How to perform the most accurate possible phase measurements

D. W. Berry,[1,2] B. L. Higgins,[3] S. D. Bartlett,[4] M. W. Mitchell,[5] G. J. Pryde,[3,*] and H. M. Wiseman[3,†]

[1]*Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1*

[2]*Centre for Quantum Computer Technology, Department of Physics, Macquarie University, Sydney 2109, Australia*

[3]*Centre for Quantum Computer Technology, Centre for Quantum Dynamics, Griffith University, Brisbane 4111, Australia*

[4]*School of Physics, University of Sydney, Sydney 2006, Australia*

[5]*ICFO-Institut de Ciencies Fotoniques, Mediterranean Technology Park, 08860 Castelldefels (Barcelona), Spain*

We present the theory of how to achieve phase measurements with the minimum possible variance in ways that are readily implementable with current experimental techniques. Measurements whose statistics have high-frequency fringes, such as those obtained from maximally path-entangled $(|N,0\rangle + |0,N\rangle)/\sqrt{2}$ ("NOON") states, have commensurately high information yield (as quantified by the Fisher information). However, this information is also highly ambiguous because it does not distinguish between phases at the same point on different fringes. We provide schemes to eliminate this phase ambiguity in a highly efficient way, providing phase estimates with uncertainty that is within a small constant factor of the Heisenberg limit, the minimum allowed by the laws of quantum mechanics. These techniques apply to NOON state and multipass interferometry, as well as phase measurements in quantum computing. We have reported the experimental implementation of some of these schemes with multipass interferometry elsewhere. Here, we present the theoretical foundation and also present some additional experimental results. There are three key innovations to the theory in this paper. First, we examine the intrinsic phase properties of the sequence of states (in multiple time modes) via the equivalent two-mode state. Second, we identify the key feature of the equivalent state that enables the optimal scaling of the intrinsic phase uncertainty to be obtained. This enables us to identify appropriate combinations of states to use. The remaining difficulty is that the ideal phase measurements to achieve this intrinsic phase uncertainty are often not physically realizable. The third innovation is to solve this problem by using realizable measurements that closely approximate the optimal measurements, enabling the optimal scaling to be preserved. We consider both adaptive and nonadaptive measurement schemes.

I. INTRODUCTION

The measurement of phase is an important task in both metrology and quantum computing. The measurement of optical phase is the basis of much precision measurement, whereas the measurement of the phase encoded in a register of qubits is vital to a broad range of quantum algorithms [1–3]. In optical phase measurement the precision is usually bound by the standard quantum limit (SQL), where the phase uncertainty is $\Theta(N^{-1/2})$ in the number of resources N [4]. On the other hand, the fundamental limit imposed by quantum mechanics is $\Theta(N^{-1})$ [5,6], often called the Heisenberg limit (HL). There have been many proposals to approach this limit. Reference [5] proposed using squeezed states in one port of an interferometer, as well as homodyne measurements, to beat the SQL. Measurements of this type have been experimentally demonstrated [7–9]. Another type of nonclassical state that has been proposed [10,11] and experimentally demonstrated [12–22] is the maximally path-entangled $(|N,0\rangle + |0,N\rangle)/\sqrt{2}$ (NOON) states. These provide the maximum phase resolution for a given photon number, although they have the problem that they do not directly provide a unique estimate of the phase.

In quantum computing, the phase, corresponding to the eigenvalue of an operator, can be estimated using Kitaev's algorithm [1], or the quantum phase estimation algorithm (QPEA) [2,3]. The QPEA is based on applying the inverse quantum Fourier transform (QFT) [23,24]. The inverse QFT, followed by a computational basis measurement, can be applied using just local measurements and control, without requiring entangling gates [25]. That simplification allows the phase measurement to be achieved optically, using only linear optics, photodetectors, and electronic feedback onto phase modulators [26]. The optical implementation could use a succession of NOON states or a succession of multiple passes of single photons (as was used in Ref. [26]). Using NOON states, or multiple passes, results in high phase sensitivity, but an ambiguous phase estimate. The role of the QPEA is to resolve this ambiguity.

The minimum uncertainty for measurements of a phase shift is $\Theta(N^{-1})$ in terms of the total number of applications of that phase shift, N, regardless of whether those applications are applied in series or in parallel [27]. In quantum optical interferometry, N is the maximum total number of passes of photons through the phase shift. In this formalism, a single pass of an N-photon NOON state and N passes of a single photon are regarded as the same number of resources. This is convenient because it enables physical systems that give mathematically identical results to be treated within a unified mathematical formalism. We emphasize that, in practice, although these resources are mathematically identical, they are not physically identical and will be useful in different situations. In particular, the resources for NOON states are used in parallel, which means that they are used within a short

*g.pryde@griffith.edu.au

†h.wiseman@griffith.edu.au

space of time. This is needed for measurements where there is a stringent time limit to the measurement, for example, due to fluctuation of the phase to be measured or decoherence of the physical system. In contrast, N passes of a single photon are using resources in series; i.e., the applications of the phase shift are sequential. This is useful for measurements of a fixed phase, where there is no intrinsic time limit, but it is required to measure the phase with minimum energy passing through the sample.

In this paper we examine the general problem of how to obtain the most accurate possible phase estimates, by efficiently eliminating ambiguities. This theory applies to general phase measurements in optics (with NOON states or multiple passes) and quantum computation, although for clarity we will primarily present the discussion in terms of NOON states. The problem with simply using the QPEA to eliminate phase ambiguities is that it produces a probability distribution with large tails, which means that the standard deviation is $\Theta(N^{-1/2})$, well short of the Heisenberg limit of $\Theta(N^{-1})$. A method of overcoming this problem was presented in Ref. [26], which used an adaptive scheme to achieve $\Theta(N^{-1})$ scaling. This work was further expanded in Ref. [28], which proved analytically that scaling at the Heisenberg limit can be achieved without needing adaptive measurements. References [26,28] demonstrated these schemes experimentally, using multiple passes of single photons. An alternative scheme based on adapting the size of the NOON states was proposed in Ref. [29]. A method of eliminating phase ambiguities in the context of quantum metrology was provided in Ref. [30].

Here, we present the theoretical foundations for Refs. [26,28], with further analytical and numerical results and some additional experimental data. First, Sec. II explains the limit to the accuracy of phase measurement in more detail. Next, in Sec. III the theoretical background to adaptive interferometric measurements and the relation to the QPEA are presented. The equivalent two-mode states when using repeated measurements are presented in Sec. IV, and it is shown that (with one exception) they have canonical phase variance scaling as the Heisenberg limit. The adaptive measurements are presented in Sec. V, where the analytical results and numerical results for large N are given. In Sec. VI it is shown that increasing the number of repetitions leads to a phase uncertainty that still scales as the SQL, rather than the Heisenberg limit. In Sec. VII the theory for some simplifications to the adaptive scheme is presented. These simplifications include a hybrid scheme as well as a nonadaptive scheme. The approach of adapting the size of the NOON state [29] is given in Sec. VIII. Finally, we present the conclusions and a table summarizing the results in Sec. IX.

II. LIMITS TO PHASE MEASUREMENT

The limit to the accuracy of phase measurements can be derived in a simple way from the uncertainty principle for phase [31],

$$\Delta\phi\Delta n \geq 1/2, \tag{2.1}$$

where the uncertainties are quantified by the square root of the variance. For a single-mode optical field, n is the photon number and the phase shift is given by the unitary $\exp(i\hat{n}\phi)$, with $\hat{n}$ as the number operator. More generally, one can consider a phase shift with $\hat{n}$ being any operator with nonnegative integer eigenvalues. The same uncertainty relation will hold regardless of the particular physical realization.

The uncertainty principle (2.1) is exact if the variance that is used for the phase is $V_H \equiv \mu^{-2} - 1$, where $\mu \equiv |\langle e^{i\hat{\phi}}\rangle|$, introduced by Holevo [32]. The Holevo variance coincides with the usual variance for a narrow distribution peaked well away from the phase cut. Here, $\hat{\phi}$ is an unbiased estimator of the phase, in the sense that $e^{i\phi} \propto \langle e^{i\hat{\phi}}\rangle$. Note that the hat notation is used to indicate a phase estimator, rather than a phase operator. If one has a biased phase estimator, one must use instead $\mu = \langle\cos(\hat{\phi} - \phi)\rangle$.

If n is upper bounded by N, then the uncertainty in n can never exceed $N/2$. This implies that the phase uncertainty is lower bounded as

$$\Delta\phi \geq 1/N. \tag{2.2}$$

Because this lower bound to the phase uncertainty may be derived from the uncertainty principle for phase, it is usually called the Heisenberg limit. This derivation was presented in terms of the standard deviation for the phase and the mean photon number in Refs. [33,34]. As explained in [33,34], that argument is not rigorous because the uncertainty relation (2.1) is not exact for the usual standard deviation, and the uncertainty in the photon number is not upper bounded by the mean photon number. The above derivation makes the Heisenberg limit rigorous by using the square root of the Holevo variance to obtain an exact uncertainty principle and using an upper limit on n.

The lower bound of $1/N$ on the phase uncertainty is not tight. The exact achievable lower bound, with optimal measurements, is

$$\Delta\phi_{HL} = \sqrt{V_H} = \tan\left(\frac{\pi}{N+2}\right) \sim \frac{\pi}{N}. \tag{2.3}$$

The asymptotic result was found in the single-mode case in Ref. [35], and the exact result was found in Ref. [36]. This bound in the case of two-mode interferometry, where N is the total number of photons, was found in Refs. [37,38]. For two-mode interferometry, n is the number of photons passing through the phase shift. An alternative scenario for phase measurement was considered in Ref. [27]. There, it was shown that the same limit holds in a general situation involving N applications of a phase shift to qubits, interspersed with unitaries.

A unified way of defining N that is independent of the physical implementation is as follows. The complete measurement scheme, including any feedback, may be represented by preparation of a pure quantum state that depends on a parameter ϕ, followed by measurement. Denote the family of such states, parametrized by ϕ, as $|\psi(\phi)\rangle$. The Fourier transform of this family of states may be taken independently of the basis and can be denoted as $|\widetilde{\psi}(s)\rangle$. Then N can be defined as the minimum size of the interval that supports $|\widetilde{\psi}(s)\rangle$. See Appendix C for detailed explanations.

Because of the way the Heisenberg limit is derived, there are a number of conditions on its validity:

(1) The error is quantified by the square root of the variance of the difference between the phase estimate and the system phase.

(2) N is the total number of applications of the phase shift (the total number of photon passes in quantum optics).

(3) There is no *a priori* information about the system phase; the prior probability distribution is flat on the interval $[0, 2\pi)$.

(4) The variance is evaluated by averaging over all possible system phases, using this flat distribution.

Essentially these conditions require the measurement to be a self-contained measurement of a completely unknown phase. As these are the conditions on the Heisenberg limit, methods to achieve the Heisenberg limit or scaling as the Heisenberg limit, should satisfy these conditions. For the measurement schemes that we describe here we are careful to ensure that all of these conditions are satisfied. It is possible to derive a similar limit on the measurement accuracy in the phase sensing regime, where it is required to measure small phase shifts [12]. That limit is also called the Heisenberg limit, although that definition of the Heisenberg limit differs from the definition used here.

The condition that the error is quantified by the square root of the variance is the most stringent condition on the error because the standard deviation can be used to place bounds on all other commonly used measures of the uncertainty [39]. On the other hand, it is possible for other measures of uncertainty to give unrealistically small values that are not meaningful. In particular, the reciprocal-peak likelihood can give an uncertainty scaling as $1/\bar{n}^2$ in the mean photon number [40]. That small uncertainty is not meaningful because it does not translate into a correspondingly small standard deviation [33].

For most other measures of uncertainty, if they are small then (even if the variance for a single measurement is large) results from separate measurements can be combined to yield an overall estimate with small variance. The problem is that this does not necessarily preserve the scaling—a fact which is often ignored in the analysis. If a single measurement uses N_1 applications of the phase shift, and there are M repetitions of the measurement, then the total resources are $N = MN_1$. The variance can then scale no better than $1/(N_1\sqrt{M}) = \sqrt{M}/N$. Sometimes scaling as $1/(N_1\sqrt{M})$ has been referred to as Heisenberg-limited scaling (for a recent example, see Ref. [41]). However, that is not strictly correct because if the number of repetitions M required increases with N_1, then the Heisenberg limit of $\Theta(1/N)$ scaling will not be obtained, as pointed out recently in Ref. [42].

Note also that the first condition requires that the error is measured between the actual system phase and the phase estimates. This prevents biased phase estimates giving unrealistically small values of the error. This condition also means that the uncertainty should not be based entirely on the probability distribution for the system phase based on the measurement results. That is, the uncertainty is the spread obtained in the phase estimates from the measurements, rather than the spread in the Bayesian probability distribution obtained from a single measurement.

The second condition is simply that the resources are quantified in the same way as in the proof of the bound. This means that the Heisenberg limit, as defined here, is not beaten by schemes such as those in Refs. [43,44], where the resources are quantified in terms of the number of interacting systems. In these papers the number of applications of the phase shift is equivalent to $\nu t \|H\|$ using their notation (see Appendix C). The variance obtained in Refs. [43,44] is not smaller than $\pi/(\nu t \|H\|)$. Note also that, provided the resources are quantified in this way, it is not possible to obtain better scaling than $1/N$ by combining parallel and serial resources. In particular, for optics, N is the number of photon passes, so a state with n photons (the parallel resource) that passes through a phase shift p times (the serial resource) is using total resources $N = np$. This means that it is not possible to obtain better scaling than $1/N$ by, for example, performing the phase measurements we propose with multiple passes of NOON states.

The third condition means that no pre-existing information about the phase can be used in the measurement. The fourth condition makes this rigorous by ruling out the possibility of just considering one system phase where the measurement is particularly accurate. This prevents the measurement from implicitly using information about the system phase. In contrast, work on phase measurement often considers the regime of small phase shifts, where the performance of the measurement is considered only for a small range of phases. This approach may be reasonable for states that are close to classical, but breaks down if all the resources are concentrated into a single nonclassical state in order to approach the Heisenberg limit. For example, if all resources are concentrated into measurement with a single NOON state, then accuracy of $\Theta(1/N)$ would be obtained only if the phase shift is smaller than $1/N$; otherwise, the phase estimate would be ambiguous. That is, for any fixed phase shift (no matter how small), it would not be possible to maintain Heisenberg-limited scaling for arbitrarily large N. The phase measurement schemes that we present here show how to eliminate phase ambiguities to achieve accuracy of $\Theta(1/N)$ without needing *any* initial knowledge of the phase.

III. THEORETICAL BACKGROUND

A. Interferometric measurements

Interferometric measurements are typically considered via the Mach-Zehnder interferometer, as in Fig. 1. Two input modes are combined at a beam splitter, after which each of the modes is subjected to a phase shift, and the two modes are recombined at a second beam splitter. The phase shift to be measured, ϕ, is in one arm and a controllable phase shift, Φ, may be added in the second arm. The first beam splitter is not necessary for the analysis, and it is more convenient to consider the state in the arms of the interferometer, i.e., in modes a and b.

