Maximum-likelihood statistics of multiple quantum phase measurements

Alistair S. Lane, Samuel L. Braunstein,* and Carlton M. Caves†
Center for Laser Studies, University of Southern California, Los Angeles, California 90089-1112
(Received 3 August 1992)

Shapiro, Shepard, and Wong [Phys. Rev. Lett. **62**, 2377 (1989)] suggested that a scheme of multiple phase measurements, using quantum states with minimum "reciprocal peak likelihood," could achieve a phase sensitivity scaling as $1/N_{\rm tot}^2$, where $N_{\rm tot}$ is the mean number of photons available for all measurements. We have simulated their scheme for as many as 240 measurements and have found optimum phase sensitivities for $3 \le N_{\rm tot} \le 120$; a power-law fit to the simulated data yields a phase sensitivity that scales as $1/N_{\rm tot}^{0.82\pm0.01}$. By using a combination of numerical and analytical techniques, we extend our results to higher values of $N_{\rm tot}$ than are accessible to our simulations; we find no evidence for phase sensitivities better than the benchmark $1/N_{\rm tot}$ sensitivity of squeezed-state interferometry. We conclude that reciprocal peak likelihood is not a good measure of phase sensitivity. We discuss other factors that are important to phase sensitivity.

PACS number(s): 03.65.Bz, 42.50.Dv, 89.70.+c

I. INTRODUCTION

Phase changes in optical-frequency fields, produced by tiny changes in optical path length, are used to make some of the most exquisitely sensitive measurements in physics. If an arbitrarily large amount of energy is available, there are no (nonrelativistic) limits on the accuracy of phase measurements. In real devices, however, practical limitations restrict how much energy can be used. Thus it is sensible physically to pose the following question: given a limit on the available energy—or, equivalently, on the available number N of photons—what limit does quantum mechanics impose on the accuracy of phase measurements?

A back-of-the-envelope answer to this question combines a heuristic number-phase uncertainty principle,

$$\Delta N \Delta\phi \gtrsim 1 , \tag{1.1}$$

with the assumption that the photon-number uncertainty ΔN is bounded by the mean photon number N,

$$\Delta N \lesssim N , \tag{1.2}$$

and thus arrives at a limit on phase uncertainty,

$$\Delta\phi \gtrsim 1/N . \tag{1.3}$$

Is there anything wrong with this simple derivation? The uncertainty principle (1.1) is not rigorous as it stands, but it can be made so without affecting the argument [1, 2]. The questionable assumption is Eq. (1.2): it is easy to devise quantum states for which the photon-number uncertainty ΔN is arbitrarily large compared to the mean number N; there is no reason to exclude such states, especially for the multiple-measurement schemes of the sort that are the subject of this paper. This scarcely demonstrates that it is possible to beat the naive limit (1.3), but it leaves the question open and stimulates further enquiry.

The naive limit (1.3) serves mainly to sharpen the physically relevant question by establishing a benchmark sensitivity: given N photons, is it possible to measure a phase with a sensitivity better than $1/N$? This paper considers a particular scheme for measuring phase, proposed by Shapiro, Shepard, and Wong (SSW) [3]. The SSW scheme assumes that "ideal phase measurements" are made on a family of quantum states that SSW obtained by maximizing the "peak likelihood," given a constraint on mean photon number. Using "reciprocal peak likelihood" as their measure of sensitivity, SSW conjectured that their scheme had a phase sensitivity that scales as $1/N^2$.

The phase probability distribution for a SSW state has broad tails running up to a narrow central peak, under which lies negligible probability. Because of the broad tails, a single sampling of the SSW phase distribution provides almost no information about the location of the central peak. SSW argued (correctly) that many samplings could determine the peak's location accurately. To perform many measurements, however, requires many copies of the SSW state and thus more photons. We *interpret* SSW's proposal as a suggestion that the phase sensitivity of *multiple* SSW phase measurements scales as $1/N^2$, where N is now understood to be the mean number of photons available for *all* the measurements.

This paper reports the results of a detailed investigation of the multiple-measurement version of the SSW scheme. Much of our effort has gone into numerical simulation of multiple SSW phase measurements. For the photon numbers accessible to our simulations, which unfortunately are much smaller than photon numbers readily available in the laboratory, we find that the SSW scheme does not beat the naive $1/N$ limit [4]. We are unable to determine numerically the sensitivity for larger photon numbers, because the required simulations lie well outside the capabilities of our computers. Nonetheless, by using a judicious combination of results from our sim-

ulations and key analytic results [5,6], we obtain sensitivities for photon numbers much higher than in our simulations, although sensitivities for arbitrarily high photon numbers remain beyond our reach.

There are two main approaches to investigating quantum limits on phase sensitivity. The first approach pays considerable attention to realizability with present technology. Pick a favorite *realizable* technique (or potentially realizable technique with extensions of present technology) for measuring phase; examples include interferometric measurements of the phase difference between two optical paths or heterodyne measurements of both field quadratures from which one derives a phase [7–11]. Then investigate the phase sensitivity obtained with various *realizable* quantum states (or potentially realizable quantum states with extensions of present technology) as input states. For example, in the case of interferometry, the sensitivity is determined by the noise properties of the light that enters the interferometer's normally unused input port [12]. If that port is "illuminated" by vacuum, the phase sensitivity is $1/\sqrt{N}$, whereas if it is illuminated by properly squeezed light, one can achieve the benchmark $1/N$ sensitivity [12–14]. Both the strength and the weakness of this first approach lie in its attention to present technology: this attention ensures that there is some connection to real measurements, but leaves no assurance that one is addressing issues of ultimate sensitivity.

The second approach attempts to get at the question of ultimate sensitivity by abandoning any pretense to realizability with present technology. Instead, assume an "ideal phase measurement"—the best phase measurement allowed by quantum mechanics—without worrying about how to realize such a measurement. Then, using some criterion for phase sensitivity, optimize the sensitivity, subject to the constraint on available number of photons, over the set of *all* input quantum states. Historically this approach has run into trouble at the first step, bedeviled by controversy over the formal description of an ideal phase measurement [15–18]. The difficulty lies in the absence of a Hermitian phase operator. By analogy with position and momentum operators, one tries to introduce a phase operator that is conjugate to the number operator $\hat{N} = \hat{a}^\dagger\hat{a}$, where $\hat{a}^\dagger$ and $\hat{a}$ are creation and annihilation operators for the field mode of interest. This attempt fails, however, because the truncation of the photon-number series at $n = 0$ prevents the number-lowering operator $\widehat{e^{i\phi}}$ from being unitary and thus prevents its argument from being i times a Hermitian phase operator conjugate to $\hat{N}$ [16,19] (hence the wide "hat" on $\widehat{e^{i\phi}}$, which emphasizes that there is no Hermitian phase operator $\hat{\phi}$). In the absence of a Hermitian phase operator, one cannot use *conventional* quantum measurement theory to define an ideal phase measurement.

Susskind and Glogower (SG) [15] nonetheless introduced the nonunitary number-lowering operator,

$$\widehat{e^{i\phi}} \equiv \sum_{n=0}^{\infty} |n\rangle\langle n+1| = \hat{a}\hat{N}^{-1/2} = (\hat{N}+1)^{-1/2}\hat{a}\,, \tag{1.4}$$

as a polar decomposition of the annihilation operator. Much of subsequent work dwells on alleged problems with the SG operator (1.4) (for a discussion see Refs. [1, 20, 21]). These problems stem from the nonunitary character of $\widehat{e^{i\phi}}$, which means that $\widehat{e^{i\phi}}$ does not commute with its Hermitian conjugate:

$$[\widehat{e^{i\phi}}, (\widehat{e^{i\phi}})^\dagger] = |0\rangle\langle 0|\,. \tag{1.5}$$

Introducing the Hermitian real and imaginary parts of $\widehat{e^{i\phi}}$,

$$\widehat{e^{i\phi}} = \widehat{\cos\phi} + i\,\widehat{\sin\phi}\,, \tag{1.6}$$

one can rewrite the commutator (1.5) as

$$[\widehat{\cos\phi}, \widehat{\sin\phi}] = \frac{i}{2}|0\rangle\langle 0|\,. \tag{1.7}$$

The question of ideal phase measurements and of their connection to the SG operator was resolved some time ago [19, 22], in work on the formal theory of quantum measurements. The question has recently been taken up once again, in work coming from two quite different directions. The most direct way to proceed [2, 3, 19, 22–24] is to follow in the path of earlier work on phase measurements [19, 22] and to notice that the SG operator has (unnormalizable) eigenstates

$$|e^{i\phi}\rangle = \sum_{n=0}^{\infty} e^{in\phi}|n\rangle\,, \tag{1.8}$$

where

$$\widehat{e^{i\phi}}|e^{i\phi}\rangle = e^{i\phi}|e^{i\phi}\rangle\,. \tag{1.9}$$

The eigenstates (1.8) have the properties of states with definite phase [25], yet they are not orthogonal because of the nonunitary character of $\widehat{e^{i\phi}}$ (they are not δ-function normalized). That the states (1.8) are not orthogonal is another way of demonstrating that there is no Hermitian phase operator. Nonetheless, the states (1.8) are (over)complete in that they provide a resolution of the identity operator:

$$\int_{\phi_0}^{\phi_0+2\pi} \frac{d\phi}{2\pi}\,|e^{i\phi}\rangle\langle e^{i\phi}| = \hat{1}\,. \tag{1.10}$$

Here ϕ_0 is a reference phase that we leave arbitrary for the moment. For the bulk of this paper, however, we choose $\phi_0 = -\pi$, so that the relevant 2π interval is centered about $\phi = 0$.

Within the formalism of generalized measurements [19, 26, 27], the completeness property (1.10) is sufficient to associate a *phase probability distribution* with any quantum state $|\psi\rangle$,

$$P_{|\psi\rangle}(\phi) \equiv \left\langle \psi \left| \left(\frac{1}{2\pi}|e^{i\phi}\rangle\langle e^{i\phi}| \right) \right| \psi \right\rangle = \frac{1}{2\pi}|\langle e^{i\phi}|\psi\rangle|^2\,. \tag{1.11}$$

This distribution, normalized on any 2π interval as

$$\int_{\phi_0}^{\phi_0+2\pi} d\phi\, P_{|\psi\rangle}(\phi) = 1\,, \tag{1.12}$$

is interpreted as giving the statistics of an ideal phase measurement. The measure $|e^{i\phi}\rangle\langle e^{i\phi}|\, d\phi/2\pi$ is an example of a positive-operator-valued measure (POVM) [19, 26, 27]. Notice that one can introduce a "phase wave function" [19, 28]

$$\psi(\phi) \equiv \frac{1}{\sqrt{2\pi}}\langle e^{i\phi}|\psi\rangle = \frac{1}{\sqrt{2\pi}}\sum_{n=0}^{\infty} e^{-in\phi}\langle n|\psi\rangle\,, \tag{1.13}$$

whose absolute square is the distribution (1.11). The completeness property (1.10) yields the inverse of this expansion,

$$\langle n|\psi\rangle = \frac{1}{\sqrt{2\pi}}\int_{\phi_0}^{\phi_0+2\pi} d\phi\, e^{in\phi}\psi(\phi)\,. \tag{1.14}$$

Another tack has been taken by Pegg, Barnett, and coworkers [1,20,21,28–35] (for early work on this approach, see [36]). They truncate the harmonic-oscillator Hilbert space to an $(s+1)$-dimensional subspace spanned by the first $s+1$ photon-number eigenstates $|0\rangle, |1\rangle, \ldots, |s\rangle$. The intention is to calculate the quantum statistics of a *Hermitian* phase operator within the finite-dimensional subspace and then to take the $s \to \infty$ limit of those statistics at the end of the calculation. To this end they introduce within the finite-dimensional subspace $s+1$ orthonormal "phase eigenstates" [25]

$$|\phi_m\rangle_{(s)} \equiv \frac{1}{\sqrt{s+1}}\sum_{n=0}^{s} e^{in\phi_m}|n\rangle \tag{1.15}$$

[cf. Eq. (1.8)], labeled by phase eigenvalues

$$\phi_m = \phi_0^{(s)} + \frac{2\pi m}{s+1}\,,\quad m = 0, 1, \ldots, s\,. \tag{1.16}$$

Here $\phi_0^{(s)}$ is a reference phase that can be left arbitrary; a convenient choice, which leaves the eigenvalues centered about $\phi = 0$, is $\phi_0^{(s)} = -s\pi/(s+1)$. The phase eigenvalues are not continuous, but they become dense in the limit $s \to \infty$.

On the finite-dimensional subspace Pegg and Barnett introduce a Hermitian "phase operator" defined by

$$\hat{\phi}_{(s)} \equiv \sum_{m=0}^{s} \phi_m |\phi_m\rangle_{(s)\ (s)}\langle\phi_m|\,, \tag{1.17}$$

and they interpret a measurement of this operator as an ideal phase measurement. It is easy to show that $e^{i\hat{\phi}_{(s)}}$ is the number-lowering operator within the finite-dimensional subspace,

$$\begin{aligned} e^{i\hat{\phi}_{(s)}} &= \sum_{m=0}^{s} e^{i\phi_m}|\phi_m\rangle_{(s)\ (s)}\langle\phi_m| \\ &= \sum_{n=0}^{s-1} |n\rangle\langle n+1| + e^{i(s+1)\phi_0^{(s)}}|s\rangle\langle 0| \end{aligned} \tag{1.18}$$

[cf. Eq. (1.4)], except for the additional term that raises $|0\rangle$ to $|s\rangle$, thus making $e^{i\hat{\phi}_{(s)}}$ unitary.

One can derive standard phase probabilities on the finite-dimensional subspace and then take the limit $s \to \infty$. Specifically, for a quantum state $|\psi\rangle$ the probability that an ideal phase measurement yields a value in an interval Δ is

$$\begin{aligned} p_{|\psi\rangle}(\phi \in \Delta) &= \lim_{s\to\infty}\left(\sum_{\phi_m\in\Delta} |_{(s)}\langle\phi_m|\psi\rangle|^2\right) \\ &= \left\langle\psi\left|\left[\lim_{s\to\infty}\left(\sum_{\phi_m\in\Delta}|\phi_m\rangle_{(s)\ (s)}\langle\phi_m|\right)\right]\right|\psi\right\rangle, \end{aligned} \tag{1.19}$$

For *physical* [37] states $|\psi\rangle$, the sum in Eq. (1.19) can be replaced by an integral

$$\begin{aligned} &\lim_{s\to\infty}\left(\sum_{\phi_m\in\Delta}|\phi_m\rangle_{(s)\ (s)}\langle\phi_m|\right) \\ &\quad\longrightarrow \lim_{s\to\infty}\left(\int_\Delta \frac{d\phi}{2\pi/(s+1)}\,|\phi_m\rangle_{(s)\ (s)}\langle\phi_m|\right) \\ &\quad= \int_\Delta \frac{d\phi}{2\pi}\,|e^{i\phi}\rangle\langle e^{i\phi}|\,, \end{aligned} \tag{1.20}$$

where the $2\pi/(s+1)$ comes from the eigenvalue spacing. The result is the same probability distribution

$$P_{|\psi\rangle}(\phi) = \frac{p_{|\psi\rangle}(\phi\in d\phi)}{d\phi} = \left\langle\psi\left|\left(\frac{1}{2\pi}|e^{i\phi}\rangle\langle e^{i\phi}|\right)\right|\psi\right\rangle \tag{1.21}$$

[cf. Eq. (1.11)] as comes from using the SG operator to generate a POVM.

Coming from different directions, the two approaches to ideal phase measurements lead to the same probability distribution. The Hermitian-phase-operator approach uses conventional quantum measurement theory, but requires a potentially tricky limit; the POVM-based approach is direct, requiring no limit, but it assumes familiarity with the theory of generalized measurements. The reader can take his or her pick, with the assurance that the two approaches, properly used, produce identical quantum phase statistics. The equivalence between the two approaches for "physical states" has been demonstrated and emphasized by Shapiro and Shepard (SS) [2] and by Hall [23].

The improved formal understanding of phase measurements does not mean that anyone knows how to realize an ideal phase measurement, but it should end disputes about the formal description and resolve questions about the status of the SG operator. What then of the alleged problems of the SG operator [1, 20, 21]? The perceived problems invariably stem from calculating, with respect to state $|\psi\rangle$, some aspect of the statistics of the Hermitian operators $\widehat{\cos\phi}$ or $\widehat{\sin\phi}$ [Eq. (1.6)] and finding that the results are not in accord with one's expectations for an ideal phase measurement. For example, it is easy to ver-

ify that the vacuum-state second moments of $\widehat{\cos\phi}$ and $\widehat{\sin\phi}$ are

$$\langle 0|(\widehat{\cos\phi})^2|0\rangle = \langle 0|(\widehat{\sin\phi})^2|0\rangle = \tfrac{1}{4}\,, \tag{1.22}$$

whereas the random-phase character of the vacuum ought to lead to second moments equal to $\frac{1}{2}$. Indeed, since the vacuum has a uniform phase distribution $P_{|0\rangle}(\phi) = 1/2\pi$, the second moments of $\cos\phi$ and $\sin\phi$, calculated with respect to $P_{|0\rangle}(\phi)$, do equal $\frac{1}{2}$. It is easy to find the general relation between the two kinds of second moments:

$$\int_{\phi_0}^{\phi_0+2\pi} d\phi\, \cos^2\phi\, P_{|\psi\rangle}(\phi) = \langle\psi|(\widehat{\cos\phi})^2|\psi\rangle + \tfrac{1}{4}|\langle 0|\psi\rangle|^2 \tag{1.23}$$

(an identical relation holds for $\sin\phi$ replacing $\cos\phi$).

Is there a puzzle here? No. The second moment of $\cos\phi$ on the left-hand side of Eq. (1.23), calculated with respect to the distribution $P_{|\psi\rangle}(\phi)$ which describes the statistics of an ideal phase measurement, naturally applies to an ideal phase measurement; in contrast, the second moment of $\cos\phi$ on the right-hand side, calculated with respect to the quantum state $|\psi\rangle$, has no *a priori* connection with a phase measurement. What could be simpler? The second moment on the right-hand side must mean something, and indeed $\langle\psi|(\widehat{\cos\phi})^2|\psi\rangle$ is the second moment of $\cos\phi$ for an ideal measurement of the Hermitian operator $\widehat{\cos\phi}$—quite a different thing from a phase measurement.

The problems with the SG operator are a matter of perception; they come from expecting, with no justification, its Hermitian real and imaginary parts to have the statistics associated with an ideal phase measurement. The SG operator has a well defined role: it generates an (over)complete set of eigenstates, from which one derives the probability distribution for an ideal phase measurement; that task accomplished, its job is over. From the phase probability distribution, one can calculate the expectation value of any function of ϕ; there is no reason to expect functions of the Hermitian real and imaginary parts of the SG operator to generate the same expectation values.

Having agreed on the description of an ideal phase measurement, we specialize to such a measurement for the remainder of this paper. Following SSW and SS, we formulate the problem of estimating an unknown phase shift in the following way. Suppose that the field mode of interest is prepared in a fiducial state

$$|\psi\rangle = \sum_{n=0}^{\infty} a_n e^{i\delta_n}|n\rangle\,, \tag{1.24}$$

where $a_n e^{i\delta_n} = \langle n|\psi\rangle$ and $a_n = |\langle n|\psi\rangle|$ is real and nonnegative. The field then propagates through a "phase shifter" that changes the mode's phase by θ. The objective is to determine the phase shift θ as accurately as possible from the result of an ideal phase measurement. Formally the phase shift corresponds to applying the phase-translation operator $e^{i\theta\hat{N}}$ to the fiducial state, the result being a phase-shifted state

$$|\psi_\theta\rangle \equiv e^{i\theta\hat{N}}|\psi\rangle = \sum_{n=0}^{\infty} e^{in\theta} a_n e^{i\delta_n}|n\rangle\,. \tag{1.25}$$

The phase distribution for the phase-shifted state,

$$P_{|\psi_\theta\rangle}(\phi) = \frac{1}{2\pi}|\langle e^{i\phi}|\psi_\theta\rangle|^2 = \frac{1}{2\pi}\left|\sum_{n=0}^{\infty} e^{-in(\phi-\theta)} a_n e^{i\delta_n}\right|^2 = P_{|\psi\rangle}(\phi-\theta) = P(\phi|\theta)\,, \tag{1.26}$$

is just the fiducial phase distribution (1.11) shifted by θ. The shorthand notation $P(\phi|\theta)$, which suppresses reference to the fiducial state, reminds one that this is a conditional probability distribution for the result ϕ of an ideal phase measurement, given the phase shift θ. We have the freedom to shift the phase of the fiducial state, and we use this freedom to put the the maximum of its distribution $P_{|\psi\rangle}(\phi)$ at $\phi = 0$. With this choice it is convenient to choose the reference phase $\phi_0 = -\pi$ [Eq. (1.10)], so that we are dealing with phases on the interval $-\pi < \phi \le \pi$, and also to restrict the phase shift to the same interval $-\pi < \theta \le \pi$.

To proceed further, we need a quantitative measure of phase sensitivity. Now that the formal description of an ideal phase measurement has been established, it is this choice that is a sensible subject for discussion and that merits further work (for a review of the issues, see [24]). Indeed, a conclusion that can be drawn from this paper is that it is not known at present what criterion assures the best phase sensitivity for multiple measurements.

