound the average information gained per estimate. Our results hold for any prior distribution of the shift parameter, and generalize to estimates of any shift generated by an operator with discrete eigenvalues.

DOI: [10.1103/PhysRevX.2.041006](https://doi.org/10.1103/PhysRevX.2.041006)

Subject Areas: Photonics, Quantum Physics, Quantum Information

## I. INTRODUCTION

In many measurement scenarios, an environmental variable acts to translate or shift a property such as the optical phase or position of a probe state. Accurate estimation of the shifted parameter allows a correspondingly accurate measurement of the environmental variable. For example, interferometric measurements of quantities ranging from temperature to gravitational-wave amplitudes rely on the estimation of an optical phase shift. An important aim of quantum metrology is to determine the fundamental bounds on the resolution of such estimates, and how these bounds scale with available resources such as energy [1–4].

Let us denote the initial probe state by the density operator  $\rho_0$ , and the generator of shifts by some Hermitian operator  $G$ . Then if the shift parameter  $\Phi$  has the value  $\phi$ , the final probe state is  $\rho_\phi = \exp(-iG\phi)\rho_0 \exp(iG\phi)$ . In the following discussion, particular attention is paid to the estimation of a *phase* shift parameter, as this is sufficient for discussing various nonlinear estimation schemes previously proposed in the literature [5–11]. The generator  $G$  in this case has integer eigenvalues, so that  $\rho_{\phi+2\pi} = \rho_\phi$ . More generally, however, our results apply to any shift generator  $G$  having a

discrete eigenvalue spectrum. This includes the atomic scheme proposed in Ref. [12], which has recently led to the first experimental demonstration of nonlinear quantum metrology [13].

Returning to an optical example, a linear phase shift of a single-mode optical probe state corresponds to  $G = N$ , where  $N$  is the photon-number operator. Similarly, for a probe state comprising  $m$  such modes, each undergoing a nonlinear quadratic phase shift, one has  $G = (N_1)^2 + \dots + (N_m)^2$  [5,6]. In cases like this, we quantify the resources  $n$  by the total mean photon number  $\sum_j \langle N_j \rangle$ . Alternatively, for a probe comprising  $n$  atomic qubits, each with a Pauli Z operator  $\sigma_z^{(j)}$ , one may consider the generator  $G = \sigma_z^{(1)} + \dots + \sigma_z^{(n)}$ , and powers thereof, corresponding to linear and nonlinear Ramsey interferometers, respectively [7,8,10,11]. Again,  $n$  quantifies the resources.

We note that this quantification of resources  $n$  is different from the  $N$  (which we will denote  $\mathcal{N}$ ) used in Refs. [14–18]. The  $n$  used here typically corresponds to the conspicuous physical resources required to generate the probe state and is what has previously been used to claim an advantage when using nonlinear interactions [5–13].

If  $\hat{\Phi}$  denotes an estimate of a shift parameter  $\Phi$  for some measurement scheme, then a standard measure [19,20] of the performance of the estimate is given by the average estimation error (called rms error in Ref. [19]),

$$\epsilon(\hat{\Phi}) := \sqrt{E[(\hat{\Phi} - \Phi)^2]}, \quad (1)$$

*Published by the American Physical Society under the terms of the [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.*

where the expectation value here is defined as

$$E[(\hat{\Phi} - \Phi)^2] \equiv \int d\phi d\hat{\phi} (\hat{\phi} - \phi)^2 p(\hat{\phi}|\phi) \wp(\phi). \quad (2)$$

Here,  $p(\hat{\phi}|\phi)$  is the probability density of the estimate conditioned on a fixed shift value  $\Phi = \phi$ , and  $\wp(\phi)$  denotes the prior probability density of the shift parameter. Measurement schemes which minimize the average estimation error, for given resources such as the average photon number or number of qubits available, are of fundamental interest in quantum metrology.

However, attention has often focused instead on minimizing a different quantity, the ‘‘local precision,’’ defined for a *fixed* value of the shift parameter,  $\Phi = \phi$ , by [21,22]

$$P_\phi(\hat{\Phi}) := \left\langle \left( \frac{\hat{\Phi}}{|d\langle \hat{\Phi} \rangle_\phi / d\phi|} - \phi \right)^2 \right\rangle_\phi, \quad (3)$$

where  $\langle \bullet \rangle_\phi$  denotes an average with respect to the conditional probability density  $p(\hat{\phi}|\phi)$ . Some proposed nonlinear measurement schemes can achieve local precisions that scale in terms of the number of resources  $n$  as, for example,  $n^{-3/2}$  [5,7,8,12,13],  $n^{-2}$  [10], or  $2^{-n}$  [9], for some value of  $\phi$ . Even so, we show below that the corresponding average estimation errors can scale no better than the usual Heisenberg scaling,  $n^{-1}$ .

For estimates that are, approximately, locally unbiased for *all* values of  $\Phi$  over some interval [23], one has

$$\int d\phi \wp(\phi) P_\phi(\hat{\Phi}) \approx \epsilon(\hat{\Phi}) \quad (4)$$

for shift parameters confined to this interval, providing a simple connection to the average estimation error. However, many phase estimates are unbiased only over very limited ranges [24], where these ranges are of widths comparable to the local precision itself. Thus, for example, while a high local precision of  $2^{-K}$  in some region may allow the  $K$ th binary digit of a phase shift to be estimated, it often does not allow the preceding digits to be estimated with any accuracy. These must either already be known (e.g., in phase-tracking [25] or phase-sensing [26] applications), which requires the prior probability distribution  $\wp(\phi)$  to be almost as narrow as the posterior distribution (after the measurement), or they must be determined using further resources. Hence, unless the phase is already very well known, the scaling of  $P_\phi(\hat{\Phi})$  may be a very poor guide to the scaling of  $\epsilon(\hat{\Phi})$ .

Indeed, whereas the local precision has a scaling lower bound set by the rms variance,  $\Delta G$ , of the generator for the probe state [21,22], the average estimation error has an asymptotically stronger (i.e., higher) lower bound, set by the entropy,  $H(G)$ , of the generator [27,28]. Thus, maximizing the variance, rather than the entropy, of  $G$ , does not typically minimize  $\epsilon(\hat{\Phi})$ . Here we further generalize and strengthen this entropic bound, in Secs. II and III, to

replace  $H(G)$  by the so-called  $G$  asymmetry of the probe state [29]. The fundamental role of this quantity is emphasized by showing that it also bounds the mutual information between the shift parameter  $\Phi$  and any estimate  $\hat{\Phi}$ . An important consequence demonstrated in Sec. IV is that, in a surprising contrast to the case of local precision, simply replacing  $G$  by some nonlinear function thereof, such as  $F = G^2$ , cannot improve the average estimation error or the information gain.