The optimal phase measurement, also called the canonical measurement, can be imagined to be performed directly on the modes in the arms, thereby also omitting the second beam splitter. The phase statistics of such a measurement can be regarded as the intrinsic phase-difference statistics of the
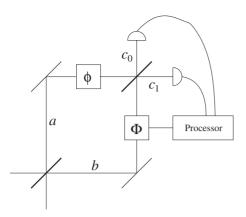
FIG. 1. The Mach-Zehnder interferometer, with the addition of a controllable phase Φ in one arm. The unknown phase to be estimated is ϕ.

state of these two modes. The joint state of modes a and b (as shown in Fig. 1) can be written as a superposition of joint number states. We make the usual assumption that the coefficients in this superposition are positive and real before application of any phase shift. If the total photon number is fixed at N, then the canonical positive operator-valued measure (POVM) is of the form [45,46]

$$F_{\text{can}}(\hat{\phi}) = \frac{1}{2\pi}|\hat{\phi}\rangle\langle\hat{\phi}|, \quad |\hat{\phi}\rangle = \sum_{n=0}^{N} e^{in\hat{\phi}}|n\rangle|N-n\rangle. \quad (3.1)$$

This POVM gives the probability density for the continuous quantity $\hat{\phi}$; that is, the probability of the measurement result being in the interval A is

$$P(\hat{\phi} \in A) = \int_A \text{Tr}[F_{\text{can}}(\hat{\phi})\rho]d\hat{\phi}. \quad (3.2)$$

This measurement is applied to modes a and b after application of the phase shift ϕ, but without the phase shift Φ (or with Φ taken to be zero). The result of the measurement, $\hat{\phi}$, is then the estimator for the unknown phase ϕ.

Except in special cases, it is not possible to perform canonical measurements with standard optical equipment (photon counters and linear optical elements such as beam splitters). To approximate canonical measurements, Refs. [38,39] use adaptive phase measurements—an idea introduced in Ref. [47]. The sequence of detections is used to obtain a Bayesian probability distribution for the phase. The phase Φ is then adjusted to minimize the expected variance after detection of the next photon. Note that the time between detections gives no information about the phase. In addition, for fixed total photon number N, the measurement operator for no detection does not alter the state. This means that the time between detections can be ignored in the analysis.

To obtain the probability distribution for the phase, it is convenient to keep a record of the un-normalized system state. The un-normalized system state after m detections, and for system phase ϕ, will be denoted as $|\psi(\vec{u}_m,\phi)\rangle$. Here, $\vec{u}_m=(u_1,\ldots,u_m)$ is the vector of m measurement results. The measurement results are taken to be $u_j=0$ or 1 to indicate

that the jth photon is detected in mode c_0 or mode c_1. The annihilation operators at the output of the final beam splitter, c_0 and c_1, are

$$c_u = [e^{i\phi}a + (-1)^u e^{i\Phi}b]/\sqrt{2}, \quad (3.3)$$

where a and b are the operators for the field at the positions shown in Fig. 1. After detection, the state is updated as

$$|\psi(\vec{u}_{m+1},\phi)\rangle = \frac{c_{u_{m+1}}}{\sqrt{N-m}}|\psi(\vec{u}_m,\phi)\rangle. \quad (3.4)$$

The probability for the sequence of measurement results is given by $P(\vec{u}_m|\phi)=\langle\psi(\vec{u}_m,\phi)|\psi(\vec{u}_m,\phi)\rangle$. (The division by $\sqrt{N-m}$ ensures that the probabilities for the different measurement results sum to 1.)

Using Bayes' theorem, the probability for the system phase given the measurement results is

$$P(\phi|\vec{u}_m) = \frac{P(\phi)P(\vec{u}_m|\phi)}{P(\vec{u}_m)}. \quad (3.5)$$

The initial probability distribution $P(\phi)$ is flat, and the probability $P(\vec{u}_m)$ is independent of the phase. Therefore, the probability distribution for the system phase is

$$P(\phi|\vec{u}_m) \propto P(\vec{u}_m|\phi) = \langle\psi(\vec{u}_m,\phi)|\psi(\vec{u}_m,\phi)\rangle. \quad (3.6)$$

Because the annihilation operators c_u contain the exponential $e^{i\phi}$, the un-normalized state and therefore the probability can be expressed in terms of powers of this exponential. This enables the probability distribution to be efficiently represented in terms of the coefficients of the powers of $e^{i\phi}$.

The unbiased phase estimate with the smallest variance is $\hat{\phi}=\arg\langle e^{i\phi}\rangle$ [39]. Here, the average is taken over ϕ using the Bayesian probability distribution. If the probability distribution $P(\phi|\vec{u}_m)$ is represented as a Fourier series in ϕ, $\langle e^{i\phi}\rangle$ is simply the coefficient of the $e^{-i\phi}$ term. Recall that the Holevo variance in the phase estimates is defined as $V_H=\mu^{-2}-1$, where $\mu=|\langle e^{i\hat{\phi}}\rangle|$. In order to ensure that the measurement is covariant, we take the initial value of Φ to be random. Then the value of the sharpness μ is given by [39]

$$\mu = \frac{1}{2\pi}\sum_{\vec{u}_m}\left|\int e^{i\phi}P(\vec{u}_m|\phi)d\phi\right|. \quad (3.7)$$

See Appendix B for a more detailed explanation.

This expression can be rewritten as an average over the sharpnesses of the individual Bayesian probability distributions

$$\mu = E[|\langle e^{i\phi}\rangle|]. \quad (3.8)$$

Here, E indicates an expectation value over the measurement results and the initial feedback phase, and the angular brackets indicate an average over ϕ. The way this expression is interpreted is that, for any actual system phase and initial feedback phase, we obtain some measurement results $\vec{u}_m$ and determine a Bayesian probability distribution for the system phase based on those measurement results.

In Refs. [38,39] an algorithm for choosing the feedback phase was introduced, which maximizes the expected sharpness of the Bayesian probability distribution after the next

detection. Maximizing the average sharpness, as in Eq. (3.8), minimizes the variance in the phase estimates. Therefore, this feedback minimizes the variance in the phase estimates after the next detection. It might be thought that, since the ultimate aim is to minimize the Holevo variance V_H, the feedback phase should be chosen to minimize the expected Holevo variance of the Bayesian probability distribution after the next detection. However, the variance we want to minimize is that in the phase estimates, not that in the Bayesian probability distributions. If we were minimizing the variance in the Bayesian probability distributions, then we would be minimizing

$$E[|\langle e^{i\phi}\rangle|^{-2} - 1]. \tag{3.9}$$

We actually want to minimize the variance in the phase estimates, which is given by

$$V_H = \{E[|\langle e^{i\phi}\rangle|]\}^{-2} - 1. \tag{3.10}$$

Thus, to minimize V_H, or indeed any monotonically decreasing function of μ, for the measurement scheme as a whole, one should aim to maximize Eq. (3.7) as in Refs. [38,39].

Alternatively, the result that one should aim to maximize Eq. (3.7) may be shown directly from Eq. (3.7). The feedback phase Φ_m for detection m is a function of the prior measurement results $u_1, \ldots, u_{m-1}$ and the initial feedback phase Φ_1. In order to maximize μ as given by Eq. (3.7), we need to optimize each of these Φ_m for each measurement record $\vec{u}_{m-1}$. We can rewrite the sum for μ in Eq. (3.7) as

$$\mu = \frac{1}{2\pi}\sum_{\vec{u}_{m-1}}\sum_{u_m=0}^{1}\left|\int e^{i\phi}P(\vec{u}_m|\phi)d\phi\right|. \tag{3.11}$$

Then for each $\vec{u}_{m-1}$ we have a sum

$$\mu_{\vec{u}_{m-1}} = \sum_{u_m=0}^{1}\left|\int e^{i\phi}P(\vec{u}_m|\phi)d\phi\right| \tag{3.12}$$

that depends only on the feedback phase Φ_m for that $\vec{u}_{m-1}$. Thus, the multivariable maximization is reduced to separate maximizations of a single variable for each $\vec{u}_{m-1}$. As there are only two terms in the sum for $\mu_{\vec{u}_{m-1}}$, there is an analytical solution for the feedback phase Φ_m that maximizes this function, as detailed in Ref. [39].

In Ref. [39] it was shown that this approach can, in theory, yield an uncertainty only slightly larger than Eq. (2.3) with the optimal input states. The problem is that these states have not yet been produced experimentally. The simplest N-photon input state to realize in practice is that with all N photons entering one port of the first beam splitter in Fig. 1 (with vacuum at the other port). In this case, with the single-pass interferometer design of Fig. 1, the variance in $\hat{\phi}$ can scale no better than $\Theta(N^{-1})$ [corresponding to an uncertainty of $\Theta(N^{-1/2})$]. This limit cannot be surpassed regardless of how effective the adaptive measurements are, or even with canonical phase measurements. To beat this scaling it is necessary to use nonclassical multiphoton states or a variation in the interferometer design that achieves the same end. The nonclassical states with the highest resolution, NOON states [10], give a Holevo variance that is formally infinite due to

the manifold ambiguity in the phase estimate. The same is true if one simply increases the number of passes through the phase shift. Practical solutions to this problem [26,28] have come from considering phase measurements in quantum computing, which are discussed in the following subsection.

B. Phase measurements in quantum computing

In quantum computing, phase estimation is used for finding an eigenvalue of a unitary operator U. If an eigenstate $|u\rangle$ of U is known, the task is to find the corresponding eigenvalue $e^{i\phi}$. A control qubit is placed in the state $(|0\rangle + |1\rangle)/\sqrt{2}$, and target qubits are placed in the state $|u\rangle$. A controlled-U^{2^k} operation (i.e., an operation that applies the operator U^{2^k} to a target qubit conditional on the state of the control qubit) then transforms the control qubit into the state $(|0\rangle + e^{i2^k\phi}|1\rangle)/\sqrt{2}$. One approach to finding the phase is the QPEA [2,3], in which one performs this procedure on identically prepared qubits for $k=0, \ldots, K$ [2]. This yields a state of the form

$$\sum_{y=0}^{2^{K+1}-1} e^{i\phi y}|y\rangle, \tag{3.13}$$

where $|y\rangle$ is a state on the $K+1$ control qubits. Provided the phase has an exact K-bit binary expansion of the form $\phi = 2\pi \times 0.a_0 \cdots a_K$ (where a_j are binary digits), performing the inverse QFT yields the state $|a_0 \cdots a_K\rangle$, so measurement in the computational basis yields the digits of the phase. If the phase is not of the form $\phi = 2\pi \times 0.a_0 \cdots a_K$, then accuracy to this number of digits (with a success probability of $1-\epsilon$) can still be achieved by increasing K to $K' = K + O(\ln(1/\epsilon))$ [2].

An alternative approach was given by Kitaev [1]. There, rather than using the inverse QFT, separate measurements are used. For each k the number of qubits used is $O(\ln(K/\epsilon))$, which enables $2^k\phi$ to be localized in one of eight subintervals of $[0, 2\pi]$ with an error probability of $\leq \epsilon/l$. By combining measurements for different values of k, the value of ϕ is determined with precision $\pi 2^{-K-2}$ and error probability of $\leq \epsilon$. In this paper we mostly take inspiration from the QPEA, but in Sec. VII B we also use ideas from Kitaev's algorithm. Note that the QPEA was, rather inaccurately, referred to as the Kitaev algorithm in Ref. [26].

Quantum computing algorithms that require phase estimation are for discrete problems, which are typically quantified by the error probability (i.e., the probability that an incorrect answer is given). For such problems, we usually require that the phase error is smaller than a certain amount in order to obtain the correct final answer. If the phase error is larger than this, its magnitude is irrelevant. We contrast this situation to the case in optics, where phase estimation is typically treated as a continuous problem for which the magnitude of any error is also important. Therefore, the quality of a phase estimation measurement is better quantified by the variance rather than a confidence interval. Also, in quantum computing it is often possible to implement the controlled unitary operation efficiently for arbitrary k, so the resources used scale as K. In contrast, for optics implementing a multiple of the phase shift requires more resources, so the total resources used scale as 2^K.

To implement quantum computing style phase measurements in quantum optics, the multiples of the phase shift can be achieved using either NOON states or multiple passes through a phase shift. The inverse QFT can be implemented by using the simplification of [25], which requires only local measurements and control.

Using a single photon as the input to a Mach-Zehnder interferometer, the state in the arms of the interferometer is $(|0\rangle+|1\rangle)/\sqrt{2}$, where $|0\rangle$ and $|1\rangle$ are the states corresponding to the photon in one arm or the other. Normally the phase shift ϕ changes the state to $(|0\rangle+e^{i\phi}|1\rangle)/\sqrt{2}$. With ν passes through the phase shift, the state is changed to $(|0\rangle+e^{i\nu\phi}|1\rangle)/\sqrt{2}$. The beam splitter then acts as a Hadamard operator, and the photodetectors give a measurement in the computational basis. With a controllable phase of Φ in the other arm, the probabilities of the measurement results are

$$P(u|\phi) = \tfrac{1}{2}[1 + (-1)^u \cos(\nu\phi - \Phi)]. \quad (3.14)$$

Alternatively, using a NOON state $(|\nu,0\rangle+|0,\nu\rangle)/\sqrt{2}$ in the arms of the interferometer, the phase shift changes the state to $(e^{i\nu\phi}|\nu,0\rangle+|0,\nu\rangle)/\sqrt{2}$. Until detection of all ν photons, no phase information is obtained, so there is no information on which to base a feedback phase. After detection of all photons, the probability of the sequence of measurement results is

$$P(\vec{u}_\nu|\phi) = \tfrac{1}{2}\{1 + (-1)^{u_1+\cdots+u_\nu}\cos[\nu(\phi-\Phi)]\}. \quad (3.15)$$

This means that the probability distribution for the phase, given the measurement results, only depends on the parity of the measurement results, i.e., whether $u_1+\cdots+u_\nu$ is odd or even. The phase information obtained is identical to that for a single photon with multiple passes, with the single measurement result u replaced with the parity and the controllable phase Φ replaced with $\nu\Phi$. From this point on the parity obtained from the NOON measurement will be represented as a single measurement result u.

To apply the phase measurements from quantum computing theory to optics, it is necessary to use multiple time modes. That is, independent single photons are used, or NOON states that are sufficiently spaced apart in time that the photons from the different NOON states can be unambiguously distinguished. The two cases are mathematically equivalent. For brevity, for the remainder of this paper, we present the analysis in terms of NOON states, but the analysis also holds for multiple passes of single photons, as well as phase measurements in quantum computing, because they are mathematically equivalent. The techniques we present can also be applied to Hamiltonian parameter estimation, as discussed in Appendix C, although there is the complication that the range of the parameter is not limited to $[0,2\pi]$. Multiple passes of single photons were used in the experimental demonstration below and in Refs. [26,28].