Summy and Pegg [33] choose phase standard deviation (square root of phase variance) of the fiducial state to characterize phase sensitivity. There are problems with the standard deviation of a quantity such as phase, which is defined on a finite interval with the end points identified, since the standard deviation depends on the choice of reference interval [19]. For a very narrow phase distribution, if the reference interval is chosen so that half the probability lies near each of the end points, then the standard deviation is of order π. For narrow distributions this problem, once recognized, is unimportant, since it can be avoided by judicious choice of the reference interval. Summy and Pegg seek the fiducial state with minimum phase standard deviation, both for an upper bound on photon number and for a constraint on mean photon number, and they find in both cases a phase standard deviation that goes as $1/N$ (see also Ref. [38]).

Hall and Fuss [39] consider a related problem in digital phase communication. They ask for the state that maximizes the amount of probability within a nonzero phase interval, they formulate the question both for an upper bound on photon number and for a constraint on mean photon number, and they present results for the case of an upper bound.

SSW adopt a different criterion, which comes from estimation theory. Given the result ϕ of an ideal phase measurement, the *maximum-likelihood* (ML) *estimate* $\theta_{\rm MLE}$ of the phase shift is the θ that maximizes the "likelihood" $P(\phi|\theta) = P_{|\psi\rangle}(\phi-\theta)$. Given our assumption that the maximum of the fiducial distribution lies at $\phi = \theta$, the ML estimate is just the result ϕ:

$$\theta_{\rm MLE} = \arg[\max_{-\pi<\theta\le\pi} P(\phi|\theta)]$$
$$= \arg[\max_{-\pi<\theta\le\pi} P_{|\psi\rangle}(\phi-\theta)] = \phi\,. \qquad (1.27)$$

SSW choose reciprocal peak likelihood, $\delta\phi = 1/P(\phi|\phi) = 1/P_{|\psi\rangle}(0)$, as their measure of phase sensitivity; thus they seek the fiducial state that maximizes the peak likelihood $P(\phi|\phi) = P(0|0)$.

The justification for choosing reciprocal peak likelihood as a measure of phase sensitivity is that if a phase distribution has most of its probability concentrated near its peak—like a Gaussian or a rectangular distribution—then the width of the distribution is approximately $\delta\phi$. What makes the SSW criterion questionable is that it only guarantees to make one point on the phase distribution very high; there is no assurance that any substantial fraction of the probability will be associated with this one point. It is quite possible for the optimal fiducial phase probability distribution to end up with a very high, very thin central peak, which sits on broad tails that encompass almost all the probability. Indeed, as noted by SSW and emphasized further by SS, this is precisely what happens.

To find the optimal state under the SSW maximum peak-likelihood criterion, one maximizes

$$P_{|\psi\rangle}(0) = \frac{1}{2\pi}\left|\sum_{n=0}^{\infty} a_n e^{i\delta_n}\right|^2 , \qquad (1.28)$$

subject to normalization,

$$1 = \sum_{n=0}^{\infty} a_n^2 , \qquad (1.29)$$

and to the constraint on mean photon number,

$$\bar{N} = \sum_{n=0}^{\infty} n a_n^2 \qquad (1.30)$$

(henceforth we let $\bar{N}$ denote the state's mean photon number). It is obvious that one should choose $\delta_n = 0$, which leaves the fiducial phase distribution symmetric about $\phi = 0$ with a maximum there. Thus the problem simplifies to maximizing

$$\sum_{n=0}^{\infty} a_n \qquad (1.31)$$

subject to the constraints. The trivial maximization [2] yields

$$a_n = \frac{A}{r+n} , \qquad (1.32)$$

where A and r are Lagrange multipliers.

The resulting optimal fiducial state does not have a finite mean photon number, so SSW and SS reformulate the optimization problem in a truncated Hilbert space where M is the highest allowed photon number ($a_n = 0$ for $n > M$). One *should* use the constraints (1.29) and (1.30) to solve for A and r in terms of M and $\bar{N}$ and then vary M, with $\bar{N}$ held fixed, to find the maximum peak likelihood. This procedure being quite difficult, however, SSW and SS adopt the simpler course of choosing $r = 1$ as a representative value; then A and M are determined by the normalization and mean-photon-number constraints.

Keeping in mind that we have not really found the state with maximum peak likelihood, we nonetheless follow SSW and SS and adopt the state

$$|\psi_{\rm SSW}\rangle = A\sum_{n=0}^{M}\frac{1}{1+n}|n\rangle \qquad (1.33)$$

as the fiducial state for the remainder of this paper. The SSW state (1.33) has a phase wave function,

$$\psi_{\rm SSW}(\phi) = \frac{A}{\sqrt{2\pi}}\sum_{n=0}^{M}\frac{e^{-in\phi}}{1+n}$$
$$= \frac{Ae^{i\phi}}{\sqrt{2\pi}}\sum_{l=1}^{M+1}\frac{e^{-il\phi}}{l} = \frac{Ae^{i\phi}}{\sqrt{2\pi}}\rho^*(\phi)\,, \qquad (1.34)$$

where for later use we introduce

$$\rho(\phi) \equiv \sum_{l=1}^{M+1}\frac{e^{il\phi}}{l}\,. \qquad (1.35)$$

The fiducial phase probability distribution is

$$P(\phi|0) = |\psi_{\rm SSW}(\phi)|^2 = \frac{A^2}{2\pi}|\rho(\phi)|^2\,, \qquad (1.36)$$

and the phase-shifted distribution is

$$P(\phi|\theta) = \frac{A^2}{2\pi}|\rho(\phi-\theta)|^2\,. \qquad (1.37)$$

The reason for investigating in detail the SSW proposal becomes apparent when one calculates the SSW reciprocal peak likelihood [2, 3]:

$$\delta\phi = \frac{1}{P(\phi|\phi)} \simeq \frac{12}{\pi(\bar{N}+1)^2}\,. \qquad (1.38)$$

If reciprocal peak likelihood were a good measure of phase sensitivity, one could claim, as SSW do, a phase sensitivity far better than the benchmark $1/\bar{N}$ sensitivity. As indicated above, however, even though the SSW distribution has a narrow central peak, it is so narrow that, for $\bar{N}$ even moderately bigger than 1, it encompasses a vanishingly small amount of probability. Essentially all the probability resides in tails, which cover the entire 2π of phase and which are essentially independent of $\bar{N}$. A single phase measurement drawn from the SSW distribution provides almost no information about the phase shift θ.

Nonetheless, the SSW distribution does have a very narrow central feature; it ought to be possible to locate that feature with extraordinary accuracy. SSW and SS suggest that by making a *sequence* of phase measurements, one can map out the phase distribution and thus locate the narrow central peak. The price one pays, of course, is that the photon-number constraint no longer applies to each measurement, but rather to the number

of photons used in the entire sequence. Taking this suggestion seriously, we have analyzed a sequence of SSW phase measurements.

Specifically, we investigate the following multiple-measurement version of SSW phase measurements. We assume that N_p field modes, which we call "pulses," are all prepared in the same SSW state, which has mean number of photons $\bar{N}$. The mean total number of photons is $N_{\rm tot} = N_p\bar{N}$. All the pulses are phase shifted by the same amount θ, and then an ideal phase measurement is performed on each. The measurements yield N_p pieces of phase data, $(\phi_1, \ldots, \phi_{N_p}) \equiv \boldsymbol{\phi}$, from which one must estimate the actual phase shift θ. Partly because SSW's work is based on ML estimation and partly because ML estimation is optimal asymptotically in N_p, we use the ML estimator for θ,

$$\theta_{\rm MLE} = \theta_{\rm MLE}(\boldsymbol{\phi}) = \arg[\max_{-\pi<\theta\le\pi} \mathcal{L}(\boldsymbol{\phi}|\theta)] , \tag{1.39}$$

where

$$\mathcal{L}(\boldsymbol{\phi}|\theta) \equiv \sum_{j=1}^{N_p} \ln P(\phi_j|\theta) \tag{1.40}$$

is the log-likelihood function. This entire process—preparing and phase shifting the pulses, collecting the phase data, and estimating θ—we call a single "experiment," because it is an idealization of a multiple-measurement laboratory experiment for determining θ. The ultimate result of an experiment is a single number, the ML estimate $\theta_{\rm MLE}$.

What the experimenter wants to know is how much confidence can be placed in the estimate $\theta_{\rm MLE}$. Thus we seek to determine the probability distribution $P(\theta_{\rm MLE}|\theta)$ of the ML estimator, given an actual phase shift θ. This probability distribution contains all the statistical information about the estimator. A convenient way to summarize important information in a single number is to give a symmetric, two-tailed $100\chi\%$ confidence interval $\Delta\theta_{\rm MLE}(\chi)$, defined by saying that the estimator has probability χ, called the confidence coefficient, to lie within an interval of width $\Delta\theta_{\rm MLE}(\chi)$ that brackets symmetrically the actual phase shift θ ($1-\chi$ is thus an error probability). Such a confidence interval is unquestionably a good measure of phase sensitivity.

We calculate estimator distributions and associated confidence intervals by a combination of numerical and analytic techniques. Our numerical work consists of simulating experiments on a computer using Monte Carlo simulation; by doing a sufficient number of experiments for particular values of $\bar{N}$ and N_p, we map out (statistically) the estimator distribution $P(\theta_{\rm MLE}|\theta)$ for those values of $\bar{N}$ and N_p. Since accurate simulations are demanding of computer time, we have been able to do a sufficient number of numerical experiments only for representative, moderate values $\bar{N} \le 5$, $N_p \le 240$, and $N_{\rm tot} \le 480$. For these representative values we find 68.26% and 95% confidence intervals. Then, for a particular value of $N_{\rm tot}$, we determine the optimal division of photons among the pulses by finding the minimum confidence interval as $\bar{N}$ and N_p are varied with $N_{\rm tot}$ held fixed. It turns out that the optimal division requires using many pulses with relatively few (mean) photons per pulse. Over the range $3 \le N_{\rm tot} \le 120$ accessible to our numerical simulations, we find the following optimized 68.26% and 95% confidence intervals:

$$\Delta\theta_{\rm MLE}(0.6826) \sim 1/N_{\rm tot}^{0.82\pm0.01} , \quad 3 \le N_{\rm tot} \le 120 , \tag{1.41a}$$

$$\Delta\theta_{\rm MLE}(0.95) \sim 1/N_{\rm tot}^{0.79\pm0.01} , \quad 3 \le N_{\rm tot} \le 120 . \tag{1.41b}$$

We supplement our numerical results by appealing to two powerful theorems from standard estimation theory and to a set of new results due to Braunstein. The two theorems from estimation theory involve a quantity called the *Fisher information* [40], which is determined by the fiducial phase distribution. The first theorem gives a lower bound, called the *Cramér-Rao lower bound*, on the variance of *any* unbiased estimator. The second theorem, called *Fisher's theorem*, shows that the ML estimator is asymptotically (in N_p) Gaussian, with variance given by the Cramér-Rao lower bound; it thus establishes that ML estimation is asymptotically efficient. Braunstein's new results [5] give the largest corrections to the asymptotic Gaussian behavior of the ML estimator and thus provide an estimate of the value of N_p at which the asymptotic behavior sets in.

Armed with these analytic tools, we are able to estimate, first, the optimal division of photons among the pulses and, second, the standard deviation of the ML estimator at this optimal operating point. The standard deviation (the square root of the variance) is nearly the same as the 68.26% confidence interval at the optimal operating point. Given analytic forms for the Fisher information and Braunstein's correction terms, we could extend these estimates to arbitrarily large numbers of photons and thus settle the question of the sensitivity of SSW measurements. Unfortunately, for the SSW phase distribution, the Fisher information and Braunstein's correction terms involve integrals that we are unable to evaluate analytically. We are forced to settle for numerical integrations that go up to $N_{\rm tot} \simeq 30\,000$. Using this approach, we find that the standard deviation of the optimized ML estimator obeys an approximate power law. For $N_{\rm tot} \simeq 100$, the slope of the power law is $\simeq -0.84$, in agreement with our simulations. For higher $N_{\rm tot}$ the slope steepens, achieving a value $\simeq -0.92$ for $N_{\rm tot} \simeq 30\,000$. Thus this approach confirms and extends the results of our numerical work. In particular, it confirms that the optimal operating point requires many pulses with relatively few (mean) photons per pulse. By combining these results with yet another semianalytic technique that gives lower bounds on the optimal phase sensitivity, we can say that multiple SSW phase measurements definitely do *not* beat the benchmark $1/N_{\rm tot}$ sensitivity for $N_{\rm tot} \lesssim 200\,000$.

What can we conclude from our investigation of SSW measurements? First, *we find no evidence for a* $1/N_{tot}^2$ *sensitivity or even for a sensitivity as good as the benchmark* $1/N_{\rm tot}$ *sensitivity of squeezed-state interferometry.*

Definitive conclusions are limited to the domain of total photon number that we are able to investigate. *Within that domain, $N_{\rm tot} \lesssim 200\,000$, we conclude that multiple SSW measurements do not achieve the $1/N_{\rm tot}^2$ phase sensitivity suggested by SSW, nor do they achieve the benchmark $1/N_{\rm tot}$ sensitivity*—although they come much closer to the benchmark sensitivity than we would have guessed at the beginning of our work. Finally, *we conclude that reciprocal peak likelihood is not a good measure of sensitivity.*

Section II focuses on the fiducial SSW phase probability distribution and presents both analytical and numerical evidence of its behavior. Sections II A–II C outline the most important properties of the SSW distribution and should be consulted by any reader of this paper. Sections II D–II F contain more detailed analytical evidence of the behavior of the SSW distribution. Section II G spells out the case for multiple SSW measurements.

Section III presents our analysis of the multiple-measurement version of SSW measurements. Section III A introduces estimation theory for multiple measurements and emphasizes the importance of the Cramér-Rao lower bound and the Fisher information. Section III B describes maximum-likelihood estimation and Fisher's efficiency theorem. Section III C presents analytic bounds and estimates for the Fisher information of the SSW distribution. Section III D details the procedures used for the numerical simulations and describes the results of the simulations. The key results of the paper are contained in Sec. III E, which presents optimized confidence intervals obtained from the simulations and from combined numerical-analytical approaches. These optimized confidence intervals characterize the sensitivity of multiple SSW phase measurements.

Section IV discusses issues raised by our work and questions for further research. In particular, it outlines difficulties to be overcome in finding the ultimate quantum limitations on phase measurements.

II. SSW PHASE PROBABILITY DISTRIBUTION

A. Definition of SSW distribution

The fiducial SSW phase probability distribution is given by Eq. (1.36). Properties of this distribution have been investigated by SSW and SS and also by Schleich, Dowling, and Horowicz [41].

The SSW distribution contains two constants: the normalization constant A and the upper bound on photon number M. These two constants are determined by the normalization condition (1.29),

$$1 = A^2 \sum_{n=0}^{M} \frac{1}{(1+n)^2} = A^2 \sum_{l=1}^{M+1} \frac{1}{l^2} , \tag{2.1}$$

and by the mean-photon-number constraint (1.30),

$$\bar{N} = A^2 \sum_{n=0}^{M} \frac{n}{(1+n)^2} = -1 + A^2 \sum_{l=1}^{M+1} \frac{1}{l} . \tag{2.2}$$

We find it convenient to work with the complex sum defined in Eq. (1.35),

$$\rho(\phi) = \sum_{l=1}^{M+1} \frac{e^{il\phi}}{l} = \sum_{n=0}^{M} \frac{e^{i(n+1)\phi}}{n+1} , \tag{2.3}$$

in terms of which the fiducial phase distribution is

$$P(\phi|0) = \frac{A^2}{2\pi} |\rho(\phi)|^2 . \tag{2.4}$$

Throughout this section we choose the reference phase so that ϕ lies in the interval $(-\pi, \pi]$.

Notice that $\rho(\phi)$ is a generating function for moments of $\hat{N}+1$:

$$\rho(\phi) = \frac{1}{A^2} \sum_{k=0}^{\infty} \frac{(i\phi)^k}{k!} \langle (\hat{N}+1)^{k+1} \rangle . \tag{2.5}$$

This expansion allows us to write

$$\langle (\hat{N}+1)^{k+1} \rangle = A^2 (-i)^k \left. \frac{d^k \rho}{d\phi^k} \right|_{\phi=0} = A^2 \sum_{l=1}^{M+1} l^{k-1} . \tag{2.6}$$

B. Constants A and M

The constant A is given directly in terms of M by the normalization condition (2.1),

$$A^2 = \left(\sum_{l=1}^{M+1} \frac{1}{l^2} \right)^{-1} . \tag{2.7}$$

For large M the sum converges rapidly to its $M \to \infty$ limit

$$\sum_{l=1}^{\infty} \frac{1}{l^2} = \zeta(2) = \frac{\pi^2}{6} . \tag{2.8}$$

Indeed, we can obtain the first correction to this limit by using

$$\begin{aligned} \frac{1}{M+2} = \int_{M+2}^{\infty} \frac{dx}{x^2} &= \sum_{l=M+2}^{\infty} \frac{1}{l(l+1)} \\ &< \sum_{l=M+2}^{\infty} \frac{1}{l^2} \\ &< \sum_{n=M+1}^{\infty} \frac{1}{n(n+1)} = \frac{1}{M+1} \end{aligned} \tag{2.9}$$

to obtain [41]

$$A^2 = \frac{6}{\pi^2} + \frac{36}{\pi^4 (M+1)} + O\left(\frac{1}{M^2} \right) . \tag{2.10}$$

The mean-photon-number constraint (2.2) can be written as [Eq. (2.6)]

$$\bar{N} + 1 = A^2 \sum_{l=1}^{M+1} \frac{1}{l} = A^2 \rho(0) , \tag{2.11}$$

and the series for $\rho(0)$ can be approximated by an asymptotic expansion [Eq. (0.131) of Ref. [42]],

$$\rho(0)=\sum_{l=1}^{M+1}\frac{1}{l}=\gamma+\ln(M+1)+\frac{1}{2(M+1)}+O\left(\frac{1}{M^2}\right), \tag{2.12}$$

where $\gamma=0.577\,215\,664\,9\ldots$ is Euler's constant. Putting these two equations together with Eq. (2.10), we find

$$\bar{N}+1=\frac{6}{\pi^2}[\gamma+\ln(M+1)]+\frac{6}{\pi^2(M+1)}\left(\frac{6}{\pi^2}[\gamma+\ln(M+1)]+\frac{1}{2}\right)+O\left(\frac{\ln M}{M^2}\right). \tag{2.13}$$

Inverting this expansion gives M in terms of $\bar{N}$:

$$M+1=e^{-\gamma}\exp\left(\frac{\pi^2}{6}(\bar{N}+1)\right)-\bar{N}-\frac{3}{2}+O\left(\frac{(\ln M)^2}{M}\right). \tag{2.14}$$

Schleich, Dowling, and Horowicz [41] have given the leading terms in Eqs. (2.13) and (2.14). Because M must be an integer, the expansion (2.14) actually yields an exact expression for M,

$$M=I_{\text{near}}\left(e^{-\gamma}\exp\left(\frac{\pi^2}{6}(\bar{N}+1)\right)-\bar{N}-\frac{5}{2}\right), \tag{2.15}$$

where $I_{\text{near}}(x)$ is the nearest integer to the real number x.

The SSW phase distribution can be characterized by a single constant, which can be taken to be either the mean number of photons $\bar{N}$ or the upper bound on photon number M, the two being related by Eq. (2.13) or by Eq. (2.15). The key thing to notice is the exponential increase of M with $\bar{N}$ [2, 41]. Even for very modest mean photon numbers, the upper bound becomes enormous ($\bar{N}=100$ gives $M\sim 10^{72}$), so the asymptotic formulas (2.10) and (2.13) can be used with confidence. In particular, even for very modest values of $\bar{N}$, A^2 takes on its limiting value $6/\pi^2$. Table I illustrates the behavior of M, $\bar{N}$, and A.

The exponential increase of the upper bound for this state might seem unimportant (a coherent state has no upper bound at all), until one calculates the photon-number variance [2]

$$\operatorname{var}(\hat{N})=\langle(\hat{N}+1)^2\rangle-(\bar{N}+1)^2=A^2(M+1)-(\bar{N}+1)^2 \tag{2.16}$$

[cf. Eq. (2.6)]. The photon-number uncertainty goes as $\sqrt{M}$ and thus also blows up exponentially as $\bar{N}$ increases. We revisit the variance (2.16) in Sec. III C, where it provides a bound on the Fisher information.

The peak likelihood can also be extracted from Eqs. (2.4) and (2.11) [2, 3]:

$$P(\phi|\phi)=P(0|0)=\frac{(\bar{N}+1)^2}{2\pi A^2}=\frac{\pi(\bar{N}+1)^2}{12}+O\left(\frac{(\ln M)^2}{M}\right)=\frac{A^2}{2\pi}[\gamma+\ln(M+1)]^2+O\left(\frac{\ln M}{M}\right). \tag{2.17}$$

The reciprocal peak likelihood thus goes as $1/\bar{N}^2$; this behavior is the source of the sensitivity claims of SSW and SS.

