Adaptive single-shot phase measurements: A semiclassical approach

H. M. Wiseman[1,2,*] and R. B. Killip[3]

[1]*Department of Physics, The University of Queensland, St. Lucia 4072, Australia*

[2]*Department of Physics, The University of Auckland, Auckland, New Zealand*

[3]*Department of Mathematics, The University of Auckland, Auckland, New Zealand*

(Received 21 January 1997)

The standard single-shot estimate for the phase of a single-mode pulse of light is the argument of the complex amplitude of the field. This complex amplitude can be measured by heterodyne detection, in which the local oscillator is detuned from the system so that all quadratures are sampled equally. Because different quadratures do not commute, such a measurement introduces noise into the phase estimate, with a variance scaling as N^{-1}, where N is the maximum photon number. This represents the shot-noise limit or standard quantum limit (SQL). Recently, one of us [H.M. Wiseman, Phys. Rev. Lett. **75**, 4587 (1995)] proposed a way to improve upon this: a real-time feedback loop can control the local oscillator phase to be equal to the estimated system phase plus $\pi/2$, so that the phase quadrature of the system is measured preferentially. The phase estimate used in the feedback loop at time t is a functional of the photocurrent from time 0 up to time t in the single-shot measurement. In this paper we consider a very simple feedback scheme involving only linear electronic elements. Approaching the problem from semiclassical detection theory, we obtain analytical results for asymptotically large photon numbers. Specifically, we are able to show that the noise introduced by the measurement has a variance scaling as $N^{-3/2}$. This is much less than the SQL variance, but still much greater than the minimum intrinsic phase variance which scales as N^{-2}. We briefly discuss the effect of detector inefficiencies and delays in the feedback loop. [S1050-2947(97)06407-X]

PACS number(s): 42.50.Dv, 42.50.Lc

I. INTRODUCTION

The division of radio broadcasting into amplitude modulation and frequency modulation channels neatly illustrates the two simplest ways to encode information into an electromagnetic field: in the intensity or in the phase. The ultimate limits to the channel capacity for any form of communication is set by quantum mechanics, and communication via the electromagnetic field is no exception to this. Of course radio broadcasting operates nowhere near any quantum limit, but in the future it may be useful to push technology to the limit where every bit of intensity, or phase information, in every available mode of the field is used.

Communication near this ultimate quantum limit (UQL) would require not only the ability to engineer states which code information with minimum error (that is, states with very well-defined intensity or phase), but also the ability to estimate the intensity or phase accurately from a *single* measurement. In communication the single-shot requirement is a matter of optimization rather than absolute necessity, since a certain amount of redundancy could be built into the communications system, by sending every pulse twice, for example. However, there are other possible applications in which it would be necessary, such as the precision measurements of weak signals. If such signals had an extraterrestrial origin then they would be essentially nonrepeatable, and so it would be desirable to make a measurement which is as good as possible on each pulse.

If the relevant information were encoded in the intensity of the pulse, then the desired measurement could be performed simply using a photodetector. In principle this can precisely determine the number of photons in the pulse. The chief practical limitation is from detector inefficiencies, which are now quite small. However, phase is a different matter. The ideal form of phase measurement is known as a canonical phase measurement [1]. Unfortunately, there is no known way to realize such a measurement and neither is there ever likely to be [2]. There are, nevertheless, imperfect techniques for phase measurements which can be realized. Heterodyne detection is one example [1]. These techniques are imperfect because they introduce excess noise into the measurement result. The variance of this introduced noise scales inversely with the photon number. This is the characteristic scaling of the shot-noise limit or standard quantum limit (SQL) of phase noise. This excess variance is far above the intrinsic phase variance (the UQL) for a state which is optimized to have a minimum phase variance. Thus a UQL communication system based on phase would not be possible with standard phase measurement schemes.