A careful analysis in Sec. V shows that nonlinearity can improve the scaling of  $\epsilon(\hat{\Phi})$  beyond  $n^{-1}$  for *iterative* implementations. These are implementations where the shift is applied on a sequence of probes of different sizes, so that  $G$  is replaced by a suitable sum of nonlinear generators. However, for a probe state comprising  $n$  qubits, it is shown that even adaptive variable-pass implementations of previously proposed nonlinear schemes can at best achieve scalings exponential in  $\sqrt{n}$  for the average estimation error. In contrast, in Sec. VI, we show that the best possible scaling for  $\epsilon(\hat{\Phi})$  is exponential in  $n$ , for both qubit and optical probes, regardless of whether the generator is linear or nonlinear. Moreover, an exponential scaling is in fact achievable via *linear* estimation schemes, if *multipass* implementations are allowed. Whether nonlinear schemes are more robust than linear schemes to the presence of loss is left as a question for future investigation.

## II. AN INFORMATION BOUND

The mutual information  $H(\hat{\Phi} : \Phi)$  between the shift parameter and its estimate is a measure of performance in its own right, quantifying the average number of bits obtained per estimate. A general upper bound for mutual information is obtained here, applicable to *any* generator  $G$  having a discrete spectrum, which will be used in Sec. III to obtain a lower bound for the average estimation error. Several useful properties of this bound are also established.

Consider a parameter  $\Phi$  with some prior distribution  $\wp(\phi)$ , and define an average prior state  $\bar{\rho} = \int d\phi \wp(\phi) \rho_\phi$ . Then, using the Holevo bound [30], one immediately has  $H(\hat{\Phi} : \Phi) \leq S(\bar{\rho}) - \int d\phi \wp(\phi) S(\rho_\phi) = S(\bar{\rho}) - S(\rho_0)$ . Here  $S(\rho) = -\text{tr}[\rho \ln \rho]$  denotes the von Neuman entropy of the state  $\rho$ . Now define

$$\mathcal{U}_G(\rho) := \sum_g \Pi_g \rho \Pi_g = \lim_{w \rightarrow \infty} \frac{1}{w} \int_0^w d\phi e^{-iG\phi} \rho e^{iG\phi}, \quad (5)$$

where  $\Pi_g$  is the projection onto the eigenspace corresponding to eigenvalue  $g$  of  $G$ , and the second equality may be checked by considering a basis diagonal in  $G$ . This map is unital, i.e., it maps the unit operator to itself. Using the nondecreasing property of von Neumann entropy under unital maps [30], together with  $\mathcal{U}_G(\bar{\rho}) = \mathcal{U}_G(\rho_0)$ , then yields the desired upper bound,

$$H(\hat{\Phi}: \Phi) \leq A_G(\rho_0) := S[\mathcal{U}_G(\rho_0)] - S(\rho_0), \quad (6)$$

for the mutual information.

The upper bound,  $A_G(\rho_0)$  in Eq. (6), may be recognized as the increase in quantum entropy due to an ideal measurement of  $G$  on the probe state, with postmeasurement state  $\mathcal{U}_G(\rho_0)$ . This entropy increase is relevant to bounding efficiencies in quantum thermodynamics [31]. More generally,  $A_G(\rho)$  represents the *asymmetry* of the state  $\rho$  with respect to a unitary group  $G$  (in this paper, the one-parameter Abelian group with Hermitian generator  $G$ ) [29]. The  $G$  asymmetry quantifies the degree to which  $\rho_0$  can break the symmetry of  $G$  (in this paper, the extent to which it carries information about the variable  $\Phi$  which is conjugate to  $G$ ) [29,32–34]. For the case where  $G$  has integer eigenvalues,  $A_G(\rho_0)$  quantifies to what extent  $\rho_0$  can act as a phase reference, an attribute clearly essential for detecting phase shifts. Note that, for  $G$  with incommensurate eigenvalue gaps, the corresponding group is noncompact, but expression (5) for  $\mathcal{U}_G$  allows one to generalize the  $G$  asymmetry (6) to this case also.

A form of Eq. (6) has been previously given for the special case of compact groups where the average prior state  $\bar{\rho}$  is symmetric with respect to the group [32–34]. For the case of a phase shift (i.e., a  $G$  with integer eigenvalues), this means a prior distribution  $\varphi(\phi)$  that is uniform over the unit circle. Equation (6) represents a generalization, for the case of one-parameter groups, to an arbitrary discrete generator  $G$  and arbitrary prior distributions of the shift parameter. Note that  $\Phi$  ranges over  $(-\infty, \infty)$  if  $e^{-iG\phi}$  is nonperiodic, corresponding to a noncompact group.

Several useful properties of  $A_G(\rho)$  are required below. First, if  $\rho$  is pure and/or  $G$  is nondegenerate, then the states  $\rho_g = \Pi_g \rho \Pi_g / p_g$ , with  $p_g = \text{tr}[\rho \Pi_g]$ , are pure and mutually orthogonal, and Eqs. (5) and (6) yield

$$A_G(\rho) = H(G|\rho) - S(\rho), \quad (7)$$

where  $H(G|\rho) = -\sum_g p_g \ln p_g$  is the entropy of the generator for state  $\rho$ . Second, one has the general bounds

$$A_{f(G)}(\rho) \leq A_G(\rho) \leq H(G|\rho), \quad (8)$$

$$A_G[\lambda\rho + (1-\lambda)\rho'] \leq \lambda A_G(\rho) + (1-\lambda)A_G(\rho'), \quad (9)$$

where  $f(G)$  is any function of  $G$ , and  $0 \leq \lambda \leq 1$ . The lower bound in Eq. (8) is saturated when  $f$  is 1:1, as may be seen by replacing  $G$  by  $f(G)$  and  $f$  by  $f^{-1}$ , while the upper bound is saturated for any pure state from Eq. (7). The convexity of  $A_G(\rho)$ , as per Eq. (9), implies that the  $G$  asymmetry is maximized for pure states.

The lower bound in (8) is obtained by noting that the eigenspaces of  $G$  are subspaces of the eigenspaces of  $f(G)$ , so that  $\mathcal{U}_G \circ \mathcal{U}_{f(G)} = \mathcal{U}_G$ , and by using the nondecreasing property of von Neumann entropy under unital maps [30] for the particular case  $\mathcal{U}_{f(G)}(\rho) \rightarrow (\mathcal{U}_G \circ \mathcal{U}_{f(G)})(\rho)$ . To obtain the upper bound, let  $|\psi\rangle$  be some purification of

$\rho$  on a tensor product of the probe Hilbert space with an ancilla  $a$ , so that  $\rho = \text{tr}_a[|\psi\rangle\langle\psi|]$ . Rewriting  $A_G(\rho)$  as  $\sum_g p_g S(\rho \| \rho_g)$ , where  $S(\rho \| \sigma) = \text{tr}[\sigma(\ln\sigma - \ln\rho)]$  denotes the relative entropy of  $\rho$  and  $\sigma$ , one then has

$$\begin{aligned} H(G|\rho) &= H(G \otimes 1 || \psi\rangle\langle\psi|) = A_{G \otimes 1}(|\psi\rangle\langle\psi|) \\ &= \sum_g p_g S[|\psi\rangle\langle\psi| \| (\Pi_g \otimes 1)|\psi\rangle\langle\psi|(\Pi_g \otimes 1)/p_g] \\ &\geq \sum_g p_g S(\rho \| \rho_g) = A_g(\rho) \end{aligned}$$

as desired, where the second equality follows from Eq. (7) and the inequality from the nonincreasing property of relative entropy under the operation of tracing over the ancilla [30]. Finally, Eq. (9) may be obtained via the representation  $A_G(\rho) = \lim_{w \rightarrow \infty} w^{-1} \int_0^w d\phi S(\rho \| \rho_\phi)$ , following from Eq. (5), and using the joint convexity property of the relative entropy [30].

