Quantum Limits on Precision Measurement of Phase

Samuel L. Braunstein
Department of Physics, Technion–Israel Institute of Technology, 32 000 Haifa, Israel
(Received 27 May 1992)

We investigate the quantum mechanical bound to how precisely we can determine a phase shift given only a constraint on the mean total number of photons available. By considering how quickly one can gain information from data analysis, we derive the sensitivity achievable (in principle) for measurements involving even highly non-Gaussian noise. Using these results we calculate the sensitivity of several recent proposals for precision phase measurement, and show that no proposal to date beats the sensitivity believed achievable by squeezed state interferometry.

PACS numbers: 03.65.Bz, 06.20.–f, 42.50.Dv, 89.70.+c

Small changes in the phase of light, detected interferometrically, have served to probe the fundamental laws of physics for over a hundred years [1, 2]. However, only recently have we begun to formulate precisely and to investigate the question of how quantum mechanics limits measurements of phase shifts. Detailed consideration of this issue has been delayed because of the lack of a consensus over how phase measurements should be treated within quantum theory.

Classically, phase is just a parameter of angular location, on phase space, of the Liouville distribution of a single-mode harmonic oscillator. Quantum mechanically the idealized description of the phase shift Φ induced in light by a *linear* dispersive optical element is given by $|\psi_\Phi\rangle = \exp(-i\Phi\hat{n})|\psi\rangle$, where $\hat{n}$ is the photon number operator, and $|\psi\rangle$ is some fiducial state of the light. That is, quantum mechanically a shift in phase is still a c-number parameter (describing a material property of our optical element at a given frequency). For many years quantum theoretical approaches [3–9] dealing with phase measurements have concentrated on the paradigm that to each classical variable there is a corresponding *natural* quantum observable. However, since Φ is a parameter it may be determined via data analysis from *any* suitable measurement having a parametric dependence on it—this admits a much broader class of measurements for the precision determination of phase shifts than does the more dogmatic traditional approach.

The simplest example of this is standard interferometry: Measurements are performed by interfering the shifted state with one (from the second interferometer arm) having a fixed phase relationship to the fiducial state, and then counting photons. The interference ensures that the measured photocount distribution $p(n|\Phi)$ has a parametric dependence on the phase shift. Mathematically, we may describe the phase as a location parameter of one state relative to another along curves in Hilbert space [10] generated by $-i\hat{n}$. The parametric dependence of the outcomes $p(\xi|\Phi)$ is affected both by which curve we access through our choice of fiducial state and by the observable $\hat{\xi}$ used. How we make these choices is part of the overall experimental design—determining the sensitivity achievable.

This description of Φ as labeling curves in Hilbert space yields several formal advantages. It forms a very general language [10], within standard quantum theory, in which to describe *all* types of precision measurements including those like time [11, 12] and phase where the usual operator approach is problematic. Further, it pinpoints the importance of inference (i.e., "classical" data analysis) as a fundamental part of precision measurements within quantum theory. We now turn to what must go into an analysis of the ultimate quantum limits to measuring phase.

Without constraints on our resources there would be no limit to the precision with which we could determine a phase shift. The least restrictive constraint we can have is one on the total average energy available, thus we will assume that there is some limit on the mean (as both an ensemble and a temporal average) total number of photons available $N_{\rm tot}$. If this energy were derived from a laser with mean power P (at frequency Ω), and if the phase shift were only stable over the time τ, we would have $N_{\rm tot} \simeq P\tau/\hbar\Omega$ ($2\pi\hbar$ is Planck's constant). Any attack on the question of ultimate efficiency must allow for schemes that split the total energy into smaller packets and for the subsequent data analysis of the multiple measurements. For "symmetric" schemes, where the N packets are identical quantum states with mean photon number $\langle\hat{n}\rangle$, this split must satisfy

$$N_{\rm tot} = \langle\hat{n}\rangle N \, .$$

We investigate such schemes where all of the measurements are of the same variable.

How well can data be analyzed? The Cramér-Rao lower bound (CRLB) [13] gives a strict lower bound for the standard deviation of *any* estimate of Φ [14], based on N independently but identically distributed observations of some variable ξ, as

$$\Delta\Phi \geq 1/\sqrt{NF} \, ,$$

when the Fisher information $F \equiv \int d\xi \, [p(\xi|\Phi)']^2/p(\xi|\Phi)$ exists (throughout, primes denote the derivative $\partial/\partial\Phi$).