C. Plots of SSW distribution

Considerable insight into the behavior of the SSW distribution can be gained by plotting the distribution and related functions for representative values of $\bar{N}$ (or M). Figure 1 plots the fiducial SSW phase distribution for $\bar{N}=0.56$, 1.0, and 3.0. At the resolution of the figure the distribution takes on an $\bar{N}$-independent form for $\bar{N}\gtrsim 3$, except for the narrow central peak, whose height—the peak likelihood—grows as given by Eq. (2.17). The $\bar{N}$-independent form we call the universal behavior of the SSW distribution. The universal behavior is a key ingredient in our later analysis, because it allows us to argue

TABLE I. Numerical values of the upper bound on photon number M, mean photon number $\bar{N}$, and normalization constant A. The first column gives M; the second column gives $\bar{N}$ calculated from Eqs. (2.7) and (2.11); the third column returns to M using Eq. (2.15), thus illustrating the exactness of that formula; the fourth column gives A calculated from Eq. (2.7); and the fifth column gives an approximate value of A calculated from Eq. (2.10).

M	$\bar{N}$	$M(\bar{N})$	A	A (approx.)
1	0.200 0	1	0.8	0.793
2	0.346 9	2	0.735	0.731
3	0.463 4	3	0.702	0.700
4	0.560 07	4	0.683 2	0.681 8
12	1.024 41	12	0.636 6	0.636 4
30	1.496 45	30	0.619 89	0.619 85
74	2.003 854	74	0.612 862	0.612 855
399	3.000 110 2	399	0.608 851 3	0.608 851 0
2 089	4.000 058 17	2 089	0.608 103 941	0.608 103 932
10 848	5.000 023 44	10 848	0.607 961 167 6	0.607 961 167 2

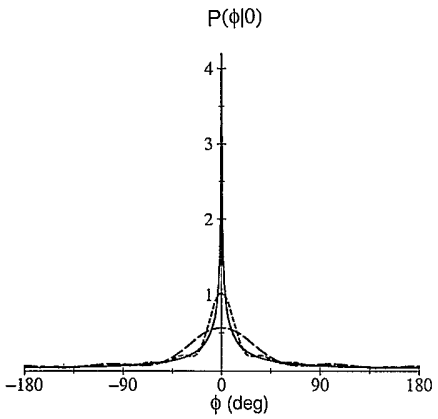

FIG. 1. Fiducial SSW phase probability distribution $P(\phi|0)$ for three values of M: $M = 4$ ($\bar{N} = 0.560$, long-dashed line), $M = 12$ ($\bar{N} = 1.024$, short-dashed line), $M = 399$ ($\bar{N} = 3.000$, solid line). At the resolution of this figure, a value of $\bar{N} \gtrsim 3$ gives a distribution that is indistinguishable from the $\bar{N} = 3$ graph, except for a higher peak at $\phi = 0$. The insensitivity to the value of $\bar{N} \gtrsim 3$, except for the height of the central peak, represents the universal behavior of the distribution.

that certain aspects of the multiple-measurement statistics are independent of $\bar{N}$. One can plot a greater range of values for $\bar{N}$ on the same graph by plotting $\ln P(\phi|0)$, as we do in Fig. 2, where the highest mean photon number is $\bar{N} = 6.0$. Figure 2 shows clearly the $\bar{N}$-independent, universal behavior of the SSW distribution for $\bar{N} \gtrsim 3$.

The reader should also note the oscillations in the distribution for lower values of $\bar{N}$. There are exactly M periods of oscillation, with an oscillation amplitude $\sim 1/M$, except near $\phi = 0$. These oscillations are a consequence of the sharp cutoff in photon number at $n = M$. Although they are present for higher values of $\bar{N}$, they cannot be seen at the resolution of Figs. 1 and 2 because of the decreasing amplitude. The central peak at $\phi = 0$ can be thought of as one of these oscillations, much larger than the rest because all the terms in the sum for $\rho(\phi)$ add constructively at $\phi = 0$. Thus the central peak has a width $\sim 1/M$ that decreases exponentially in the mean photon number. Dowling [43] has suggested replacing the sharp photon-number cutoff of the SSW state with an exponential cutoff. The resulting phase probability distribution is quite similar to the SSW distribution, but has no oscillations outside the central peak.

The cumulative probability,

$$C(\phi) \equiv \int_{-\pi}^{\phi} d\phi' \, P(\phi'|0) \, , \tag{2.18}$$

is plotted for four values of $\bar{N}$ in Fig. 3. The even symmetry of the fiducial SSW distribution $P(\phi|0)$ about $\phi = 0$ implies that $C(\phi) = 1 - C(-\phi)$. The cumulative probability tends to smooth out the oscillations in $P(\phi|0)$. It also displays very clearly the universal behavior: for $\bar{N} \gtrsim 2$ the cumulative probability assumes a universal shape. The growth in the peak likelihood does not affect $C(\phi)$, because the narrow central peak in $P(\phi|0)$ encloses a vanishingly small amount of probability $\sim \bar{N}^2/M \sim (\ln M)^2/M$.

The oscillations in the SSW distribution are accentuated by taking derivatives. Figure 4 plots $d \ln P(\phi|0)/d\phi$ for $\bar{N} = 1.5$ ($M = 30$). The M oscillations are a consequence of the highest Fourier components in the sum for $\rho(\phi)$. Figure 5 plots $d^2 \ln P(\phi|0)/d\phi^2$ for $\bar{N} = 1.5$

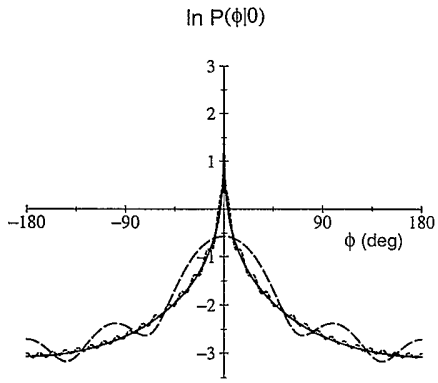

FIG. 2. Natural logarithm of the fiducial SSW phase probability distribution $\ln P(\phi|0)$ for four values of M: $M = 4$ ($\bar{N} = 0.560$, long-dashed line), $M = 30$ ($\bar{N} = 1.496$, short-dashed line), $M = 399$ ($\bar{N} = 3.000$, solid line), $M = 56\,299$ ($\bar{N} = 6.000$, solid line). The $\bar{N} = 3$ and 6 plots are indistinguishable at the resolution of this figure (that $\bar{N} = 3$ does not have as high a peak at $\phi = 0$ is obscured), thus displaying the universal behavior of the SSW distribution outside the central peak. Even the $\bar{N} = 1.5$ plot is the same as the $\bar{N} = 3$ and 6 plots outside the central peak, except for oscillations that make it appear to be braided onto the universal behavior.

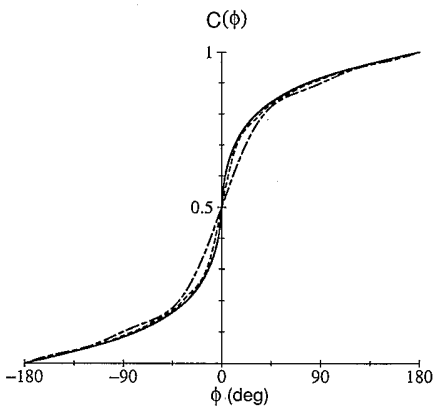

FIG. 3. Cumulative probability $C(\phi)$ [Eq. (2.18)] for four values of M: $M = 4$ ($\bar{N} = 0.560$, long-short-dashed line), $M = 12$ ($\bar{N} = 1.024$, short-dashed line), $M = 74$ ($\bar{N} = 2.004$, solid line), $M = 10\,848$ ($\bar{N} = 5.000$, solid line). The $\bar{N} = 2$ and 5 plots are indistinguishable at the resolution of this figure; at this resolution, the cumulative probability has a universal shape for $\bar{N} \gtrsim 2$, reflecting the universal behavior of the underlying SSW distribution.

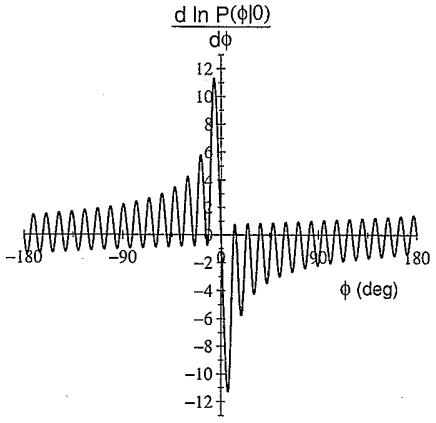

FIG. 4. First derivative of $\ln P(\phi|0)$ for $M = 30$ ($\bar{N} = 1.496$). Notice that there are M oscillations in the derivative, all of roughly unit amplitude except for the central oscillation.

($M = 30$). The two large positive peaks surrounding the large negative valley at $\phi = 0$ are what give rise to the sharp central peak in the fiducial SSW phase probability distribution $P(\phi|0)$. Although the amplitude of these central features grow rapidly, as M^2, they are, like all the other oscillations, only $\sim 1/M$ wide. Consequently, the central peak of the probability distribution is also $\sim 1/M$ wide, its height increasing only logarithmically in M.

D. Large-M analytic approximations

To clarify further the generic behavior of the SSW phase distribution, we derive in this subsection large-M analytic approximations to the distribution. As we learn in Sec. II B, the exponential increase of M with $\bar{N}$ makes these approximations quite good even for very modest mean photon numbers. The main task is to find analytic approximations for the complex function $\rho(\phi)$ [Eq. (2.3)],

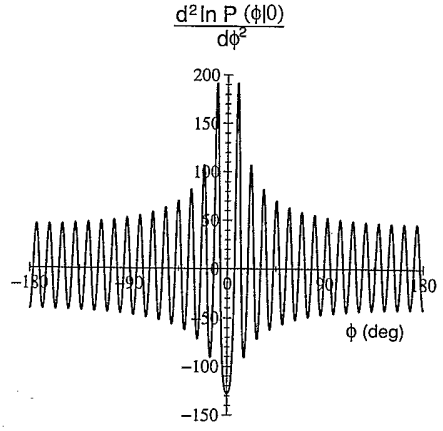

FIG. 5. Second derivative of $\ln P(\phi|0)$ for $M = 30$ ($\bar{N} = 1.496$). The outlying oscillations have amplitude $\sim M$, while the sharp central M-shaped feature has amplitude $\sim M^2$, giving the narrow central peak in the fiducial SSW phase probability distribution.

from which $P(\phi|0)$ follows.

To get a first approximation, which describes the universal behavior outside the central peak, we write $\rho(\phi)$ as

$$\rho(\phi) = \sum_{l=1}^{\infty} \frac{e^{il\phi}}{l} - \sum_{l=M+2}^{\infty} \frac{e^{il\phi}}{l} = -\ln(1-e^{i\phi}) + R(M) , \tag{2.19}$$

where

$$\ln(1 - e^{i\phi}) = \ln[2\sin(|\phi|/2)] \mp i\frac{\pi - |\phi|}{2} \tag{2.20}$$

(the upper sign applies for $\phi > 0$, and the lower sign for $\phi < 0$), and where the remainder term is defined by

$$R(M) \equiv -\sum_{l=M+2}^{\infty} \frac{e^{il\phi}}{l} = -e^{i\phi} \sum_{n=M+1}^{\infty} \frac{e^{in\phi}}{n+1} . \tag{2.21}$$

Combining the two forms of the remainder term allows us to write

$$\begin{aligned}(1 - e^{-i\phi})R(M) &= \frac{e^{i(M+1)\phi}}{M+1} - \sum_{n=M+1}^{\infty} \frac{e^{in\phi}}{n(n+1)} \\ &= \frac{e^{i(M+1)\phi}}{M+2} - \sum_{l=M+2}^{\infty} \frac{e^{il\phi}}{l(l+1)} .\end{aligned} \tag{2.22}$$

The right-hand side of Eq. (2.22) is absolutely convergent, its absolute value bounded above by $2/(M+2)$. This allows us to bound $R(M)$ by

$$|R(M)| \le \frac{1}{(M+2)\sin(|\phi|/2)} \le \frac{\pi}{|\phi|(M+2)} , \qquad -\pi < \phi \le \pi . \tag{2.23}$$

The result is an approximation,

$$\rho(\phi) = -\ln(1 - e^{i\phi}) + O\left(\frac{1}{|\phi|M}\right) , \tag{2.24}$$

which, when used to find the phase distribution (2.4), gives [41]

$$P(\phi|0) = P_U(\phi) + O\left(\frac{1 + |\ln(|\phi|/\pi)|}{|\phi|M}\right) . \tag{2.25}$$

Here we define a new distribution

$$\begin{aligned}P_U(\phi) &= \frac{A^2}{2\pi} \left|\ln(1 - e^{i\phi})\right|^2 \\ &= \frac{A^2}{2\pi} \left[\{\ln[2\sin(|\phi|/2)]\}^2 + \left(\frac{\pi - |\phi|}{2}\right)^2\right] ,\end{aligned} \tag{2.26}$$

where A^2 can be replaced by $6/\pi^2$ within the accuracy of the approximation. The approximations (2.24) and (2.25), which are independent of $\bar{N}(M)$, are good outside the central peak; indeed, $P_U(\phi)$ gives the universal, $\bar{N}$-independent, $M \to \infty$ behavior of the SSW distribution. These approximations miss entirely the oscilla-

tions outside the central peak—i.e., they have smoothed derivatives, except for the singularity at $\phi = 0$—because the amplitude of the oscillations goes to zero as $M \to \infty$.

By writing

$$\ln(1 - e^{i\phi}) = \ln|\phi| + O(|\phi|^2) \mp i\frac{\pi - |\phi|}{2} \tag{2.27}$$

[cf. Eq. (2.20)], we can derive a small-angle approximation to the universal distribution:

$$P_U(\phi) = \frac{A^2}{2\pi}(\ln|\phi|)^2 + \frac{\pi A^2}{8} - \frac{1}{4}A^2|\phi| + O(|\phi|^2[1 + |\ln(|\phi|/\pi)|]) , \tag{2.28}$$

where again we can replace A^2 by $6/\pi^2$ within the overall approximation. Equation (2.28) describes the small-angle runup to the central peak.

A crude but informative approximation to the entire SSW distribution arises from putting a flat cap on the universal distribution (2.26)—i.e., from using $P_U(\phi)$ for $|\phi| \geq 1/M$ and using $P_U(1/M)$ for $|\phi| \leq 1/M$. This crude approximation gives the right peak likelihood $P_U(1/M) \simeq (A^2/2\pi)(\ln M)^2 \sim \bar{N}^2/2\pi A^2$. The probability under the flat cap is $\simeq (A^2/\pi)[(\ln M)^2/M] \sim (1/\pi A^2)(\bar{N}^2/M)$, which gives an estimate of the vanishingly small probability under the central peak.

We can improve our analytic approximation to $\rho(\phi)$ by applying the Euler-MacLaurin summation formula to the series for $R(M)$. For a function $F(x)$ the appropriate form of the Euler-MacLaurin formula reads

$$\sum_{l=M+2}^{\infty} F(l) = \int_{M+1}^{\infty} dx\, F(x) - \frac{F(M+1) + F(\infty)}{2} + \sum_{k=1}^{\infty} \frac{B_{2k}}{(2k)!} \frac{d^{2k-1}F(x)}{dx^{2k-1}}\bigg|_{x=M+1}^{\infty} \tag{2.29}$$

[see Eq. (3.6.28) of Ref. [44]], where the coefficients B_{2k} are Bernoulli numbers. Letting $F(x) = e^{ix\phi}/x$ allows us to write $R(M)$ as an integral plus corrections that can be converted to an asymptotic expansion in $1/M$. We can evaluate the first term in the asymptotic expansion by applying the Euler-MacLaurin formula to the function $G(x) = e^{ix\phi}e^{-\epsilon x}/(M+1)$, which has the same correction terms to first order in $1/M$ ($\epsilon > 0$ is a convergence parameter that is set to zero at the end of the calculation). Putting all this together, we find

$$\begin{aligned} R(M) &= -\sum_{l=M+2}^{\infty} \frac{e^{il\phi}}{l} = -\int_{M+1}^{\infty} dx\, \frac{e^{ix\phi}}{x} + \int_{M+1}^{\infty} dx\, \frac{e^{ix\phi}e^{-\epsilon x}}{M+1} - \sum_{l=M+2}^{\infty} \frac{e^{il\phi}e^{-\epsilon l}}{M+1} + O\left(\frac{1}{M^2}\right) \\ &= -E_1[-i(M+1)\phi] - \frac{e^{i(M+1)\phi}}{M+1}\left(\frac{1}{i\phi} - \frac{1}{1-e^{-i\phi}}\right) + O\left(\frac{1}{M^2}\right) , \end{aligned} \tag{2.30}$$

where

$$E_1(z) \equiv \int_z^{\infty} dt\, \frac{e^{-t}}{t} \tag{2.31}$$

is the exponential integral (see Chap. 5 of Ref. [44]). Putting Eq. (2.30) into Eq. (2.19) yields an approximate form for $\rho(\phi)$, valid to order $1/M$ over the entire range of ϕ. The $O(1/M)$ correction in Eq. (2.30) describes the oscillations in the SSW distribution.

We can use our new approximation for $\rho(\phi)$ in two ways. For small angles the series expansion of the exponential integral,

$$E_1(z) = -\gamma - \ln z - \sum_{k=1}^{\infty} \frac{(-z)^k}{kk!} , \tag{2.32}$$

gives

$$\rho(\phi) = \gamma + \ln(M+1) + \frac{e^{i(M+1)\phi}}{2(M+1)} + O(M|\phi|) + O\left(\frac{1}{M^2}\right) , \tag{2.33}$$

in agreement with Eq. (2.12). The asymptotic expansion of the exponential integral,

$$E_1(z) \sim \frac{e^{-z}}{z}\left(1 - \frac{1}{z} + \frac{2}{z^2} - \frac{6}{z^2} + \cdots\right) , \tag{2.34}$$

gives an improved approximation to $\rho(\phi)$ outside the central peak,

$$\rho(\phi) = -\ln(1 - e^{i\phi}) + \frac{e^{i(M+1)\phi}}{M+1}\frac{1}{1-e^{-i\phi}} + O\left(\frac{1}{|\phi|^2M^2}\right) \tag{2.35}$$

[cf. Eq. (2.24)], with the new $1/M$ term describing oscillations on top of the universal distribution (2.26).

E. Gaussian fit to SSW distribution

Another useful analytic result is a straightforward Gaussian fit to the central peak of the SSW distribution. This fit comes from matching to second order the Taylor expansion of the logarithm of the SSW distribution to the corresponding expansion for a Gaussian function. The resulting (unnormalized) fit is [41]

$$P_G(\phi) = P(0|0)e^{-\phi^2/2\Sigma^2} , \tag{2.36}$$

where the half-width of the Gaussian function is given by

$$\begin{aligned} \Sigma &= \left[-\left(\frac{d^2 \ln P(\phi|0)}{d\phi^2}\right)\bigg|_{\phi=0}\right]^{-1/2} \\ &= \frac{\sqrt{(\bar{N}+1)/A^2}}{M+1}\left(1 - \frac{2A^2}{\bar{N}+1} + \frac{1}{M+1}\right)^{-1/2} \\ &= \frac{1}{M+1}\frac{\gamma + \ln(M+1)}{[\gamma - 2 + \ln(M+1)]^{1/2}}\left[1 + O\left(\frac{1}{M}\right)\right] \\ &\sim \frac{(\ln M)^{1/2}}{M} . \end{aligned} \tag{2.37}$$

The extremely narrow width of the Gaussian fit is consistent with the foregoing discussion. Schleich, Dowling, and Horowicz [41] have drawn attention to the very narrow width of the Gaussian fit.

A reasonable way to estimate the area under the central peak is to find the area under the Gaussian fit:

$$\int_{-\pi}^{\pi} d\phi\, P_G(\phi) = P(0|0)\Sigma\sqrt{2\pi} = \frac{A^2}{\sqrt{2\pi}} \frac{[(\bar{N}+1)/A^2]^{5/2}}{M+1} \left(1 - \frac{2A^2}{\bar{N}+1} + \frac{1}{M+1}\right)^{-1/2}$$
$$= \frac{A^2}{\sqrt{2\pi}} \frac{1}{M+1} \frac{[\gamma + \ln(M+1)]^3}{[\gamma - 2 + \ln(M+1)]^{1/2}} \left[1 + O\left(\frac{1}{M}\right)\right] \sim \frac{A^2}{\sqrt{2\pi}} \frac{(\ln M)^{5/2}}{M} . \tag{2.38}$$

Again this estimate is consistent with the foregoing discussion and shows that *the area under the central peak of the SSW distribution decreases nearly exponentially with mean photon number.* Nearly all the probability is found under the universal part of the SSW distribution, even for very modest values of $\bar{N}$.