NOON states are often said to provide super-resolution of phase due to the multiple peaks in the phase distribution obtained in the interval $[0,2\pi]$. If a measurement is performed with just a *single* NOON state, there is a problem in that it is not possible to distinguish which of these peaks corresponds to the phase. This results in the Holevo phase

variance being infinite. To obtain a useful phase estimate, the phase information from a NOON state needs to be combined with other phase information.

The QPEA, with the QFT implemented according to the scheme of [25], provides a method to combine the phase information from different NOON states to yield an unambiguous measurement of the phase. Initially the controllable phase Φ is set to be zero. The first measurement is taken with a NOON state with $\nu=2^K$. Provided the system phase has an exact K-bit binary expression of the form $\phi=2\pi \times 0.a_0 \cdots a_K$, this measurement yields the value of a_K as the first measurement result u_1. Given this measurement result, the feedback phase is adjusted by $u_1\pi/2^K$ such that $\phi-\Phi =2\pi \times 0.a_0 \cdots a_{K-1}$. Then the next measurement with $\nu =2^{K-1}$ yields the next digit a_{K-1}, and so forth. In this way, all digits a_k are determined with certainty.

This procedure gives a measurement described by a POVM of the form

$$F(\hat{\phi}_l) = |\hat{\phi}_l\rangle\langle\hat{\phi}_l|, \quad |\hat{\phi}_l\rangle = \frac{1}{\sqrt{N+1}}\sum_{n=0}^{N} e^{in\hat{\phi}_l}|n,N-n\rangle, \quad (3.16)$$

where $N=2^{K+1}-1$ and $\hat{\phi}_l=\pi l/2^K$, for $l=0,\ldots,N$. The state $|n,N-n\rangle$ is labeled by the total number of photons in each arm. That is, $|n,N-n\rangle=|n\rangle\otimes|N-n\rangle$, where

$$|n\rangle \equiv |n_0\rangle \otimes \cdots \otimes |n_K\rangle, \quad (3.17)$$

and where n_k are the binary digits of n, and similarly for $|N-n\rangle$.

Since the QPEA gives a phase estimate with K bits of accuracy, at the cost of 2^K resources, it would seem that it should enable phase variance near the Heisenberg limit. Surprisingly, this is not the case. In fact, it gives a variance above the SQL, as we explain in the next section.

IV. EQUIVALENT STATES FOR MULTIPLE TIME MODES

To assist in determining the phase variance yielded by phase measurement algorithms, we introduce the concept of equivalent states. In general, consider two different systems with respective sets of basis states $\{|\chi_n\rangle\}$ and $\{|\xi_n\rangle\}$, such that a phase shift ϕ adds a factor of $e^{in\phi}$ to each basis state. States in the different systems that are identical except for the basis may be regarded as equivalent. That is, $\sum_n \psi_n |\chi_n\rangle$ is equivalent to $\sum_n \psi_n |\xi_n\rangle$. The canonical measurements are identical except for the basis states used, and the distributions obtained for the phase are identical.

In order to analyze a state with multiple time modes, we determine the state with a single time mode that it is equivalent to. We call this the two-mode equivalent state because it has two spatial modes and just a single time mode. The sequence of K NOON states with the photon number increasing from 1 to 2^K by powers of 2 is

$$|\psi_K\rangle = \frac{1}{2^{(K+1)/2}}(|2^K,0\rangle + |0,2^K\rangle) \otimes \cdots \otimes (|1,0\rangle + |0,1\rangle).$$

$$(4.1)$$

This state is changed under the phase shift to

$$|\psi_K(\phi)\rangle = \frac{1}{2^{(K+1)/2}}(e^{i2^K\phi}|2^K,0\rangle + |0,2^K\rangle) \otimes \cdots \otimes (e^{i\phi}|1,0\rangle$$

$$+ |0,1\rangle). \tag{4.2}$$

Expanding the product, there is a sum of orthogonal states such as $|2^K,0\rangle \otimes |0,2^{K-1}\rangle \otimes \cdots \otimes |1,0\rangle$. Each state has a coefficient of $e^{in\phi}$, where n is the sum of the photon numbers in the first modes from each pair of modes. State (4.1) is therefore equivalent to, in terms of its phase properties, an equally weighted superposition state with one time mode,

$$\frac{1}{2^{(K+1)/2}}\sum_{n=0}^{N}|n,N-n\rangle, \tag{4.3}$$

with $N = 2^{K+1} - 1$.

In terms of these equivalent two-mode states, the POVM for the QPEA can be written as in Eq. (3.16), with the states now being the equivalent two-mode states. This POVM is close to a canonical measurement, except that the possible phase measurement results are restricted to $\hat{\phi}_l = \pi l/2^K$. To obtain a canonical measurement, the initial feedback phase Φ_1 can be taken to be random, rather than zero. For a particular value of this initial feedback phase, the POVM is the same as that in Eq. (3.16), except with $\hat{\phi}_l = \Phi_1 + \pi l/2^K$. Using random Φ_1, all $\hat{\phi}$ are included in the POVM, so the measurement becomes the canonical phase measurement as in Eq. (3.1).

Using the canonical measurement, the value of μ for this equal superposition state is $\mu = N/(N+1) = 1 - 2^{-(K+1)}$. Therefore, the Holevo phase variance is

$$V_H = \frac{1}{(1 - 2^{-(K+1)})^2} - 1 = \frac{2}{N} + \frac{1}{N^2}. \tag{4.4}$$

This scales as $1/N$, which is the SQL rather than the expected Heisenberg limit. To see how this is compatible with the precision of the QPEA, one must examine the probability distribution for the phase estimate. Using $|n\rangle$ to denote $|n,N-n\rangle$, this distribution is given by

$$P(\hat{\phi}) = \frac{1}{2\pi}\left|\left(\sum_{n=0}^{N}e^{-in\hat{\phi}}\langle n|\right)\left(\frac{1}{\sqrt{N+1}}\sum_{n=0}^{N}e^{in\phi}|n\rangle\right)\right|^2$$

$$= \frac{\sin^2[(N+1)(\hat{\phi}-\phi)/2]}{2\pi(N+1)\sin^2[(\hat{\phi}-\phi)/2]}. \tag{4.5}$$

Clearly, this distribution has a peak (around ϕ) with a width $O(N^{-1})$, as expected for scaling at the Heisenberg limit. But its variance scales as the SQL because of its high tails. This is most easily seen using the Collett phase variance $V_C \equiv 2(1-\mu)$ [48]. The Collett variance coincides with the Holevo variance whenever either is small [39]. Using $x = \hat{\phi} - \phi$ for the phase error, we have

$$V_C = \langle 4\sin^2[(\hat{\phi}-\phi)/2]\rangle = \frac{2}{\pi(N+1)}\int_{-\pi}^{\pi}\sin^2(Nx/2)dx. \tag{4.6}$$

That is, the tails are so high that the variance integral is finite only because of the finite range of the phase. The integral

evaluates to $2/(N+1)$, the same SQL scaling as found above for V_H.

We have already seen that the QPEA realizes a canonical phase measurement. Therefore, if it fails to attain the Heisenberg limit, the fault must lie with state (4.3). One approach is to replace state (4.3) with a more general state of the form

$$\sum_{n=0}^{N}\psi_n|n,N-n\rangle. \tag{4.7}$$

Because the measurement is canonical, provided the coefficients ψ_n are optimal the minimum phase uncertainty (2.3) could be obtained.

Unfortunately the minimum phase uncertainty state [35–38] is not separable between the different time modes and would therefore be extremely difficult to create. However, it is not necessary for the state to be a minimum phase uncertainty state in order to obtain a phase variance *scaling* as the Heisenberg limit. For a state $\sum_{n=0}^{N}\psi_n|n,N-n\rangle$, the phase variance V_C is given by

$$2(1 - |\langle e^{i\hat{\phi}}\rangle|) = \sum_{n=-1}^{N}(\psi_n - \psi_{n+1})^2, \tag{4.8}$$

where ψ_{-1} and ψ_{N+1} are defined to be zero. Provided the state coefficients vary relatively smoothly, the maximum value of ψ_n should be of order $1/\sqrt{N}$, and the successive differences should be of order $N^{-3/2}$. This implies that $(\psi_n - \psi_{n+1})^2$ should be of order N^{-3}, so the variance should scale at the Heisenberg limit of N^{-2}. This reasoning shows that the restrictions on the properties of the state are quite weak if one wishes to obtain scaling at the Heisenberg limit.

The equally weighted state does not give phase variance scaling at the Heisenberg limit because there is a large jump, of order $1/\sqrt{N}$, between ψ_{-1} and ψ_0 and between ψ_N and ψ_{N+1}. In order to obtain the required smoothness, one approach is to use multiple copies of the state. The resulting equivalent two-mode state is then the repeated convolution of the original state (in terms of the squared state coefficients). The convolution has the general property of smoothing functions, and therefore can be expected to yield scaling at the Heisenberg limit. In particular we can prove the following theorem.

Theorem 1. The state

$$|\psi\rangle = \frac{1}{(N_K+1)^{M/2}}\left(\sum_{n=0}^{N_K}e^{in\phi}|n,N_K-n\rangle\right)^{\otimes M} \tag{4.9}$$

has a canonical phase variance of $\Theta(\ln N_K/N_K^2)$ for $M=2$ and $\Theta(1/N_K^2)$ for $M>2$.

This theorem means that, for fixed M greater than 2, the canonical phase variance scales as $1/N^2$, where $N = N_K M$. The quantity M is the number of copies of the state. The way we achieve a state equivalent to Eq. (4.9) is by using M copies of each of the individual NOON states. In applying this theorem, we take $N_K = 2^{K+1} - 1$, although that is not required for the theorem. A similar behavior with number of copies was found in Ref. [49] for canonical multimode phase estimation (modulo π) on multiple copies of a squeezed vacuum state, where $M > 4$ copies were required to attain the

Heisenberg limit. There, the optimal number of copies, in terms of minimizing the canonical Holevo variance for a fixed mean photon number, was eight; in our case, it is three (see Fig. 5).

Proof. Here, we give a nonrigorous derivation of the scaling using a continuous approximation. For the rigorous proof of the theorem without using this approximation, see Appendix A. State (4.9) is equivalent to the state

$$\frac{1}{(N_K + 1)^{M/2}} \sum_{n=0}^{N_K} e^{in\phi} \sqrt{f_M(n)} |n, N_K - n\rangle, \qquad (4.10)$$

where $f_M(n)$ is the number of combinations of values of $n_1, n_2, \ldots, n_M$ that sum to n. This quantity can be regarded as the number of points in a hyperplane perpendicular to a line running between opposite corners of a hypercubic lattice of dimension M.

In the continuous approximation, the value of $f_M(n)$ is the area of the cross section of the hypercube and is equal to $n^{M-1}/(M-1)!$ for $n \leq N_K$. The contribution to the variance V_C for $n \leq N_K$ is then, in the continuous approximation,

$$\frac{1}{(N_K + 1)^M} \int_0^{N_K} \left(\frac{d\sqrt{f_M(n)}}{dn} \right)^2 dn$$

$$= \frac{1}{(N_K + 1)^M} \int_0^{N_K} \frac{n^{M-3}}{4[(M-2)!]^2} dn. \qquad (4.11)$$

There is a clear difference between the cases $M=2$ and $M>2$. For $M>2$, the integral gives an expression proportional to N_K^{M-2}, so overall the expression scales as $\Theta(1/N_K^2)$. On the other hand, for $M=2$, the integrand is proportional to $1/n$, rather than a positive power of n or a constant. The divergence at $n=0$ can be ignored because the continuous approximation breaks down. However, the integral yields $\ln N_K$ from the upper bound, which means that the expression overall scales as $\Theta(\ln N_K/N_K^2)$.

The contribution to V_C for $N - N_K \leq n \leq N$ is the same. For $N_K < n < N - N_K$, we use the fact that $f'_M(n) \leq N_K^{M-2}$ and $f_M(n) \geq N_K^{M-1}/(M-1)!$ to find that the integrand is upper bounded by $(M-1)! N_K^{M-3}$. This means that the contribution to V_C for $N_k < n < N - N_k$ is $O(1/N_K^2)$ for both $M=2$ and $M>2$. Hence, the canonical phase variance, in the continuous approximation, is $\Theta(\ln N_K/N_K^2)$ for $M=2$ and $\Theta(1/N_K^2)$ for $M>2$. ☐

V. ADAPTIVE MEASUREMENTS FOR MULTIPLE TIME MODES

A. Deriving the recurrence relation

The drawback to using these multiple copies of the NOON states is that the canonical measurements can no longer be achieved exactly using adaptive measurements. On the other hand, it is possible to approximate the canonical measurements for larger numbers of repetitions, M. The idea is to use a generalization of the adaptive phase measurements for a single time mode. As in that case, at each step the feedback phase Φ is chosen to minimize the expected variance after the next detection.

One starts with performing M measurements, one on each of M NOON states, each with photon number $\nu = 2^K$ for some integer K. Next, measurements are performed on M NOON states with $\nu = 2^{K-1}$, and one continues this sequence on NOON states with $\nu = 2^k$ for $k = K, K-1, \ldots, 1$. This is similar to the optical implementation of the QPEA, except at each stage one uses M NOON states of each size rather than 1. There is a subtlety in using the feedback algorithm of Ref. [38], in that when measurements have been performed with NOON states down to size $\nu = 2^k$, the phase is only known modulo $2\pi/2^k$. To address this, rather than maximizing $|\langle e^{i\hat{\phi}} \rangle|$, the quantity that is maximized is $|\langle e^{i2^k\hat{\phi}} \rangle|$.

Using the same adaptive scheme as before, numerical testing indicates that the phase variances scale as the Heisenberg limit for $M > 3$. For small values of M the measurements are a poor approximation of the canonical measurements; for $M=3$ the variance scales as $N^{-3/2}$, and for $M=2$ the variance scales as N^{-1}.

The feedback can be achieved in an efficient way when performing measurements on these states. Because the individual NOON states are not entangled with each other, measurements on one do not affect the states of the others, except for the normalization when considering the un-normalized state. Therefore, it is only necessary to keep track of the evolution of the normalization. Recall that the normalization gives the probability distribution for that sequence of detection results and is proportional to the probability distribution for the phase.

At each stage the probability distribution for the total measurement results is obtained by multiplying the probability for the current measurement result by the preceding probability distribution

$$P(\vec{u}_m | \phi) = P(u_m | \phi) P(\vec{u}_{m-1} | \phi). \qquad (5.1)$$

The probability distribution for the current measurement result, with a NOON state with $\nu = 2^k$, is

$$P(u_m | \phi) = \tfrac{1}{2}\{1 + (-1)^{u_m} \cos[2^k(\phi - \Phi_m)]\}. \qquad (5.2)$$

Provided the measurements are made on the largest NOON states first, the overall probability distribution can only contain different powers of $e^{i2^k\phi}$. The probability distribution may therefore be stored using just the coefficients of the different powers. That is, the values of $p_j^{(k)}(\vec{u}_m)$ in the expansion

$$P(\vec{u}_m | \phi) = \sum_j p_j^{(k)}(\vec{u}_m) e^{ij2^k\phi} \qquad (5.3)$$

are stored.