Equations (6) and (8) imply, in particular, that the mutual information is bounded by the entropy of the generator for the probe state, i.e.,

$$H(\hat{\Phi}: \Phi) \leq H(G|\rho_0). \quad (10)$$

Thus, for example, for a generator having  $d$  distinct eigenvalues, no more than  $\ln d$  nats, i.e.,  $\log_2 d$  bits, of information can be extracted per probe state about the value of the shift parameter. (Note that “nats” are the units of mutual information when evaluated using the natural logarithm base  $e$ , whereas “bits” are the units corresponding to using the logarithm base 2 [35].)

### III. BOUNDS FOR RESOLUTION OF SHIFT PARAMETERS

#### A. Average estimation error

A strong bound for the average estimation error in Eq. (1) may be derived analogously to weaker bounds obtained in Refs. [28,36], i.e., by combining a quantum upper bound—such as Eq. (6)—for the mutual information with the classical lower bound [35,37]

$$H(\hat{\Phi}: \Phi) \geq H(\Phi) - \frac{1}{2} \ln[2\pi e \epsilon(\hat{\Phi})], \quad (11)$$

where  $H(\Phi) = -\int \varphi(\phi) \ln[\varphi(\phi)] d\phi$  denotes the entropy of the prior probability density  $\varphi(\phi)$  for  $\Phi$ . This lower bound is well known from rate-distortion theory and follows from the inequality chain [37]  $H(\hat{\Phi}: \Phi) = H(\Phi) - H(\Phi|\hat{\Phi}) = H(\Phi) - H(\Phi - \hat{\Phi}|\hat{\Phi}) \geq H(\Phi) - H(\Phi - \hat{\Phi}) \geq H(\Phi) - \frac{1}{2} \ln[2\pi e \epsilon(\hat{\Phi})]$ , where  $H(A|B)$  denotes the conditional entropy  $H(AB) - H(B)$ .

In particular, the combination of Eqs. (6) and (11) immediately yields the fundamental bound

$$\epsilon(\hat{\Phi}) \geq (2\pi e)^{-1/2} e^{H(\Phi)} e^{-A_G(\rho_0)} \quad (12)$$

for the average estimation error, for any discrete generator  $G$ . This bound both strengthens and generalizes previous

entropic bounds in the literature [17,27,28,36]. For example, Nair [28] and Yuen [36] use weaker upper bounds for the mutual information, corresponding to replacing  $A_G(\rho_0)$  in Eq. (12) by the quantum channel capacity under a fixed photon-number constraint. Hall and coworkers have previously obtained bounds in a different manner, based on entropic uncertainty relations, which correspond to replacing  $A_G(\rho)$  in Eq. (12) by the upper bound in Eq. (8) [17,27] (and, alternatively, by the upper bound in Eq. (7) for nondegenerate generators [27]), and replacing  $e^{H(\Phi)}$  by  $1/q_{\max}$ , where  $q_{\max}$  denotes the maximum value of  $\wp(\phi)$  [27].

Note that our bound (12) is applicable to iterative schemes, including adaptive ones, where the measurement performed on some probe-state components is dependent (in practice, through additional known phase rotations) on the outcomes of earlier measurements on other components [4]. This applicability holds because such a measurement scheme is formally equivalent to first applying shift generators  $G_1, G_2, \dots$  to respective probe components (e.g., qubits or optical modes), corresponding to applying the total generator  $G = G_1 + G_2 + \dots$ , and then performing the measurements sequentially (and adaptively).

### B. Local precision

A bound for the local precision in Eq. (3) follows via the quantum Cramer-Rao inequality [2,3] and has the form [21,22]

$$P_\phi(\hat{\Phi}) \geq (2\Delta G)^{-1}, \quad (13)$$

where  $\Delta G$  denotes the rms deviation of the (total) generator  $G$  for the probe state. Note that, taking the averages of  $\nu$  independent estimates, one obtains the usual statistical enhancement factor of  $1/\sqrt{\nu}$  for both of the bounds in (12) and (13). This will not be discussed further here, other than to remark that, although the latter bound for  $P_\phi(\hat{\Phi})$  may be asymptotically achievable as  $\nu \rightarrow \infty$  [21,22,24,25], this does not imply anything about the achievability of the corresponding bound for  $\epsilon(\hat{\Phi})$ .

### C. Comparisons

For generators with integer eigenvalues, i.e., phase-shift generators, the scaling of  $\epsilon(\hat{\Phi})$  with the exponentiated  $G$  asymmetry  $e^{-A_G(\rho_0)}$  in Eq. (12) implies a scaling with the rms error  $\Delta G$  which is at least as strong as that for  $P_\phi(\hat{\Phi})$  in Eq. (13) (ignoring multiplicative constants of order unity). This is a consequence of the inequality chain

$$e^{-A_G(\rho)} \geq e^{-H(G|\rho)} \geq (2\pi e)^{-1/2} [(\Delta G)^2 + 1/12]^{-1/2} \quad (14)$$

for such generators, where the first inequality follows from Eq. (8) and the second is well known [37,38]. For the case of a completely unknown phase shift, with  $\wp(\phi) = 1/(2\pi)$ , Eqs. (12) and (14) yield the asymptotic lower

bound  $\epsilon(\hat{\Phi}) \geq (e\Delta G)^{-1}$  for the average estimation error, which is comparable to the lower bound (13) for local precision. However, importantly, the bound in Eq. (12) is significantly more powerful, as we now show.

Consider, for example, a probe comprising  $n$  qubits, and generator  $G = \sigma_z^{(1)} + \dots + \sigma_z^{(n)}$ . The pure probe state  $(|z, z, \dots, z\rangle + |-z, -z, \dots, -z\rangle)/\sqrt{2}$ , where  $|\pm z\rangle$  denotes the eigenstates of  $\sigma_z$ , then gives the maximum possible value  $\Delta G = n$  in Eqs. (13) and (14). That is why this state, equivalent to a NOON state or GHZ state [39], is often considered in quantum metrology. However, the corresponding  $G$  asymmetry follows via Eq. (7) as only  $A_G(\rho_0) = H(G|\rho_0) = \ln 2$ , implying via Eq. (12) that the average estimation error does not decrease at all as a function of  $n$ . The only way an average estimation error scaling as  $n^{-1}$  would be possible from this state would be if there were sufficient prior information, that is, from Eq. (12), if  $-H(\Phi) + A_G(\rho_0)$  were of order  $\ln n$ . But since  $A_G(\rho_0) = \ln 2$ , this means  $-H(\Phi) \sim \ln n$  itself. Hence, the amount of prior information about the parameter to be estimated would already be sufficient to locate it with the precision achievable by the measurement.