F. Fourier series for SSW distribution

The Fourier series for the SSW distribution can be found by applying the change-of-variables formula

$$\sum_{l,m=1}^{L} f_{lm} = \sum_{j=1}^{L} f_{jj} + \sum_{k=1}^{L-1} \sum_{j=k/2+1}^{L-k/2} \left(f_{j+k/2,j-k/2} + f_{j-k/2,j+k/2} \right) \tag{2.39}$$

to the double series for the SSW distribution,

$$P(\phi|0) = \frac{A^2}{2\pi} \sum_{l,m=1}^{M+1} \frac{e^{i(l-m)\phi}}{lm} . \tag{2.40}$$

The result is a Fourier cosine series

$$P(\phi|0) = \frac{1}{2\pi} + \sum_{k=1}^{M} F_k(M) \cos k\phi , \tag{2.41}$$

where the Fourier coefficients $F_k(M)$ are given by

$$F_k(M) = \frac{A^2}{\pi} \sum_{j=k/2+1}^{M+1-k/2} \frac{1}{j^2 - (k/2)^2} . \tag{2.42}$$

Notice that the highest Fourier coefficient, $F_{k=M}(M) = A^2/\pi(M+1)$, decreases as $1/M$.

The Fourier series can be integrated immediately to give the cumulative probability (2.18),

$$C(\phi) = \int_{-\pi}^{\phi} d\phi'\, P(\phi'|0) = \frac{\phi+\pi}{2\pi} + \sum_{k=1}^{M} \frac{F_k(M)}{k} \sin k\phi . \tag{2.43}$$

Differentiation of the Fourier series yields

$$\frac{dP(\phi|0)}{d\phi} = -\sum_{k=1}^{M} k F_k(M) \sin k\phi \tag{2.44}$$

and

$$\frac{d^2P(\phi|0)}{d\phi^2} = -\sum_{k=1}^{M} k^2 F_k(M) \cos k\phi . \tag{2.45}$$

Since the terms in the series for $F_k(M)$ decrease monotonically with j, we can estimate the Fourier coefficients from the following integral bounds for any monotonically decreasing function $f(x)$:

$$\int_r^{s+1} dx\, f(x) \le \sum_{j=r}^{s} f(j) \le f(r) + \int_r^s dx\, f(x) . \tag{2.46}$$

Our case corresponds to $r = k/2+1$, $s = M+1-k/2$, and

$$f(x) = \frac{A^2}{\pi} \frac{1}{x^2 - (k/2)^2} , \tag{2.47}$$

so the bounds take the explicit form

$$\frac{A^2}{k\pi} \ln\left(\frac{(M-k+2)(k+1)}{M+2}\right) \le F_k(M) \le \frac{A^2}{k\pi} \left[\frac{k}{k+1} + \ln\left(\frac{(M-k+1)(k+1)}{M+1}\right)\right] . \tag{2.48}$$

Since the logarithms vary relatively slowly, one expects that $F_k(M)$ should be roughly proportional to $1/k$. This expectation is borne out by the plots of Fourier coefficients for $\bar{N} = 7.0$ in Fig. 6. The highest Fourier coeffi-

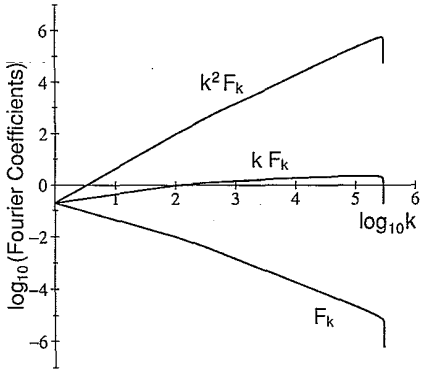

FIG. 6. Log-log plot of Fourier coefficients for the fiducial SSW phase probability distribution and its first two derivatives, F_k, kF_k, and k^2F_k, vs Fourier order k for $M = 291\,334$ ($\bar{N} = 7.000$).

cients $F_{k\simeq M}(M)$ have an amplitude that goes roughly as $1/M$, in agreement with our previous comments. Thus, as M increases, the oscillations in the SSW distribution, while becoming progressively more rapid, also become progressively smaller (see Figs. 1 and 2).

The Fourier coefficients $kF_k(M)$ for the first derivative of $P(\phi|0)$ are roughly constant with k, and the Fourier coefficients $k^2F_k(M)$ for the second derivative of $P(\phi|0)$ are roughly linear in k (see Fig. 6). The result is that the oscillations do not die away in the first derivative of the SSW distribution, regardless of the value of M, and they grow in the second derivative, with amplitude that increases roughly linearly with M. Around $\phi = 0$ the Fourier components interfere constructively to create the central features in the SSW probability distribution and its derivatives.

G. Need for multiple measurements

It should be evident at this point that most of the probability in the SSW distribution lies in the broad tails and that these tails are essentially independent of the mean number of photons. One way to summarize these conclusions is to give confidence intervals for the SSW distribution for various values of M. In Table II we tabulate symmetric, two-tailed confidence intervals for the SSW distribution. The symmetric, two-tailed $100\chi\%$ confidence interval $\Delta\phi(\chi)$ (confidence coefficient $\chi = 1 - \alpha$) is defined to be the width of the interval, symmetrically bracketing the central peak, that contains probability χ, corresponding to error probability α. Formally the definition reads

$$\begin{aligned}\Delta\phi(\chi) = \Delta\phi(1-\alpha) &\equiv C^{-1}(1-\alpha/2) - C^{-1}(\alpha/2) \\ &= C^{-1}\left(\frac{1+\chi}{2}\right) - C^{-1}\left(\frac{1-\chi}{2}\right) \\ &= 2C^{-1}\left(\frac{1+\chi}{2}\right), \end{aligned} \tag{2.49}$$

where $C^{-1}(q)$ is the angle ϕ such that $C(\phi) = q$ [Eq. (2.18)]. The last form of the confidence interval follows from the even symmetry of the SSW cumulative probability.

We use the 68.26% confidence interval ($\alpha = 0.3174$) in Table II and throughout this paper because of its familiarity from Gaussian statistics: the symmetric, two-tailed 68.26% confidence interval for a Gaussian distribution is the full width of the Gaussian function—i.e., two standard deviations. We also report the 95% confidence interval ($\alpha = 0.05$), which for a Gaussian function is nearly twice the width of the 68.26% confidence interval—i.e., nearly four standard deviations. Table II also gives the commonly used 99% confidence interval ($\alpha = 0.01$). Note that the 99% confidence interval encompasses almost all of the 360° of available phase no matter how many photons are put into the SSW state. At $M = 0$ the SSW state is the vacuum state $|0\rangle$, with random phase described by $P(\phi|0) = 1/2\pi$. Setting an upper bound of only one photon ($M = 1$, $\bar{N} = 0.2$) actually gives the *smallest* confidence intervals for low error probabilities. This perhaps surprising result is a consequence of the oscillations in the SSW distribution, which conspire to make the probability distribution for $M = 1$ dip below all the others, including the universal distribution, out at the edges. As M increases, the confidence intervals converge to $\bar{N}$-independent universal values characteristic of the universal behavior of the SSW distribution.

The sizes of the universal confidence intervals confirm that most of the probability lies in the $\bar{N}$-independent tails. The sizes also indicate that a single phase measurement using the SSW state provides almost no information and certainly is incapable of locating the very narrow central peak. Since it does not help to make $\bar{N}$ larger, the only hope for the SSW state, as SSW and SS emphasize, is to make many phase measurements on identical states and to combine the data from all these measurements in order to find the location of the central peak. The appropriate energy constraint for such a multiple-measurement scheme involves the mean total number of photons used for all the measurements. In Sec. III we analyze multiple measurements governed by the SSW distribution.

III. MULTIPLE SSW PHASE MEASUREMENTS

A. Multiple measurements and the Cramér-Rao lower bound

Our first task is to formulate the description of a single "experiment" that consists of N_p SSW phase mea-

TABLE II. Symmetric, two-tailed $100\chi\%$ confidence intervals $\Delta\phi(\chi)$ (in deg) for the fiducial SSW phase probability distribution $P(\phi|0)$.

M	$\bar{N}$	$\Delta\phi(0.6826)$ (deg)	$\Delta\phi(0.95)$ (deg)	$\Delta\phi(0.99)$ (deg)
0	0.000	245.736	342.000	356.400
1	0.200	156.057	288.329	342.282
4	0.560	103.257	312.022	351.378
12	1.024	97.822	302.285	349.160
30	1.496	95.271	300.413	347.907
74	2.004	93.414	299.806	347.812
399	3.000	92.567	299.387	347.702
2089	4.000	92.405	299.314	347.686
10848	5.000	92.374	299.300	347.683

surements, all aimed at determining the same phase shift θ. We imagine that there are N_p field modes, which we call "pulses." Each pulse is prepared in the fiducial SSW state $|\psi_{\rm SSW}\rangle$ of Eq. (1.33), which has mean number of photons $\bar{N}$. Each pulse is then phase shifted by angle θ, as in Eq. (1.25). On each pulse one performs an ideal phase measurement, its statistics described by the conditional probability $P(\phi|\theta)$ of Eq. (1.37). Thus a single experiment produces a set of N_p phases, $\phi_1,\ldots,\phi_{N_p}$, the results of the N_p phase measurements. We abbreviate the N_p pieces of phase data as a vector $\boldsymbol{\phi} = (\phi_1,\ldots,\phi_{N_p})$. Since the N_p measurements are independent, the probability distribution for the data $\boldsymbol{\phi}$, given true phase shift θ, is a product of probability distributions,

$$P(\boldsymbol{\phi}|\theta) = \prod_{j=1}^{N_p} P(\phi_j|\theta) = \prod_{j=1}^{N_p} \frac{A^2}{2\pi}|\rho(\phi_j-\theta)|^2 . \quad (3.1)$$

Throughout this section we choose the reference phase so that ϕ is defined on the interval $(-\pi,\pi]$. The energy constraint must now be expressed in terms of the mean total number of photons, $N_{\rm tot} = N_p\bar{N}$, for the entire set of pulses.

An ideal phase measurement leaves the field in a SG eigenstate (1.8), which has infinite energy. Thus an ideal phase measurement (like an ideal position measurement, which leaves the system in a position eigenstate) is an idealization, which can only be realized approximately. Moreover, since a pulse is left in an SG eigenstate, further measurements on a single pulse do not yield additional information about the phase shift θ.

Given the data $\boldsymbol{\phi}$, the experimenter must estimate the true phase shift θ. In general, the estimator θ_E can be any function of $\boldsymbol{\phi}$:

$$\theta_E = \theta_E(\boldsymbol{\phi}) = \theta_E(\phi_1,\ldots,\phi_{N_p}) . \quad (3.2)$$

We can derive formally the probability distribution $P(\theta_E|\theta)$ of the estimator, given a true phase shift θ:

$$P(\theta_E|\theta) = \int d\boldsymbol{\phi}\, P(\boldsymbol{\phi}|\theta)\delta(\theta_E-\theta_E(\boldsymbol{\phi})) . \quad (3.3)$$

The probability distribution $P(\theta_E|\theta)$ contains all the statistics of the estimator. From it we can calculate, for example, the mean value of the estimator,

$$E[\theta_E] = \int d\theta_E\,\theta_E P(\theta_E|\theta) = \int d\boldsymbol{\phi}\,\theta_E(\boldsymbol{\phi})P(\boldsymbol{\phi}|\theta) \quad (3.4)$$

(we use $E[X]$ to denote the expected (mean) value of X, given true phase shift θ), and the variance of the estimator,

$$\begin{aligned}\mathrm{var}(\theta_E) &= E\left[(\theta_E - E[\theta_E])^2\right] \\ &= \left(\int d\theta_E\,\theta_E^2 P(\theta_E|\theta)\right) - (E[\theta_E])^2 \\ &= \left(\int d\boldsymbol{\phi}\,[\theta_E(\boldsymbol{\phi})]^2 P(\boldsymbol{\phi}|\theta)\right) - (E[\theta_E])^2 . \quad (3.5)\end{aligned}$$

Notice that both $E[\theta_E]$ and $\mathrm{var}(\theta_E)$ are functions of the true phase shift θ.

A fundamental theorem in statistics [40] places a lower bound, called the *Cramér-Rao lower bound* (CRLB), on the variance of *any* estimator θ_E when the true value is θ:

$$\mathrm{var}(\theta_E) \geq \frac{\left(\frac{dE[\theta_E]}{d\theta}\right)^2}{E\left[\left(\frac{\partial\mathcal{L}(\boldsymbol{\phi}|\theta)}{\partial\theta}\right)^2\right]} . \quad (3.6)$$

Here

$$\mathcal{L} = \mathcal{L}(\boldsymbol{\phi}|\theta) \equiv \ln P(\boldsymbol{\phi}|\theta) \quad (3.7)$$

is the log-likelihood function. The proof of the CRLB is so simple that it can be sketched in one long sentence: if we let

$$\mathcal{T} = \mathcal{T}(\boldsymbol{\phi}|\theta) \equiv \frac{\partial\mathcal{L}(\boldsymbol{\phi}|\theta)}{\partial\theta} , \quad (3.8)$$

then differentiating with respect to θ the normalization condition for $P(\boldsymbol{\phi}|\theta)$ gives

$$E[\mathcal{T}] = 0 , \quad (3.9)$$

and differentiating with respect to θ the expected value (3.4) of the estimator gives

$$\begin{aligned}\left(\frac{dE[\theta_E]}{d\theta}\right)^2 &= (E[\theta_E\mathcal{T}])^2 = (E[(\theta_E - E[\theta_E])\mathcal{T}])^2 \\ &\leq \mathrm{var}(\theta_E)E[\mathcal{T}^2] . \quad (3.10)\end{aligned}$$

The final inequality, which gives the CRLB, is the Schwarz inequality in the form $(E[XY])^2 \leq E[X^2]E[Y^2]$.

The form of the CRLB just presented is general, applying to any probability distribution $P(\boldsymbol{\phi}|\theta)$ that satisfies mild differentiability conditions and to any estimator θ_E. For the case we deal with here, we can simplify the CRLB in several ways. First, we are interested in *unbiased* estimators, for which the expected value of the estimator returns the true value, i.e.,

$$E[\theta_E] = \theta . \quad (3.11)$$

For an unbiased estimator the numerator in the CRLB is 1. We specialize to an unbiased estimator in all that follows. For a quantity such as phase, defined on a finite interval with the interval's end points identified, there are subtleties associated with the notion of an unbiased estimator. We discuss these subtleties briefly in Sec. III B.

A second simplification is that for the case we deal with here, the independence of the separate measurements means that the log-likelihood function (3.7) is a sum of contributions from the various measurements,

$$\mathcal{L}(\boldsymbol{\phi}|\theta) = \sum_{j=1}^{N_p} \ln P(\phi_j|\theta) , \quad (3.12)$$

so the denominator in the CRLB becomes a sum of contributions from the various pulses. Thus the CRLB simplifies to

$$\mathrm{var}(\theta_E) \geq \frac{1}{N_p F(\theta)} , \tag{3.13}$$

where

$$\begin{aligned} F(\theta) &= \int_{-\pi}^{\pi} d\phi\, P(\phi|\theta) \left(\frac{\partial \ln P(\phi|\theta)}{\partial \theta} \right)^2 \\ &= \int_{-\pi}^{\pi} d\phi\, \frac{1}{P(\phi|\theta)} \left(\frac{\partial P(\phi|\theta)}{\partial \theta} \right)^2 \\ &= -\int_{-\pi}^{\pi} d\phi\, P(\phi|\theta) \frac{\partial^2 \ln P(\phi|\theta)}{\partial \theta^2} \end{aligned} \tag{3.14}$$

is called the *Fisher information* [40].

A third simplification is a consequence of our phase-estimation problem. Any distribution $P(\phi|\theta)$ (including the SSW distribution) obtained by phase shifting a fiducial state, as in Eq. (1.26), is translationally invariant—i.e., it is a function only of $\phi - \theta$—and is periodic with period 2π. For any distribution with these properties we can replace $\partial/\partial\theta$ by $-\partial/\partial\phi$ throughout the definition (3.14) of the Fisher information; furthermore, for any such distribution the Fisher information is independent of θ. Thus the Fisher information for translationally invariant, 2π-periodic distributions can be calculated using the fiducial distribution $P(\phi|0)$:

$$\begin{aligned} F &= \int_{-\pi}^{\pi} d\phi\, P(\phi|0) \left(\frac{d \ln P(\phi|0)}{d\phi} \right)^2 \\ &= \int_{-\pi}^{\pi} d\phi\, \frac{1}{P(\phi|0)} \left(\frac{dP(\phi|0)}{d\phi} \right)^2 \\ &= -\int_{-\pi}^{\pi} d\phi\, P(\phi|0) \frac{d^2 \ln P(\phi|0)}{d\phi^2} . \end{aligned} \tag{3.15}$$

The Fisher information is sensitive to sharp structures in $P(\phi|\theta)$. For example, the Fisher information for a Gaussian function is the inverse of the variance of the Gaussian function. An estimator that achieves the Fisher-information lower bound in the CRLB is called an *efficient* estimator. For a Gaussian function the mean of the data is an efficient estimator, but there is no guarantee, in general, that an efficient estimator can be found. In Sec. III C we consider the behavior of the Fisher information for the SSW distribution and find that F is roughly linear in M. This behavior shows that the SSW Fisher information is sensitive to the narrow central peak and that, if there were an efficient estimator for small N_p, SSW measurements would have extraordinary sensitivity, exponentially small in the mean photon number. We have no evidence one way or the other as to the existence of an efficient estimator for the SSW distribution for small N_p, so in the next subsection we do the best we know how: we specialize to maximum-likelihood estimation, which is efficient asymptotically in N_p.

B. Maximum-likelihood estimation

From the point of view of the experimenter, the true phase shift θ is unknown, so he or she assigns it a uniform *prior* probability distribution $P(\theta) = 1/2\pi$. With this choice the *posterior* distribution for the true value θ, given the experimental data ϕ, is

$$P(\theta|\phi) = \frac{P(\phi|\theta)(1/2\pi)}{P(\phi)} . \tag{3.16}$$

Maximum-likelihood estimation estimates the parameter θ as the value θ_{MLE} that is most likely, given the experimental data. In the case of a uniform prior distribution, Eq. (3.16) shows that maximum-likelihood estimation is the same as choosing θ_{MLE} to be the value of θ that maximizes $P(\phi|\theta)$, given the data ϕ. Maximizing a function being the same as maximizing its logarithm, we can write

$$\theta_{\mathrm{MLE}}(\phi) \equiv \arg\Big[\max_{\theta_0 < \theta \leq \theta_0 + 2\pi} \mathcal{L}(\phi|\theta) \Big] , \tag{3.17}$$

where $\mathcal{L}$ is the log-likelihood function defined in Eq. (3.12), and where θ_0 is a reference phase that determines the 2π interval in which one searches for the maximum of the 2π-periodic function $\mathcal{L}(\phi|\theta)$.

The special symmetries of the SSW distribution impress themselves on the probability distribution $P(\theta_{\mathrm{MLE}}|\theta)$ of the ML estimator. Translational invariance and 2π periodicity imply immediately that $P(\theta_{\mathrm{MLE}}|\theta)$ is also translationally invariant—i.e., a function only of $\theta_{\mathrm{MLE}} - \theta$—and periodic with period 2π. Moreover, the even symmetry of $P(\phi|\theta)$ about $\phi = \theta$ implies that $P(\theta_{\mathrm{MLE}}|\theta)$ is even about $\theta_{\mathrm{MLE}} = \theta$ (indeed, for the SSW distribution, the ML estimator distribution is even about a *peak* at the true value).

It is important to note that the mean, variance, and higher moments of $P(\theta_{\mathrm{MLE}}|\theta)$ depend on the choice of reference interval relative to the true phase shift. In particular, the ML estimator for the SSW distribution is unbiased *only if the reference interval is centered about the true value* [i.e., $\theta_0 = \theta - \pi$ in Eq. (3.17)]; for other choices, the ML estimator is quite obviously biased in general. What this tells us is that the only reasonable choice of reference interval is the one centered about the peak at the true value. Only for this choice do the moments give meaningful information about the estimator distribution.

Is there a problem then with the whole procedure, since moments of the ML estimator depend on the choice of reference interval relative to an unknown true phase shift? It is worth disspelling this notion. The essentially experimental procedure of evaluating the ML estimator from the data, as in Eq. (3.17), is independent of the true phase shift, as it must be. Having got the ML estimator, the experimenter enquires about the confidence to be placed in the estimate. To answer the experimenter's question, we must investigate the statistics of the ML estimator, given an *assumed* true phase shift. For our answer to be meaningful, we choose a reference interval centered about the assumed true phase shift, in which case the ML estimator is unbiased. The difficulty does not lie in ML estimation; rather it lies in our attempt to quantify the error in ML estimation. Throughout the following, we implement this discussion in the following way: to generate the statistics of the ML estimator, we assume a true phase shift $\theta = 0$—i.e., we choose $\theta = 0$

on the right-hand side of Eq. (3.3); to evaluate $\theta_{\mathrm{MLE}}(\phi)$, we adopt a reference interval centered about $\theta = 0$—i.e., we choose $\theta_0 = -\pi$ in Eq. (3.17).