The coefficients are updated as

$$p_j^{(k)}(\vec{u}_m) = \frac{1}{2}\left\{ p_j^{(k)}(\vec{u}_{m-1}) + \frac{(-1)^{u_m}}{2}[p_{j-1}^{(k)}(\vec{u}_{m-1})e^{-i2^k\Phi_M} + p_{j+1}^{(k)}(\vec{u}_{m-1})e^{i2^k\Phi_M}] \right\}. \qquad (5.4)$$

When examining measurements for the next lower value of ν, given by $\nu = 2^{k-1}$, the coefficients are expanded out by a factor of 2. That is, the probability distribution is written as

$$P(\vec{u}_m|\phi) = \sum_j p_j^{(k-1)}(\vec{u}_m)e^{ij2^{k-1}\phi}, \qquad (5.5)$$

where $p_{2j}^{(k-1)}(\vec{u}_m) = p_j^{(k)}(\vec{u}_m)$ and $p_{2j+1}^{(k-1)}(\vec{u}_m) = 0$.

In addition, it is only necessary to keep track of a limited number of $p_j^{(k)}(\vec{u}_m)$. At the end of the measurements, the quantity of importance is $p_1^{(0)}(\vec{u}_{M(K+1)})$. This can depend only on $p_j^{(k)}(\vec{u}_m)$ for j in the range $-2M, \ldots, 2M$, and only these coefficients need be recorded. In fact, only coefficients for $j = 0, \ldots, 2M$ need be recorded because those for negative j are the complex conjugate of those for positive j.

To illustrate these ideas, consider the first measurement for $\nu = 2^k$, where the first few coefficients are

$$p_0^{(k)}(\vec{u}_m) = p_0^{(k+1)}(\vec{u}_m),$$

$$p_1^{(k)}(\vec{u}_m) = 0,$$

$$p_2^{(k)}(\vec{u}_m) = p_1^{(k+1)}(\vec{u}_m),$$

$$p_3^{(k)}(\vec{u}_m) = 0, \qquad (5.6)$$

where $m = M(K-k)$. After the next detection, the first few coefficients are changed to

$$p_0^{(k)}(\vec{u}_{m+1}) = p_0^{(k+1)}(\vec{u}_m)/2,$$

$$p_1^{(k)}(\vec{u}_{m+1}) = (-1)^{u_{m+1}}[p_0^{(k+1)}(\vec{u}_m)e^{-i2^k\Phi_{m+1}}$$
$$+ p_1^{(k+1)}(\vec{u}_m)e^{i2^k\Phi_{m+1}}]/4,$$

$$p_2^{(k)}(\vec{u}_{m+1}) = p_1^{(k+1)}(\vec{u}_m)/2. \qquad (5.7)$$

The feedback phase is chosen to maximize the average of $|p_1^{(k)}(\vec{u}_{m+1})|$ over the two values of u_{m+1}. For both measurement results,

$$|p_1^{(k)}(\vec{u}_{m+1})| = \tfrac{1}{4}|p_0^{(k+1)}(\vec{u}_m)e^{-i2^{k+1}\Phi_{m+1}} + p_1^{(k+1)}(\vec{u}_m)|. \quad (5.8)$$

Because $p_0^{(k)}(\vec{u}_m)$ is real and positive (the probability distribution is real and positive), the feedback phase that maximizes this average is

$$\Phi_{m+1} = -2^{-(k+1)}\arg[p_1^{(k+1)}(\vec{u}_m)]. \qquad (5.9)$$

We now investigate this solution for various values of M.

B. Analytical result for $M=1$

In the case $M = 1$, Eq. (5.9) yields a recurrence relation that can be solved, yielding [50]

$$p_0^{(k)}(\vec{u}_{K-k+1}) = 2^{k-K-1},$$

$$p_1^{(k)}(\vec{u}_{K-k+1}) = \frac{2^{K-k+1}-1}{2^{2(K-k+1)}}\exp\left(-i\pi\sum_{l=1}^{K-k+1}u_l/2^{K-k+1-l}\right). \qquad (5.10)$$

The feedback phase obtained is

$$\Phi_{K-k+1} = 2\pi\sum_{l=1}^{K-k}u_l/2^{K+1-l} = 2\pi\times 0.00\cdots0u_{K-k}u_{K-k-1}\cdots u_1, \qquad (5.11)$$

where the expression following "$\times$" is a binary expansion. For each measurement result u_l the feedback phase is adjusted by the appropriate fraction of 2π. This feedback is identical to that obtained for the linear optics implementation of the QPEA described in Sec. III. In addition, summing over $|p_1^{(k)}(\vec{u}_{K-k+1})|$ at the end of the measurement gives $\mu = 1-2^{-(K+1)}$. This yields exactly the same variance as in Eq. (4.4).

C. Analytical result for $M=2$

It is also possible to analytically determine the variance for $M = 2$. In this case, the values of p_0 and p_1 after the second measurement for a given k are

$$p_0^{(k)}(\vec{u}_{m+2}) = p_0^{(k)}(\vec{u}_m)/4 + \frac{(-1)^{u_{m+1}}}{8}\mathrm{Re}[p_0^{(k)}(\vec{u}_m)e^{i2^k(\Phi_{m+2}-\Phi_{m+1})}$$
$$+ p_2^{(k)}(\vec{u}_m)e^{i2^k(\Phi_{m+2}+\Phi_{m+1})}],$$

$$p_1^{(k)}(\vec{u}_{m+2}) = \frac{(-1)^{u_{m+1}}}{8}[p_0^{(k)}(\vec{u}_m)e^{-i2^k\Phi_{m+1}} + p_2^{(k)}(\vec{u}_m)e^{i2^k\Phi_{m+1}}]$$
$$+ \frac{(-1)^{u_{m+2}}}{8}[p_0^{(k)}(\vec{u}_m)e^{-i2^k\Phi_{m+2}} + p_2^{(k)}(\vec{u}_m)e^{i2^k\Phi_{m+2}}]. \qquad (5.12)$$

Taking the absolute value of p_1 yields

$$p_1^{(k)}(\vec{u}_{m+2}) = \tfrac{1}{8}|e^{i2^{k-1}(\Phi_{m+2}-\Phi_{m+1})}$$
$$+ (-1)^{u_{m+2}-u_{m+1}}e^{i2^{k-1}(\Phi_{m+2}-\Phi_{m+1})}|$$
$$\times|p_0^{(k)}(\vec{u}_m) - p_2^{(k)}(\vec{u}_m)e^{i2^k(\Phi_{m+2}+\Phi_{m+1})}|. \qquad (5.13)$$

Regardless of whether the feedback phase (5.9) is assumed for Φ_{m+1}, or both Φ_{m+1} and Φ_{m+2} are maximized over, the solution is

$$2^k(\Phi_{m+2} - \Phi_{m+1}) = \pi/2,$$

$$2^k(\Phi_{m+2} + \Phi_{m+1}) = -\arg[p_2^{(k)}(\vec{u}_m)] + \pi/2. \qquad (5.14)$$

This means that the feedback scheme minimizes the variance two detections in advance, rather than just one. Independent of the detection results, the absolute value obtained is

$$|p_1^{(k)}(\vec{u}_{m+2})| = \frac{\sqrt{2}}{8}\sqrt{[p_0^{(k)}(\vec{u}_m)]^2 + |p_2^{(k)}(\vec{u}_m)|^2}. \qquad (5.15)$$

Using this feedback also gives

$$p_0^{(k)}(\vec{u}_{m+2}) = p_0^{(k)}(\vec{u}_m)/4. \qquad (5.16)$$

This again yields a recurrence relation that can be solved. The result is

$$p_0^{(k)}(\vec{u}_{m+2}) = 2^{2(k-K-1)},$$

$$\left|p_1^{(k)}(\vec{u}_{m+2})\right| = 2^{2(k-K-1)}\sqrt{1 - 2^{-k-K-1}}. \quad (5.17)$$

At the end of the measurement, summing over $\left|p_1^{(0)}(\vec{u}_{2(K+1)})\right|$ gives $\mu = \sqrt{1 - 2^{-(K+1)}}$. This yields the Holevo variance in this case as

$$V_\text{H} = \frac{1}{2^{K+1} - 1} = \frac{2}{N}. \quad (5.18)$$

Hence, in this case, the Holevo phase variance is *exactly* $2/N$. This is almost identical to the result for $M=1$, where the variance is approximately $2/N$. This means that, despite the intrinsic state having relatively good phase properties [with the variance scaling as $(\ln N)/N^2$], the adaptive measurements are far less accurate than the canonical measurement. Furthermore, because the feedback phases minimize the variance two detections in advance, it seems unlikely that a better feedback scheme would yield significantly more accurate measurements.

D. Numerical results for $M > 2$

For larger values of M, more accurate measurements are obtained, but we have not found exact analytical results. It is possible to obtain exact results for smaller values of K simply by systematically calculating the final values of $p_1^{(0)}$ for every possible combination of measurement results. For larger values of K this approach is no longer feasible, but instead the variance can be estimated by generating random system phases, then generating measurement results according to their probability of occurring. There are then two alternative ways of estimating the phase variance. One is to use the mean of the final values of $\left|p_1^{(0)}\right|$ as an estimate of μ. The other is to use the variance in the final phase estimates relative to the system phase. Numerically it is found that using the values obtained for $\left|p_1^{(0)}\right|$ gives a more accurate estimate of the phase variance.

The results for $M=1-6$ are shown in Fig. 2. The variances are multiplied by N^2, so that the Heisenberg limit appears as a horizontal line. The lines for $M=1$ and 2 are not scaling as the Heisenberg limit and are indistinguishable on this plot for N above about 10. The variances in these cases scale as $1/N$, and the numerically obtained variances agree with the analytical predictions to within the numerical precision. The lines for $M=4-6$ are clearly scaling as $1/N^2$. The line for $M=3$ has intermediate scaling and is consistent with an asymptotic scaling of $1/N^{3/2}$.

Although the results for $M=4-6$ scale as the Heisenberg limit, they have different scaling constants. The variances for these cases, multiplied by N^2, are shown on a linear scale in Fig. 3. The smallest scaling constant of about 23 is obtained for $M=5$.

VI. EFFECT OF INCREASING THE NUMBER OF REPETITIONS

In order to obtain higher precision, there are two parameters that can be increased. The value of K can be increased
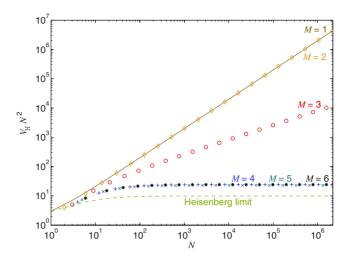

FIG. 2. (Color online) The Holevo phase variance multiplied by N^2 for $M=1-6$. The results for $M=1$ are shown as the solid line, for $M=2$ as the diamonds, for $M=3$ as the circles, for $M=4$ as the pluses, for $M=5$ as the crosses, and for $M=6$ as the asterisks. The Heisenberg limit is also shown as the dashed line for comparison.

or the number of repetitions, M, can be increased. Scaling at the Heisenberg limit can only be obtained by increasing K, not by increasing M because the state coefficients do not gradually increase from zero in the latter case. Instead they are negligible for most values of n and suddenly rise to significant values near the center (for example, see Fig. 4).

The result can be shown rigorously using the uncertainty relation (in terms of the variance) $V_\text{H}(\phi)V(n) \geq 1/4$ [32], where $V(n)$ is the variance for the photon number. To determine the number variance, take $p_M(n) = f_M(n)/(N_K+1)$ and use the recurrence relation (see Appendix A)

$$f_M(n) = \sum_{k=0}^{N_K} f_{M-1}(n-k). \quad (6.1)$$

This yields the recurrence relation $\langle n \rangle_M = \langle n \rangle_{M-1} + N_K/2$ [where $\langle n \rangle_M = \Sigma_n n p_M(n)$]. This gives $\langle n \rangle_M = MN_K/2$. The
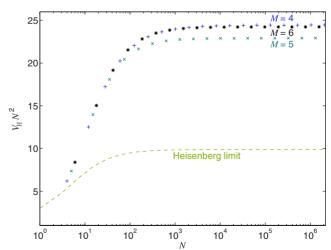

FIG. 3. (Color online) The Holevo phase variance multiplied by N^2 for $M=4-6$. The results for $M=4$ are shown as the pluses, for $M=5$ as the crosses, and for $M=6$ as the asterisks. The Heisenberg limit is also shown as the dashed line for comparison.
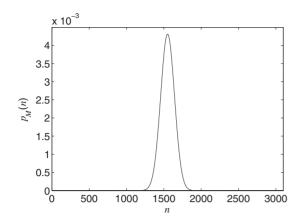
FIG. 4. The squares of the state coefficients for the equivalent two-mode state for $K=4$ and $M=100$.

same approach for $\langle n^2 \rangle_M$ gives the recurrence relation

$$\langle n^2 \rangle_M = \langle n^2 \rangle_{M-1} + N_K(2N_K+1)/6 + N_K \langle n \rangle_{M-1}. \quad (6.2)$$

Solving this gives $\langle n^2 \rangle_M = MN_K(N_K+2)/12 + M^2 N_K^2/4$, so $V(n) = MN_K(N_K+2)/12$. Using the uncertainty relation, we obtain the lower bound on the phase variance,

$$V_H(\phi) \geq \frac{3}{MN_K(N_K+2)}. \quad (6.3)$$

Thus, with fixed K, the variance can only scale as $1/M$, not as $1/M^2$. On the other hand, because the number variance scales as the square of N_K, the Heisenberg limit can be achieved when increasing K, as we have already seen.

The predicted phase variance obtained using the adaptive measurement scheme with $K=4$ is plotted as a function of M in Fig. 5. Experimental results using multiple passes are also shown in this figure and agree with the theoretical predictions. These experimental results were obtained using an ap-
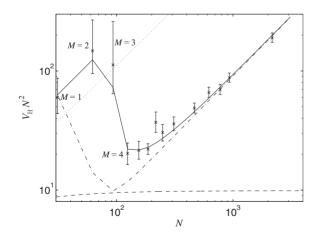

FIG. 5. The theoretical predictions and experimental results for adaptive measurements with $K=4$ and a range of values of M (given by $M=N/31$). The solid line is the predictions for the adaptive scheme, the dashed-dotted line is the intrinsic variance, and the crosses and error bars are for the experimental data. The SQL for single passes is shown as the dotted line, and the dashed line is the Heisenberg limit.

paratus and methodology identical to that described in Ref. [26]. When increasing M, the variance at first decreases rapidly and approaches the Heisenberg limit. Then, as M is increased further, the variance no longer decreases at the same rate as the Heisenberg limit and instead scales as the SQL for single passes.