It is thus apparent that the lower bounds (12) and (13) can exhibit markedly different behavior, with the former bound having an asymptotically stronger scaling in general. It follows that probe states generating optimal scaling for  $P_\phi(\hat{\Phi})$ , obtained by maximizing  $\Delta G$  under various constraints for some value of  $\phi$ , do not necessarily correspond to optimal bounds for  $\epsilon(\hat{\Phi})$ . Since it is the latter quantity that has direct operational significance for the performance of the estimate, this finding has crucial implications for some nonlinear estimation schemes proposed in the literature, as will be seen below.

## IV. RESOLUTION PUZZLE: NONLINEARITY VS $G$ ASYMMETRY

### A. Probes comprising $n$ qubits

Several nonlinear phase-estimation schemes have been proposed for Ramsey interferometry, based on a probe state comprising  $n$  qubits [7–11]. For example, defining  $J_z := \sigma_z^{(1)} + \dots + \sigma_z^{(n)}$ , the generator  $(J_z)^q$  has been considered in Refs. [7,8,11] for  $q = 2, 3, \dots$ , and the generator  $nJ_z$  in Ref. [10]. A nonlinearity equivalent to  $J_z^2$ , although defined in terms of a fixed number of bosons shared between two modes, was considered in Ref. [5]. Furthermore, the generators  $H$  and  $A$  defined via

$$H + iA = (\sigma_x^{(1)} + i\sigma_y^{(1)}) \otimes \dots \otimes (\sigma_x^{(n)} + i\sigma_y^{(n)}) \quad (15)$$

were considered in Ref. [9].

Now, the linear generator  $G = J_z$  has  $n + 1$  distinct eigenvalues,  $-n, -n + 2, \dots, n$ , and hence  $A_G(\rho) \leq H(G|\rho) \leq \ln(n + 1)$ . It follows immediately from Eq. (12) (see also Eq. (22) of Ref. [27]), that the corresponding average estimation error can scale no better than



$(n+1)^{-1}$  with qubit number, corresponding asymptotically to the Heisenberg scaling limit of  $n^{-1}$ .

However, noting from Eq. (8) that  $A_{f(G)}(\rho) \leq A_G(\rho)$ , for any function  $f$ , precisely the same conclusion follows for the nonlinear generators  $G = (J_z)^q$  and  $G = nJ_z$ . That is, the average estimation error cannot achieve better than Heisenberg scaling for these generators. Moreover, the nonlinear generators  $G = H$  and  $G = A$  of Eq. (15) do not even allow the possibility of Heisenberg scaling, as they each have only three distinct eigenvalues, 0 and  $\pm 2^{n-1}$  [40].

The above results are in stark contrast to the best possible scalings of local precision for these schemes, which improve on Heisenberg scaling, with  $n^{-q}$  for  $G = (J_z)^q$  [5,7,8,11];  $n^{-2}$  for  $G = nJ_z$  [10]; and  $2^{-n}$  for  $G = H$  or  $A$  [9]. This difference in scalings immediately raises a conundrum: How can nonlinearity improve the local precision yet not the average estimation error?

This puzzle may be further deepened by noting that the probe states yielding *optimal* local precisions are generally an equally weighted superposition of two orthogonal eigenstates of  $G$ , corresponding to the maximum and minimum eigenvalues of the generator [7,8,10,11]. Thus,  $A_G(\rho_0) = \ln 2$  for such a probe state, implying that the average estimation error cannot decrease with  $n$  at all, as discussed in Sec. III C above. Indeed, Eq. (10) implies that no more than 1 bit of information about the phase shift can be gained via such an “optimal” probe state.

### B. Probes comprising optical modes

An analogous puzzle holds for optical probes. As a simple example, if  $G = N$  is the number operator for a single-mode field, then  $H(G|\rho) \leq \ln(e\langle N+1 \rangle)$ , implying from Eqs. (8) and (12) that the average estimation error can scale no better than  $\langle N+1 \rangle^{-1}$ . (See also Refs. [17,27,28,36].) But for any nonlinear generator  $G = f(N)$ ,  $A_G(\rho) \leq A_N(\rho) \leq H(N|\rho)$  from Eq. (8). Hence, the same scaling bound also applies to nonlinear generators for single-mode fields.

In contrast, nonlinearity can significantly enhance the local precision. For example, choosing the coherent probe state  $\rho_0 = |\alpha\rangle\langle\alpha|$  and nonlinear generator  $G = N^2$ ,  $P_\phi(\hat{\Phi})$  can scale as  $\langle N \rangle^{-3/2}$  for large  $\langle N \rangle$  [5]. For this case, the photon-number distribution is Poissonian, which is well approximated by a Gaussian distribution for large  $\langle N \rangle$ . Thus, using Eq. (7),  $A_G(\rho_0) = H(N^2|\rho_0) = H(N|\rho_0) \approx (1/2) \ln(2\pi e\langle N \rangle)$ , implying via Eq. (12) that the corresponding average estimation error can decrease with  $\langle N \rangle$ . However, it cannot scale even as well as the Heisenberg limit,  $\langle N \rangle^{-1}$  but rather is lower bounded by the standard quantum-limit scaling,  $\langle N \rangle^{-1/2}$ .

These examples again lead to the puzzle that, while nonlinearity can improve the scaling of the local precision, it cannot, by itself, influence the scaling of the average estimation error. This raises the question: Can

nonlinear schemes offer any advantage over linear schemes?

## V. PUZZLE RESOLUTION: ITERATIVE SCHEMES

It has been seen that a simple replacement of a generator by a nonlinear function thereof cannot lead to an improved scaling of the average estimation error, in marked contrast to the situation regarding the local precision. However, a careful analysis shows that, with a suitable sum of nonlinear generators, the bounds in Eqs. (12) and (6) allow for an enhanced scaling of  $\epsilon(\hat{\Phi})$ , and that this enhanced scaling could, plausibly, be achievable by adaptive measurements. This analysis resolves the above puzzle to some degree. Significantly, however, the scaling of the average estimation error does not necessarily achieve the same scaling as the local precision.

As a first example, let  $G(l)$  denote the nonlinear generator  $(J_z)^2$ , for  $l$  qubits, and let  $\rho(l)$  denote an equally weighted superposition of two eigenstates of  $G(l)$ , corresponding to its minimum and maximum eigenvalues (i.e., to 0 and  $l^2$  if  $l$  is even, and 1 and  $l^2$  if  $l$  is odd). Now consider the total generator and corresponding composite probe state defined by

$$G_{\text{it}} := \sum_{k=1}^K G(n_k), \quad \rho_0 := \otimes_{k=1}^K \rho(n_k), \quad (16)$$

with  $n_k := \lceil 2^{(k-1)/2} \rceil$  (where  $\lceil x \rceil$  denotes the smallest integer not less than  $x$ ). Since the phase shift generated by  $G(l)$  has period approximately equal to  $2\pi/l^2$ , this ensures that the phase shift generated by  $G(n_k)$ , on the probe state component  $\rho(n_k)$ , has period approximately equal to  $(2\pi)/2^{k-1}$ .