The importance of ML estimation comes from a theorem [40], due to Fisher, that shows that ML estimation is efficient asymptotically in N_p. Specifically, given some mild restrictions on the underlying probability distribution $P(\phi|\theta)$, the theorem states that

$$\frac{\theta_{\mathrm{MLE}} - \theta}{1/\sqrt{N_p F(\theta)}} \tag{3.18}$$

is a standard normal random variable as $N_p \to \infty$. In other words, the estimator distribution $P(\theta_{\mathrm{MLE}}|\theta)$ is asymptotically a Gaussian function with mean $E[\theta_{\mathrm{MLE}}] = \theta$ and variance $1/N_p F(\theta)$:

$$P(\theta_{\mathrm{MLE}}|\theta) \underset{N_p\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi[1/N_p F(\theta)]}} \times \exp\left(-\frac{(\theta_{\mathrm{MLE}} - \theta)^2}{2[1/N_p F(\theta)]}\right). \tag{3.19}$$

Notice that the theorem implies that ML estimation is *asymptotically* unbiased, even in cases where it is biased for small N_p. More importantly, comparison with the CRLB (3.13) shows that ML estimation is asymptotically as good as or better than any estimation technique. The theorem is proved by first expanding $\partial\mathcal{L}(\phi|\theta)/\partial\theta$, evaluated at the true phase shift, about the ML estimate θ_{MLE}, where the derivative is zero by definition, and then applying the central-limit theorem to the expansion.

We have simulated multiple SSW phase measurements with ML estimation of the phase shift. We summarize the results of our simulations by giving symmetric, two-tailed $100\chi\%$ confidence intervals $\Delta\theta_{\mathrm{MLE}}(\chi)$ for the ML estimator, because confidence intervals are more informative than the variance for markedly non-Gaussian statistics. It is important to stress that the CRLB gives a lower bound on the *variance* of the ML estimator; although confidence intervals are loosely related to the variance, we have no similar lower bound on them. The two are connected asymptotically, however, by the above theorem. Since the ML estimator is asymptotically Gaussian, the two-tailed 68.26% confidence interval limits to twice the square root of the variance (two standard deviations) and thus is bounded by the asymptotic CRLB. One of the aims of our simulations is to determine the behavior of the ML estimator distribution—and its associated confidence intervals—in the preasymptotic regime.

To make a general comparison of asymptotic ML confidence intervals with the CRLB, we find it convenient to introduce a one-standard-deviation CRLB width

$$\sigma_{\mathrm{CRLB}} \equiv \frac{1}{\sqrt{N_p F}}. \tag{3.20}$$

Since the ML estimator is asymptotically Gaussian, its confidence intervals limit to Gaussian confidence intervals. We can express this limit as

$$\Delta\theta_{\mathrm{MLE}}(\chi) \underset{N_p\to\infty}{\longrightarrow} n_\sigma(\chi)\sigma_{\mathrm{CRLB}}, \tag{3.21}$$

where $n_\sigma(\chi)$ is the number of Gaussian standard deviations within a two-tailed Gaussian confidence interval with confidence coefficient χ. To put it differently, $n_\sigma(\chi)$ is the $100\chi\%$ confidence interval for a unit-standard-deviation Gaussian function:

$$n_\sigma(\chi) = 2C_G^{-1}\left(\frac{1+\chi}{2}\right) = 2\sqrt{2}\,\mathrm{erf}^{-1}(\chi) \tag{3.22}$$

[$n_\sigma(0.6826) = 2.00$; $n_\sigma(0.95) = 3.92$]. Here C_G^{-1} denotes the inverse of the cumulative probability for a unit-standard-deviation Gaussian function, and erf^{-1} denotes the inverse of the error function.

The Fisher information for the SSW distribution is sensitive to the narrow central peak, so the efficiency theorem for ML estimation means that given enough pulses to get into the asymptotic regime, one can probe the location of that peak. This not surprising conclusion is our motivation for considering multiple SSW measurements. The key question can now be formulated: how many pulses are required to get into the asymptotic regime? On the answer to this question rides the ultimate sensitivity of multiple SSW measurements. Our numerical simulations of multiple SSW measurements answer this question for small values of $\bar{N} \le 5$, $N_p \le 240$, and $N_{\mathrm{tot}} \le 480$. Before turning to the simulations, however, we consider in more detail the behavior of the Fisher information.

C. Fisher information for SSW distribution

Table III gives the Fisher information F of the SSW distribution for a few representative, small values of $\bar{N}$. The Fisher information is calculated by numerical integration of the first form in Eq. (3.15). For these small values of $\bar{N}$ the Fisher information increases roughly linearly with M, although it is not hard to discern that the increase is actually somewhat faster than linear. Assuming provisionally that $F \sim M$, we can compare the Fisher information for the SSW distribution with that for a Gaussian function. For a Gaussian function the Fisher information is the inverse of the variance—i.e., the inverse of the square of the width of the distribution. In contrast, the width of the central peak of the SSW distribution is $\sim 1/M$, so the SSW Fisher information is of order the inverse of the width. Thus the SSW Fisher information is sensitive to the very narrow central feature of the distribution, but not so sensitive as the Fisher information for a Gaussian function of the same width; the number of pulses needed for the CRLB to approach the width of the central peak is $N_p \sim M$. Still, if $F \sim M$, then F is exponentially large in the mean photon number, which would lead to extraordinary phase sensitivity were there an efficient estimator for small N_p.

To draw solid conclusions from our numerical simulations, we need to know at least a bit about the asymptotic behavior of F for large M. We suspect that it would be extremely difficult to investigate the asymptotic behavior numerically, because we conjecture that the $F \sim M$ behavior is the leading term multiplying an asymptotic expansion in $1/\ln M \sim 1/\bar{N}$. Thus, in contrast to the quantities investigated in Sec. II, to access the asymp-

TABLE III. SSW Fisher information F and two-standard-deviation, $N_p = 1$ CRLB width (in deg) for various values of $M(\bar{N})$. The Fisher information F is calculated numerically using the first form in Eq. (3.15). The CRLB width in column 4 [cf. Eq. (3.20)] should be compared with the 68.26% confidence interval $\Delta\phi(0.6826)$ tabulated in Table II.

M	$\bar{N}$	F	$2\sigma_{\mathrm{CRLB},N_p=1} = 2/\sqrt{F}$ (deg)
1	0.2	0.400 0	181.19
4	0.560	2.064	79.77
12	1.024	7.722	41.24
30	1.496	22.44	24.19
74	2.004	61.83	14.57
399	3.000	379.3	5.88
2 089	4.000	2 127.3	2.48
10 848	5.000	11 497.4	1.07

totic behavior of the Fisher information would require truly enormous values of M. For the larger values of $\bar{N}$ in Table III, the numerical integrations are already hampered by the fine-scale oscillations in the derivative of the SSW distribution. Fortunately, we need only very weak bounds on the asymptotic behavior of F. On the upper side we derive a strict upper bound on F, which establishes that asymptotically F cannot increase faster than linearly in M. On the lower side we argue that F does increase linearly in M asymptotically, with fractional corrections of order $1/\ln M$.

The strict upper bound on F can best be derived by manipulating the variance of photon number $\hat{N}$ for an arbitrary pure state $|\psi\rangle$, which has phase wave function $\psi(\phi)$ defined by Eq. (1.13). The derivation is identical to Cohen's [45] manipulations of the momentum variance in terms of the position wave function, except that Cohen does not make the connection to Fisher information. The first step is to write the phase wave function as

$$\psi(\phi) = R(\phi)\exp\left(-i\int_{-\pi}^{\phi} d\phi'\,\mathcal{N}(\phi')\right) , \tag{3.23}$$

where $R(\phi) = |\psi(\phi)| = \sqrt{P(\phi|0)}$ is the square root of the phase probability distribution $P(\phi|0) = P_{|\psi\rangle}(\phi)$ [cf. Eq. (1.26)], and where $\mathcal{N}(\phi)$ is a species of "local photon number" at phase ϕ, analogous to the "local momentum" introduced by Cohen [45]. Now, using the correspondence $\hat{N} \leftrightarrow i(d/d\phi)$, we write the photon-number mean and second moment as

$$\langle\hat{N}\rangle = \int_{-\pi}^{\pi} d\phi\,\psi^*(\phi) i\frac{d}{d\phi}\psi(\phi) = \langle\mathcal{N}\rangle , \tag{3.24}$$

$$\langle\hat{N}^2\rangle = \int_{-\pi}^{\pi} d\phi\left|\frac{d\psi(\phi)}{d\phi}\right|^2 = \frac{1}{4}F + \langle\mathcal{N}^2\rangle . \tag{3.25}$$

In these expressions the averages of $\mathcal{N}$ and $\mathcal{N}^2$ are calculated with respect to $R^2(\phi) = P(\phi|0)$, and

$$F = 4\int_{-\pi}^{\pi} d\phi\left(\frac{dR(\phi)}{d\phi}\right)^2 = \int_{-\pi}^{\pi} d\phi\,\frac{1}{P(\phi|0)}\left(\frac{dP(\phi|0)}{d\phi}\right)^2 \tag{3.26}$$

is the Fisher information (3.15) associated with estimating a phase shift θ applied to fiducial state $|\psi\rangle$. Converting to photon-number variance, we find

$$\mathrm{var}(\hat{N}) = \langle\hat{N}^2\rangle - \langle\hat{N}\rangle^2 = \frac{1}{4}F + \mathrm{var}(\mathcal{N}) , \tag{3.27}$$

where $\mathrm{var}(\mathcal{N})$ is the variance of the local photon number.

This way of writing the photon-number variance allows us to place a strict upper bound on the Fisher information,

$$F = 4\,\mathrm{var}(\hat{N}) - 4\,\mathrm{var}(\mathcal{N}) \le 4\,\mathrm{var}(\hat{N}) , \tag{3.28}$$

valid for any fiducial state. Specialized to the SSW case, we have the upper bound

$$F \le 4A^2(M+1) - 4(\bar{N}+1)^2 \le 4A^2(M+1) \simeq \frac{24}{\pi^2}(M+1) \tag{3.29}$$

[cf. Eq. (2.16)], which shows that F cannot grow faster than linearly in M asymptotically.

We derive a crude approximation to F by employing the analytic forms for the SSW distribution from Sec. II. We begin by using the even symmetry of the SSW distribution to write the Fisher information (3.15) in the form

$$F = 2\int_0^{\pi} d\phi\,P(\phi|0)\left(\frac{d\ln P(\phi|0)}{d\phi}\right)^2 . \tag{3.30}$$

We now choose an angle ξ/M, where ξ is to be of order unity, so that ξ/M lies at the transition between the central peak and the universal behavior outside the central peak. For $0 \le \phi < \xi/M$, we approximate the SSW distribution by the Gaussian fit $P_G(\phi)$ of Sec. II E; in doing so, we make fractional errors of order $(\xi/M)^2/\Sigma^2 \sim \xi^2/\ln M$ in the integrand of the Fisher information. For $\phi \ge \xi/M$, we approximate the SSW distribution by the universal distribution $P_U(\phi)$ of Eq. (2.26); in doing so, we make fractional errors of order $1/\xi\ln(M/\xi)$ in the integrand of the Fisher information.

Using these approximations, we write the Fisher integral (3.30) as the sum of two contributions,

$$F = F_1 + F_2 , \tag{3.31}$$

$$F_1 = 2\int_0^{\xi/M} d\phi\, P_G(\phi)\left(\frac{d\ln P_G(\phi)}{d\phi}\right)^2 , \tag{3.32}$$

$$F_2 = 2\int_{\xi/M}^{\pi} d\phi\, P_U(\phi)\left(\frac{d\ln P_U(\phi)}{d\phi}\right)^2 , \tag{3.33}$$

the first from within the central peak and the second from the universal behavior leading up to the central peak. Plugging the Gaussian fit (2.36) into Eq. (3.32) yields

$$F_1 \sim \frac{A^2}{\pi} M(2\ln M)^{3/2}\int_0^{\xi/\sqrt{2\ln M}} dx\, x^2 e^{-x^2} \sim \frac{\xi^3 A^2}{3\pi} M , \tag{3.34}$$

where the fractional errors introduced by approximating $P(0|0) \sim (A^2/2\pi)(\ln M)^2$ and in evaluating the integral are again of order $\xi^2/\ln M$. The second contribution F_2 can be evaluated by using the leading term in the small-angle approximation (2.28) to $P_U(\phi)$, which introduces fractional errors of order $1/(\ln M)^2$, and by extending the integral to infinity (with negligible error):

$$F_2 = \frac{4A^2}{\pi}\int_{\xi/M}^{\infty}\frac{d\phi}{\phi^2} = \frac{4A^2}{\xi\pi} M . \tag{3.35}$$

The final result is that under these approximations the Fisher information,

$$F \sim \frac{A^2}{\pi} M\left(\frac{\xi^3}{3} + \frac{4}{\xi}\right) , \tag{3.36}$$

is composed of roughly equal contributions from within the central peak and from the universal behavior leading up to the central peak. Both contributions are linear in M, with fractional errors of order $1/\ln M$. This suggests that F is linear in M asymptotically and that Eq. (3.36) is the leading term multiplying an asymptotic expansion in $1/\ln M$. If this conjecture is correct, then the asymptotic behavior of F does not show up until enormous values of M.

D. Monte Carlo simulations

We have used Monte Carlo simulation on a computer to build up the statistics of the ML estimator for multiple SSW phase measurements. The underlying probability for a simulation is the fiducial SSW phase distribution $P(\phi|0)$ for some value of $M(\bar{N})$. We specialize (without loss of generality) to a true phase shift $\theta = 0$ and accordingly adopt a reference interval $(-\pi, \pi]$ for ϕ. A single "experiment" within the simulation consists of drawing N_p phases $\phi_1, \ldots, \phi_{N_p}$ from the distribution $P(\phi|0)$—these are the results of ideal phase measurements on N_p pulses—and then determining the ML estimate $\theta_{\rm MLE}$ as the maximum of the log-likelihood function (3.12). The formal definition of the ML estimate is given by Eq. (3.17); in accordance with the discussion following Eq. (3.17) we choose a reference interval $(-\pi, \pi]$ for $\theta_{\rm MLE}$. The simulation is made up of a large number N_e of experiments, from which we build up the statistics of the ML estimator.

We have run simulations based on two independently prepared computer codes. Results from the two codes agree within the accuracies of the simulations. The first of the codes (prepared by Braunstein) was used to run a limited number of very-high-accuracy simulations for $\bar{N} = 2$. We do not describe the procedures used in nor present any results from the Braunstein code here. The second code (prepared by Lane) was used to run simulations for representative values of $\bar{N} \le 5$. We describe in some detail the computer algorithms used in the Lane code. Before doing so, however, we give a general discussion, applicable to any estimator, of how to extract the estimator distribution and confidence intervals from the results of the simulation.

1. *Confidence intervals*

The result of a simulation is a data set of N_e phase-shift estimates, $\theta_1, \theta_2, \ldots, \theta_{N_e}$, drawn from the estimator probability distribution $P(\theta_E|0)$. From this data one wishes to determine the estimator statistics and, ultimately, to evaluate confidence intervals for the estimator. The accuracy of these efforts improves as N_e increases.

Confidence intervals can be extracted directly from the cumulative estimator probability,

$$C_E(\theta_E) \equiv \int_{-\pi}^{\theta_E} d\theta'_E\, P(\theta'_E|0) , \tag{3.37}$$

where the even symmetry of $P(\theta_E|0)$ implies that $C_E(\theta_E) = 1 - C_E(-\theta_E)$. The symmetric, two-tailed $100\chi\%$ confidence interval (confidence coefficient $\chi = 1-\alpha$) is defined as in Eq. (2.49):

$$\Delta\theta_E(\chi) = \Delta\theta_E(1-\alpha) \equiv C_E^{-1}(1-\alpha/2) - C_E^{-1}(\alpha/2) . \tag{3.38}$$

An approximation to the cumulative probability can be extracted reliably by linear interpolation between the discrete results of a simulation. (This is in contrast to the estimator probability distribution, for which some smoothing operation, akin to binning, would be required to obtain an approximate form from discrete data points.) The procedure we use is the following. Given the phase-shift estimates θ_j, ordered so that $\theta_1 \le \theta_2 \le \cdots \le \theta_{N_e}$, the sample cumulative estimator probability at the discrete data points can be defined by

$$C_s(\theta_j) = \frac{j}{N_e+1} , \quad j = 0, 1, \ldots, N_e, N_e+1 , \tag{3.39}$$

where we add two points $\theta_0 = -\pi$ and $\theta_{N_e+1} = \pi$ at the boundaries, and where the denominator $N_e + 1$ gives proper normalization. This discrete-point cumulative probability is extended by linear interpolation to a continuous sample cumulative probability on the interval $[-\pi, \pi]$:

$$C_s(\theta) = C_s(\theta_j) + \frac{\theta-\theta_j}{\theta_{j+1}-\theta_j}[C_s(\theta_{j+1}) - C_s(\theta_j)] , \quad \theta_j \le \theta \le \theta_{j+1} . \tag{3.40}$$

The sample inverse map from cumulative probabilities to angles, denoted here by $\theta_s(C)$, also follows from linear interpolation between the data points:

$$\theta_s(C) = \theta_j + [(N_e+1)C - j](\theta_{j+1} - \theta_j) ,$$

$$\frac{j}{N_e+1} \le C \le \frac{j+1}{N_e+1} . \quad (3.41)$$

The two-tailed $100\chi\%$ sample confidence interval is defined by analogy with Eq. (3.38):

$$\Delta\theta_E(\chi) \equiv \theta_s(1-\alpha/2) - \theta_s(\alpha/2) = \theta_s\left(\frac{1+\chi}{2}\right) - \theta_s\left(\frac{1-\chi}{2}\right) . \quad (3.42)$$

Notice that the sample cumulative probability (3.40) is not even because of the statistical nature of the data set (although its average over many simulations is even); for the same reason the sample confidence interval (3.42) is not symmetric about $\theta = 0$, except on the average over many simulations.

The same statistical fluctuations mean that the sample confidence interval (3.42) fluctuates from one simulation to the next. We need to quantify the likely error in the sample confidence interval in terms of a confidence level. To do this, consider the actual angular interval whose width is the confidence interval with confidence coefficient $\chi = 1 - \alpha$. During a particular simulation some number J of the phase-shift estimates fall within this true interval, so it is assigned cumulative probability J/N_e (end-point effects that change the numerator and denominator by ± 1 are unimportant when N_e is large). Each experiment in the simulation is a Bernoulli trial, with probability χ for the estimate to fall within the true interval. Thus the probability for J points in the interval is a binomial distribution with mean $N_e\chi$ and variance $N_e\chi\alpha$. When N_e is large, the binomial distribution can be approximated by a Gaussian (a rule of thumb often used is that the Gaussian approximation is fairly good for $N_e\chi \gtrsim 5$ and $N_e\alpha \gtrsim 5$; for the tightest confidence interval we use, $\alpha = 0.01$, this would require $N_e \gtrsim 500$, which is satisfied by almost all of our simulations). In this Gaussian approximation, the cumulative probability assigned to the true interval also has a Gaussian distribution, with mean χ and variance $\chi\alpha/N_e$.

The upshot of this discussion is that at the $100\chi_c\%$ confidence level (χ_c not to be confused with χ), we can say that the cumulative probability assigned to the true interval lies within upper and lower bounds defined by

$$\chi_\pm = \chi \pm \frac{1}{2} n_\sigma(\chi_c)\sqrt{\frac{\chi\alpha}{N_e}} , \quad (3.43)$$

where $n_\sigma(\chi_c)$ is the number of Gaussian standard deviations within a symmetric, two-tailed Gaussian confidence interval with confidence coefficient χ_c [cf. Eq. (3.22)]. It is useful to rewrite $\chi_\pm$ in terms of a fractional error f on the error probability α:

$$\chi_\pm = 1 - \alpha(1 \mp f) = \chi \pm \alpha f . \quad (3.44)$$

Comparison of Eqs. (3.43) and (3.44) gives a fractional error

$$f = \frac{1}{2} n_\sigma(\chi_c)\sqrt{\frac{\chi}{\alpha N_e}} \quad (3.45)$$

at the $100\chi_c\%$ confidence level. To double the fractional accuracy on the error probability at given confidence levels for the estimator and for its error requires doing four times as many experiments in a simulation.

In our simulations we use a 95% confidence level for the error in the confidence interval; the fractional error at this confidence level is [$n_\sigma(0.95) = 3.92$]

$$f = 1.96\sqrt{\frac{\chi}{\alpha N_e}} . \quad (3.46)$$

For a 95% estimator confidence interval ($\chi = 0.95$), we need to do $N_e = 811$ experiments to achieve a fractional error $f = 0.3$. Most of our simulations had more experiments than this, although for the largest values of N_p some of the simulations were done with fewer experiments. Most of the simulations had $N_e = 4\,226$ experiments, corresponding to a 99% estimator confidence interval ($\chi = 0.99$) with fractional error $f = 0.3$, or $N_e = 9\,508$ experiments, corresponding to $f = 0.2$ for $\chi = 0.99$.