VII. SIMPLIFICATIONS OF THE ADAPTIVE MEASUREMENTS

The adaptive scheme described above is complicated to implement in practice due to the calculation required to determine the feedback phase. This calculation is non-Markovian, in the sense that it depends in a nontrivial way on all the prior measurement results. For this reason it is useful to develop schemes using simplified (Markovian) feedback, or no feedback. Two alternative schemes were proposed and experimentally demonstrated in Ref. [28]. Here, we provide further motivation and theoretical analysis of these alternative measurement schemes.

A. Hybrid scheme

One scheme is based on a simple modification of the QPEA. The QPEA is Markovian because the *increment* in the feedback phase depends only on the immediately preceding result. As discussed above, the QPEA gives a canonical measurement, but the state is equivalent to an equal superposition and thus has poor phase properties. It yields a probability distribution for the phase with tails that drop off slowly and give the dominant contribution to the phase variance. The idea behind hybrid measurements is to use additional "standard" measurements (single passes of single photons) to reduce the size of the tails of the distribution. The hybrid measurements are still Markovian, because they do not use any feedback beyond what is used in the QPEA. However, these measurements do not scale at the Heisenberg limit, and instead the variance scales at best as close to $1/N^{3/2}$. In particular, the result is given by the following theorem.

Theorem 2. The canonical phase variance of the state

$$|\psi_{N_K,M}\rangle = \frac{1}{2^{M/2}\sqrt{N_K+1}} \left(\sum_{n=0}^{N_K} e^{in\phi} |n, N_K-n\rangle \right) \otimes (|0,1\rangle + e^{i\phi}|1,0\rangle)^{\otimes M}, \quad (7.1)$$

scales as $\Omega(N^{-3/2})$, where $N=N_K+M$.

Here, Ω is the standard notation for a lower bound on the scaling. In applying this theorem, we take $N_K=2^{K+1}-1$ to be the number of photons used in the QPEA and M to be the number of photons used in the standard interferometry.

Proof. To obtain the equivalent two-mode state for M repetitions of measurements with single photons we apply the recurrence relation (6.1). Repeatedly applying this recurrence relation gives

$$f_M(n) = \binom{M}{n}. \quad (7.2)$$

This result can also be obtained from the fact that this number is the number of ways of choosing n ones from M bits.
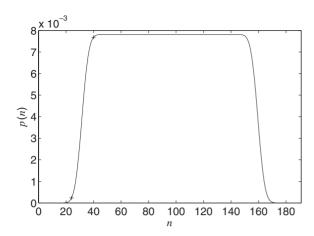
FIG. 6. The squares of the state coefficients for the equivalent two-mode state for hybrid measurements with $N_K = 127$ and $M = 64$. The pluses show the positions of n_- and n_+.

Then applying the recurrence relation again with the flat distribution gives

$$f(n) = \sum_{k=0}^{N_K} \binom{M}{n-k}. \tag{7.3}$$

This gives the probability distribution for n,

$$p(n) = \frac{1}{2^M(N_K+1)} \sum_{k=0}^{N_K} \binom{M}{n-k}. \tag{7.4}$$

Evaluating the moments for this probability distribution gives

$$\langle n \rangle = (N_K + M)/2,$$

$$\langle n^2 \rangle = \frac{M}{4}(M+1) + \frac{N_K M}{2} + \frac{N_K}{6}(1 + 2N_K). \tag{7.5}$$

This gives the variance in n as $(3M + 2N_K + N_K^2)/12$. As above, using the uncertainty relation for the Holevo variance yields

$$V_H(\hat{\phi}) \geq \frac{3}{3M + 2N_K + N_K^2}. \tag{7.6}$$

This relation implies that, if $N_K = O(N^{1/2})$, then the SQL is obtained. To obtain better scaling than the SQL, it is necessary for N_K to increase with N more rapidly than $\sqrt{N}$.

This result on its own does not show that the intrinsic variance cannot scale better than $N^{-3/2}$, and it is necessary to also examine other features of the equivalent two-mode state. The problem with the equivalent two-mode state is that the state coefficients rapidly increase, then there is a wide flat region (for example, see Fig. 6), rather than the gradual increase to a maximum, which is required.

To put a bound on the phase variance, consider the value of $p(n)$ for $n = n_- = \lfloor M/2 - \sqrt{M} \rfloor$ and $n = n_+ = \lfloor M/2 + \sqrt{M} \rfloor$. These values are chosen as points between which $p(n)$ varies rapidly (see Fig. 6). Because $M < N$, and N_K increases more rapidly than $\sqrt{N}$, asymptotically we must have $N_K > 2\sqrt{M}$. Using this, and Hoeffding's inequality [51], we obtain

$$p(n_+) = \frac{1}{2^M(N_K+1)} \sum_{k=n_+-N_K}^{n_+} \binom{M}{k} \geq \frac{1}{2^M(N_K+1)} \sum_{k=n_-}^{n_+} \binom{M}{k}$$

$$> (1 - 2e^{-2})/(N_K + 1), \tag{7.7}$$

$$p(n_-) = \frac{1}{2^M(N_K+1)} \sum_{k=n_--N_K}^{n_-} \binom{M}{k} \leq e^{-2}/(N_K + 1). \tag{7.8}$$

In addition, it is easily verified that $n_+ - n_- \leq 4\sqrt{M/3}$, so

$$\sum_{n=n_-}^{n_+-1} [\sqrt{p(n+1)} - \sqrt{p(n)}]^2 \geq \frac{[\sqrt{p(n_+)} - \sqrt{p(n_-)}]^2}{n_+ - n_-}$$

$$\geq \frac{(\sqrt{1 - 2e^{-2}} - e^{-1})^2}{4(N_K+1)\sqrt{M/3}}$$

$$= \frac{1}{2\kappa(N_K+1)\sqrt{M}}, \tag{7.9}$$

where $\kappa \approx 4.9$. The points n_- and n_+ are those between which the state coefficients rapidly increase. There is also a second region for large n where the state coefficients rapidly decrease. The symmetry of the distribution implies that the sum over that region has the same lower bound. Using Eq. (4.8), and the fact that $V_H \geq V_C$, the variance for sufficiently large N is lower bounded by

$$V_H(\hat{\phi}) \geq \frac{1}{\kappa(N_K+1)\sqrt{M}}. \tag{7.10}$$

The smallest possible value for this lower bound is obtained for both N_K and M of order N, which gives

$$V_H(\hat{\phi}) = \Omega(N^{-3/2}). \tag{7.11}$$

$\square$

In Ref. [28] it was proven that measurements could yield a variance scaling as $O(\sqrt{\ln N} N^{-3/2})$. The lower-bound scaling here is the same, except for a small $\sqrt{\ln N}$ factor. The method used for the proof in [28] was to consider an analysis of the data where, if the phase estimates from the single-photon measurements and the QPEA differed by too large an amount, the phase estimate from the single-photon measurements was used, and otherwise the phase estimate from the QPEA was used.

The general idea behind the proof of [28] is that the phase estimate from single-photon measurements has a negligible probability of having error larger than $O(1/\sqrt{M})$. Therefore, if the phase estimates from the single-photon measurements and the QPEA differ by more than $O(1/\sqrt{M})$, it is almost certainly the QPEA phase estimate that is wrong, so the phase estimate from the single-photon measurements should be used. The variance then is of order $1/M$, but because it occurs with probability $\sqrt{M}/N_K$, the total contribution to the variance is $O(1/(N_K\sqrt{M}))$. If the phase estimates do agree to within $O(1/M)$, then the contribution to the variance is again of order $1/(N_K\sqrt{M})$.

The additional factor of $\sqrt{\ln N}$ comes about because it is not quite true that the probability of the single-photon measurements having error larger than $O(1/\sqrt{M})$ can be ignored.
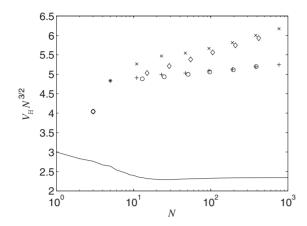
FIG. 7. The variance for hybrid measurements. The solid line is the canonical phase variance for the equivalent two-mode state. The crosses and pluses are for $M = 2^K$, and increments in Φ of π/M and $\pi/2$, respectively. The diamonds and circles are for values of M that give the minimum value of $V_H N^2$, and increments in Φ of π/M and $\pi/2$, respectively.

FIG. 8. The variance for nonadaptive measurements with M optimized as a function of K, but independent of k. The crosses are for 2^{20} samples and the pluses are for 2^{25} samples.

This probability must scale down with M to prevent it giving a contribution to the variance larger than $O(1/(N_K \sqrt{M}))$. To ensure that this probability is sufficiently small, the size of the error should be $O(\sqrt{(\ln M)/M})$. This leads to the additional $\sqrt{\ln N}$ factor in the final result.

The analytical results show that there is a lower bound on the variance for hybrid measurements of $\Omega(N^{-3/2})$ and an upper bound of $O(\sqrt{\ln N} N^{-3/2})$. The scaling must be close to $N^{-3/2}$, but these bounds leave open the question of whether there is an additional logarithmic factor. The hybrid measurements have been simulated for values of K up to 8, and the results are shown in Fig. 7. The variances shown are multiplied by $N^{3/2}$ to make the scaling clearer. For the results shown the value of $V_H N^{3/2}$ increases with N, suggesting that the variance is slightly larger than $O(N^{-3/2})$.

In Fig. 7, results are shown for increments in Φ of π/M and $\pi/2$. The variance for the single-photon measurements on their own is slightly smaller with the π/M increments than with the $\pi/2$ increments. However, for the hybrid measurements, using increments of $\pi/2$ yields significantly better results. Results are also shown for $M = 2^K$, as well as for where the value of M has been adjusted to minimize $V_H N^{3/2}$. The motivation for using $M = 2^K$ is that the analytical results suggest that $M \approx N/3$ is optimal. Numerically it was found that slightly higher values of M gave slightly better results, particularly for Φ increments of π/M.

B. Nonadaptive scheme

It might be expected that approaching the Heisenberg limit requires an adaptive measurement. Normally the variance would be approximately the sum of the canonical phase variance and an additional variance due to the measurement technique [36], and nonadaptive measurements would introduce a variance scaling as the SQL [39]. Despite this, it is still possible to achieve the same scaling as the Heisenberg limit with a nonadaptive technique [28]. The reason why this is possible is that the nonadaptive measurement with mul-
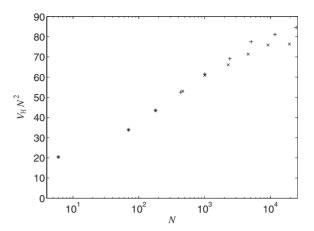
tiple time modes has a different POVM from the nonadaptive measurement on a single time mode.

A new feature of the nonadaptive technique of [28] is that the number of repetitions, M, is now a function of the values of K and k. The largest value of M is for a single-photon NOON state, and the number of repetitions is decreased as the size of the NOON state (or number of passes in Ref. [28]) is increased. The reason for changing the number of repetitions is essentially that the role of the measurements for smaller values of k is to distinguish between the multiple phase estimates provided by larger values of k. If the measurements for smaller values of k do not distinguish between these phase estimates accurately, then the resulting error is large. On the other hand, because the resource cost for the smaller values of k is small, the cost of repeating these measurements is low. It is therefore better to repeat these measurements more often, in order to prevent these large errors.

If the number of repetitions is not changed with k, then the numerical results indicate that the scaling of the variance is not as good as $1/N^2$. The numerical results are shown in Fig. 8. The value of $V_H N^2$ increases approximately linearly in $\ln N$, indicating that the scaling is as $(\ln N)/N^2$. For these results the value of M was increased with K, even though it was independent of k. It was found that the best value of M increased nearly linearly in K, from about 10 for $K=2$ to 24 for $K=9$. The primary contribution to the variance was from low-probability results with large error. This meant that extremely large numbers of samples need to be used to obtain accurate estimates of the variance. Calculations with 2^{20} samples yielded lower estimates of the variance than those for 2^{25} samples because they did not sufficiently sample the low-probability results.

The reason why M needs to be increased with K in the nonadaptive case can be understood in the following way. For simplicity, consider just the measurements with $k=0$ (a single-photon NOON state), and assume that the measurements with larger values of k are sufficient to narrow down the estimates of the phase to ϕ and $\phi + \pi$. The measurements with $k > 0$ measure the phase modulo π, so they cannot distinguish between these alternatives. They do not give the phase exactly modulo π, but assuming that they do can only

decrease the phase variance. The measurements with $k=0$ could distinguish between the two alternatives if they were known.

In the adaptive case, the phase is known accurately (although not perfectly) modulo π at the stage the measurements with $k=0$ are performed. The feedback phase Φ can therefore be chosen so as to efficiently distinguish between the two alternatives. However, in the *nonadaptive* case, the values of Φ used are completely independent of the actual phase. In the case where there is a fixed value of M, there is a nonzero probability that the measurement results will be more consistent with $\phi+\pi$ rather than ϕ, so there is a π error in the phase estimate.

In particular, the probability that all the measurement results with $k=0$ are consistent with $\phi+\pi$ rather than ϕ is

$$P_{\mathrm{error}} = \prod_{m=1}^{M} \frac{1}{2}[1 - |\cos(\phi - \Phi_m)|]. \qquad (7.12)$$

There will be other combinations of measurement results that are also more consistent with $\phi+\pi$, but the probability for this combination on its own provides a lower bound on the overall probability that the final phase estimate is $\phi+\pi$. For some particular values of the system phase this probability will be zero, but averaging over all system phases gives a nonzero error probability.

This lower bound on the probability of a measurement result that gives a π error in the phase estimate means that, if nonadaptive measurements are performed with a fixed value of M, the phase variance cannot decrease below some minimum value. More generally, if the value of M is allowed to depend on K (but still be independent of k), then there need not be a fixed lower bound on the variance, but because of Eq. (6.3) the variance can scale no better than $M(N)/N^2$. As M must increase with K (and therefore N) in order to remove the lower bound on the variance, the variance does not scale at the Heisenberg limit.

For the nonadaptive scheme given in Ref. [28] the variation of M used was $M(K,k)=2+3(K-k)$. Numerically it was found that this variation of M gave the smallest asymptotic variance that could be reliably calculated. The value of $V_{\mathrm{H}}N^2$ was approximately 40.5 for values of N up to more than 10^7 (see Fig. 9). This corresponds to an uncertainty of approximately 2.03 times the Heisenberg limit.

Alternative values of M given by $M(K,k)=4+2(K-k)$ did give slightly smaller values of $V_{\mathrm{H}}N^2$ for large N. However, calculations in this case were not reliable because the variance depends heavily on results with large error and small probability. Using 2^{20} samples gave significantly smaller values of $V_{\mathrm{H}}N^2$ in most cases, except for one value which was much larger. When the number of samples was increased to 2^{25} the value of $V_{\mathrm{H}}N^2$ increased to approximately the same as for $M(K,k)=2+3(K-k)$ for N up to about 10^5. For larger values of N the variance was still smaller, but it is likely that the number of samples is still too small to obtain a reliable result.