The basic idea is that the  $k$ th bit in a binary expansion of  $\Phi/(2\pi)$  is estimated from the  $k$ th component of the probe state. Note that it is impossible to obtain more than 1 bit from each component of the probe state, i.e., more than  $\ln 2$  nats, as a consequence of Eq. (6) and the property  $A_{G(l)}[\rho(l)] = H[G(l)|\rho(l)] = \ln 2$ . This is the idea behind the famous quantum phase-estimation algorithm [30] for linear phase shifts, subsequently generalized in Refs. [14–16]. In practice, to achieve the best scaling for the average estimation error, it may be necessary to use  $M > 1$  copies of each component of the probe state to estimate each bit accurately, when combined with an adaptive measurement sequence [14,15]. Counterintuitively, it is the least significant ( $K$ th) bit of  $\Phi/(2\pi)$  that should be determined first, to allow the optimal measurement of the  $(K-1)$ th bit, and so on, up to the most significant bit.

The total number of qubits required in the above setup is  $n = M \sum_k n_k \approx M(2^{K/2} - 1)/(\sqrt{2} - 1)$ . Furthermore, there are  $2^K$  distinct eigenvalues of  $G$ , of the form approximately equal to  $\sum_k b_k 2^{k-1}$  for  $b_k = 0$  or 1, where these have a uniform distribution for  $\rho_0$ . Taking into

account the  $M$  copies of  $G_{\text{it}}$  and  $\rho_0$ , the corresponding generator  $G = G_{\text{it}}^{(1)} + \dots + G_{\text{it}}^{(M)}$  has eigenvalues ranging from 0 to  $M(2^K - 1)$ , implying a  $G$  asymmetry  $A_G(\otimes^M \rho_0) \leq H(G|\otimes^M \rho_0) \leq \ln[M2^K]$ . (Note that for  $M > 1$  the distribution of  $G$  over  $\otimes^r \rho_0$  is not uniform, and this upper bound is not tight.) Thus Eq. (12) yields the following lower bound for the average estimation error,

$$\epsilon(\hat{\Phi}) \geq \frac{(2\pi/e)^{1/2}}{M[1 + (\sqrt{2} - 1)n/M]^2} \approx \frac{(2\pi/e)^{1/2} M}{(3 - 2\sqrt{2}) n^2}, \quad (17)$$

in the worst-case scenario of a completely random phase shift. Note that this bound is compatible with the scaling expected for a scheme that determines the first  $K$  bits of  $\hat{\Phi}$ , giving  $\epsilon(\hat{\Phi}) \leq (2\pi)/2^{K+1} \sim (M/n)^2$ . In other words, for  $M$  large enough for this bitwise estimation scheme to work, we would expect the scaling with  $n$  in Eq. (17) to be attainable.

Thus, this adaptive scheme demonstrates the possibility of an asymptotic  $n^{-2}$  scaling for the average estimation error. This is the same scaling (up to a constant factor) as for the optimal local precision for the generator  $G(n)$  [7,11]. Furthermore, an  $n^{-q}$  asymptotic scaling can be obtained for an analogous adaptive scheme based on the nonlinear generator  $(J_z)^q$ , with

$$\epsilon(\hat{\Phi}) \geq \frac{c_q M^{q-1}}{n^q} \quad (18)$$

for a phase shift random over  $[0, 2\pi)$ , where  $c_q$  increases exponentially with  $q$ . Analogous results may be obtained for iterative implementations of the schemes in Refs. [5,6].

However, a correspondence of scalings between  $P_\phi(\hat{\Phi})$  and  $\epsilon(\hat{\Phi})$  does not hold more generally. For example, let  $G(l)$  instead denote the nonlinear generator  $H + 2^{l-1}$  for  $l$  qubits, with  $H$  as in Eq. (15) but with  $l$  in place of  $n$ , and with the additive constant being chosen to simplify eigenvalue counting. Furthermore, let  $\rho(l)$  denote an equally weighted superposition of the two eigenstates corresponding to the minimum and maximum eigenvalues 0 and  $2^l$  of  $G(l)$  (noting that such superpositions include the separable states  $|z, z, \dots, z\rangle$  and  $|-z, -z, \dots, -z\rangle$  [9,40]). Successive estimation of  $l$  binary digits then corresponds to  $n_k = k - 1$  in Eq. (16), requiring a total qubit number  $n = MK(K - 1)/2 \approx MK^2/2$ . One has  $A_G(\otimes^M \rho_0) \leq \ln(M2^K)$  as before, yielding the lower bound

$$\epsilon(\hat{\Phi}) \geq (2\pi/e)^{1/2} M^{-1} 2^{-\sqrt{2n/M}} \quad (19)$$

for the case of a completely random phase shift. Thus the scaling with  $n$  is considerably worse than the  $2^{-n}$  scaling of the local precision for the corresponding single-generator scheme [9].

This last result demonstrates that the local precision does not necessarily characterize the performance of the average estimation error even for adaptive implementations.

It follows that comparisons between various schemes, whether linear or nonlinear, should be made on the basis of the operationally significant quantity,  $\epsilon(\hat{\Phi})$  in Eq. (1), rather than  $P_\phi(\hat{\Phi})$  in Eq. (3).

## VI. LINEAR SCHEMES WITH OPTIMAL SCALING

### A. Probes comprising $n$ qubits

Since any generator  $G$  for  $n$  qubits has at most  $2^n$  distinct eigenvalues, it follows from Eqs. (7) and (8) that  $A_G(\rho_0) \leq H(G|\rho_0) \leq n \ln 2$ , with equality for a probe state that is an equally weighted superposition of the corresponding eigenstates. Hence, from Eq. (12), the best possible scaling for the average estimation error satisfies

$$\epsilon(\hat{\Phi}) \geq (2\pi e)^{-1/2} e^{H(\Phi)} 2^{-n}. \quad (20)$$

Estimation schemes having an exponential scaling linear in  $n$  are therefore of fundamental interest. Note, per Eq. (19), that such a scaling is *not* attained via an adaptive implementation of the nonlinear scheme in Ref. [9], despite the local precision scaling as  $2^{-n}$  for this scheme.

It is possible that nonlinear schemes exist with an exponential scaling in  $n$  for  $\epsilon(\hat{\Phi})$ . Here we show that, surprisingly, such a scaling can be achieved with *linear* generators. The extra ingredient that makes this possible is to allow for multiple (and varied) applications of the phase shift prior to measurement.

In particular, following the ideas in Higgins *et al.* [14], consider a probe system comprising  $m$  unentangled qubits, each in the state  $|+\rangle = (| - z\rangle + |z\rangle)/\sqrt{2}$ , where the  $k$ th qubit is subjected to  $2^{k-1}$  applications of a linear phase shift generated by  $(1 + \sigma_z^{(k)})/2$ . In Ref. [14], this was achieved experimentally via multiple passes through a medium. Another possibility would be to suitably increase interaction times of the qubits with the phase-shift medium. The total generator and the probe state therefore have the forms

$$G_{\text{it}} = \sum_{k=1}^K 2^{k-1} (1 + \sigma_z^{(j)})/2, \quad \rho_0 = \otimes^K |+\rangle\langle +|. \quad (21)$$

The total phase shift of the  $k$ th qubit thus has period  $2\pi/2^{k-1}$ , and so can be used to estimate the  $k$ th bit of  $\Phi/(2\pi)$ , in the adaptive manner explained above [14].