For Gaussian noise this bound is easily achieved for all

N by using the mean of the data as the estimator, so the sensitivity after N measurements improves as $1/\sqrt{N}$ over the first. Coherent and squeezed state interferometry have essentially Gaussian noise in the limit that intensity fluctuations can be approximated by a quadrature phase amplitude. For coherent state interferometry, where a single measurement has sensitivity $1/2\sqrt{\langle\hat{n}\rangle}$, N measurements will give

$$\Delta\Phi = \frac{1}{2\sqrt{\langle\hat{n}\rangle N}} = \frac{1}{2N_{\rm tot}^{0.5}} ,$$

which is independent of the split in our resources; this is the shot noise limit. Squeezed state interferometry is believed [15] to be able to achieve a single measurement sensitivity scaling as $1/\langle\hat{n}\rangle$. The multiple measurement sensitivity would then be $\Delta\Phi \simeq 1/\langle\hat{n}\rangle\sqrt{N}$, which for fixed $N_{\rm tot}$ is optimized by using as few pulses as possible. Ideally for $N = 1$ we attain

$$\Delta\Phi \simeq 1/N_{\rm tot} ,$$

with a multiplicative constant of $O(1)$; we ignore such constants in the comparison of very different schemes, and ask simply for the scaling law.

For non-Gaussian noise the CRLB cannot be achieved, in general, for any method of data analysis. Thus, we must concentrate on specific methods. We use here the method called maximum likelihood (ML) estimation; other methods are unlikely to be significantly more efficient. For the N data points $\xi_1, \xi_2, \ldots, \xi_N$, each distributed via $p(\xi|\Phi)$, Bayes' theorem says that the "likelihood" for Φ will be

$$p(\Phi|\xi_1,\xi_2,\ldots,\xi_N) \propto \prod_{i=1}^{N} p(\xi_i|\Phi) ,$$

if we have no initial prejudice. The ML estimate is the value of Φ that maximizes this likelihood function, and typical measures of our confidence in this value are the standard deviation $\Delta\Phi$ or the 68% confidence interval $(\Delta\Phi)_{68\%}$. Further, Fisher's theorem [13] tells us that as $N \to \infty$ the likelihood function approaches a Gaussian distribution with standard deviation $\Delta\Phi \to 1/\sqrt{NF}$. Thus ML estimation achieves the CRLB for large enough data sets, and so is asymptotically efficient.

Figure 1 gives a schematic picture of the convergence of ML estimation for non-Gaussian noise based on the CRLB (dashed line) and Fisher's theorem. The schematic convergence of the ML estimator (solid line) shows two regimes: asymptotic behavior matching the CRLB and a preasymptotic fast convergence. Where should the split in our resources be made based on this picture? Let us parametrize the split by s with $N = sN_{\rm tot}$ and $\langle\hat{n}\rangle = 1/s$.

Asymptotic regime.—If the Fisher information scales as $F \simeq \langle\hat{n}\rangle^f$, then the asymptotic ML sensitivity will scale as

$$\Delta\Phi \simeq s^{(f-1)/2}/\sqrt{N_{\rm tot}} .$$

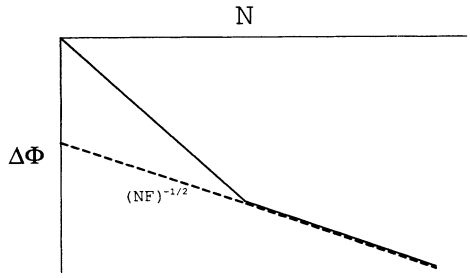

FIG. 1. Log-log plot of $\Delta\Phi$ vs N showing (solid line) the schematic convergence of the ML estimator with increasing sample size for *fixed* $\langle\hat{n}\rangle$ (and hence fixed F). The location of the "knee," where the fast preasymptotic convergence of non-Gaussian noise turns into the asymptotic behavior approaching $1/\sqrt{NF}$ of the CRLB (dashed line), gives the optimal split in $N_{\rm tot} = \langle\hat{n}\rangle N$ for multiple measurement schemes.

If (i) $f > 1$, then the optimal split (inside this regime) will be at the "knee" between the two regimes; (ii) $f = 1$, then the sensitivity is independent of the split so we may choose the knee; and (iii) $f < 1$, then this scheme is worse than coherent state interferometry so we ignore it.