Using other functions for $M(K,k)$ did give slightly better results for small values of K. For $K=1$, the functional dependence that gave the smallest value of $V_{\mathrm{H}}N^2$ was $M(K,k)=1$
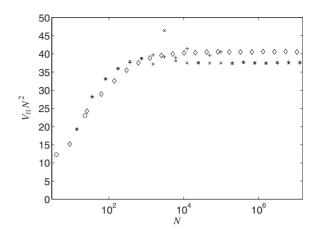

FIG. 9. The variance for nonadaptive measurements with M as a function of both K and k. The results for $M(K,k)=2+3(K-k)$ are shown as diamonds and the results for $M(K,k)=4+2(K-k)$ are shown as both the pluses and crosses. The crosses are for 2^{20} samples and the pluses are for 2^{25} samples. The circles are $M(K,k)=1+K-k$ for $K=1$ and $M(K,k)=1+4(K-k)$ for $K=2$.

$+K-k$. For $K=2$, the best values of M were given by $M(K,k)=1+4(K-k)$. The results for these cases are also shown in Fig. 9.

VIII. ADAPTING THE SIZE OF THE NOON STATE

Up to this point we have considered a fixed sequence of NOON states, and the only parameter which was varied adaptively was the feedback phase. It is also possible to consider adjusting the size of the NOON states used according to the measurement results. The most obvious way of doing this is to select the NOON state that minimizes the expected variance after the next detection. The problem with this approach is that it will select large NOON states that give small variance but result in a greater use of resources. As the aim is to obtain the smallest possible scaling constant for $V_{\mathrm{H}}N^2$, a better approach is to choose the NOON state that minimizes this quantity.

To determine results for a specific value of N, the sizes of the NOON states were restricted such that the total resources did not exceed N. The resulting variances obtained are shown as the crossed circles in Fig. 10. For the larger values of N ($2^{11}=2048$ and above), it was only feasible to use 2^{10} samples, so these results have low accuracy. Nevertheless, it is clear that using this approach yields a significantly higher variance than simply using adaptive measurements with a fixed sequence of NOON states. In addition, the scaling of the variance appears to have increased to $(\ln N)/N^2$.

It is possible to improve on this adaptive scheme by using a radically different approach [29]. There are three main features of this approach:

(1) The phase estimation begins with 100 measurements with single photons and no feedback (similar to the hybrid technique introduced in Sec. VII A).

(2) After the first step, the feedback phase is chosen such that the expected probabilities for the two measurement results are equal.
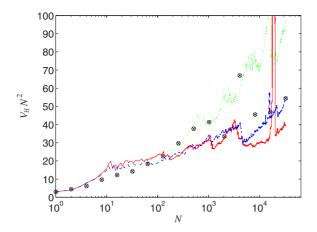
FIG. 10. (Color online) The variance for measurements where the size of the NOON states are adapted. The crossed circles are where the size is chosen to minimize the expected value of $V_H N^2$ after the next detection, and the sizes are chosen such that a particular value of N is obtained. The continuous lines are where all intermediate values of N are sampled. The dotted green line is where $V_H N^2$ is minimized, the solid red line is where $[\langle S \rangle - \ln(2\pi)]/\ln N$ is minimized, and the dashed blue line is where $V_H N^2$ is minimized, but the feedback of [29] is used for the phase.

(3) After the first step, the size of the NOON state is chosen to minimize the expected entropy of the phase distribution after the next measurement.

More specifically, the size of the NOON states is chosen to minimize $\langle S \rangle / \ln N$, where N is the number of resources after the next detection. The motivation for dividing by $\ln N$ can be understood in the following way. For a probability distribution, the negative of the entropy approximately corresponds to the information that may be communicated using a distribution of that form. As the distribution narrows by a factor of 2, the phase is effectively known to an additional bit, and the entropy decreases by 1. Therefore, the entropy is approximately proportional to the logarithm of the phase uncertainty. The phase uncertainty varies as a power of N, so the entropy should be approximately proportional to the logarithm of N. The best phase information for a given amount of resources can therefore be obtained by minimizing $\langle S \rangle / \ln N$.

The calculations given in Ref. [29] also used some additional simplifications to increase the speed. Rather than searching all possible values of the size of the next NOON state (which could be many thousands), only about 30 were searched. Denoting the number of resources used so far by N, only sizes up to $\lceil N/3 \rceil$ were searched, and in step sizes of $\lceil N/100 \rceil$. In addition, each run was used to obtain phase estimates for each intermediate number of resources. There is a complication in doing this because many of the NOON states have $\nu > 1$, so there are not phase estimates for all values of N. To obtain phase estimates for all N, the estimates for these skipped values were approximated by taking the phase estimate from the next lower value of N that was not skipped. This gives samples for all intermediate values of N and cannot underestimate the variance. Even with these simplifications, the phase variance scaled approximately as $40/N^2$.

On further investigation, we have found that the feedback scheme of Ref. [29] can be further simplified and provide

more accurate phase estimates. One simplification is in the approximation used for the entropy. It is not practical to calculate the entropy exactly, but it can be accurately approximated from the moments of the distribution. Reference [29] used a higher-order approximation involving the third- and fourth-order moments [the approximation given in Eq. (10) of [29] is for the negative of the entropy]. The calculation can be simplified by using the simpler approximation

$$S \approx \tfrac{1}{2}[1 + \ln(2\pi) + \ln V_C], \qquad (8.1)$$

Numerically it is found that this simpler approximation gives accuracy of the phase measurements similar to that using the higher-order approximation.

It was also found that the variance could be reduced if the quantity that was minimized was of the form $[\langle S \rangle + C]/\ln N$, where C is some constant. One value that gave particularly good results was $C = -\ln(2\pi)$ [which just corresponds to changing the sign on $\ln(2\pi)$ in Eq. (8.1) before dividing by $\ln N$]. In Ref. [29] the measurement scheme was started with 100 single-photon measurements, in order to obtain a narrowly peaked distribution. The drawback to this approach is that it increases the variance for total photon numbers of a similar order. The majority of the resources need to be used in NOON states with high photon numbers in order to obtain results near the Heisenberg limit. To reduce this problem, the number of single-photon measurements was reduced to ten.

In Fig. 10 the results from three different approaches are shown. The first scheme described in this section, except with the sampling scheme of Ref. [29], is shown as the dotted line. This sampling appears to give slightly worse results in this case as compared to the results where a fixed maximum value was used. For larger values of N, $V_H N^2$ is typically above 60. The results where $[\langle S \rangle - \ln(2\pi)]/\ln N$ was minimized are also shown in this figure. This seemed to give the best results out of the alternative variations of the scheme of Ref. [29] that were tested, with variances significantly below those where $V_H N^2$ was minimized.

The third alternative that is shown in Fig. 10 is that where $V_H N^2$ is minimized when choosing the size of the NOON state, but the feedback phase is chosen to ensure that the probabilities of the two detection results are equal (as in Ref. [29]). This can be achieved simply by replacing $[\langle S \rangle + C]/\ln N$ with $\langle S \rangle + 2 \ln N$. This scheme also gives better results than the first scheme and gives variances close to those for the scheme minimizing $[\langle S \rangle - \ln(2\pi)]/\ln N$.

Even though the schemes where the sizes of the NOON states are adapted are far more flexible than the schemes with a fixed sequence of NOON states, they do not appear to give any better results. For most methods tested the variance obtained is far greater than that for a fixed sequence of NOON states. Even for the best method tested, the variances obtained were simply comparable to those for the fixed sequence of NOON states and did not give any improvement. Because the fixed sequence of NOON states is far simpler to implement, it would be preferable to use for most applications.

TABLE I. Scalings of the phase variance for the various phase estimation schemes considered in this paper. The column "NOON sequence" shows the number of photons in each state in the sequence, in temporal order (which is important for adaptive measurements). In all but the last scheme, these photon numbers are powers of 2 between 1 (a single-photon state) and 2^K. Two variances are shown: the intrinsic variance (i.e., for a canonical phase measurement) and the variance for the actual measurement scheme. The former is undefined for the final scheme because the sequence of states there is not predetermined. Rather than exact variances, only the scaling of the variance with N, the total number of photons, is shown. Θ, O, and Ω are standard Bachmann-Landau notation; Θ means the asymptotic scaling, while O and Ω mean upper and lower bounds, respectively, on the asymptotic scaling. The symbol $\widetilde{O}$ indicates that multiplying factors scaling as powers of $\log(N)$ are ignored. We use an asterisk, such as in Θ^*, to indicate that the scaling is only proven numerically. For each result we give the section where the result is discussed, as well as a citation to previous work (if it exists). Finally, the "Rating" column gives an easy guide to the performance of the various phase estimation schemes. The stars have the following interpretation: ∗—variance equal to, or a constant multiplier greater than, the SQL of $1/N$ asymptotically; ∗∗—variance scaling as the SQL, but smaller by a constant multiplier; ∗∗∗—variance scaling roughly as the geometric mean of the SQL, and the HL of π^2/N^2 asymptotically; ∗∗∗∗—variance scaling almost as well as the HL; and ∗∗∗∗∗—variance equal to, or a constant multiplier times greater than, the HL.

Scheme	NOON sequence	Intrinsic var.	Reference	Estimate var.	Rating	Reference
Single photons	$1, 1, \ldots, 1$	$\Theta(1/N)$		$\Theta(1/N)$	∗	
QPEA: $M = 1$	$2^K, 2^{K-1}, \ldots, 1$	$\Theta(1/N)$	[26]; Sec. IV	$\Theta(1/N)$	∗	[26]; Sec. V
Generalized QPEA $M = 2$	2^K twice, $\ldots$, 1 twice	$\Theta(\ln N/N^2)$	Sec. IV	$\Theta(1/N)$	∗	[26]; Sec. V
Generalized QPEA $M = 3$	2^K thrice, $\ldots$, 1 thrice	$\Theta(1/N^2)$	Sec. IV	$\Theta^*(1/N^{3/2})$	∗∗∗	Sec. V
Generalized QPEA $M \geq 4$	2^K M-fold, $\ldots$, 1 M-fold	$\Theta(1/N^2)$	Sec. IV	$\Theta^*(1/N^2)$	∗∗∗∗∗	[26]; Sec. V
Generalized QPEA K fixed	2^K M-fold, $\ldots$, 1 M-fold	$\Theta(2^{-2K}/N)$	Sec. VI	$\Theta^*(2^{-2K}/N)$	∗∗	Sec. VI
Nonadaptive, best $M(K)$	$2^K, \ldots, 1$, each $M(K)$-fold	$\Theta(M(K)/N^2)$	Sec. VII	$\Theta^*(\ln N/N^2)$	∗∗∗∗	Sec. VII
Nonadaptive, best $M(K,k)$	$\{2^k\}$, each $M(K,k)$-fold	$\Theta(1/N^2)$	[28]; Sec. VII	$\Theta(1/N^2)$	∗∗∗∗∗	[28]; Sec. VII
Hybrid QPEA+singles	$2^K, 2^{K-1}, \ldots, 1, 1, \ldots, 1$	$\Omega(1/N^{3/2})$	Sec. VII	$O(\sqrt{\ln N}/N^{3/2})$	∗∗∗	[28]; Sec. VII
Nonbinary adaptive	Adaptively chosen	N/A		$\widetilde{O}^*(1/N^2)$	∗∗∗∗	[29]; Sec. VIII

IX. CONCLUSIONS

NOON states achieve the highest phase resolution possible for a given photon number but do not provide an unambiguous phase estimate. By combining measurements with a range of NOON states with different photon numbers, it is possible to achieve a phase resolution at the Heisenberg limit and eliminate the ambiguity in the phase estimate. As well as being applicable to NOON states, these schemes can also be performed using single photons, by passing them through the phase shift multiple times, as experimentally demonstrated in Refs. [26,28].

In this paper we have considered many different measurement schemes involving sequences of NOON states, some of which give a variance scaling at the Heisenberg limit, and some of which do not. Below, we discuss these results, but for the benefit of the reader we present a summary of all the key results in this paper in Table I.

Using a sequence of NOON states with photon numbers that are decreasing powers of 2, an elegant adaptive optical phase measurement can be performed by implementing the quantum phase estimation algorithm from quantum computing theory [2,3]. This provides a canonical measurement of the phase, which is a remarkable result because normally it is impossible to perform a canonical phase measurement using linear optics. The canonical measurement is possible because the photons are separated into distinct time modes.

This approach does not provide variance scaling as the Heisenberg limit because the underlying state has poor phase properties and gives a significant probability for large phase errors. Reduced phase variance could be achieved using states that are entangled between the different time modes. A more practical alternative is to use multiple copies of the state, which also provides reduced canonical phase variance. The canonical measurements can no longer be performed, but for four or more copies it is possible to accurately approximate the canonical phase measurement using adaptive measurements. The minimum variance is then obtained using five copies.

Perhaps surprisingly, it is also possible to achieve the Heisenberg limit using nonadaptive measurements. If the same number of copies of each size of NOON state is used, then the phase uncertainty scales as $\sqrt{\ln N}/N$. In order to achieve the Heisenberg limit, the number of copies must be a function of the size of the NOON state. In particular, the number of copies must be of order unity (e.g., 2) for the largest NOON state and should increase by a constant number (e.g., 3) at each step as one halves the size of the NOON state.

An alternative to fixing the number of copies of a given NOON state is to consider schemes which choose the size of the next NOON state to be used adaptively. This approach can yield a variance close to the Heisenberg limit. The obvious approach, which is to choose a NOON state which minimizes the variance multiplied by N^2, provides scaling slightly worse than the Heisenberg limit. However, it is possible to obtain scaling much closer to the Heisenberg limit using an alternative scheme where the entropy of the phase distribution is minimized instead. This approach was first presented in Ref. [29]. Here, we have simplified and improved that scheme in a number of ways.

Some other alternatives that might be expected to provide phase measurements at the Heisenberg limit do not. This can be conveniently proven by examining canonical measurements for the equivalent two-mode state. One such alternative is hybrid measurements, where the QPEA is supplemented with single-photon detections, to reduce the probability of large phase errors. This approach yields a canonical phase uncertainty scaling as $1/N^{3/4}$, and the uncertainty in the phase estimate from the hybrid measurements scales slightly worse. Another alternative that does not yield the Heisenberg limit is simply increasing the number of repetitions, rather than the maximum size of the NOON state (or, equivalently, the number of passes). While this gives results which remain below the standard quantum limit of $1/\sqrt{N}$ for phase precision, they *scale* in exactly the same way as the SQL.