There is one case in which the SQL for phase measurements can be simply surpassed; that is if one assumes that before starting the measurement one already knows the phase of the system to be φ within a small uncertainty $\delta\varphi \ll 1$. In this case one can obtain an estimate of the system phase in which the excess variance is much smaller than that of the SQL by using homodyne detection. Homodyne detection involves passing the mode to be measured through a 50/50 beam splitter, in order to combine it with an intense field (treated classically) called the local oscillator. By choosing the local oscillator phase Φ to be equal to $\varphi+\pi/2$, the difference photocurrent from the two output ports of the beam splitter yields a measurement of the phase quadrature $X_\Phi = ae^{-i\Phi}+a^\dagger e^{i\Phi}$, where a is the annihilation operator of

*Electronic address: wiseman@physics.uq.edu.au

 © 1997 The American Physical Society

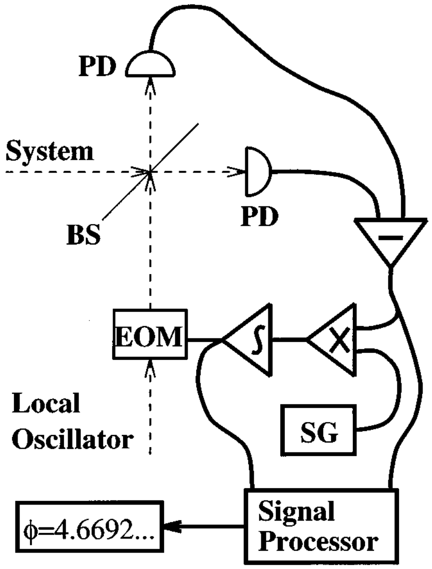

FIG. 1. Diagram for the experimental apparatus for making an adaptive phase measurement. Thin dashed lines indicate light rays and the thin continuous line labeled BS represents a 50/50 beam splitter. Medium lines represent electro-optic devices: photodetectors (PD) and an electro-optic phase modulator (EOM). Thick lines represent electrical components: a subtractor, a multiplier, an integrator, a signal generator (SG), a signal processor, and a digital read out giving the measured value of $\phi\in[0,2\pi)$. The necessity for these particular electrical elements alone is a consequence of the feedback algorithm explained in Sec. II.

the mode to be measured. Also assuming also that the amplitude r of the field is reasonably well defined (with $r\gg 1,\delta r$) and is also known prior to the measurement, then the phase of the field is very well approximated by

$$\phi\approx\varphi+\frac{ae^{-i\Phi}+a^{\dagger}e^{i\Phi}}{2r}=\varphi+\frac{ae^{-i\varphi}-a^{\dagger}e^{i\varphi}}{2ri}. \qquad (1.1)$$

However, by making these assumptions, one is really removing the problem from the realms of phase measurements. A true phase measurement ideally should not rely on any prior knowledge of the amplitude of the field, and certainly should not rely upon any prior knowledge of that phase. The heterodyne measurements mentioned above are true phase measurements in this sense, although they are not particularly accurate measurements.

While not being a true phase measurement, the phase quadrature measurement by homodyne detection suggests how it may be possible to construct a true phase measurement, which should be superior in accuracy to a heterodyne measurement (and so surpass the SQL). Rather than measuring a quadrature of predefined phase, the phase of the local oscillator could be adjusted during the course of the measurement to measure the *estimated* phase quadrature of the system by homodyne detection. Here the estimated phase of the system would have to be inferred from the photocurrent record *so far* from the *single* pulse. That is to say, the local oscillator phase would be continuously adjusted by a feedback loop to be in a quadrature with the estimated system phase over the course of a single measurement (see Fig. 1). This novel idea of *adaptive* single-shot measurements was proposed recently by one of us [5], but it turns out that it has been proposed, in a different context, at least once before [6]. It is a sort of quantum feedback which is quite different from that investigated previously (see Ref. [7] for a review) in that it has no effect on the evolution of the system.

In Ref. [5] it was shown numerically that an adaptive phase measurement does surpass the SQL for states with large photon numbers. However, it is very difficult with numerical results to determine just how much better an adaptive measurement is, because the system size must be extremely large to obtain valid scaling laws. An analytical result was obtained in Ref. [5], but it pertained to a system which contained at most one photon. In this case, the adaptive technique is as good as a canonical phase measurement, whereas the standard technique is definitely inferior. This result for a single-photon field is obviously of little practical use for communication.