Preasymptotic regime.—If we suppose that the sensitivity here scales as

$$\Delta\Phi \simeq \frac{1}{\langle\hat{n}\rangle^a N^b} = \frac{1}{s^{b-a}N_{\rm tot}^a} ,$$

then (i) for $b \geq a$, the knee again is optimal; and (ii) for $b < a$, the optimal split is at $N = 1$, in which case the multiple measurement strategy fails to be interesting.

These scaling arguments suggest that, as long as we beat shot noise, either the optimal split occurs at the knee or the multiple measurement strategy should be abandoned. It is possible that the preasymptotic ML convergence has a more complicated behavior, perhaps with multiple knees, but we only expect this to occur if the probability distribution for observations $p(\xi|\Phi)$ has intermediate scales (i.e., scales other than those which describe the knee location).

Braunstein [16] has calculated the first $O(1/N)$ correction to Fisher's theorem as

$$(\Delta\Phi)^2 = \frac{1}{NF}\left[1 + \frac{N_{\rm var}}{N} + O\left(\frac{1}{N^2}\right)\right] , \qquad (1)$$

where the quantity $N_{\rm var}$ is a functional of the distribution of observations $p(\xi|\Phi)$. We see that the knee between the asymptotic CRLB behavior and the preasymptotic fast convergence will occur at $N \simeq N_{\rm var}$. Thus, the optimum multiple measurement sensitivity will occur at $N \simeq N_{\rm var}$ with

$$\frac{1}{N_{\rm var}F} \lesssim (\Delta\Phi)^2 \lesssim \frac{2}{N_{\rm var}F} , \qquad (2)$$

unless it is beaten by a single measurement using all the resources, or by a split at some intermediate knee in the preasymptotic ML convergence. In any case this range of

sensitivities is attainable even if it fails to be optimum.

Our formulation of phase measurement is not prejudiced, so if we learned how to make ideal "phase" measurements then we could apply the above multiple measurement strategy to them as well. In this case the distribution of observations is expected to be given by the Susskind-Glogower (SG) [5] phase distribution

$$p(\phi|\Phi) = p(\phi - \Phi) = \frac{1}{2\pi}\left|\sum_{m=0}^{\infty} e^{-im\phi}\langle m|\psi_\Phi\rangle\right|^2 ,$$

expanded here in the number state basis, and $|\psi_\Phi\rangle$ is the state measured. In this case, since Φ is a "translation" parameter, the knee location simplifies [16] to

$$N_{\rm var} = \frac{2}{F^2}\int d\phi \left(\frac{p''^2}{p} - \frac{p'^4}{3p^3}\right) - 2 , \qquad (3)$$

for unbiased estimation. Further, we see that for *single* measurements the ML uncertainty equals the ideal phase measurement uncertainty, i.e., $\Delta\Phi = \Delta\phi$. Thus, we can make comparisons with *single* measurement schemes [17, 18] which have obtained $\Delta\phi \simeq 1/N_{\rm tot}$; indeed, the multiplicative constants appear to beat squeezed state interferometry, but may not in practice if there are hidden costs to ideal phase measurements.

Several recent proposals [19–21] based on ideal phase measurements have claimed a sensitivity for measuring phase, using the multiple measurement strategy, scaling as

$$1/N_{\rm tot}^2 ,$$

in terms of the available resources (with this scaling we could pay almost any hidden cost): (i) Shapiro, Shepard, and Wong (SSW) [19] obtained the family of states

$$|\psi_\Phi\rangle \propto \sum_{m=0}^{M} \frac{e^{im\Phi}}{m+r}|m\rangle ,$$

where M is a cutoff, and they somewhat arbitrarily fixed $r = 1$; (ii) Shapiro and Shepard (SS) [20] modified the original SSW proposal by allowing $r > 0$ (we restrict our discussion to $0 < r < 1$); and (iii) Dowling [21] modified SSW differently by introducing a smooth cutoff μ,

$$|\psi_\Phi\rangle \propto \sum_{m=0}^{\infty} \frac{e^{(i\Phi - 1/\mu)m}}{m+1}|m\rangle .$$

The SG distributions for each of these proposals have similar features: A flat distribution (from $-\pi$ to π) gives way to a scale free logarithmic singularity as ϕ approaches Φ, and at a scale determined by the cutoff this "singularity" is softened into a rounded cap. There are therefore only two scales appearing in each of these distributions. The first determines when the deviation from a flat distribution can be seen; however, the knee corresponding to *this* scale will occur with sensitivity $\Delta\Phi \simeq 1$ since the behavior is dominated by the random ML estimation of a uniform distribution. The cutoff gives the second scale, and sets the size of $N_{\rm var}$ and F by describing the finest structures that appear in the SG distributions. From our earlier scaling arguments, therefore, we conclude that the optimum multiple measurement sensitivity for these proposals will occur at the knee $N \simeq N_{\rm var}$.