In this paper we have demonstrated this latter result experimentally by keeping the maximum number of passes fixed while increasing the number of photons used. This demonstration helps us to illustrate why gravitational wave interferometers do not approach the limit of $1/N$ for phase precision even though they use multiple passes: in such interferometers the number of passes is *fixed* at its maximum value (which is typically very large), while the number of "repetitions" (the number of photons in the coherent state which is used) is also made very large. This contrasts with all of the schemes that do scale as $1/N$, in which the number of repetitions is of order unity for the largest pass number (or the largest NOON state), and in which the number of passes (or the size of the NOON state) must be varied across a large range of values from one to its maximum value.

APPENDIX A: PROOF OF THEOREM 1

Here, we give the full proof of Theorem 1, which gives the scaling of the canonical phase variance for repetitions of the measurements. In general $f_M(n)$ can be obtained by the recurrence relation

$$f_M(n) = \sum_{k=0}^{N_K} f_{M-1}(n-k). \tag{A1}$$

Note that this is a discrete convolution of f_{M-1} with the uniform distribution. This recurrence relation comes about because the number of combinations of the values of $n_1, n_2, \ldots, n_M$ may be obtained, for each value of n_1, by determining the number of combinations of $n_2, \ldots, n_M$ that sum to $n-n_1$. The simplest example is $M=1$, where $f_1(n)=1$ for $0 \le n \le N_K$ and $f_1(n)=0$, otherwise. For $M=2$, evaluating the sum yields

$$f_2(n) = \begin{cases} n+1, & 0 \le n \le N_K \\ 2N_K - n + 1, & N_K \le n \le 2N_K \\ 0, & \text{otherwise}. \end{cases} \tag{A2}$$

A crucial point is that the increment between successive values of f_M is bounded. The recurrence relation implies

$$|f_M(n+1) - f_M(n)| \le (N_K + 1)\max_k |f_{M-1}(k+1) - f_{M-1}(k)|. \tag{A3}$$

Because $|f_2(n+1) - f_2(n)| \le 1$, this means that

$$|f_M(n+1) - f_M(n)| \le (N_K + 1)^{M-2}. \tag{A4}$$

The function is symmetric, so $f_M(n) = f_M(N-n)$. It is possible to determine the general value of $f_M(n)$ for $n \le N_K$. The solution is

$$f_M(n) = \binom{n+M-1}{M-1} = \frac{(n+M-1)!}{n!\,(M-1)!}. \tag{A5}$$

To prove this, first note that it is correct for $M=2$. Next, use the recurrence relation to give

$$\begin{aligned} f_M(n) - f_M(n-1) &= \sum_{k=0}^{N_K} f_{M-1}(n-k) - \sum_{k=0}^{N_K} f_{M-1}(n-k-1) \\ &= \sum_{k=0}^{N_K} f_{M-1}(n-k) - \sum_{k=1}^{N_K+1} f_{M-1}(n-k) \\ &= f_{M-1}(n) - f_{M-1}(n-N_K-1). \end{aligned} \tag{A6}$$

For $n \le N_K$, the argument of $f_M(n-N_K-1)$ is negative, so this term is zero. If the solution is correct for $M-1$ and for $n-1$, then

$$\begin{aligned} f_M(n) &= f_M(n-1) + f_{M-1}(n) \\ &= \frac{(n+M-2)!}{(n-1)!\,(M-1)!} + \frac{(n+M-2)!}{n!\,(M-2)!} \\ &= (n+M-1)\frac{(n+M-2)!}{n!\,(M-1)!} \\ &= \frac{(n+M-1)!}{n!\,(M-1)!}. \end{aligned} \tag{A7}$$

Using the formula gives, for $n < N_K$,

$$f_M(n+1) - f_M(n) \le (n+2)^{M-2}, \tag{A8}$$

$$f_M(n+1) + f_M(n) \ge \frac{(n+2)^{M-1}}{(M-1)!}. \tag{A9}$$

For $M > 2$, this gives

$$\frac{[f_M(n+1) - f_M(n)]^2}{[f_M(n+1) + f_M(n)]} \le (M-1)!\,(N_K+1)^{M-3}. \tag{A10}$$

Due to symmetry, the same result holds for $n > N - N_K$. This part of the derivation also holds for $n=-1$ or N. For $M=2$ or 1, a decreasing function of n is obtained, so n cannot be replaced with N_K for the upper bound.

The distribution for $f_M(n)$ cannot have a nontrivial minimum (the trivial minimum is zero at the bounds). That is, it cannot decrease with n, then increase. This can be shown by induction in the following way. If and only if $f_M(n)$ has a nontrivial minimum, then there exist n, n', $\delta n > 0$, and $\delta n' > 0$ such that $n' < n$, $f_M(n) > f_M(n - \delta n)$, and $f_M(n') < f_M(n' - \delta n')$. That would imply $f_{M-1}(n) > f_{M-1}(n - N_K - \delta n)$ and $f_{M-1}(n') < f_{M-1}(n' - N_K - \delta n')$. However, that would imply that $f_{M-1}(n)$ has a nontrivial minimum. As we know that is not the case for $M = 1$ and 2, using induction shows that $f_M(n)$ cannot have a nontrivial minimum for any M.

This result implies that $f_M(n)$ increases to a maximum at the center, then decreases. In particular, for $N_K \le n \le N - N_K$, $f_M(n) \ge f_M(N_K)$. Now we have $f_M(N_K) \ge (N_K + 1)^{M-1}/(M-1)!$, so

$$\frac{[f_M(n+1) - f_M(n)]^2}{[f_M(n+1) + f_M(n)]} \le (M-1)! \, (N_K + 1)^{M-3}, \quad \text{(A11)}$$

for $N_K < n \le N - N_K$. Therefore, overall

$$\sum_{n=-1}^{N_K} \frac{[f_M(n+1) - f_M(n)]^2}{[f_M(n+1) + f_M(n)]} \le M! \, (N_K + 1)^{M-2}. \quad \text{(A12)}$$

The variance V_C for the state is given by

$$2(1 - |\langle e^{i\hat\phi} \rangle|) = \frac{1}{(N_K+1)^M} \sum_{n=-1}^{N} \left[\sqrt{f_M(n+1)} - \sqrt{f_M(n)} \right]^2$$

$$\le \frac{1}{(N_K+1)^M} \sum_{n=-1}^{N} \frac{[f_M(n+1) - f_M(n)]^2}{[f_M(n+1) + f_M(n)]}$$

$$\le \frac{M!}{(N_K+1)^2} = \frac{M! \, M^2}{(N+M)^2}. \quad \text{(A13)}$$

Hence, for $M > 2$, the variance scales as the Heisenberg limit of $1/N^2$.

In the case $M = 2$, the bound for $n < N_K$ is

$$\frac{[f_M(n+1) - f_M(n)]^2}{[f_M(n+1) + f_M(n)]} \le \frac{1}{n+1}. \quad \text{(A14)}$$

Taking the sum from $n = 0$ to $N_K - 1$ gives

$$\sum_{n=0}^{N} \frac{[f_M(n+1) - f_M(n)]^2}{[f_M(n+1) + f_M(n)]} \le \int_0^{N_K} \frac{dn}{n+1} = \ln(N_K + 1). \quad \text{(A15)}$$

This approach therefore yields an upper bound on the variance that scales as $(\ln N)/N^2$, rather than the Heisenberg limit of $1/N^2$. In fact, this is the actual scaling for these states. Evaluating the exact value of V_C gives

$$2(1 - |\langle e^{i\hat\phi} \rangle|) = 2\left(1 - \frac{2}{(N_K+1)^2} \sum_{n=1}^{N_K} \sqrt{n(n+1)}\right)$$

$$= \frac{2\ln N}{N^2} + \frac{c}{N^2} + O(N^{-3}). \quad \text{(A16)}$$

where $c \approx 6.5949$.

APPENDIX B: DETAILS OF SHARPNESS FORMULA

Here, we give further details of the meaning of Eq. (3.7) and its interpretation as an average over the sharpnesses of the individual Bayesian probability distributions obtained. A subtlety that has been omitted in the discussion in the main text is that the probability distribution for the measurement results is now a function of both the system phase and the initial feedback phase and should therefore be written as $P(\vec{u}_m | \phi, \Phi_1)$. The other feedback phases are all chosen deterministically based on the measurement results and Φ_1 and can be omitted here. We use the notation Φ_j to indicate the feedback phase before the jth detection. The value of μ may then be written explicitly as

$$\mu = \left| \frac{1}{2\pi} \int d\Phi_1 \sum_{\vec{u}_m} P(\vec{u}_m | \phi, \Phi_1) e^{i\hat\phi} \right|. \quad \text{(B1)}$$

Because the phases are relative, the probability of the measurement results depends only on the difference between the system phase and the first feedback phase. That is,

$$P(\vec{u}_m | \phi, \Phi_1) = P(\vec{u}_m | \phi - \Phi_1, 0). \quad \text{(B2)}$$

Similarly the probability distribution for the system phase will depend on Φ_1 but will depend only on the difference between the two; that is,

$$P(\phi | \vec{u}_m, \Phi_1) = P(\phi - \Phi_1 | \vec{u}_m, 0). \quad \text{(B3)}$$

In the remainder of this appendix we use the probabilities in this form and omit the argument of 0.

The later feedback phases, as well as the phase estimate $\hat\phi$, depend on the measurement results and the value of Φ_1. Because the phases are relative, these phases should only depend in a linear way on the initial feedback phase. That is,

$$\hat\phi(\vec{u}_m, \Phi_1) = \hat\phi(\vec{u}_m, 0) + \Phi_1. \quad \text{(B4)}$$

Using this, and changing variables to $\aleph = \phi - \Phi_1$, we find that μ becomes

$$\mu = \left| \frac{1}{2\pi} \int d\aleph \sum_{\vec{u}_m} P(\vec{u}_m | \aleph, 0) e^{i[\hat\phi(\vec{u}_m, 0) - \aleph]} \right|. \quad \text{(B5)}$$

In the following analysis we need to consider a set of measurement results given an actual system phase ϕ and an initial feedback phase Φ_1, as well as the Bayesian probability distribution for the system phase given those measurement results. Since we effectively have two system phases, the actual system phase and the dummy variable for the system phase in the Bayesian probability distribution, we use the symbol ϕ_B for the variable in the Bayesian probability distribution. We also use $\aleph_B \equiv \phi_B - \Phi_1$. Then the optimal value of $\hat\phi$ is

$$\hat{\phi}(\vec{u}_m, \Phi_1) = \Phi_1 + \arg \int e^{i\aleph_B} P(\aleph_B | \vec{u}_m) d\aleph_B$$

$$= \arg \int e^{i(\aleph_B + \Phi_1)} P(\aleph_B | \vec{u}_m) d\aleph_B$$

$$= \arg \int e^{i\phi_B} P(\phi_B - \Phi_1 | \vec{u}_m) d\phi_B = \arg\langle e^{i\phi_B}\rangle,$$
(B6)

where the angular brackets indicate that the average is taken for that value of $\vec{u}_m$ and Φ_1. The integral for μ then becomes

$$\mu = \frac{1}{2\pi} \sum_{\vec{u}_m} \left| \int e^{i\aleph} P(\vec{u}_m | \aleph) d\aleph \right|.$$
(B7)

In this expression the integral over $\aleph$ can be an integral over the actual system phase for fixed Φ_1 or an integral over the feedback phase Φ_1 for fixed ϕ. For Eq. (3.7) in the main text, we have taken $\Phi_1 = 0$ so $\aleph = \phi$.

This expression can be rewritten as

$$\mu = \frac{1}{2\pi} \int d\Phi_1 \sum_{\vec{u}_m} P(\vec{u}_m | \phi_B - \Phi_1) \left| \int e^{i\aleph} P(\aleph | \vec{u}_m) d\aleph \right|.$$
(B8)

The interpretation of the variables in this expression now changes. We now interpret this expression as the sharpness of the Bayesian probability distribution for the system phase given measurement results $\vec{u}_m$ averaged over the initial feedback phase Φ_1 and measurement results $\vec{u}_m$, for a specific system phase. To make this clearer, we can relabel the variables as below,

$$\mu = \frac{1}{2\pi} \int d\aleph \sum_{\vec{u}_m} p(\vec{u}_m | \aleph) \left| \int e^{i\aleph_B} P(\aleph_B | \vec{u}_m) d\aleph_B \right|$$

$$= \frac{1}{2\pi} \int d\aleph \sum_{\vec{u}_m} p(\vec{u}_m | \aleph) \left| \int e^{i\phi_B} P(\phi_B - \Phi_1 | \vec{u}_m) d\phi_B \right|$$

$$= E[|\langle e^{i\phi_B}\rangle|].$$
(B9)

E indicates an expectation value over the measurement results and $\aleph = \phi - \Phi_1$. This is independent of the actual system phase and may therefore be regarded as an average over the measurement results and Φ_1. As above the angular brackets indicate an average over the system phase in the Bayesian probability distribution ϕ_B. In the main text we simply call this ϕ, as we do not have the actual system phase and system phase for the Bayesian probability distribution in the same expression.

APPENDIX C: GENERAL MEASURE OF RESOURCES

As explained in Sec. II, we adopt the general definition of N in a system-independent way as the size of the smallest interval supporting the Fourier transform of the family of states $|\psi(\phi)\rangle$. Here, $|\psi(\phi)\rangle$ is interpreted as follows. The entire measurement scheme may be regarded as the preparation of a pure quantum state, followed by a unitary operation that

depends on ϕ, and finally a measurement. The preparation of the state and the ϕ-dependent unitary operation may be grouped together into preparation of the ϕ-dependent state $|\psi(\phi)\rangle$. In general there will be restrictions on both the input state (for example, a limited photon number) and the unitary operation (for example, a limited number of passes through the phase shift) that will limit the possible states $|\psi(\phi)\rangle$.

Taking the Fourier transform of the family of states $|\psi(\phi)\rangle$ yields

$$|\tilde{\psi}(s)\rangle = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-is\phi} |\psi(\phi)\rangle d\phi.$$
(C1)

The states $|\tilde{\psi}(s)\rangle$ are not (necessarily) normalized. We consider the case where restrictions on the measurement scheme imply that the family $|\tilde{\psi}(s)\rangle$ has support on a finite interval $[N_{\min}, N_{\max}]$, and so

$$|\psi(\phi)\rangle = \frac{1}{\sqrt{2\pi}} \int_{N_{\min}}^{N_{\max}} e^{is\phi} |\tilde{\psi}(s)\rangle ds.$$
(C2)

We define N as the smallest value of $N_{\max} - N_{\min}$.

There are two main cases that we can consider:

(1) $|\tilde{\psi}(s)\rangle$ has support on a discrete set $\{s_j\} \in [N_{\min}, N_{\max}]$, with every $s_i - s_j$ being integer valued, and

(2) $|\tilde{\psi}(s)\rangle$ may be nonzero for any $s \in [N_{\min}, N_{\max}]$.

In the first case, $|\psi(\phi)\rangle$ is periodic in ϕ with period 2π. Then ϕ is a phase, and the variance may be measured via the Holevo variance. In the second case there is no periodicity imposed on $|\psi(\phi)\rangle$. The appropriate measure of the variance of ϕ is then the usual variance.