In this work we are concerned with obtaining analytical results for adaptive phase measurements of fields with large intensities. The approach adopted is to consider measurements on coherent states with large coherent amplitudes. This yields asymptotic results which can easily be generalized to states other than coherent states. As well as being analytically tractable, the approach using coherent states is much simpler conceptually than that used in Ref. [5]. That is because a coherent field can be treated semiclassically: the field itself evolves deterministically and the noise in the measurement is due to the shot noise of the photoelectric effect: a constant classical driving field causes the ejection of a photoelectron at Poisson-distributed times. Thus the results of Sec. II of this paper may be understood using only semiclassical concepts. The results of Sec. III, which generalize those of Sec. II for phase-optimized states, require a small knowledge of quantum estimation theory which is summarized in Sec. III. Section IV concludes with a discussion of experimental practicality.

II. SEMICLASSICAL PHASE ESTIMATION

A. Semiclassical photodetection theory

We wish to consider phase measurements of a single-mode pulse of the electromagnetic field. Let us consider this pulse to be close to a plane wave with a transverse area $\mathcal{A}$ and let us fix our spatial position to be that of a presumed perfect photodetector covering the area $\mathcal{A}$. Then, since we are using a semiclassical argument, this pulse will produce at the detector position a classical time-varying electric field of the form

$$E(t)=\left(\frac{2\hbar\omega u(t)}{\epsilon_0\mathcal{A}c}\right)^{1/2}\mathrm{Re}[\alpha e^{-i\omega t}], \qquad (2.1)$$

where α is a complex number and $u(t)$ is a real and positive mode function which is normalized so that

$$\int_0^T u(t)dt=1, \qquad (2.2)$$

where T is some total time which is necessarily much greater than ω^{-1}, so that the pulse can be essentially monochromatic. This is essential in order for quantum effects to provide the limit to the phase uncertainty. If the characteristic spectral width of the pulse $\Gamma \sim \int_0^T |u'(t)| dt \gtrsim T^{-1}$ is too large then the phase uncertainty will be dominated by the uncertainty Γ in the frequency which gives $\delta\phi \sim \Gamma/\omega$. In all that follows we assume this uncertainty to be negligible.

In Eq. (2.1) the complex amplitude α is dimensionless and defined in such a way that if all of the energy of the field were converted to photoelectrons (as will occur at our perfect detector), the rate of photoelectron production would be, to a very good approximation,

$$\lambda(t) = \frac{\epsilon_0 \overline{E^2(t)} \mathcal{A} c}{\hbar \omega} = |\alpha|^2 u(t), \tag{2.3}$$

where the bar over $E^2(t)$ indicates an average over many optical cycles. This rate is derived from the power of the beam $\epsilon_0 \overline{E^2(t)} \mathcal{A} c$, which is derived from its energy density $\epsilon_0 \overline{E^2(t)}$. Given the normalization (2.2), and the independence of the photoelectron-production events, the total number of photoelectrons will be a Poisson-distributed number n with mean $|\alpha|^2$.

This method of detection obviously yields no information about the phase of α. To do this requires mixing the system with a local oscillator of known phase at a beam splitter. For simplicity, we assume that the local oscillator has the same mode function $u(t)$ as the system, and a much larger intensity. Specifically, we assume the local oscillator electric field to be given by [9]

$$E_{\rm LO}(t) = \left(\frac{2\hbar \omega u(t)}{\epsilon_0 \mathcal{A} c} \right)^{1/2} \beta \; \mathrm{Re}[e^{-i\omega t + i\Phi(t)}], \tag{2.4}$$

where β is a dimensionless real number and the local oscillator phase $\Phi(t)$ is a function which is arbitrary but slowly varying compared to ωt. We now put the system pulse into one port of a 50/50 beam splitter, and the local oscillator into the other. If the two output ports are covered by perfect photodetectors then the rate of photodetection in those ports is