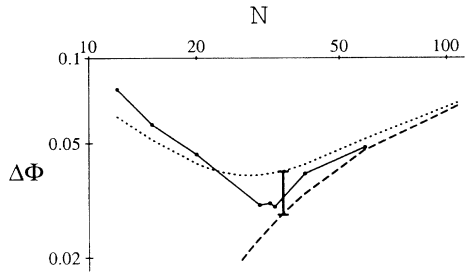

FIG. 2. Log-log plot of $\Delta\Phi$ vs N for *fixed* $N_{\rm tot} = 60$ for the SSW proposal. The connected dots show Monte Carlo calculations of $(\Delta\Phi)_{68\%}$ (from Refs. [22, 23]). The dashed and dotted curves show the CRLB behavior and the $O(1/N)$ correction in Eq. (1), respectively. The giant "error bar" between these theoretical curves shows the range $\left[1/\sqrt{N_{\rm var}F}, \sqrt{2}/\sqrt{N_{\rm var}F}\right]$ predicted from the knee location $N \simeq N_{\rm var}$.

Figures 2 and 3 compare $\Delta\Phi$ of the CRLB (dashed line) and the $O(1/N)$ correction of Eq. (1) (dotted line) with the $(\Delta\Phi)_{68\%}$ calculated from Monte Carlo simulation (joined dots) [22,23] for the SSW proposal. Near the Gaussian asymptotic regime the standard deviation and 68% confidence intervals are almost equal, so this comparison is sensible. Figure 2 makes this comparison for fixed $N_{\rm tot} = 60$; the giant "error bar" plotted between the dotted and dashed curves at $N = N_{\rm var}$ shows the range predicted by Eq. (2). Figure 3 shows all the optimum $(\Delta\Phi)_{68\%}$ calculated in the Monte Carlo simulations, and the ranges (dashed to dotted lines) as predicted by Eq. (2). These results give us confidence in both the scaling arguments and which scales are relevant.

Figure 4 shows a log-log plot of $\Delta\Phi = 1/\sqrt{N_{\rm var}F}$ vs $N_{\rm tot}$, which is the lower "bound" in Eq. (2) for all three

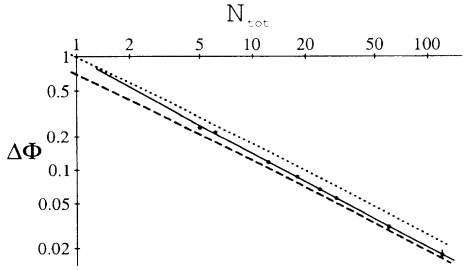

FIG. 3. Log-log plot of $\Delta\Phi$ vs $N_{\rm tot}$ for the optimal split in $N_{\rm tot} = \langle\hat{n}\rangle N$ for the SSW proposal. The solid line is the best fit through the optimum $(\Delta\Phi)_{68\%}$ in Monte Carlo simulations (from Refs. [22, 23]). The dashed and dotted curves show the lower and upper "bounds" for the range in Eq. (2).

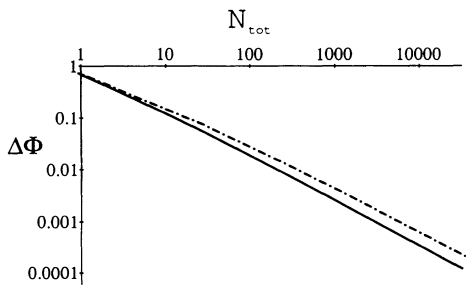

FIG. 4. Log-log plot of $\Delta\Phi = 1/\sqrt{N_{\rm var}F}$ vs $N_{\rm tot}$, showing the convergence of ML estimation for the optimal resource split as predicted by the knee location $N_{\rm var}$. The solid line is the SSW proposal ($r = 1$), and the SS proposal for $r = 0.1$ and 0.01; the dot-dashed line is that for the Dowling proposal.