To get upper and lower bounds on the estimator confidence interval at the $100\chi_c\%$ level, we map the upper and lower probability bounds $\chi_\pm$ to corresponding confidence intervals using the sample inverse map (3.41), as in Eq. (3.42):

$$\Delta\theta_E(\chi_\pm) \equiv \theta_s\left(\frac{1+\chi_\pm}{2}\right) - \theta_s\left(\frac{1-\chi_\pm}{2}\right) . \quad (3.47)$$

2. *Computer algorithms*

Because of the complexity of the SSW probability distribution—in particular, its poor behavior from a numerical standpoint—we describe briefly in this subsection the operation of the Lane computer code and the checks made to ensure integrity of the results. All coding was done in FORTRAN using double-precision arithmetic for the real variables. The code can be divided into two sections: (i) generation of the fiducial SSW probability distribution and its cumulative probability and (ii) simulation of the experiments (the simulation code could be applied to an arbitrary fiducial probability distribution).

The fiducial SSW distribution was generated from the exact series expressions [Eqs. (2.4), (2.7), and (2.3)]. For $\bar{N} \gtrsim 2$, it turned out to be more efficient to store the relevant distributions in tables rather than calculating them directly each time they were needed. Table-lookup and direct-calculation simulations yielded identical results within the estimated errors.

To generate the measured phases $\phi_1, \ldots, \phi_{N_p}$ for a particular experiment, one first generates N_p points $c_1, \ldots, c_{N_p}$ on the unit interval, drawn from a uniform distribution. (Random-number generation of these points was seeded by a randomly generated integral seed, which itself was seeded by an incrementing integer, so that no two experiments had the same seed.) These N_p points are a sample of the cumulative probability $C(\phi)$

[Eq. (2.18)]. To get the measured phases, one maps to phase using the inverse of the cumulative probability. This was done by finding the root ϕ_j of the equation $C(\phi_j) - c_j = 0$ (the final stage of the root finding employed an adaptation of an algorithm ZBRENT [46]). For smaller $\bar{N}$, the cumulative probability used in the root finding was often calculated directly from the Fourier expansion (2.43); otherwise the cumulative probability was found from a table generated by numerical integration. [The tabulated cumulative probability was scaled to make $C(0) = 0.5$, thus ensuring an unbiased mapping to phase, but the scale correction was always negligible (of order 10^{-12}). Symmetry of the tables guaranteed that the full integral was one.] The numbers $\phi_1, \ldots, \phi_{N_p}$ so derived are a random sample from the fiducial SSW phase probability distribution; i.e., they simulate N_p measurements phase made by our imaginary experimental apparatus.

Once the measured phases ϕ have been generated, the log-likelihood function $\mathcal{L}(\phi|\theta) = \ln P(\phi|\theta)$ is maximized with respect to θ to find the ML estimate for that experiment. The maximization requires many calculations of $\ln P(\phi_j|\theta)$. For $\bar{N} \lesssim 2$, many simulations were done using direct calculation of the log probability, but for the higher values of $\bar{N}$, the log probability was found by interpolating from a table to avoid spending the extra time necessary to calculate $P(\phi)$ directly from the exact sum form with its $M + 1$ terms. The table lookups used a quadratic interpolation method, using first and second derivatives (basically a Taylor expansion); the table was increased in size to guarantee a certain preselected interpolation accuracy (typically 10^{-8}).

Maximization of the log likelihood is particularly difficult for the SSW distribution because of the narrow central peak. For N_p small enough to be in the preasymptotic regime (which applies to most of our simulations), the log likelihood tends to have many high, narrow peaks, each associated with one or more of the measured phases. The maximization was performed by taking each measured phase ϕ_j, bracketing it, expanding the bracket until a local maximum was found, and then converging to that maximum to the desired accuracy (using a modification of the routine BRENT [46]). The size of the initial bracket in either direction from ϕ_j was chosen to be the lesser of (i) the standard deviation of the Gaussian fit [Eq. (2.37)] and (ii) half the distance to the nearest measured phase. This choice was made to avoid missing any local maxima. Because the time to run a single experiment increased roughly as N_p^2 (running time for an entire simulation roughly proportional to $N_e N_p^2$), the code was changed to effect a speedup for some of the later simulations. The change consisted of ignoring measured phases whose initial bracket was contained in a previously (expanded and) searched bracket. Many simulations had been completed before this change, but comparisons showed that the change did not affect the results of the simulations within the estimated errors. After finding all local maxima, they were compared to find the global maximum and, hence, the ML estimate.

The ML estimates $\theta_1, \ldots, \theta_{N_p}$ for each simulation were saved on disk. Later they were analyzed by a separate program, which calculated the mean, the median (which was usually closer to 0 than the mean), the sample standard deviation, and the confidence intervals of interest, with their upper and lower bounds, as outlined in the preceding subsection.

Another check was to run the simulation for $N_p = 1$, where the ML estimator has the fiducial SSW distribution, because $\theta_{\mathrm{MLE}}(\phi_1) = \phi_1$. The simulations gave the same confidence intervals as the fiducial SSW probability distribution $P(\phi|0)$ within the estimated errors. Finally we also used the symbolic program MATHEMATICA [47] to check the cumulative probabilities for the SSW distribution. This check, done symbolically for $\bar{N} \leq 3$ and using MATHEMATICA's high-precision arithmetic for $\bar{N} \leq 4$, yielded excellent agreement with the cumulative probabilities calculated by other methods.

3. *Results of simulations*

The results of our simulations for $\bar{N} = 1$, 1.5, 2, 3, 4, and 5, summarized as 68.26% confidence intervals $\Delta\theta_{\mathrm{MLE}}(0.6826)$, are plotted in Fig. 7. These simulations represent about 7 months computation on a Sun-4 workstation. Two features are worthy of note. The first is the change of slope in the $\bar{N} = 1$, 1.5, and 2 curves as N_p becomes large enough that the confidence intervals reach the CRLB. The region of N_p where the confidence interval has limited to the CRLB we refer to as the "asymptotic regime"; smaller values of N_p we refer to as the "preasymptotic regime." The transition between the two regimes we call the "knee." Notice how the knee for $\bar{N} = 1$ occurs at a smaller value of N_p than the knee for $\bar{N} = 2$. Notice also that our simulations do not go to high enough N_p to see the asymptotic CRLB behavior for $\bar{N} \geq 3$. The second feature is the absence of much change in the confidence interval as $\bar{N}$ increases in the

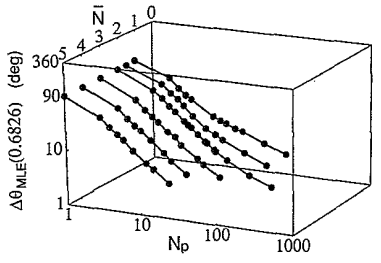

FIG. 7. Three-dimensional log-log plot of 68.26% confidence interval $\Delta\theta_{\mathrm{MLE}}(0.6826)$ (in deg) vs number of pulses N_p for six values of $M(\bar{N})$: $M = 12$ ($\bar{N} = 1.024$), $M = 30$ ($\bar{N} = 1.496$), $M = 74$ ($\bar{N} = 2.004$), $M = 399$ ($\bar{N} = 3.000$), $M = 2089$ ($\bar{N} = 4.000$), $M = 10848$ ($\bar{N} = 5.000$). To aid the eye, the confidence intervals for each value of $\bar{N}$ are connected by straight-line segments to give the curves shown. This figure includes confidence intervals from all our simulations for these values of $\bar{N}$. Other simulations not shown (having different values of $\bar{N}$) were run as part of the optimization procedure described in Sec. III E 1. They are omitted here to avoid cluttering the figure. Error bars for the simulations shown are also suppressed.

preasymptotic regime (with N_p held fixed).

These features are clearer—at the expense of obscuring the curves for different values of $\bar{N}$—in Fig. 8, where the curves for the first five integral values of $\bar{N}$ are laid on top of each other. The CRLB lines for $\bar{N} = 1$, 2, and 3 are also shown on the figure. One can see clearly the transition to the asymptotic straight-line CRLB for $\bar{N} = 1$ and 2; for these two values of $\bar{N}$ (and also for $\bar{N} = 1.5$) our simulations go to large enough N_p to be comfortably into the asymptotic regime. Furthermore, Fig. 8 shows very clearly a universal $\bar{N}$-independent preasymptotic behavior for $\bar{N} \geq 2$, which appears in the log-log plot as a nearly straight line with a slope of about -1.

The following picture emerges from Figs. 7 and 8. For $N_p = 1$ the 68.26% confidence interval for the ML estimator is the same as the 68.26% confidence interval for the fiducial SSW distribution (see column 3 of Table II) and covers about a quarter of the 360° of phase. As N_p increases from 1, there is an initial "finite-phase regime," in which the confidence interval is nearly constant. The finite-phase regime, which for the 68.26% confidence interval runs only from $N_p = 1$ to 3, is a consequence of the finite 360° of phase, which limits the size of confidence intervals.

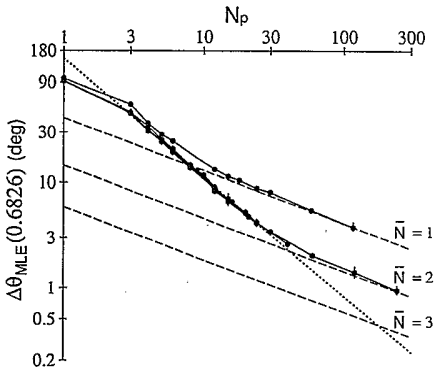

FIG. 8. Log-log plot of 68.26% confidence interval $\Delta\theta_{\rm MLE}(0.6826)$ (in deg) vs number of pulses N_p for five values of $M(\bar{N})$: $M = 12$ ($\bar{N} = 1.024$), $M = 74$ ($\bar{N} = 2.004$), $M = 399$ ($\bar{N} = 3.000$), $M = 2089$ ($\bar{N} = 4.000$), $M = 10848$ ($\bar{N} = 5.000$). The calculated confidence intervals are represented by dots; error bars give 95% confidence upper and lower bounds on the confidence intervals [Eq. (3.47)]. The error bars generally lie within or extend just outside the dots, except at the highest values of N_p, where the number of experiments in a simulation is limited by the considerable computer time it takes to do a single experiment. Nonetheless, even at the highest values of N_p, the simulations have enough accuracy to be meaningful. For each value of $\bar{N}$ the confidence intervals are connected by straight-line segments to give the curves shown. The three dashed lines give the two-standard-deviation CRLB, $2\sigma_{\rm CRLB} = 2/\sqrt{N_p F}$ [Eq. (3.20)] (in deg), for $\bar{N} = 1$, 2, and 3. For $\bar{N} = 1$ and 2, but not for higher values, the simulations go to high enough N_p to see the transition to the asymptotic regime where the confidence intervals approach the CRLB. In the preasymptotic regime the curves for $\bar{N} \geq 2$ lie on top of one another, reflecting the universal $\bar{N}$-independent preasymptotic behavior discussed in the text. The dotted line, a regression fit to the $\bar{N} = 3$, 4, and 5 data points (excluding $N_p = 1$), characterizes the preasymptotic behavior in this region of N_p.

Following the finite-phase regime is the preasymptotic regime, where the confidence intervals display $\bar{N}$-independent, nearly power-law behavior as N_p varies. The preasymptotic behavior occurs when there are too few pulses in the experiment to locate the narrow central peak of the SSW distribution; the N_p measurements are then probing the SSW distribution outside the central peak. For $\bar{N} \gtrsim 2$ the SSW distribution outside the central peak is essentially independent of $\bar{N}$; this we called the universal behavior of the SSW distribution in Sec. II. It is not surprising then that for $\bar{N} \gtrsim 2$ the estimator confidence intervals display a corresponding universal behavior in the preasymptotic regime.

As N_p increases further, the nearly straight preasymptotic line bends until it matches onto the straight line of the CRLB (corresponding to a $N_p^{-1/2}$ power law) in the asymptotic regime. The preceding discussion suggests that the knee between the two regimes should occur when there are sufficiently many pulses to probe the central peak of the SSW distribution, of width $\sim 2/M$. Thus the transition should occur when the 68.26% confidence interval itself becomes of order $2/M$. Essentially what is happening is that N_p becomes large enough that a substantial fraction of the phase results lie within the central peak, which can be approximated by a Gaussian. Thus the statistics of the ML estimator become Gaussian, as predicted by Fisher's efficiency theorem of Sec. III B.

We can formulate a crude description of this picture, to be refined as we go along. In the preasymptotic regime the 68.26% confidence interval (if measured in radians) goes roughly as

$$\Delta\theta_{\rm MLE}(0.6826) \sim \frac{2}{N_p}\,, \tag{3.48}$$

which corresponds to an order unity confidence interval (in radians) for small N_p, and which in a log-log plot gives a straight line with slope -1. The crude description is thus roughly consistent with the numerical results. Refining it means, first, using the results of the simulations to determine precisely the intercept and particularly the slope of the preasymptotic line and, second, investigating whether there are corrections to power-law behavior which cannot be discerned from the limited numerical results. For the present, however, the key feature captured in Eq. (3.48) is the universal $\bar{N}$-independent aspect of the preasymptotic behavior.

In the asymptotic regime the 68.26% confidence interval is given by the CRLB:

$$\Delta\theta_{\rm MLE}(0.6826) = 2\sigma_{\rm CRLB} = \frac{2}{\sqrt{N_p F}}\,. \tag{3.49}$$

Equating Eqs. (3.48) and (3.49) to find the transition between the two regimes gives a knee at $N_p \sim F$. If we assume, as the discussion in Sec. III C suggests, that the Fisher information goes as $F \sim M$, then we locate the knee at $N_p \sim M$, with a confidence interval at the knee $\Delta\theta_{\rm MLE}(0.6826) \sim 2/M$, in accordance with the preceding discussion.

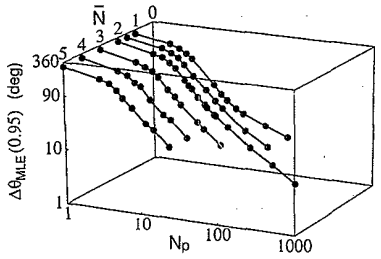

FIG. 9. Three-dimensional log-log plot of 95% confidence interval $\Delta\theta_{\rm MLE}(0.95)$ (in deg) vs number of pulses N_p for the same six values of $M(\bar{N})$ as in Fig. 7. The comments in the caption of Fig. 7 apply here.

The gross features of this crude description remain valid, even after we refine its details. It indicates why our simulations were unable to reach the asymptotic regime for $\bar{N} > 2$. The number of pulses to reach the knee grows linearly with M and hence exponentially with $\bar{N}$. Since the execution time of our codes for a single experiment was roughly proportional to N_p^2, it was hopeless to reach the asymptotic regime except for the smallest values of $\bar{N}$. The more refined model and its implications are discussed in detail in the following subsection.

Our simulations also give 95% estimator confidence intervals, which are plotted in Fig. 9. The basic features of this plot—an initial finite-phase regime, a (nearly straight) preasymptotic line, and an asymptotic CRLB regime—are familiar from the 68.26% confidence intervals. The initial finite-phase regime is longer, extending to about $N_p = 5$, because 95% confidence intervals are initially considerably wider than 68.26% confidence intervals (see column 4 of Table II) and thus bear the imprint of the finite 360° of phase out to larger N_p. The preasymptotic behavior is again universal for $\bar{N} \gtrsim 2$. The knee transition to the asymptotic CRLB regime can be discerned for $\bar{N} = 1$ and 2, where our simulations extend comfortably into the asymptotic regime. The knee occurs at a larger value of N_p than for 68.26% confidence intervals because the 95% confidence intervals are more sensitive to the wings of the estimator distribution, which become Gaussian only for larger values of N_p. To put it another way, larger values of N_p are required before a 95% confidence interval lies comfortably within the Gaussian central peak of the SSW distribution.

Figure 9 does not give error bars, but the error bars are small enough that the 95% confidence intervals are reliable. In the subsequent analysis we concentrate on 68.26% confidence intervals, but we give results for 95% confidence intervals where appropriate.

E. Phase sensitivity of multiple SSW measurements

1. *Optimized numerical results: Low $N_{\rm tot}$*

Up until now we have presented the results of our numerical simulations as confidence intervals that are functions of two variables, the mean number of photons per pulse $\bar{N}$ and the number of pulses N_p. What we are really interested in, however, is the *optimal* performance, given a constraint on the mean total number of photons $N_{\rm tot}$. In this subsection we translate the numerical results into optimized confidence intervals that are a function only of $N_{\rm tot}$. The first step of the translation is to sort the simulations into groups, each of which represents a fixed value of $N_{\rm tot}$. The simulations within each group correspond to different ways of dividing up the total photon-number resources, represented by $N_{\rm tot}$, between the number of pulses N_p and the number of photons per pulse $\bar{N}$. Because M and N_p have only integral values, it is not possible to make all the simulations in each group have *exactly* the same value of $N_{\rm tot}$, but they all have the same value to within a small deviation.

The resulting constant-$N_{\rm tot}$ curves for the 68.26% confidence intervals, plotted as a function of N_p, are displayed in Fig. 10. (We do not plot here the corresponding graphs for 95% confidence intervals, but similar conclusions apply to them.) The small-N_p parts of the curves show the same universal preasymptotic behavior as in Fig. 8. Even though $\bar{N}$ varies along each curve, corresponding to hopping from one curve to another in Fig. 8, that variation makes essentially no difference because the essence of the universal behavior in Fig. 8 is its independence of $\bar{N}$. The preasymptotic behavior in Fig. 10 is closely described by a straight line through all the preasymptotic points. As N_p increases in Fig. 10, there is a transition from the universal preasymptotic behavior to the asymptotic, constant-$N_{\rm tot}$ CRLB. Except for very small values of $N_{\rm tot}$, this transition is sudden, standing out like a check mark. What changes as $N_{\rm tot}$ increases is the location of the transition to asymptotic behavior.

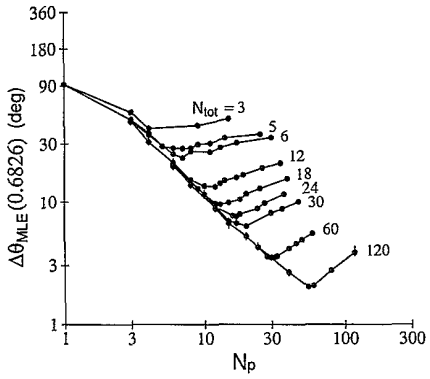

FIG. 10. Log-log plot of 68.26% confidence intervals $\Delta\theta_{\rm MLE}(0.6826)$ (in deg) vs number of pulses N_p for selected values of $N_{\rm tot}$. Error bars give 95% confidence upper and lower bounds on the confidence intervals [Eq. (3.47)]. Points with the same $N_{\rm tot}$ are joined by line segments to give the curves shown. Because M and N_p have only integral values, there are slight variations of $N_{\rm tot}$ within each curve. This can lead to noticeable oscillations in the curves for the smaller values of $N_{\rm tot}$. The small-N_p parts of the curves show the universal preasymptotic behavior. For the higher values of $N_{\rm tot}$, the preasymptotic line turns up sharply when it reaches the constant-$N_{\rm tot}$ CRLB. For each $N_{\rm tot}$ there is an optimum number of pulses that gives the minimum confidence interval.

The transition stands out clearly in Fig. 11, which plots the 68.26% confidence intervals for $N_{\rm tot} = 60$ along with the CRLB for $N_{\rm tot} = 60$. As the number of pulses increases, with a corresponding decrease in mean number of photons per pulse to maintain $N_{\rm tot} = 60$, the confidence-interval width at first decreases until it approaches the CRLB and then increases along the CRLB. The location of the knee transition, at which the ML estimator becomes approximately Gaussian as the confidence interval approaches the CRLB, gives the optimal division of resources between number of pulses and mean number of photons per pulse.

We have located the optimal operating point—i.e., the location of the minimum in the confidence interval—and determined the optimized (minimum) confidence interval for the values of $N_{\rm tot}$ in Fig. 10 for both 68.26% and 95% confidence intervals. After an initial set of simulations, the optimal operating point was first estimated using a quadratic regression fit. The minimum was then located more accurately by doing additional simulations at and around the initial estimate. Because of the time required to do sufficiently accurate additional simulations, the location was not always pinned down to a particular M-N_p pair, especially for 95% confidence intervals. It is important to note, however, that we are mainly interested in the minimum value of the confidence interval, not its exact location in M-N_p. We believe that these optimized confidence intervals have been determined sufficiently accurately.

The optimized 68.26% and 95% confidence intervals are plotted as functions of $N_{\rm tot}$ in Fig. 12. Table IV lists the plotted points, together with the values of N_p and M at which the minima were found. The table shows clearly how, after an initial settling in, the number of photons per pulse increases only slowly with increasing total photon number, whereas the number of pulses (and M) increase roughly linearly. Figure 12 shows that for each of the confidence intervals the optimized points lie nicely on a straight line. The slopes of these lines characterize the improvement in sensitivity with $N_{\rm tot}$ in the low-$N_{\rm tot}$ regime that is accessible to our simulations.