$$\lambda_\pm = \tfrac{1}{2} u(t) |\alpha \pm \beta e^{i\Phi(t)}|^2. \tag{2.5}$$

In the desired limit $\beta \gg |\alpha|$, the rate of photodetections at the two output ports will be dominated by the local oscillator. Consider a time interval $[t, t+\delta t)$ where δt is very small compared to the time over which the pulse shape changes $\sim u(t)/|u'(t)|$, but very large compared to the mean time between detections $\sim [\beta^2 u(t)]^{-1}$. The first condition allows us to treat $u(t)$ as a constant over that interval, so that the number of photodetections in each detector will be a Poisson process with mean $\lambda_\pm \delta t$. The second condition makes this mean much greater than one so that the Poisson process can be approximated by a Gaussian

$$\delta n_\pm(t) = \lambda_\pm(t)\delta t + \sqrt{\lambda_\pm(t)}\, \delta W_\pm(t), \tag{2.6}$$

where $\delta W_\pm(t)$ represent independent Gaussian random variables of mean zero and variance δt.

The signal of interest is the difference between the photocurrents at the two ports. This can be defined rigorously in terms of the noncommuting limits

$$I(t) = \lim_{\delta t \to 0} \lim_{\beta \to \infty} \frac{\delta n_+(t) - \delta n_-(t)}{\beta \delta t}$$
$$= u(t) 2\, \mathrm{Re}[\alpha e^{-i\Phi(t)}] + \sqrt{u(t)}\, \xi(t), \tag{2.7}$$

where $\xi(t) = \lim_{\delta t \to 0} [\delta W_+(t) - \delta W_-(t)]/\delta t$ is a Gaussian white-noise term [10] with the autocorrelation function

$$\langle \xi(t)\xi(t') \rangle = \delta(t-t'). \tag{2.8}$$

Thus the signal photocurrent is proportional to the quadrature of the system with phase $\Phi(t)$, plus the shot noise due to the local oscillator.

B. Heterodyne

A true phase measurement must determine the phase of the system with equal accuracy, regardless of the value of that phase. The standard way to do this is to sample all quadratures equally. This can be achieved by heterodyne detection, where the local oscillator phase $\Phi(t)$ is given by $\Phi(0) - t\Delta$. Here Δ is the detuning of the local oscillator from the system, which is much less than ω, but much greater than the characteristic pulse bandwidth Γ. This ensures that over the course of the pulse, the phase of the quadrature changes sufficiently rapidly for all quadratures to be measured with equal accuracy. This is to be contrasted with homodyne detection where the local oscillator is resonant with the system so that Φ is a constant and only one quadrature is measured. While this latter measurement is certainly phase sensitive, it can only be used to estimate the phase if the system phase and amplitude are approximately known beforehand, as explained in the Introduction.

Substituting $\Phi(t) = \Phi(0) - t\Delta$ into Eq. (2.7), we see that the heterodyne photocurrent has a deterministic part which varies sinusoidally with frequency Δ under the envelope $u(t)$. The amplitude of these oscillations is proportional to $|\alpha|$, while their phase is proportional to $\arg(\alpha) - \Phi(0)$. The complex amplitude α can therefore be estimated by taking the complex Fourier transform of the photocurrent at the appropriate frequency Δ. That is, we need to take the integral of the photocurrent over the time interval $[0,T)$, multiplied by the kernel $\exp(-it\Delta)$

$$A = \int_0^T dt\, I(t) e^{i[\Phi(0) - t\Delta]}. \tag{2.9}$$

1. General formulas

The integral of the photocurrent which we desire can be written in a more generally applicable way as

$$A = \int_0^T dt\, I(t) e^{i\Phi(t)}. \tag{2.10}$$

This can be evaluated as

$$A = \alpha - \alpha^* B + i\sigma, \tag{2.11}$$

where we have used Eq. (2.2) and defined another integral

$$B = -\int_0^T dt\, e^{2i\Phi(t)} u(t) \tag{2.12}$$

and a random variable

$$\sigma = \int_0^T e^{i\Phi(t) - i\pi/2} \sqrt{u(t)}\, dW(t), \tag{2.13}$$

where $dW(t) = \xi(t)dt$ is an infinitesimal Wiener increment [10]. It is easy to verify that this random variable satisfies

$$\langle \sigma \rangle = 0\,; \quad \langle |\sigma|^2 \rangle = 1\,; \quad \langle \sigma^2 \rangle = \langle B \rangle. \tag{2.14}$$