The first thing to notice is that the states $|\tilde{\psi}(s)\rangle$ may be taken to be orthogonal without affecting the minimum variance that may be obtained. This can be proven as follows. Writing these states in some basis of orthogonal states $|\zeta\rangle$,

$$|\psi(\phi)\rangle = \frac{1}{\sqrt{2\pi}} \int_{N_{\min}}^{N_{\max}} e^{is\phi} \int_{-\infty}^{\infty} \psi(s, \zeta) |\zeta\rangle d\zeta ds.$$
(C3)

Now we can construct the modified state $|\psi'(\phi)\rangle$ by adding an additional subspace with orthogonal states $|s\rangle$,

$$|\psi'(\phi)\rangle = \frac{1}{\sqrt{2\pi}} \int_{N_{\min}}^{N_{\max}} e^{is\phi} |s\rangle \int_{-\infty}^{\infty} \psi(s, \zeta) |\zeta\rangle d\zeta ds.$$
(C4)

This modified state is not necessarily normalized. Performing a measurement of this additional subspace via projections $|\theta\rangle\langle\theta|$, with

$$|\theta\rangle = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{is\theta} |s\rangle ds,$$
(C5)

yields

$$\langle\theta|\psi'(\phi)\rangle = \frac{1}{2\pi} \int_{N_{\min}}^{N_{\max}} e^{is(\phi-\theta)} \int_{-\infty}^{\infty} \psi(s, \zeta) |\zeta\rangle d\zeta ds.$$
(C6)

This state is unchanged from that in Eq. (C3), except for a shift in ϕ of θ. Thus, the state $|\psi'(\phi)\rangle$ cannot give any less information about ϕ than $|\psi(\phi)\rangle$. In addition, the state

$|\psi'(\phi)\rangle$ is of the form (C2), except $|\widetilde{\psi}(s)\rangle$ has been replaced with the orthogonal states

$$|s\rangle \int_{-\infty}^{\infty} \psi(s,\zeta)|\zeta\rangle d\zeta. \qquad (C7)$$

Hence, the states $|\widetilde{\psi}(s)\rangle$ may be replaced with orthogonal states with no loss of phase information. Thus, in examining the limits to estimation of ϕ, we can assume the states $|\widetilde{\psi}(s)\rangle$ are mutually orthogonal with no loss of generality.

In the case where the frequencies $\{s_j\}$ are restricted to differ by integer values, then $|\widetilde{\psi}(s)\rangle$ becomes a series of delta functions in s. The integral (C2) becomes a sum

$$|\psi(\phi)\rangle = e^{iN_{\min}\phi}\sum_{s=0}^{N} e^{is\phi}\widetilde{\psi}(s)|s\rangle, \qquad (C8)$$

where

$$\widetilde{\psi}(s)|s\rangle = \frac{1}{\sqrt{2\pi}}\int_{s-\epsilon}^{s+\epsilon} |\widetilde{\psi}(s')\rangle ds', \qquad (C9)$$

for $1 > \epsilon > 0$. Assuming the states $|\widetilde{\psi}(s)\rangle$ are mutually orthogonal, the states $|s\rangle$ are mutually orthogonal as well. This state is then identical to that for a phase shift on a single mode with a maximum of N photons, and it has a phase uncertainty lower bounded by Eq. (2.3), which we repeat here,

$$\Delta\phi_{HL} = \sqrt{V_H} = \tan\left(\frac{\pi}{N+2}\right) \sim \frac{\pi}{N}. \qquad (C10)$$

In the case where $|\widetilde{\psi}(s)\rangle$ is nonzero for any value of the frequency $s \in [N_{\min}, N_{\max}]$, we may find the bound on the variance as follows. Without loss of generality, the family of states can be given by

$$|\psi(\phi)\rangle = \int_0^N e^{is\phi}\widetilde{\psi}(s)|s\rangle, \qquad (C11)$$

where here we have normalized $\{|s\rangle\}$ as $\langle s|s'\rangle = \delta(s-s')$. In this continuous case the system is equivalent to position and momentum, with the frequency s equivalent to position and ϕ equivalent to momentum (provided we use units where $\hbar = 1$). It is therefore clear that the optimal measurement for ϕ is a projection in the basis

$$|\phi\rangle = \int_{-\infty}^{\infty} e^{is\phi}|s\rangle. \qquad (C12)$$

This yields a phase variance of

$$V = \int_{-\infty}^{\infty}\phi^2|\psi(\phi)|^2 d\phi - \left(\int_{-\infty}^{\infty}\phi|\psi(\phi)|^2 d\phi\right)^2, \quad (C13)$$

where $\psi(\phi)$ is the inverse Fourier transform of $\widetilde{\psi}(s)$. This result is for ϕ initially completely unknown, with a flat prior distribution over the whole real line.

If we apply a phase shift of $\Delta\phi$ to the state, then this simply changes the mean value of ϕ but leaves the variance

unchanged. Therefore, in considering the problem of minimizing the variance, we may consider states with mean ϕ equal to zero without loss of generality. The minimum uncertainty state then follows from the analogy of position and momentum. The bounds of 0 and N on s are equivalent to an infinite square well potential of size N. The minimum energy of a state in this infinite square well is $E = \pi^2/(2N^2)$, where we have taken $\hbar = 1$ and the mass equal to 1. Because ϕ is equivalent to the momentum, and we are taking the mean ϕ to be equal to zero, $V = 2E$, so the minimum variance is $V = \pi^2/N^2$. This gives the lower bound on the uncertainty of

$$\Delta\phi_{HL} = \sqrt{V} = \frac{\pi}{N}. \qquad (C14)$$

This result may also be found by taking the limit of the discrete case.

The case where the power of $e^{i\phi}$ can only take integer values is relevant to interferometry and to the gate formalism of Ref. [27]. For photons, each pass of a single photon through the phase shift gives a multiplication by a factor of $e^{i\phi}$, so $N_{\max}$ corresponds to an upper limit on the number of photon passes for the measurement scheme. It is not possible to have a negative number of photon passes, and normally there is no lower bound on the number of photon passes, so $N_{\min} = 0$ and N is the maximum number of photon passes. In cases where there is a minimum number of photon passes, then N can also be taken to be the difference between the minimum and the maximum numbers of photon passes and yield a tighter bound on the phase uncertainty.

In the gate formalism of Ref. [27] each application of the phases shift is a gate of the form

$$\begin{bmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{bmatrix}, \qquad (C15)$$

applied to a qubit. Each application gives multiplication by a factor of $e^{i\phi}$ for the $|1\rangle$ basis state, and no multiplication for the $|0\rangle$ basis state. Therefore, an upper bound on the power of $e^{i\phi}$, $N_{\max}$, is given by the number of applications of the phase shift. The lower bound is zero, so N is simply the number of applications of the phase shift. In particular cases the details of the measurement scheme may mean that there are tighter bounds on the power. For example, for some gates the qubit may be initialized to $|1\rangle$, in which case $N_{\min}$ could be taken to be greater than zero. As in the case of photons, this would yield a tighter bound on the phase uncertainty.

The third application we discuss is that of metrology of a Hamiltonian with an unknown parameter ϕ, describable as $H_\phi = \phi H$. We consider the estimation of ϕ by probing the Hamiltonian for a total time T. This is the total time that systems evolve under the Hamiltonian, which can include parallel or serial evolution of systems. The maximum power of $e^{i\phi}$ that can be obtained is $-\lambda_{\min}T$, where $\lambda_{\min}$ is the minimum eigenvalue of H. The minimum power that can be obtained is $-\lambda_{\max}T$, where $\lambda_{\max}$ is the maximum eigenvalue of H. The difference between the upper and the lower bounds on the power of $e^{i\phi}$ is therefore $N = \|H\|T$, where $\|H\|$ is the difference between the maximum and the minimum eigenvalues of H.

In the case of metrology of a Hamiltonian, the time intervals can be any real numbers, so the powers of $e^{i\phi}$ can take any real values within the bounds. Therefore, ϕ may be measured as any real number, rather than modulo 2π, and the bound on the phase uncertainty is exactly $\pi/N = \pi/(\|H\|T)$ and is measured by the standard deviation. In Ref. [44] the total probe time is νt because there are ν independent probes of time t. The limit to the phase measurement precision there is again π/N, with $N = \nu t\|H\|$. Note that Ref. [44] uses the terminology "fundamental limit" for the minimum possible phase uncertainty and uses the terminology "Heisenberg limit" in a different sense.

[1] A. Y. Kitaev, Electronic Colloquium on Computational Complexity **3**, 3 (1996).

[2] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, Proc. R. Soc. London A **454**, 339 (1998).

[3] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2000).

[4] The notation Θ means the asymptotic scaling, as opposed to the notation O, which means an upper bound on the asymptotic scaling.

[5] C. M. Caves, Phys. Rev. D **23**, 1693 (1981).

[6] B. Yurke, S. L. McCall, and J. R. Klauder, Phys. Rev. A **33**, 4033 (1986).

[7] M. Xiao, L. A. Wu, and H. J. Kimble, Phys. Rev. Lett. **59**, 278 (1987).

[8] P. Grangier, R. E. Slusher, B. Yurke, and A. LaPorta, Phys. Rev. Lett. **59**, 2153 (1987).

[9] K. Goda, O. Miyakawa, E. E. Mikhailov, S. Saraf, R. Adhikari, K. McKenzie, R. Ward, S. Vass, A. J. Weinstein, and N. Mavalva, Nat. Phys. **4**, 472 (2008).

[10] B. C. Sanders, Phys. Rev. A **40**, 2417 (1989).

[11] J. J. Bollinger, W. M. Itano, D. J. Wineland, and D. J. Heinzen, Phys. Rev. A **54**, R4649 (1996).

[12] H. Lee, P. Kok, and J. P. Dowling, J. Mod. Opt. **49**, 2325 (2002).

[13] J. G. Rarity, P. R. Tapster, E. Jakeman, T. Larchuk, R. A. Campos, M. C. Teich, and B. E. A. Saleh, Phys. Rev. Lett. **65**, 1348 (1990).

[14] E. J. S. Fonseca, C. H. Monken, and S. Pádua, Phys. Rev. Lett. **82**, 2868 (1999).

[15] K. Edamatsu, R. Shimizu, and T. Itoh, Phys. Rev. Lett. **89**, 213601 (2002).

[16] P. Walther, J.-W. Pan, M. Aspelmeyer, R. Ursin, S. Gasparoni, and A. Zeilinger, Nature (London) **429**, 158 (2004).

[17] M. W. Mitchell, J. S. Lundeen, and A. M. Steinberg, Nature (London) **429**, 161 (2004).

[18] D. Leibfried *et al.*, Nature (London) **438**, 639 (2005).

[19] H. S. Eisenberg, J. F. Hodelin, G. Khoury, and D. Bouwmeester, Phys. Rev. Lett. **94**, 090502 (2005).

[20] F. W. Sun, B. H. Liu, Y. F. Huang, Z. Y. Ou, and G. C. Guo, Phys. Rev. A **74**, 033812 (2006).

[21] T. Nagata, R. Okamoto, J. L. O'Brien, K. Sasaki, and S. Takeuchi, Science **316**, 726 (2007).

[22] K. J. Resch, K. L. Pregnell, R. Prevedel, A. Gilchrist, G. J. Pryde, J. L. O'Brien, and A. G. White, Phys. Rev. Lett. **98**, 223601 (2007).

[23] D. Coppersmith, IBM Research Report No. RC19642, 1994 (unpublished).

[24] P. W. Shor, *Proceedings of the 35th Annual Symposium on Foundations of Computer Science* (IEEE Computer Society Press, Los Alamitos, CA, 1994), p. 124.

[25] R. B. Griffiths and C. S. Niu, Phys. Rev. Lett. **76**, 3228 (1996).

[26] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde, Nature (London) **450**, 393 (2007).

[27] W. van Dam, G. M. D'Ariano, A. Ekert, C. Macchiavello and M. Mosca, Phys. Rev. Lett. **98**, 090501 (2007).

[28] B. L. Higgins, D. W. Berry, S. D. Bartlett, M. W. Mitchell, H. M. Wiseman, and G. J. Pryde, New J. Phys. **11**, 073023 (2009).

[29] M. W. Mitchell, Proc. SPIE **5893**, 589310 (2005).

[30] S. Boixo and R. D. Somma, Phys. Rev. A **77**, 052320 (2008).

[31] W. Heitler, *Quantum Theory of Radiation* (Oxford University Press, London, 1954).

[32] A. S. Holevo, in *Quantum Probability and Applications to the Quantum Theory of Irreversible Processes*, Springer Lecture Notes in Mathematics Vol. 1055, edited by L. Accardi, A. Frigerio, and V. Gorini (Springer, Berlin, 1984), p. 153.

[33] A. S. Lane, S. L. Braunstein, and C. M. Caves, Phys. Rev. A **47**, 1667 (1993).

[34] Z. Y. Ou, Phys. Rev. Lett. **77**, 2352 (1996).

[35] G. S. Summy and D. T. Pegg, Opt. Commun. **77**, 75 (1990).

[36] H. M. Wiseman and R. B. Killip, Phys. Rev. A **56**, 944 (1997).

[37] A. Luis and J. Peřina, Phys. Rev. A **54**, 4564 (1996).

[38] D. W. Berry and H. M. Wiseman, Phys. Rev. Lett. **85**, 5098 (2000).

[39] D. W. Berry, H. M. Wiseman, and J. K. Breslin, Phys. Rev. A **63**, 053804 (2001).

[40] J. H. Shapiro, S. R. Shepard, and N. C. Wong, Phys. Rev. Lett. **62**, 2377 (1989).

[41] D. Meiser and M. J. Holland, New J. Phys. **11**, 033002 (2009).

[42] L. Pezzé and A. Smerzi, Phys. Rev. Lett. **100**, 073601 (2008).

[43] A. Luis, Phys. Lett. A **329**, 8 (2004).

[44] S. Boixo, A. Datta, S. T. Flammia, A. Shaji, E. Bagan, and C. M. Caves, Phys. Rev. A **77**, 012317 (2008).

[45] B. C. Sanders and G. J. Milburn, Phys. Rev. Lett. **75**, 2944 (1995).

[46] B. C. Sanders, G. J. Milburn, and Z. Zhang, J. Mod. Opt. **44**, 1309 (1997).

[47] H. M. Wiseman, Phys. Rev. Lett. **75**, 4587 (1995).

[48] M. J. Collett, Phys. Scr. **T48**, 124 (1993).

[49] E. Bagan, A. Monras, and R. Munoz-Tapia, Phys. Rev. A **78**, 043829 (2008).

[50] For this solution the arg function in Eq. (5.9) must be taken to give values in the interval $(-2\pi, 0]$, so Φ is positive.

[51] W. Hoeffding, J. Am. Stat. Assoc. **58**, 13 (1963).