If, as for the nonlinear schemes above,  $M$  copies are used to estimate each bit accurately, then the total number of qubits required is  $n = MK$ . Following the style of arguments used in the preceding section gives a lower bound on the average estimation error of

$$\epsilon(\hat{\Phi}) \geq (2\pi e)^{-1/2} e^{H(\Phi)} M^{-1} 2^{-n/M}. \quad (22)$$

In this case, it appears that minimizing  $M$  could make a big improvement to the precision, and for  $M = 1$  the ultimate scaling (20) might be achievable. However, it must be remembered that Eq. (22) is merely a lower bound. Moreover, from the arguments in the preceding section, it

is only for  $M$  sufficiently large that we expect these bitwise estimation schemes to attain the scaling with  $n$  of the lower bounds.

Luckily, in this case, we can compare this bound to the actual performance of the best-known adaptive schemes for an initially completely random phase, as this has been studied extensively. These studies were done using the Holevo variance  $V_H(\hat{\Phi})$  [41] rather than the average estimation error, but when these are small (as here, for large  $n$ ),  $V_H(\hat{\Phi}) \leq \epsilon(\hat{\Phi})^2 \leq (\pi/2)^2 V_H(\hat{\Phi})$  [42]. In terms of scaling with  $n$ , the best performance is indeed for  $M = 1$ , which corresponds to the quantum phase-estimation algorithm [30] and yields [15]

$$\epsilon(\hat{\Phi}) = c \times 2^{-n/2}. \quad (23)$$

(The constant  $c \approx 1.18$  can be evaluated by performing the integral of the distribution of phase estimates, Eq. (4.5) of Ref. [15].) Although this is not identical scaling to (20), it is still exponential in  $n$ , unlike (19). For  $M = 2, 3$ , and  $4$ , the performance scales as  $2^{-n/4}$ , while, for  $M \geq 4$ , it scales as  $2^{-n/M}$ , achieving the lower-bound scaling in Eq. (22) as expected for sufficiently large  $M$ .

Note that, in terms of  $\mathcal{N} = M(2^{K+1} - 1)$ , the number of qubit passes through the phase shift (which is the resource considered in Refs. [14–17]), the change in scaling as  $M$  increases appears quite different. For  $M = 1$  and  $M = 2$ , the scaling is  $\mathcal{N}^{-1/2}$ ; for  $M = 3$ , it is  $\mathcal{N}^{-3/4}$ ; and, for  $M \geq 4$ , it is  $\mathcal{N}^{-1}$ . We emphasize that counting resources as above, in terms of the number of qubits  $n$ , is necessary to enable consistent comparison with the nonlinear schemes considered in Refs. [7–11]. Recently, some other papers have also considered the number of qubits (or, more strictly, the number of qubit measurements) as a resource [43,44]. However, as these studies were motivated by qubit-gate characterization in solid-state quantum computing, they imposed the constraint that the qubit measurement basis is fixed. Under this constraint, the only thing that can be chosen adaptively is the number of times the phase shift is applied to a given qubit prior to measurement. In Ref. [43], numerical evidence was presented that, using a locally optimal (“greedy”) adaptive algorithm, a scaling of approximately  $2^{-0.1n}$  is achievable. In Ref. [44], an analytical argument was given suggesting that a scaling of approximately  $2^{-0.16n}$  should be achievable. Neither achieves the scaling (23) of schemes that allows adaptive, controlled qubit rotations prior to measurement.

### B. Probes comprising optical modes

For an optical probe containing  $m$  orthogonal modes, let  $N_m$  denote the photon number of the  $m$ th mode, and  $N$  denote the total photon number  $N_1 + \dots + N_m$ . The entropy of any generator  $G = f(N_1, \dots, N_m)$  is bounded above by the joint entropy of  $N_1, \dots, N_m$  (since the distribution of  $G$  is a coarse graining of the joint distribution), yielding via Eq. (8)

$$\begin{aligned} A_G(\rho) &\leq H(G|\rho) \leq H(N_1, \dots, N_m|\rho) \\ &\leq m \ln \left[ 1 + \frac{\langle N \rangle}{m} \right] + \langle N \rangle \ln \left[ 1 + \frac{m}{\langle N \rangle} \right] \end{aligned} \quad (24)$$

$$\leq \ln e^{m+\langle N \rangle}. \quad (25)$$

The second line follows from standard statistical mechanics techniques, and the third line from the monotonic convergence of  $(1 + x/y)^y$  to  $e^x$  as  $y$  increases.

Using Eq. (12), the average estimation error therefore has the fundamental lower bound

$$\epsilon(\hat{\Phi}) \geq (2\pi e)^{-1/2} e^{H(\Phi)} e^{-m} e^{-\langle N \rangle}, \quad (26)$$

for any generator that is a function of  $N_1, \dots, N_m$ . Such functions include the total photon number,  $N = N_1 + \dots + N_m$ , in particular [36], but also include, for example, the nonlinear generators  $N^2$  and  $(N_1)^2 + \dots + (N_m)^2$ . Similarly, using Eq. (6), one has the fundamental upper bound  $m + \langle N \rangle$  for the mutual information  $H(\hat{\Phi}: \Phi)$ .

It follows that estimation schemes with an exponential scaling linear in the average photon number are of fundamental interest. Furthermore, a linear scheme, analogous to the one above for  $n$  qubits, is sufficient to obtain such a scaling (although with a different coefficient). In particular, the  $M = 1$  linear-multipass scheme of Higgins *et al.* [14], equivalent to the quantum phase-estimation algorithm [30], is precisely such a scheme, involving  $m = 2K$  modes, with each pair of modes in the superposition state  $(|0\rangle|1\rangle - |1\rangle|0\rangle)/\sqrt{2}$ . Hence,  $K = \langle N \rangle = m/2$ , and, from Eq. (23),  $\epsilon(\hat{\Phi}) \approx 1.18 \times e^{-K \ln(2)/2}$  asymptotically, which is consistent with Eq. (26) as here  $m + \langle N \rangle = 3K$ . Again, it should be noted, as for the qubit case above, the measure of resources considered here is the total mean photon number  $\langle N \rangle$  required for the scheme, rather than the number of photon passes  $\mathcal{N}$  through the phase-shift medium as in Ref. [14]. Again, we use  $\langle N \rangle$  to enable comparison with various linear and nonlinear estimation schemes [5,6]. Moreover, photon number is a natural measure of interest to consider, as it characterizes the energy resources required for a given optical scheme.

## VII. DISCUSSION

The average estimation error and mutual information have been shown to satisfy the general entropic bounds in Eqs. (6) and (12), for any shift generator having a discrete spectrum, and for any prior distribution of the shift parameter. While, for phase-shift generators, the  $G$  asymmetry can be bounded above in terms of the variance of the generator, via Eq. (14), the  $G$  asymmetry is typically much less than this upper bound. Hence, the average estimation error can scale very differently to the local

precision in Eq. (13), in terms of available resources such as number of qubits or total-input photon number.

Indeed, somewhat surprisingly, a simple replacement of a linear generator by some nonlinear function thereof may have no effect on the average estimation error yet lead to a marked improvement of the local precision. Furthermore, while such scaling differences can disappear for iterative estimation schemes, this is not always the case.