2. Specific results

The immediately preceding formulas are true for all functions $\Phi(t)$. For the present case of heterodyne detection, we can go further and find

$$B = -\int_0^T dt\, e^{2i\Phi(0) - 2it\Delta} u(t) \sim \Gamma/\Delta, \tag{2.15}$$

where Γ is the spectral width of the pulse as before. In the appropriate limit $\Delta \gg \Gamma$, the integral B vanishes. The same argument cannot be used to show that σ vanishes because $\xi(t)$ is δ correlated and so has a characteristic time which is always much shorter than Δ^{-1}. However, it is evident from Eq. (2.14) that $\langle \sigma^2 \rangle$ vanishes. This result, together with the other results in Eq. (2.14), completely characterize σ, because it is a Gaussian random variable [being the sum of independent Gaussian random variables $\xi(t)dt$]. Thus for heterodyne detection, the complete measurement is characterized by the complex number

$$A = \alpha + i\sigma, \tag{2.16}$$

where σ is a phase-independent complex Gaussian random variable satisfying $\langle |\sigma|^2 \rangle = 1$.

To estimate the phase of the field, we take the argument of the result A. We are interested in the high-intensity limit $|\alpha| \gg 1$, which would be the most useful for communication. In this limit, the noise term σ is small compared to α. Thus it is possible to treat it as a perturbation to the phase measurement. In other words, the measured phase is given by

$$\phi_{\text{het}} = \arg A = \arg(\alpha) + \text{Im}(i\sigma/\alpha) + O(1/|\alpha|^2). \tag{2.17}$$

Obviously the most likely phase result is $\arg(\alpha)$, as desired. Without loss of generality we can take α to be real. Then the most likely phase is $\phi_{\text{het}} = 0$ and the uncertainty in the phase estimate is determined by the variance

$$\langle \phi_{\text{het}}^2 \rangle \simeq (2\alpha)^{-2} \langle (\sigma + \sigma^*)^2 \rangle = \tfrac{1}{2}\, \alpha^{-2}, \tag{2.18}$$

where it is not difficult to show that the next higher-order term is of order α^{-4}. Thus for large α, the prepared phase of the coherent state can be estimated quite accurately, with an uncertainty of order α^{-1}.

C. Adaptive mark I

In this section we introduce an adaptive scheme to measure the phase. As explained in the Introduction, the guiding principle is that the local oscillator phase be adjusted over the course of the measurement to be equal to the estimated system phase plus $\pi/2$ in order to get information about the estimated phase quadrature of the system. That is to say, we set

$$\Phi(t) = \hat{\varphi}(t) + \pi/2. \tag{2.19}$$

Here the hat on $\hat{\varphi}(t)$ indicates that it is an estimator of the system phase. It is not an operator. The estimate is made on the basis of the measurement result obtained so far, over the course of the detection from time 0 to time t. This implies that it must be some functional of the photocurrent $\{I(s):s \in [0,t)\}$. Using the formal apparatus of quantum measurement theory, one of us has shown [8] that the full photocurrent $\{I(s):s \in [0,t)\}$ is not relevant, but rather only the two complex functionals

$$A_t = \int_0^t ds\, I(s) e^{i\Phi(s)}, \tag{2.20}$$

$$B_t = -\int_0^t ds\, e^{2i\Phi(s)} u(s), \tag{2.21}$$

which is a considerable simplification. For $t = T$ these integrals A_T, B_T are the results A and B already introduced. Thus, they arise naturally in the semiclassical picture, and it is not necessary to understand the theory of Ref. [8] to follow the argument presented here.

The crucial question is what to choose for $\hat{\varphi}(t)$. In this work we choose

$$\hat{\varphi}(t) = \arg A_t\,. \tag{2.22}$$

This is motivated by the following considerations.