proposals: The SSW (i.e., $r = 1$) and SS proposals with $r = 0.1$ and 0.01 are virtually indistinguishable (solid line); and the Dowling proposal (dot-dashed line). At $N_{\rm tot} \simeq 30\,000$ each of these proposals has a rate of convergence which scales as

$$\Delta\Phi \simeq 1/N_{\rm tot}^{0.9} ,$$

a sensitivity much smaller than the original prediction of $1/N_{\rm tot}^2$. These results are particularly surprising for the small r values of the SS proposal. In the limit $r \to 0$ the corresponding SG distribution is flat, and so it has been argued [24] that multiple measurement schemes are completely insensitive in this limit. This argument relies on the uniformity of this limit; however, as Fig. 4 shows there is *no* degradation in the performance of ML estimation over the range $r = 1$ to 0.01.

Phase is just a c-number parameter, describing the location of a state on a curve in Hilbert space. Any attack on the question of whether there is a quantum limit to how efficiently we can measure phase shifts must allow for schemes that split the total energy available into smaller packets and for the subsequent data analysis of the multiple measurements. For symmetric schemes we have derived the sensitivity of maximum likelihood estimation near the optimal energy split. We considered several such schemes based on the "quantum theory of phase measurement" for states having highly non-Gaussian noise. For $N_{\rm tot} \lesssim 30\,000$ these schemes do *not* beat the single measurement sensitivity of squeezed state interferometry.

The Lady Davis Fellowship Trust funded this work, and the author thanks C. M. Caves for many discussions.

[1] A. A. Michelson and E. W. Morley, Am. J. Sci. **34**, 333 (1887).
[2] A. Abramovici *et al.*, Science **256**, 325 (1992).
[3] P. A. M. Dirac, Proc. R. Soc. London A **114**, 243 (1927).
[4] W. Heitler, *The Quantum Theory of Radiation* (Oxford, Oxford, 1954), p. 65.
[5] L. Susskind and J. Glogower, Physics (Long Island City, N.Y.) **1**, 49 (1964).
[6] P. Carruthers and M. M. Nieto, Rev. Mod. Phys. **40**, 411 (1968).
[7] J.-M. Lévy-Leblond, Ann. Phys. (N.Y.) **101**, 319 (1976).
[8] S. M. Barnett and D. T. Pegg, J. Mod. Opt. **36**, 7 (1989).
[9] D. T. Pegg and S. M. Barnett, Phys. Rev. A **39**, 1665 (1989).
[10] J. Samuel and R. Bhandari, Phys. Rev. Lett. **60**, 2339 (1988).
[11] A. Peres, Am. J. Phys. **48**, 552 (1980).
[12] J. Anandan and Y. Aharonov, Phys. Rev. Lett. **65**, 1697 (1990).
[13] H. Cramér, *Mathematical Methods of Statistics* (Princeton Univ. Press, Princeton, NJ, 1946), pp. 500–504.
[14] Measures of uncertainty for angular coordinates may have trouble with the 2π periodicity for broad distributions; however, for precision measurements, where typically $\Delta\Phi \ll 1$, we can circumvent this by rotating the coordinate edges out of the distribution's "bulk."
[15] B. Yurke, S. L. McCall, and J. R. Klauder, Phys. Rev. A **33**, 4033 (1986).
[16] S. L. Braunstein, J. Phys. A **25**, 3813 (1992).
[17] G. S. Summy and D. T. Pegg, Opt. Commun. **77**, 75 (1990).
[18] M. J. W. Hall and I. Fuss, Quantum Opt. **3**, 147 (1991).
[19] J. H. Shapiro, S. R. Shepard, and N. W. Wong, Phys. Rev. Lett. **62**, 2377 (1989).
[20] J. H. Shapiro and S. R. Shepard, Phys. Rev. A **43**, 3795 (1991).
[21] J. P. Dowling, Opt. Commun. **86**, 119 (1991).
[22] S. L. Braunstein, A. S. Lane, and C. M. Caves, Phys. Rev. Lett. **69**, 2153 (1992).
[23] A. S. Lane, S. L. Braunstein, and C. M. Caves, Phys. Rev. A (to be published).
[24] Z. Hradil and J. H. Shapiro, Quantum Opt. **4**, 31 (1992).