It follows that the optimal scaling of the local precision, for a given value of the shift parameter, should be treated with some caution. As noted in relation to Eq. (4), the local precision is a direct measure of the average rms error only over an interval for which the corresponding estimate is (approximately) unbiased. However, many estimators in the literature are unbiased only over a very small interval, similar in magnitude to the local precision itself. In such a case (which is relevant, for example, in phase-tracking and phase-sensing applications [25,26]), for the optimal scaling to be achievable, the amount of prior information required about the shift parameter is typically so great that the estimate itself can only extract up to 1 bit of further information, irrespective of the number of resources  $n$ . Examples of this phenomenon have been given in Secs. III C and IV

It is concluded from the above discussion that meaningful comparisons between various estimation schemes are most easily made on the basis of the operationally significant quantity,  $\epsilon(\hat{\Phi})$  in Eq. (1), rather than  $P_\phi(\hat{\Phi})$  in Eq. (3). Alternatively, if  $P_{\phi'}(\hat{\Phi})$  for some  $\phi'$  is used, then it should be supplemented by the interval over which this estimate is (approximately) locally unbiased, i.e., the interval for which the local precision of the estimate corresponds to the actual rms error of the estimate,  $\langle(\hat{\Phi} - \phi)^2\rangle_\phi$ . Note that the width of this interval also bounds the width of any “sweet spot” for which  $P_\phi(\hat{\Phi}) = P_{\phi'}(\hat{\Phi})$ , and hence bounds the width of the prior phase-probability density required to ensure that a precision of  $P_{\phi'}(\hat{\Phi})$  is actually achieved via measurement. Per Eq. (4), the estimation error averaged over this prior distribution will then be (approximately) equal to  $P_{\phi'}(\hat{\Phi})$ .

Universal lower bounds for the average estimation error have also been given, in terms of the number of qubits or photons available, in Eqs. (20) and (26). Like Eq. (12), these bounds are independent of the form of the generator, and hence apply equally well to both linear and nonlinear schemes, including multipass schemes. They imply that it is impossible to achieve a scaling better than exponential, in terms of qubit number, and in terms of input photon number plus number of modes, respectively. Furthermore, exponential scalings *can* be attained by multipass linear estimation schemes, as has indeed been shown experimentally for optical phase shifts [14]. It follows that, in terms of the best possible scaling that can be achieved relative to qubit or input photon number, nonlinear schemes offer no

fundamental advantage over linear schemes if multiple passes are possible.

However, a practical advantage of nonlinear schemes may be a greater robustness to loss. For example, multipass linear schemes of the type discussed above will be highly sensitive to loss due to multiple (or longer) interactions with the phase-shift medium. Thus it would be interesting to find alternative physical implementations of the generator in Eq. (21) and its optical analogue. Furthermore, while only shot-noise scaling is achievable for simple linear schemes in the presence of loss [1,45,46] (including for the average estimation error [28]), there is evidence that this may not be the case for nonlinear schemes [6]. Hence, further investigation is required, including the determination of fundamental scaling bounds for lossy schemes analogous to Eqs. (20) and (26).

It would also be of interest to investigate the degree to which results generalize to the case of a generator with a continuous spectrum, such as spatial translations generated by a momentum operator. For example, the fundamental bounds (20) and (26) are universal, since  $n \ln 2$  and  $(m + \langle N \rangle) \ln e$  are respective upper bounds for mutual information, following via the Holevo bound. The possibility of other generalizations is supported by results such as a universal Heisenberg-type scaling for the average estimation error, in terms of  $\langle |G| \rangle$ , which holds for both discrete *and* continuous generators [27]. Furthermore, weaker measurement-dependent entropic bounds on the average estimation error, given in Ref. [27], may prove helpful. Note that some differences are to be expected regarding linearity vs nonlinearity for the continuous case, since, for example, the property  $H(f(G)|\rho) \leq H(G|\rho)$  no longer holds in general.

## ACKNOWLEDGMENTS

We thank D. W. Berry, R. Nair, J. A. Vaccaro and M. Zwierz for helpful discussions. This work was supported by the ARC Discovery Project DP0986503.

- 
- [1] V. Giovannetti, S. Lloyd, and L. Maccone, *Advances in Quantum Metrology*, *Nature Photon.* **5**, 222 (2011).
  - [2] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
  - [3] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
  - [4] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control* (Cambridge University Press, Cambridge, England, 2010).
  - [5] A. Luis, *Nonlinear Transformations and the Heisenberg Limit*, *Phys. Lett. A* **329**, 8 (2004).
  - [6] J. Beltran and A. Luis, *Breaking the Heisenberg Limit with Inefficient Detectors*, *Phys. Rev. A* **72**, 045801 (2005).
  - [7] S. Boixo, S. T. Flammia, C. M. Caves, and J. M. Geremia, *Generalized Limits for Single-Parameter Quantum Estimation*, *Phys. Rev. Lett.* **98**, 090401 (2007).