(1) It is suggested by the above analysis for heterodyne detection.

(2) As shown by one of us [5], it gives the best possible result if the system has at most one photon.

(3) As will be shown, it gives the feedback algorithm

$$d\Phi(t) = \frac{I(t)dt}{\sqrt{\int_0^t u(s)ds}}, \tag{2.23}$$

which should be easy to implement experimentally because it is linear in the instantaneous photocurrent $I(t)$.

(4) As will be shown, it can be approximately solved analytically.

Before proceeding further, it is convenient to introduce a new time variable

$$v = \int_0^t u(s)ds, \tag{2.24}$$

which is a monotonic function of physical time t [because $u(t)$ is assumed non-negative] which maps $[0,T]$ into $[0,1]$. In terms of this variable, the photocurrent is equal to

$$I(v)dv = 2\,\text{Re}[\alpha e^{i\Phi(v)}]dv + dW(v), \tag{2.25}$$

where $dW(v)=\xi(t)\sqrt{u(t)}dt$ is an infinitesimal Wiener increment [10] obeying the Ito rule $[dW(v)]^2=dv$. Thus the photocurrent itself also obeys $[I(v)dv]^2=dv$.

Now the complex measurement result A_v (2.20) which is to be used for the phase estimate is defined by the initial condition $A_0=0$ and the Ito stochastic differential equation

$$dA_v=e^{i\Phi(v)}I(v)dv. \tag{2.26}$$

Using the adaptive algorithm (2.22) implies that

$$dA_v=i\frac{A}{|A|}I(v)dv. \tag{2.27}$$

This is a nonlinear complex stochastic differential equation which is best treated by changing variables to $|A|_v^2$ and $\hat{\varphi}_v=\arg(A_v)$. Using the Ito calculus [10] we first find

$$d|A|_v^2=A_v^*(dA_v)+(dA_v^*)A_v+(dA_v^*)(dA_v)=dv, \tag{2.28}$$

so that $|A_v|^2=v$. Substituting this result into Eq. (2.27) gives

$$d\hat{\varphi}_v=\mathrm{Im}[d\ln A_v]=\mathrm{Im}\left[\frac{dA_v}{A_v}-\frac{(dA_v)^2}{2A_v^2}\right]=\frac{I(v)dv}{\sqrt{v}}. \tag{2.29}$$

Thus the total solution is

$$A_v=\sqrt{v}\exp\left[i\int_0^v\frac{I(v)dv}{\sqrt{v}}\right], \tag{2.30}$$

and at the end of the measurement

$$A=e^{i\hat{\varphi}}=\exp\left[i\int_0^1\frac{I(v)dv}{\sqrt{v}}\right]. \tag{2.31}$$

Since the local oscillator phase is given by $\Phi(v)=\hat{\varphi}_v+\pi/2$, we have $d\Phi=d\hat{\varphi}$. Thus

$$d\Phi(v)=\frac{I(v)dv}{\sqrt{v}}, \tag{2.32}$$

which is the simple algorithm quoted above Eq. (2.23). This feedback procedure is represented in Fig. 1. The instantaneous photocurrent $I(t)$ is multiplied by a function proportional to $[\int_0^t u(s)ds]^{-1/2}$ generated by a signal generator. The result is then integrated and the resulting current produces in an electro-optic modulator a proportional phase shift Φ. Using the expression (2.25) for the photocurrent, this algorithm gives the following nonlinear stochastic differential equation for the phase estimate $\hat{\varphi}_v$:

$$d\hat{\varphi}_v=v^{-1/2}[-2\alpha\sin\hat{\varphi}_v dv+dW(v)], \tag{2.33}$$

where we have again set α to be real for convenience, and have used the relation $\Phi(v)=\hat{\varphi}_v+\pi/2$.