Linear regression, to get the straight-line fit to the optimized 68.26% confidence intervals, yields $\log_{10}\Delta\theta^{\circ}_{\rm MLE}(0.6826) = 2.017 - 0.825\log_{10}N_{\rm tot}$. The optimized confidence intervals are subject to uncertainties due to the statistical nature of the simulations and to errors in locating the optimal operating point. We do not attempt to assess directly the uncertainties introduced by these two sources of error. Instead we assume that the random part of these errors (or any others) can be described by Gaussian fluctuations in the confidence intervals (but not in $N_{\rm tot}$). With this assumption, well-known formulas from statistics [49] give one-standard-deviation uncertainties of ± 0.006 in the intercept and ± 0.004 in the slope of the straight-line fit. The errors in locating the optimal operating point also introduce a systematic bias toward overestimating the confidence intervals. This

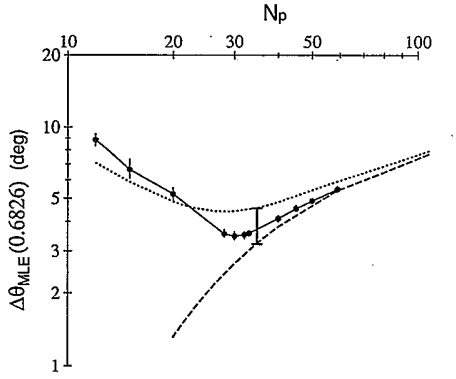

FIG. 11. Log-log plot of 68.26% confidence interval $\Delta\theta_{\rm MLE}(0.6826)$ (in deg) vs number of pulses N_p for $N_{\rm tot} = 60$; the connected dots are the results of our Monte Carlo simulations. Error bars represent 95% confidence upper and lower bounds for the confidence interval. The dashed curve plots the two-standard-deviation CRLB, $2/\sqrt{N_pF}$ (in deg), and the dotted curve plots the $O(1/N_p)$ correction of Eq. (3.73). The large error bar between these two curves shows the range of optimized sensitivities from $2/\sqrt{N_{p,\rm CK}F}$ to $2\sqrt{2/N_{p,\rm CK}F}$ [Eq. (3.75)] at the predicted knee position $N_p = N_{p,\rm CK}$ [Eq. (3.74)].

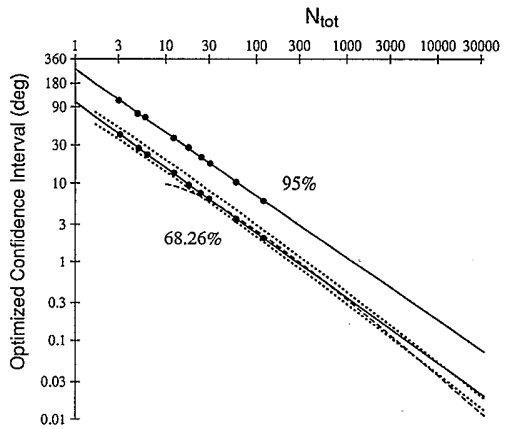

FIG. 12. Log-log plot of optimized 68.26% and 95% confidence intervals (in deg) vs actual mean total number of photons $N_{\rm tot}$ (see Table IV). The calculated confidence intervals are represented by dots; 95% confidence upper and lower bounds on the simulation confidence intervals [Eq. (3.47)] are also plotted but lie inside the dots. In all cases, the dots show confidence intervals from actual simulations (not estimated optimal values) and are plotted against the actual simulation value of $N_{\rm tot}$, thereby eliminating artificial fluctuations due to the requirement that N_p and M be integers. For the 68.26% confidence interval the straight-line regression fit (solid line) is $\log_{10}\Delta\theta^{\circ}_{\rm MLE}(0.6826) = 2.017 - 0.825\log_{10}N_{\rm tot}$, and for the 95% confidence interval the straight-line regression fit (solid line) is $\log_{10}\Delta\theta^{\circ}_{\rm MLE}(0.95) = 2.426 - 0.791\log_{10}N_{\rm tot}$. The dashed line gives the optimized 68.26% confidence interval estimated using Eqs. (3.53) and (3.61) [with c_F evaluated at $\bar{N} = 2$; cf. Eq. (3.64)], which assumes a power-law behavior in the preasymptotic regime, as discussed in Sec. III E 2. The dotted lines give the bounds of Eq. (3.75) on the standard deviation of the ML estimator (doubled and converted to degrees to match the two-tailed confidence intervals plotted from the Monte Carlo simulations), as discussed in Sec. III E 3.

TABLE IV. Optimized 68.26% and 95% confidence intervals (in deg) as a function of the mean total number of photons $N_{\rm tot}$ together with the optimal, or near optimal, division of photons into pulses, specified by the photon-number cutoff M, the mean number of photons per pulse $\bar{N}$, and the number of pulses N_p. Due to fluctuations in the simulations and to the fact that the division of photons is not always fully optimized, the last digit for the optimized confidence interval is not precise. Because N_p and M must be integers, the actual value of $N_{\rm tot} = N_p\bar{N}$ is slightly different from the (integral) value for which optimization is sought. This variation is more serious for the smaller values of $N_{\rm tot}$.

	$N_{\rm tot}$	$\Delta\theta_{\rm MLE}$ (deg)	M	$\bar{N}$	N_p	$N_{\rm tot}$ (actual)
68.26%	3	40.5	7	0.78	4	3.12
	5	27.5	6	0.72	7	5.01
	6	23.0	9	0.89	7	6.23
	12	13.4	12	1.02	12	12.29
	18	9.5	24	1.38	13	17.89
	24	7.6	26	1.42	17	24.13
	30	6.3	30	1.50	20	29.93
	60	3.5	74	2.00	30	60.12
	120	2.0	101	2.18	55	120.14
95%	3	109	1	0.20	15	3.00
	5	75	2	0.35	14	4.86
	6	68	2	0.35	17	5.90
	12	37	6	0.72	17	12.16
	18	28.0	9	0.89	20	17.80
	24	21.4	12	1.02	24	24.59
	30	17.8	12	1.02	30	30.73
	60	10.4	22	1.33	45	59.87
	120	6.0	30	1.50	80	119.72

bias primarily affects the intercept (leading to an overestimate), rather than the slope, which is our main concern. Our experience with the data suggests that ± 0.01 is a reasonable, perhaps slightly optimistic, estimate of the uncertainties.

The same procedure can be used to get a straight-line fit to the optimized 95% confidence intervals. The optimal operating point was not located as precisely for 95% confidence intervals, owing to the larger number of pulses at that operating point (it takes more pulses to reach the asymptotic regime), but the optimized 95% confidence intervals still lie remarkably well on a straight line over the numerically accessible region of $N_{\rm tot}$.

The results for both 68.26% and 95% confidence intervals can be summarized as follows:

$$\Delta\theta_{\rm MLE}(0.6826) = \frac{104^\circ \pm 2^\circ}{N_{\rm tot}^{0.82\pm0.01}}\,, \quad 3 \le N_{\rm tot} \le 120\,, \tag{3.50}$$

$$\Delta\theta_{\rm MLE}(0.95) = \frac{267^\circ \pm 5^\circ}{N_{\rm tot}^{0.79\pm0.01}}\,, \quad 3 \le N_{\rm tot} \le 120\,. \tag{3.51}$$

The significance of the slopes is clear. Over the range $3 \le N_{\rm tot} \le 120$ that is accessible to our simulations, the phase sensitivity of multiple SSW measurements is, in principle, superior to the $1/\sqrt{N_{\rm tot}}$ sensitivity of coherent states, but is *inferior* to the benchmark $1/N_{\rm tot}$ sensitivity of squeezed-state interferometry.

2. *Extrapolation of numerical results: Intermediate* $N_{\rm tot}$

In the preceding subsection we investigated numerically the optimal sensitivity of multiple SSW phase measurements. As $N_{\rm tot}$ increases, the number of pulses required to reach the optimal operating point increases roughly linearly. In contrast, the time required to do the simulations increases like N_p^2, so direct Monte Carlo simulations are unable to reach past relatively small $N_{\rm tot}$. If, however, motivated by the results of the simulations, we assume a form for the universal preasymptotic behavior of the ML confidence intervals, we can extend our results to much higher values of $N_{\rm tot}$. Although the precise form of the preasymptotic behavior is unknown, we argue as follows.

The argument is best visualized in terms of a log-log plot of ML confidence intervals vs N_p, with $\bar{N}$ (M) held fixed, as in the plots of 68.26% confidence intervals in Fig. 8. The two scales in the SSW distribution, both evident in Fig. 8, are the finite 360° of phase, which shows up at the transition from the initial finite-phase regime to the preasymptotic regime, and the width of the central peak, which shows up at the transition from the preasymptotic regime to the asymptotic CRLB regime. It is reasonable to suppose that the universal preasymptotic behavior, whatever its form, is a consequence of probing the universal small-angle runup [Eq. (2.28)] to the central peak of the SSW distribution. The $(A^2/2\pi)(\ln|\phi|)^2$ dependence of this small-angle runup, which contains no additional scales, suggests a scale-free behavior for the ML confidence intervals in the preasymptotic regime. The simplest scale-free behavior appears as a straight line in a log-log plot, and indeed a straight line is perfectly consistent with our numerical results in the preasymptotic regime. Unfortunately, this argument is by no means ironclad; it cannot rule out corrections to a straight line, such as logarithmic corrections, and as we show below, corrections are, in fact, required by our numerical results. Nonetheless, since logarithmic corrections take effect very slowly, assuming straight-line preasymptotic behavior should provide a substantial extension of our numerical results.

If we assume straight-line preasymptotic behavior, then for fixed $N_{\rm tot}$ it is always a good strategy to move down the preasymptotic line by increasing the number of pulses and decreasing the number of photons per pulse. Pursuing this strategy, one eventually encounters the asymptotic straight-line behavior of the CRLB. In our simulations a further increase in the number of pulses, for fixed $N_{\rm tot}$, yields poorer sensitivity, so the optimal sensitivity occurs at the knee transition from the preasymptotic behavior to the asymptotic CRLB. This conclusion holds quite generally, provided only that the Fisher information increases faster than $\bar{N}$—a proviso that, given the discussion in Sec. III C, seems quite secure. Thus, with the assumption of a preasymptotic straight line that has an abrupt knee transition to the asymptotic straight

line of the CRLB, the problem of finding the optimal operating point is reduced to finding the intersection of these two straight lines. We call the intersection the *line-intercept knee* and denote its position by $N_{p,\mathrm{LIK}}$. The optimized confidence interval follows immediately from a determination of the line-intercept knee.

To put this argument into practice, write the preasymptotic line for the $100\chi\%$ ML confidence interval as

$$\log_{10}\Delta\theta^{\circ}_{\mathrm{MLE}}(\chi) = I(\chi) - S(\chi)\log_{10} N_p\ , \tag{3.52}$$

where the intercept $I(\chi)$ and slope $S(\chi)$ are to be determined from the simulations. Preasymptotic slopes and intercepts for the 68.26% and 95% confidence intervals, determined from the simulations, are tabulated in Table V. The table gives slopes and intercepts for separate values of $\bar{N}$ and also for the $\bar{N} = 3$, 4, and 5 data taken together, which we take as representative of the universal preasymptotic behavior. The power law for the confidence interval (in radians) is

$$\Delta\theta_{\mathrm{MLE}}(\chi) = \frac{n_\sigma(\chi)K(\chi)}{N_p^S}\ , \tag{3.53}$$

where

$$K(\chi) = \frac{10^{I(\chi)}}{n_\sigma(\chi)}\frac{\pi}{180} \tag{3.54}$$

and $n_\sigma(\chi)$ is the number of Gaussian standard deviations within a two-tailed Gaussian confidence interval with confidence coefficient χ.

The asymptotic ML estimator is Gaussian with standard deviation

$$\sigma_{\mathrm{CRLB}} = \frac{1}{\sqrt{N_p F(\bar{N})}} \tag{3.55}$$

TABLE V. Power-law fits to the 68.26% and 95% MLE confidence intervals in the preasymptotic regime, specified by the intercept I and slope S in a log-log plot [Eq. (3.52)], for various values of $\bar{N}$ separately and for $\bar{N} = 3$, 4, and 5 collectively. Also given are line-intercept knee positions $N_{p,\mathrm{LIK}}$, determined directly from the intersection of the preasymptotic power law with the asymptotic CRLB [Eqs. (3.54) and (3.57)].

	M	$\bar{N}$	I	S	$N_{p,\mathrm{LIK}}$
68.26%	12	1.024	2.15	0.94	16
	30	1.496	2.27	1.19	19
	74	2.004	2.23	1.17	39
	399	3.000	2.16	1.10	200
	2089	4.000	2.21	1.16	562
	10848	5.000	2.20	1.15	2256
	≥ 399	3, 4, and 5	2.18	1.13	N/A
95%	12	1.024	3.29	1.47	27
	30	1.496	3.17	1.38	50
	74	2.004	3.19	1.42	78
	399	3.000	3.09	1.31	322
	2089	4.000	3.11	1.33	810
	10848	5.000	3.03	1.27	3319
	≥ 399	3, 4, and 5	3.08	1.30	N/A

[Eq. (3.20)], where we indicate explicitly the dependence of the Fisher information $F(\bar{N})$ on the mean number of photons per pulse. The corresponding $100\chi\%$ confidence interval in the asymptotic regime, being $n_\sigma(\chi)$ standard deviations wide, is given by

$$\Delta\theta_{\mathrm{MLE}}(\chi) = \frac{n_\sigma(\chi)}{\sqrt{N_p F(\bar{N})}}\ . \tag{3.56}$$

Equating the preasymptotic power law (3.53) and the asymptotic CRLB power law (3.56) locates the line-intercept knee at

$$N_{p,\mathrm{LIK}} = [K^2 F(\bar{N})]^{1/(2S-1)}\ , \tag{3.57}$$

where $\bar{N} = N_{\mathrm{tot}}/N_{p,\mathrm{LIK}}$. Line-intercept knee positions determined directly from Eqs. (3.54) and (3.57) are listed in Table V.

To solve for $N_{p,\mathrm{LIK}}$ in terms of N_{tot}, we need to know how the Fisher information varies with $\bar{N}$ (or M). The numerical results in Table III and the discussion in Sec. III C suggest that $F(\bar{N})$ increases roughly linearly with the photon-number cutoff M. We describe this roughly linear increase by writing

$$F(\bar{N}) = c_F(M)M\ , \tag{3.58}$$

where $c_F(M) \lesssim 24/\pi^2$ [Eq. (3.29)] is a slowly varying function of M. Putting this form of the Fisher information into Eq. (3.57) and taking a logarithm yields

$$\ln M = (2S-1)\ln N_{p,\mathrm{LIK}} - 2\ln K - \ln c_F\ . \tag{3.59}$$

The next step is to write $\ln M$ in terms of $\bar{N}$. In doing so, it is sufficient to use the first term in the expansion (2.13), with the result that

$$\bar{N} = \frac{N_{\mathrm{tot}}}{N_{p,\mathrm{LIK}}} = \frac{6}{\pi^2}[\,(2S-1)\ln N_{p,\mathrm{LIK}} + \gamma - 2\ln K - \ln c_F] - 1\ . \tag{3.60}$$

This equation gives N_{tot} in terms of the position $N_{p,\mathrm{LIK}}$ of the line-intercept knee. It can be inverted iteratively. One iteration is sufficient for our purposes, the result being an approximate knee position

$$N_{p,\mathrm{LIK}} \simeq \frac{N_{\mathrm{tot}}}{\frac{6}{\pi^2}[(2S-1)\ln N_{\mathrm{tot}} + \gamma - 2\ln K - \ln c_F] - 1}\ . \tag{3.61}$$

As N_{tot} increases, the additive constants and the slowly varying $\ln c_F(M)$ contribution in the denominator of Eq. (3.61) become unimportant. Indeed, in the remainder of this subsection, we neglect them entirely, because our chief aim is qualitative understanding rather than quantitative accuracy. This last approximation leads to a line-intercept knee position (optimal operating point),

$$N_{p,\mathrm{LIK}} \simeq \frac{\pi^2}{6(2S-1)}\frac{N_{\mathrm{tot}}}{\ln N_{\mathrm{tot}}}\ , \tag{3.62}$$

and to a mean number of photons per pulse at the opti-

mal operating point,

$$\bar{N} \simeq \frac{6(2S-1)}{\pi^2} \ln N_{\rm tot} \,. \tag{3.63}$$

Equations (3.62) and (3.63) give the optimal division of resources between photons per pulse and number of pulses. They are in qualitative agreement with the numerical results presented in Sec. III E 1: the number of photons per pulse grows very slowly—logarithmically—with $N_{\rm tot}$ and the number of pulses grows linearly in $N_{\rm tot}$ with a logarithmic correction.

The optimized confidence interval follows immediately upon substituting the line-intercept knee position (3.62) into the preasymptotic power law (3.53):

$$\Delta\theta_{\rm MLE}(\chi) \simeq n_\sigma(\chi) K(\chi) \left(\frac{6(2S-1)}{\pi^2} \frac{\ln N_{\rm tot}}{N_{\rm tot}} \right)^S . \tag{3.64}$$

This formula is an approximate power law, modified by a slow logarithmic variation. The corresponding slowly varying "effective" slope in a log-log plot, defined by

$$S_{\rm eff}(\chi) \equiv -\frac{d \ln \Delta\theta_{\rm MLE}(\chi)}{d \ln N_{\rm tot}} = S(\chi) \left(1 - \frac{1}{\ln N_{\rm tot}} \right) , \tag{3.65}$$

increases with $N_{\rm tot}$, ultimately reaching the preasymptotic slope $S(\chi)$. The optimized 68.26% confidence interval using the full form of $N_{p,\rm LIK}$ [Eq. (3.61)] (with a representative value for c_F chosen at $\bar{N} = 2$ where $\ln c_F = -0.2$) in the preasymptotic confidence interval equation (3.53) is plotted as a dashed line in Fig. 12, where it may be compared with the numerically optimized confidence intervals. The dashed curve gives a reasonable approximation for $N_{\rm tot}$ bigger than about 30, at which point the value of $\bar{N}$ found from the numerical optimization is about 2. For large $N_{\rm tot}$ the dashed curve starts curving downward as the effective slope increases.

The slope S determined from the $\bar{N} \geq 3$ simulations is 1.13 for the 68% confidence interval and 1.30 for the 95% confidence interval. Thus, given the straight-line extrapolation for the preasymptotic behavior, both confidence intervals beat the $1/N_{\rm tot}$ benchmark sensitivity for sufficiently high $N_{\rm tot}$. The numerically determined slopes themselves indicate, however, that the straight-line extrapolation must break down, because the steeper slope of the 95% confidence interval means that, for sufficiently high $N_{\rm tot}$, the 95% confidence intervals would become smaller than the 68.26% confidence intervals, which is clearly impossible. Since the optimized confidence intervals are Gaussian confidence intervals in the asymptotic regime, we can estimate crudely the $N_{\rm tot}$ where the straight-line extrapolation breaks down—more precisely, the $N_{\rm tot}$ beyond which it certainly cannot be valid—by finding where the 95% and 68.26% confidence intervals are related by a Gaussian ratio, i.e., by $\Delta\theta_{\rm MLE}(0.95) = 1.96\,\Delta\theta_{\rm MLE}(0.6826)$. Using the form (3.64) for the optimized confidence intervals locates the breakdown point at

$$\ln\left(\frac{N_{\rm tot}}{\ln N_{\rm tot}} \right) = \frac{a(0.95) - a(0.6826) - \ln 1.96}{S(0.95) - S(0.6826)} , \tag{3.66}$$

where

$$a(\chi) \equiv \ln[n_\sigma(\chi) K(\chi)] + S(\chi) \ln\left(\frac{6}{\pi^2} [2S(\chi) - 1] \right) . \tag{3.67}$$

Inserting the $\bar{N} \geq 3$ slopes and intercepts from Table V, one finds

$$N_{\rm tot} \simeq 200\,000 \tag{3.68}$$

for the breakdown point. For the 68.26% confidence interval, this corresponds to a line-intercept knee position $N_{p,\rm LIK} \simeq 22\,000$ [Eq. (3.62)], to mean number of photons per pulse $\bar{N} \simeq 9$ [Eq. (3.63)], to effective slope $S_{\rm eff} \simeq 1.04$ [Eq. (3.65)], and to optimized confidence interval $\Delta\theta_{\rm MLE}(0.6826) \simeq 0.002°$ [Eq. (3.64)]. Notice that $\Delta\theta_{\rm MLE}(0.6826) N_{\rm tot} \simeq 7\,\text{rad}$, so that, although the effective slope is slightly greater than 1, the optimized confidence interval given the straight-line approximation does not beat the benchmark sensitivity even at the breakdown point.

One could try to pin down further the validity of the straight-line extrapolation by analyzing less widely separated confidence coefficients. Rather than study this question further using an extrapolation known to be only approximate, we address the optimization question for large $N_{\rm tot}$ directly by using a general knee-locating technique [5, 6] that is valid for all photon numbers.

3. *Analytic techniques: High $N_{\rm tot}$*

In the preceding subsection we derived an analytic approximation to the optimized confidence intervals as a function of the mean total number of photons $N_{\rm tot}$. This approximation was based on assuming a power-law preasymptotic behavior for the confidence intervals. Although it extended our results beyond the values of $N_{\rm tot}$ investigated by our numerical simulations, it was shown to be internally inconsistent for sufficiently high values of $N_{\rm tot}$. In this subsection we use an entirely different semianalytic approach, which is based on new results in estimation theory due to Braunstein [5]. Using this approach, we can estimate optimized confidence intervals without resorting to an assumed form for the preasymptotic behavior of the confidence interval. Instead, this approach uses certain fundamental properties of the fiducial SSW probability distribution, which may be encapsulated in expectation values of various products of derivatives.