- [8] S. Choi and B. Sundaram, *Bose-Einstein Condensate as a Nonlinear Ramsey Interferometer Operating beyond the Heisenberg Limit*, *Phys. Rev. A* **77**, 053613 (2008).
- [9] S. M. Roy and S. L. Braunstein, *Exponentially Enhanced Quantum Metrology*, *Phys. Rev. Lett.* **100**, 220501 (2008).
- [10] S. Boixo, A. Datta, M. J. Davis, S. T. Flammia, A. Shaji, and C. M. Caves, *Quantum Metrology: Dynamics versus Entanglement*, *Phys. Rev. Lett.* **101**, 040403 (2008).
- [11] S. Boixo, A. Datta, S. T. Flammia, A. Shaji, E. Bagan, and C. M. Caves, *Quantum-Limited Metrology with Product States*, *Phys. Rev. A* **77**, 012317 (2008).
- [12] M. Napolitano and M. W. Mitchell, *Nonlinear Metrology with a Quantum Interface*, *New J. Phys.* **12**, 093016 (2010).
- [13] M. Napolitano, M. Koschorreck, B. Dubost, N. Behbood, R. J. Sewell, and M. W. Mitchell, *Interaction-Based Quantum Metrology Showing Scaling beyond the Heisenberg Limit*, *Nature (London)* **471**, 486 (2011).
- [14] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde, *Entanglement-Free Heisenberg-Limited Phase Estimation*, *Nature (London)* **450**, 393 (2007).
- [15] D. W. Berry, B. L. Higgins, S. D. Bartlett, M. W. Mitchell, G. J. Pryde, and H. M. Wiseman, *How to Perform the Most Accurate Possible Phase Measurements*, *Phys. Rev. A* **80**, 052114 (2009).
- [16] B. L. Higgins, D. W. Berry, S. D. Bartlett, M. W. Mitchell, H. M. Wiseman, and G. J. Pryde, *Demonstrating Heisenberg-Limited Unambiguous Phase Estimation without Adaptive Measurements*, *New J. Phys.* **11**, 073023 (2009).
- [17] M. J. W. Hall, D. W. Berry, M. Zwierz, and H. M. Wiseman, *Universality of the Heisenberg Limit for Estimates of Random Phase Shifts*, *Phys. Rev. A* **85**, 041802(R) (2012).
- [18] M. Tsang, *Ziv-Zakai Error Bounds for Quantum Parameter Estimation*, *Phys. Rev. Lett.* **108**, 230401 (2012).
- [19] H. L. Van Trees, *Detection, Estimation and Modulation Theory Part I* (John Wiley, New York, 2001), Sec. 2.4.
- [20] For the case of phase estimates, the range of the error  $\hat{\Phi} - \Phi$  is defined to be  $[-\pi, \pi)$  to remove the phase-reference ambiguity on the unit circle, implying that  $0 \leq \epsilon(\hat{\Phi}) \leq \pi$ .
- [21] S. L. Braunstein and C. M. Caves, *Statistical Distance and the Geometry of Quantum States*, *Phys. Rev. Lett.* **72**, 3439 (1994).
- [22] S. L. Braunstein, C. M. Caves, and G. J. Milburn, *Generalized Uncertainty Relations: Theory, Examples, and Lorentz Invariance*, *Ann. Phys. (N.Y.)* **247**, 135 (1996).
- [23] An estimate is locally unbiased, for a fixed phase-shift value  $\Phi = \phi$ , if  $\langle \hat{\Phi} \rangle = \phi$  and  $d\langle \hat{\Phi} \rangle/d\phi = 1$ .
- [24] D. W. Berry, M. J. W. Hall, M. Zwierz, and H. M. Wiseman, *Optimal Heisenberg-Style Bounds for the Average Performance of Arbitrary Phase Estimates*, [arXiv:1209.3547v1](https://arxiv.org/abs/1209.3547v1).
- [25] G. A. Durkin and J. P. Dowling, *Local and Global Distinguishability in Quantum Interferometry*, *Phys. Rev. Lett.* **99**, 070801 (2007).
- [26] G. Y. Xiang, B. L. Higgins, D. W. Berry, H. M. Wiseman, and G. J. Pryde, *Entanglement-Enhanced Measurement of a Completely Unknown Optical Phase*, *Nature Photon.* **5**, 43 (2010).
- [27] M. J. W. Hall and H. M. Wiseman, *Heisenberg-Style Bounds for Arbitrary Estimates of Shift Parameters Including Prior Information*, *New J. Phys.* **14**, 033040 (2012).
- [28] R. Nair, *Fundamental Limits on the Accuracy of Optical Phase Estimation from Rate-Distortion Theory*, [arXiv:1204.3761v1](https://arxiv.org/abs/1204.3761v1).
- [29] J. A. Vaccaro, F. Anselmi, H. M. Wiseman, and K. Jacobs, *Tradeoff between Extractable Mechanical Work, Accessible Entanglement, and Ability to Act as a Reference System, Under Arbitrary Superselection Rules*, *Phys. Rev. A* **77**, 032114 (2008).
- [30] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2000).
- [31] S. Lloyd, *Quantum-Mechanical Maxwell's Demon*, *Phys. Rev. A* **56**, 3374 (1997).
- [32] J. A. Vaccaro, in *Proceedings of the 8th International Conference on Quantum Communication, Measurement and Computing*, edited by O. Hirota, J. H. Shapiro, and M. Sasaki (NICT, Tokyo, 2006), pp. 421–424.
- [33] G. Gour, I. Marvian, and R. W. Spekkens, *Measuring the Quality of a Quantum Reference Frame: The Relative Entropy of Frameness*, *Phys. Rev. A* **80**, 012307 (2009).
- [34] J. A. Vaccaro, *Particle-Wave Duality: A Dichotomy between Symmetry and Asymmetry*, *Proc. R. Soc. A* **468**, 1065 (2011).
- [35] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley, New Jersey, 2006), 2nd ed., Corollary to Theorem 8.6.6.
- [36] H. P. Yuen, in *Proceedings of the Workshop on Squeezed States and Uncertainty Relations*, NASA Conference Publication No. 3135 (NASA, Goddard Space Flight Center, MD, 1992), pp. 13–21 ([http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19920012804\\_1992012804.pdf](http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19920012804_1992012804.pdf)).
- [37] P. Harremoës, *Probability and Information—Occam's Razor in Action* (unpublished); see also <http://www.math.ku.dk/~moes/bog4.pdf>, ch. 12.
- [38] W. H. Mow, *A Tight Upper Bound on Discrete Entropy*, *IEEE Trans. Inf. Theory* **44**, 775 (1998).
- [39] P. Kok, H. Lee, and J. P. Dowling, *Creation of Large-Photon-Number Path Entanglement Conditioned on Photodetection*, *Phys. Rev. A* **65**, 052104 (2002); D. Leibfried, M. D. Barrett, T. Schaetz, J. Britton, J. Chiaverini, W. M. Itano, J. D. Jost, C. Langer, and D. J. Wineland, *Toward Heisenberg-Limited Spectroscopy with Multiparticle Entangled States*, *Science* **304**, 1476 (2004).
- [40] It is straightforward to verify that  $(|z, \dots, z\rangle \pm | -z, \dots, -z\rangle)/\sqrt{2}$  and  $(|z, \dots, z\rangle \pm i| -z, \dots, -z\rangle)/\sqrt{2}$  are eigenstates of  $H$  and  $A$ , respectively, in Eq. (15), with eigenvalues  $\pm 2^{n-1}$ . Hence, since the sum of the squares of the eigenvalues of  $H$  and  $A$  is given by  $\text{tr}[H^2 + A^2] = \text{tr}[(H + iA)(H + iA)^\dagger] = \{\text{tr}[(\sigma_x + i\sigma_y)(\sigma_x - i\sigma_y)]\}^n = 4^n$ , these eigenvalues are nondegenerate and all remaining eigenvalues vanish.
- [41] H. M. Wiseman and R. B. Killip, *Adaptive Single-Shot Phase Measurements: A Semiclassical Approach*, *Phys. Rev. A* **56**, 944 (1997).
- [42] D. W. Berry, H. M. Wiseman, and J. K. Breslin, *Optimal Input States and Feedback for Interferometric Phase Estimation*, *Phys. Rev. A* **63**, 053804 (2001).

- 
- [43] A. Sergeevich, A. Chandran, J. Combes, S. D. Bartlett, and H. M. Wiseman, *Characterization of a Qubit Hamiltonian Using Adaptive Measurements in a Fixed Basis*, [Phys. Rev. A \*\*84\*\*, 052315 \(2011\)](#).
- [44] C. Ferrie, C. E. Granade, and D. G. Cory, *How to Best Sample a Periodic Probability Distribution, or on the Accuracy of Hamiltonian Finding Strategies*, [Quantum Inf. Process. \*\*11\*\* \(2012\)](#).
- [45] S. Knysh, V. N. Smelyanskiy, and G. A. Durkin, *Scaling Laws for Precision in Quantum Interferometry and the Bifurcation Landscape of the Optimal State*, [Phys. Rev. A \*\*83\*\*, 021804\(R\) \(2011\)](#).
- [46] B. M. Escher, R. L. de Matos Filho, and L. Davidovich, *General Framework for Estimating the Ultimate Precision Limit in Noisy Quantum-Enhanced Metrology*, [Nature Phys. \*\*7\*\*, 406 \(2011\)](#).