In order to attempt a solution of Eq. (2.33) it is convenient to change time variables once again to $x=2\sqrt{v}$, which is a monotonic mapping of [0,1] onto [0,2]. In terms of this variable

$$d\hat{\varphi}_x=-2\alpha\sin\hat{\varphi}_x dx+\sqrt{2/x}\,dW(x). \tag{2.34}$$

This equation is formally equivalent to the highly damped Brownian motion of a particle in a periodic potential $V(\hat{\varphi})\propto-2\alpha\cos\hat{\varphi}$, in which the temperature varies as the reciprocal of the time [10]. Clearly, for short times, the phase estimate will vary wildly, as the amount of noise in this equation diverges as $1/x$ as $x\to 0$, while the size of the deterministic term is constant. Thus the initial condition for this equation (the phase which one would guess on the basis of no information whatsoever) is immaterial as it becomes randomized immediately. This is why $\hat{\varphi}_0$ was not included in the formal solutions for A [(2.30) and (2.31)]. The physical reason for this divergence is that at short times one has very little information on which to base a phase estimate, so it is not surprising that the estimate is unstable. As time increases the noise term reduces, and for times $x\gg\alpha^{-1}$ (where $\alpha\gg 1$ is the regime of interest), the deterministic term becomes much larger than the noise term. Thus the phase will settle towards one of the minima of the potential, namely, $\hat{\varphi}=2n\pi$ for n an integer.

At the end of the pulse (when $x=2$), the obvious number to pick as the result of the phase measurement is the phase estimate currently in use by the adaptive algorithm (2.22), that is

$$\phi_{\mathrm{I}}=\hat{\varphi}\equiv\arg A, \tag{2.35}$$

which is the same as that used for heterodyne detection. To precisely evaluate the accuracy of this estimate, it would be necessary to find the solution $\hat{\varphi}_x$ of Eq. (2.34) for $x=2$. Unfortunately, it is not possible to solve this equation exactly because of the nonlinearity of the deterministic term. However, as argued above, for some time x_1, being finitely greater than α^{-1} but finitely less than 2, the phase $\hat{\varphi}_x$ will come to lie near $2n\pi$ for n an integer. We choose $n=0$ without loss of generality and linearize Eq. (2.34) around $\hat{\varphi}_x=0$. The result, which will be valid for $x_1\leqslant x\leqslant 2$, is

$$d\hat{\varphi}_x=-2\alpha\hat{\varphi}_x dx+\sqrt{2/x}\,dW(x), \tag{2.36}$$

which has the solution

$$\hat{\varphi}=\hat{\varphi}_{x_1}e^{2\alpha(x_1-2)}+\int_{x_1}^2 e^{2\alpha(x-2)}\sqrt{2/x}\,dW(x). \tag{2.37}$$

The variance for the mark I phase estimate $\phi_{\mathrm{I}}=\hat{\varphi}$ is therefore

$$\langle\phi_{\mathrm{I}}^2\rangle=\langle\hat{\varphi}_{x_1}^2\rangle e^{-4\alpha(2-x_1)}+\int_0^{2-x_1}dy\,e^{-4\alpha y}\frac{1}{1-y/2}. \tag{2.38}$$

Now the integrand in this integral is easily bounded using the following relations (which are valid in the range of integration):

$$\exp(\tfrac{1}{2}y)\leqslant\frac{1}{1-y/2}\leqslant\exp\left(\frac{\ln 2-\ln x_1}{2-x_1}y\right). \tag{2.39}$$

Hence, from the finiteness of $\langle \hat{\varphi}_{x_1}^2 \rangle$, x_1 and $2-x_1$ we finally obtain in the limit of large α

$$\langle \phi_{\rm I}^2 \rangle = \frac{1}{4\alpha} + O(\alpha^{-2}) + O[e^{-4\alpha(2-x_1)}]. \qquad (2.40)$$

To leading order this is independent of x_1 and $\hat{\varphi}_{x_1}$, which justifies our approach. Note that although it is small for large α, this variance is larger than that from heterodyne detection by a factor of α. That is to say, the excess noise of the mark I adaptive phase measurement scheme is far above the SQL.