Fisher's theorem [40] (see Sec. III B) tells us that, asymptotically in N_p, the ML estimator $\theta_{\rm MLE}$ has a Gaussian distribution with the CRLB standard deviation

$$\sigma_{\rm CRLB} = 1/\sqrt{N_p F} \,. \tag{3.69}$$

This is not enough information, however, to find optimal sensitivities because we do not know the knee position (optimal operating point) where this asymptotic behav-

ior becomes effective. The approach in this subsection allows us to locate the knee approximately from a calculation of the first $O(1/N_p)$ correction to Fisher's theorem.

The details of how this correction is calculated appear elsewhere [5], but the idea is to find asymptotic solutions of the likelihood equation

$$0 = \frac{\partial \mathcal{L}}{\partial \theta} = \frac{\partial}{\partial \theta} \sum_{j=1}^{N_p} \ln P(\phi_j|\theta) \tag{3.70}$$

[cf. Eq. (3.12)]. The solutions of this equation yield ML estimates. A Taylor expansion of this equation about the true value θ_0 of the parameter has the form

$$0 = \mathcal{L}'|_{\theta_0} + \delta\theta\, \mathcal{L}''|_{\theta_0} + \tfrac{1}{2}(\delta\theta)^2 \mathcal{L}'''|_{\theta_0} + \cdots , \tag{3.71}$$

where $\delta\theta = \theta - \theta_0$. This equation must be solved statistically. Fisher's theorem is obtained by treating $\ln P(\phi|\theta)$ as a statistical quantity and taking the limit $N_p \to \infty$; in this case the central-limit theorem can be applied directly to obtain

$$\sqrt{N_p F}\, \delta\theta \underset{N_p \to \infty}{\to} N(0,1) , \tag{3.72}$$

where $N(0,1)$ denotes a standard normal random variable with zero mean and unit variance.

Corrections to Fisher's theorem must be calculated in a less extreme limit of N_p. We must include corrections to the central-limit theorem for finite N_p and also extra terms in the Taylor expansion (3.71). When the parameter of interest is a so-called translation parameter—i.e., the underlying probability distribution $P(\phi|\theta) = P(\phi - \theta)$ is translationally invariant—and when the probability distribution leads to unbiased ML estimation, then the first correction to the variance of the ML estimator is

$$\mathrm{var}(\delta\theta) = \frac{1}{N_p F}\left[1 + \frac{N_{p,\mathrm{CK}}}{N_p} + O\left(\frac{1}{N_p^2}\right)\right] , \tag{3.73}$$

where

$$N_{p,\mathrm{CK}} = \left[\frac{2}{F^2}\int d\phi \left(\frac{(P'')^2}{P} - \frac{(P')^4}{3P^3}\right)\right] - 2 \tag{3.74}$$

is a "corrected" knee position. Equation (3.73) shows that $N_{p,\mathrm{CK}}$ describes the scale at which the asymptotic behavior of Fisher's theorem becomes effective. In other words, $N_p \simeq N_{p,\mathrm{CK}}$ is an approximate knee position for the transition from preasymptotic behavior to asymptotic CRLB behavior. Table VI gives the Fisher information F and this corrected knee position $N_{p,\mathrm{CK}}$ for various values of $\bar{N}$. Both rise roughly linearly with the photon-number cutoff M.

Near the knee $N_p \simeq N_{p,\mathrm{CK}}$, the variance of the ML estimator is bounded roughly by [6]

$$\frac{1}{N_{p,\mathrm{CK}} F} \lesssim \mathrm{var}(\delta\theta) \lesssim \frac{2}{N_{p,\mathrm{CK}} F} . \tag{3.75}$$

The left-hand side of this inequality is the CRLB, whereas the right-hand side comes from Eq. (3.73). Furthermore, for the multiple-measurement SSW scheme, we have seen that the position of the knee transition gives the optimal operating point and, hence, that the sensitivity at the knee gives the optimum sensitivity. We conclude that Eq. (3.75) bounds the optimum sensitivity that can be obtained by the SSW scheme. (In fact, we believe that this same argument holds for a broad class of multiple-measurement schemes [6], not just one that uses the SSW quantum state.)

We proceed by choosing a value of $\bar{N}$ and evaluating numerically the Fisher information (3.15) and the corrected knee position (3.74). For representative values of $\bar{N}$, the last three columns of Table VI list the corrected knee position $N_{p,\mathrm{CK}}$, the corresponding total photon number $N_{\mathrm{tot}} = \bar{N} N_{p,\mathrm{CK}}$, and the corresponding two-standard-deviation CRLB lower bound. These figures confirm again that the optimal division of resources uses many pulses with only a relatively few photons per pulse. Strictly speaking, the CRLB is a bound on the standard deviation, but the plots of 68.26% confidence intervals in Fig. 8 show that, for SSW measurements, it also acts as a lower bound on the corresponding confidence interval. Thus the two-standard-deviation CRLB in Table VI can also be viewed as an effective lower bound on the 68.26% confidence interval.

TABLE VI. For various values of the photon-number cutoff M (and corresponding mean number of photons per pulse $\bar{N}$), the Fisher information F, the knee position $N_{p,\mathrm{CK}}$ [Eq. (3.74)] determined from the corrections to the asymptotic CRLB behavior, the corresponding mean total number of photons $N_{\mathrm{tot}} = \bar{N} N_{p,\mathrm{CK}}$, and the two-standard-deviation CRLB (in deg) at the corrected knee position.

M	$\bar{N}$	F	$N_{p,\mathrm{CK}}$	N_{tot}	$2/\sqrt{N_{p,\mathrm{CK}} F}$ (deg)
4	0.560	2.064	6.07	3.400	32.38
12	1.024	7.722	11.79	12.08	12.01
30	1.496	22.44	24.18	36.18	4.919
46	1.733	36.42	34.79	60.29	3.219
74	2.004	61.83	52.92	106.0	2.003
399	3.000	379.3	253.0	759.0	0.369 9
2089	4.000	2127.3	1251.3	5005.3	0.070 24
10848	5.000	11497.4	6314.2	31571.1	0.013 45

To find the optimal operating point—i.e., the optimal division of resources between photons per pulse and number of pulses—for a preselected value of $N_{\rm tot}$ requires further numerical work to zero in on the particular value of $N_{\rm tot}$. We have carried out this additional work for $N_{\rm tot} = 60$; the results are summarized in Fig. 11, where they can be compared directly with results from the numerical simulations. The upper and lower bounds of Eq. (3.75) are indicated by a large "error bar," plotted at the appropriate knee position $N_{p,\rm CK}$. The knee position $N_{p,\rm CK}$ is slightly larger than the optimal operating point determined from the simulations, but the confidence-interval minimum found from the simulations lies within the range determined by the upper and lower bounds (3.75). Higher-order corrections to Fisher's theorem might be required to get a better estimate of the optimal operating point, but the present approach appears to be sufficient to find optimized confidence intervals with reasonable accuracy.

Figure 12 summarizes in a single log-log plot our results for optimized confidence intervals as a function of the mean total number of photons $N_{\rm tot}$. The optimized 68.26% confidence intervals found from our Monte Carlo simulations are plotted as dots, along with a best-fit (solid) straight line. The optimized 68.26% confidence intervals that follow from the straight-line extrapolation for the preasymptotic behavior [Eq. (3.64)] are plotted as a dashed line, and the upper and lower bounds on the 68.26% confidence intervals predicted by the Fisher-correction approach of this subsection [Eq. (3.75)] are plotted as dotted lines.

It is clear that the Fisher-correction approach yields reliable estimates of the optimized 68.26% confidence intervals. It reveals a steepening of the effective slope so that for $N_{\rm tot} \sim 30\,000$, the rate of convergence of the 68.26% confidence interval is given by

$$\Delta\theta \simeq 1/N_{\rm tot}^{0.92}\,, \quad N_{\rm tot} \sim 30\,000\,. \tag{3.76}$$

The straight-line extrapolation formula developed in Sec. III E 2 is also in good agreement with the numerical simulations and, moreover, with the results of the Fisher-correction approach well beyond the domain accessible to the simulations. For $N_{\rm tot} \gtrsim 5000$, however, the straight-line extrapolation results curve downward away from the results from the more objective Fisher-correction technique. Thus, for large $N_{\rm tot}$, the straight-line extrapolation yields too good a sensitivity and can only be used to set lower bounds on the sensitivity. This is consistent with the discussion in Sec. III E 2, which indicated that straight-line extrapolation of the preasymptotic behavior must break down for sufficiently high $N_{\rm tot}$.

The Fisher-correction technique of this subsection, based on the $O(1/N_p)$ correction to Fisher's theorem, is not limited to an analysis of the SSW scheme. It can be applied to multiple-measurement schemes based on other fiducial quantum states [6]. Thus the bounds of Eq. (3.75) might be important in deriving general sensitivity bounds on multiple-measurement schemes.

IV. DISCUSSION

The extraordinary $1/N_{\rm tot}^2$ sensitivity claims of Shapiro, Shepard, and Wong [3] were based on using reciprocal peak likelihood as the measure of sensitivity. In our analysis of a multiple-measurement version of SSW phase measurements, *we find no evidence for the* $1/N_{\rm tot}^2$ *sensitivity—or even for a sensitivity as good as the benchmark* $1/N_{\rm tot}$ *phase sensitivity of squeezed-state interferometry—and thus we conclude that reciprocal peak likelihood is not a good measure of phase sensitivity.* Moreover, by using the straight-line extrapolation approach of Sec. III E 2 to set a lower bound on phase sensitivity, *we conclude that for* $N_{\rm tot} \lesssim 200\,000$ *the multiple-measurement SSW scheme does not achieve a sensitivity as good as the benchmark* $1/N_{\rm tot}$ *sensitivity.*

Analysis of SSW measurements is more than just an empty exercise, however, most importantly because it emphasizes that *any investigation of ultimate quantum limits on measurements must include the possibility of dividing up the photon-number resources among many separate measurements.* Indeed, the sensitivity of multiple SSW measurements is, to us, surprisingly good, given the rather unlikely fiducial distribution, with its high central peak that encompasses an apparently negligible amount of probability.

One way to investigate further the relation of a high central peak to multiple-measurement sensitivity is to consider other SSW-style distributions. Recall that SSW did not find the state with maximum peak likelihood—they chose $r = 1$ as a representative value in Eq. (1.32)—so one might learn something by considering other values of r. For example, Hradil and Shapiro [48] have shown that in the limit $r \to 0$ the fiducial distribution converges in the mean to a uniform distribution, even though it still has a high central peak. Although this $r \to 0$ distribution would seem to be an almost laughable candidate for a good multiple-measurement sensitivity, one must be careful. It could be that the high central peak is still important for multiple measurements even though it gets drained of all probability. Indeed, Braunstein [6] has shown that for values of r as small as 0.01, the multiple-measurement sensitivity is as good as for $r = 1$, confirming that intuition about multiple measurements is sorely lacking.

Before getting too excited about multiple measurements, however, it is useful to indicate what is involved in achieving the optimal sensitivity for SSW measurements. To get an idea of what is going on, it is sufficient to use the straight-line extrapolation results of Sec. III E 2 in their crudest form, even though they are known to fail for sufficiently high $N_{\rm tot}$. Suppose one wants to measure a phase shift that is constant for a time τ and that one is willing to expend $N_{\rm tot} \gg 1$ photons on average, corresponding to average power $P = N_{\rm tot}/\tau$. To use the photon resources optimally, one can divide the available time into $N_p \sim N_{\rm tot}/\ln N_{\rm tot}$ pulses, each lasting a time $\tau_p = \tau/N_p$. One ends up making many measurements with a very wide bandwidth $B_p = 1/\tau_p = N_p/\tau \sim (N_{\rm tot}/\ln N_{\rm tot})/\tau$ (the same conclusion about bandwidth follows if one multiplexes in frequency instead of time).

For the SSW case, there is always the possibility of an enormous peak power

$$P_{\text{peak}} = M/\tau_p \sim N_p/\tau_p = N_p^2/\tau \sim P\,\frac{N_{\text{tot}}}{(\ln N_{\text{tot}})^2}\,, \qquad (4.1)$$

if there are M photons in a single pulse. Although the probability per pulse for this peak power is very small, the enormous number of pulses means that the probability for a peak power within a factor of two of P_{peak} is essentially unity.

As discussed by Hall [24], the best quantitative measure of measurement sensitivity seems to be confidence-interval width. Once one allows multiple measurements, however, evaluating confidence intervals requires choosing an estimator and then determining its statistics. Except for a few trivial probability distributions, it is not known in general what is the best estimator, although ML estimation is known to be asymptotically efficient. As this paper illustrates, however, determining the statistics of the ML estimator can be a difficult job. There must be a better way. What one would really like is a simple quantitative indicator of multiple-measurement sensitivity—for example, a single number that can be calculated from the fiducial probability distribution.

A key feature of the SSW distribution seems to be the narrow central peak. This peak is characterized in a useful and direct way for multiple measurements by the Fisher information, so maximizing the Fisher information, subject to a constraint on mean photon number, would seem to be a good idea. Unfortunately, one cannot find the ultimate sensitivity from the Fisher information alone, because one does not know how many measurements are required to reach the asymptotic behavior of Fisher's theorem. Finding out directly requires numerical simulations like those in this paper. If the distribution is spiky, the numerical work is likely to be as time consuming as the work reported here.

An equally important feature of the SSW distribution seems to be the runup to the central peak. We speculate that the form of this runup determines the preasymptotic behavior of the confidence intervals and thus the location of the knee transition to asymptotic Fisher behavior. Unfortunately, we have no clue how to relate the form of the runup to the corresponding form of the preasymptotic behavior.

On the question of finding the knee transition, however, the recent work of Braunstein [5, 6], starting from corrections to the asymptotic Fisher behavior, offers hope. The Fisher-correction technique reported in Sec. III E 3 allows one to locate the transition to asymptotic Fisher behavior at the corrected knee position $N_{p,\text{CK}}$ of Eq. (3.74) and thus (perhaps) to find optimum sensitivities like those in Eq. (3.75). This approach suggests maximizing the product $N_{p,\text{CK}}F$, subject to a constraint on the mean total number of photons $N_{\text{tot}} = \bar{N}N_{p,\text{CK}}$.

Another hopeful approach has been outlined by Jones [50]. He characterizes the sensitivity of multiple measurements in terms of the mutual information of communications theory, which incorporates multiple measurements naturally. In the language of phase measurements, the mutual information quantifies the amount of information that the measured phase data provides about the phase shift. Jones [51] has recently applied this approach to SSW phase measurements.

At present we cannot formulate in a simple way the question of ultimate phase sensitivity, because the question becomes entangled in difficult issues of classical estimation theory for multiple measurements. Nonetheless, our work shows clearly that there are surprises lurking in multiple-measurement schemes: on account of the large tails in the fiducial SSW probability distribution, we did not expect the SSW scheme to achieve even the $1/\sqrt{N}$ limit for phase measurements using coherent states, but it beats that limit resoundingly, if the photons are divided up carefully among many measurements. Any investigation of the ultimate phase sensitivity must allow for the "divide-and-conquer" strategy that divides up the available photons among many measurements of the same phase datum.

ACKNOWLEDGMENTS

This work was supported in part by the Office of Naval Research (Contract No. N00014-88-K-0042 and Grant No. N00014-91-J-1167). S.L.B. appreciated the support of the Lady Davis Fellowship Trust in writing up this work, and A.S.L. thanks S. M. Barnett and S. M. Tan for helpful discussions.

* Present address: Physics Department, Technion-Israel Institute of Technology, 32000 Haifa, Israel.
† Present address: Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM 87131-1156.

[1] D. T. Pegg and S. M. Barnett, Phys. Rev. A **39**, 1665 (1989).
[2] J. H. Shapiro and S. R. Shepard, Phys. Rev. A **43**, 3795 (1991) (cited in text as SS); see also J. H. Shapiro, in *The Workshop on Squeezed States and Uncertainty Relations*, edited by D. Han, Y. S. Kim, and W. W. Zachary, NASA Conference Publication 3135 (NASA, Washington, DC, 1992), p. 107.
[3] J. H. Shapiro, S. R. Shepard, and N. W. Wong, Phys. Rev. Lett. **62**, 2377 (1989) (cited in text as SSW).
[4] S. L. Braunstein, A. S. Lane, and C. M. Caves, Phys. Rev. Lett. **69**, 2153 (1992).
[5] S. L. Braunstein, J. Phys. A **25**, 3813 (1992).
[6] S. L. Braunstein, Phys. Rev. Lett. **69**, 3598 (1992).
[7] J. H. Shapiro and S. S. Wagner, IEEE J. Quantum Electron. **QE-20**, 803 (1984).
[8] Z. Hradil, Phys. Lett. A **146**, 1 (1990).
[9] J. W. Noh, A. Fougères, and L. Mandel, Phys. Rev. Lett. **67**, 1426 (1991).
[10] J. W. Noh, A. Fougères, and L. Mandel, Phys. Rev. A **45**, 424 (1992).
[11] M. Freyberger and W. Schleich, Phys. Rev. A (to be published).

[12] C. M. Caves, Phys. Rev. D **23**, 1693 (1981).
[13] R. S. Bondurant and J. H. Shapiro, Phys. Rev. D **30**, 2548 (1984).
[14] B. Yurke, S. L. McCall, and J. R. Klauder, Phys. Rev. A **33**, 4033 (1986).
[15] L. Susskind and J. Glogower, Physics (Long Island City, N.Y.) **1**, 49 (1964).
[16] P. Carruthers and M. M. Nieto, Rev. Mod. Phys. **40**, 411 (1968).
[17] J.-M. Lévy-Leblond, Ann. Phys. (N.Y.) **101**, 319 (1976).
[18] J. Bergou and B.-G. Englert, Ann. Phys. (N.Y.) **209**, 479 (1991).
[19] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
[20] D. T. Pegg and S. M. Barnett, Europhys. Lett. **6**, 483 (1988).
[21] S. M. Barnett and D. T. Pegg, J. Mod. Opt. **36**, 7 (1989).
[22] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic, New York, 1976), especially pp. 243–248.
[23] M. J. W. Hall, Quantum Opt. **3**, 7 (1991).
[24] M. J. W. Hall, J. Mod. Opt. (to be published).
[25] R. Loudon, *The Quantum Theory of Light*, 1st ed. (Oxford University Press, Oxford, 1973), especially pp. 143 and 144.
[26] K. Kraus, *States, Effects, and Operations: Fundamental Notions of Quantum Theory* (Springer, Berlin, 1983).
[27] E. B. Davies, *Quantum Theory of Open Systems* (Academic, London, 1976).
[28] J. A. Vaccaro and D. T. Pegg, J. Mod. Opt. **37**, 17 (1990).
[29] J. A. Vaccaro and D. T. Pegg, Opt. Commun. **70**, 529 (1989).
[30] S. M. Barnett, S. Stenholm, and D. T. Pegg, Opt. Commun. **73**, 314 (1989).
[31] S. M. Barnett and D. T. Pegg, Phys. Rev. A **41**, 3427 (1990).
[32] J. A. Vaccaro and D. T. Pegg, Phys. Rev. A **41**, 5156 (1990).
[33] G. S. Summy and D. T. Pegg, Opt. Commun. **77**, 75 (1990).
[34] S. M. Barnett and D. T. Pegg, Phys. Rev. A **42**, 6713 (1990).
[35] D. T. Pegg, J. A. Vaccaro, and S. M. Barnett, J. Mod. Opt. **37**, 1703 (1990).
[36] V. P. Popov and V. S. Yarunin, Vestn. Leningr. Univ. **22**, 7 (1973).
[37] In Refs. [20] and [1] Pegg and Barnett investigate conditions for the validity of the limit $s \to \infty$. They find that the limit does exist for what they called *physical* states—states for which all moments of $\hat{N}$ are finite.
[38] A. Bandilla, H. Paul, and H. H. Ritze, Quantum Opt. **3**, 267 (1991).
[39] M. J. W. Hall and I. Fuss, Quantum Opt. **3**, 147 (1991).
[40] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, NJ, 1946), especially pp. 500–504.
[41] W. P. Schleich, J. P. Dowling, and R. J. Horowicz, Phys. Rev. A **44**, 3365 (1991).
[42] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 4th ed. (Academic, New York, 1965).
[43] J. P. Dowling, Opt. Commun. **86**, 119 (1991).
[44] *Handbook of Mathematical Functions*, Natl. Bur. Stand. Appl. Math. Ser. No. 55, edited by M. Abramowitz and I. A. Stegun (U.S. GPO, Washington, DC, 1964).
[45] L. Cohen, Found. Phys. **20**, 1455 (1990).
[46] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, Cambridge, 1989).
[47] MATHEMATICA is a trade mark of Wolfram Research, Inc.
[48] Z. Hradil and J. H. Shapiro, Quantum Opt. **4**, 31 (1992).
[49] See, for example, R. E. Walpole, *Introduction to Statistics* (MacMillan, New York, 1974), Chap. 10.
[50] K. R. W. Jones, Ann. Phys. (N.Y.) **207**, 140 (1991).
[51] K. R. W. Jones, Phys. Scr. (to be published).