D. Adaptive mark II

If the above result (2.38) were the end of the story, then it would be a sad ending indeed for adaptive phase measurements. Fortunately, it is not the end of the story because a minor modification of the above measurement scheme yields a result which is far better, instead of being far worse, than a standard phase measurement. This modification is simply to change the final phase estimate ϕ, while keeping the adaptive algorithm precisely the same. To derive this improved phase estimate, it is instructive first to examine why the mark I phase estimate $\phi_{\rm I} = \arg A$ is so bad. Recall that for heterodyne detection we had $A = \alpha + i\sigma$, where σ was a noise term, so that $\phi = \arg A$ made good sense then. But for the more general case

$$A = \alpha - \alpha^* B + i\sigma, \qquad (2.41)$$

as already stated [Eq. (2.11)], and $B \neq 0$ for the adaptive measurement. This indicates that the second integral B should be taken into account in determining the final phase estimate ϕ. Note that σ is not a measurement result which is available to the experimenter; it is the shot noise which cannot be separated from the signal unless α is known. But it is the phase of α which we are trying to estimate so α cannot be assumed known. The only available results are the two integrals B and A, the first involving only local oscillator phase (which is an experimentally controlled parameter), and the second involving the measured photocurrent as well.

What we desire is some function of A and B which is proportional to α plus a noise term, so that its argument would be a suitable estimate of the phase. The simplest such function is

$$A + BA^* = \alpha(1 - |B|^2) + i(\sigma - B\sigma^*). \qquad (2.42)$$

In terms of the time variable v, the result B is given by

$$B = -\int_0^1 dv\, e^{2i\Phi(v)}, \qquad (2.43)$$

from which it is obvious that its absolute value is less than unity. Thus, the argument of $A + BA^*$ is, ignoring the noise term, $\arg\alpha$. We therefore choose as our mark II phase estimate the function

$$\phi_{\rm II} = \arg(A + BA^*). \qquad (2.44)$$

This choice can also be justified from a more sophisticated argument using quantum measurement theory [8].

To evaluate the accuracy of this mark II measurement, we once again assume α to be real and positive without loss of generality. Since

$$A = \exp(i\hat{\varphi}), \qquad (2.45)$$

we have from Eq. (2.41)

$$B = 1 + \alpha^{-1}[i\sigma - \exp(i\hat{\varphi})]. \qquad (2.46)$$

Recall that this variable determines the phase-dependent moments of σ

$$\langle \sigma^2 \rangle = \langle B \rangle = 1 - \alpha^{-1} \langle \exp(i\hat{\varphi}) \rangle. \qquad (2.47)$$

Now it was shown in the preceding section that $\langle \hat{\varphi}^2 \rangle \sim \alpha^{-1} \ll 1$. Thus $\langle \exp(i\hat{\varphi}) \rangle = 1 + O(\alpha^{-1})$, and we have

$$\langle \sigma^2 \rangle = 1 - \alpha^{-1} + O(\alpha^{-2}). \qquad (2.48)$$

This, coupled with the fact that $\langle |\sigma|^2 \rangle = 1$, indicates that the imaginary part of σ is at most $O(\alpha^{-1/2})$, unlike in heterodyne detection where the real and imaginary parts of σ are both of order unity. Also unlike in heterodyne detection, σ is not necessarily a Gaussian random variable. Although it is given by the sum of Gaussian random variables

$$\sigma = i\int_0^1 e^{i\Phi(v)} dW(v), \qquad (2.49)$$

these are not independent because $\Phi(v)$ depends on $dW(v')$ for $v' < v$.

From Eqs. (2.44) – (2.46), the mark II phase estimate can be written as

$$\phi_{\rm II} = \arg[2\cos\hat{\varphi} - \alpha^{-1} + \alpha^{-1}\exp(-i\hat{\varphi})i\sigma] \qquad (2.50)$$

$$= \operatorname{Im}\ln\left[1 - \frac{1}{2\alpha\cos\hat{\varphi}} + \frac{i\sigma}{2\alpha} - \frac{\sigma\tan\hat{\varphi}}{2\alpha}\right]. \qquad (2.51)$$

Now since $\langle \hat{\varphi}^2 \rangle \sim \alpha^{-1}$, $\hat{\varphi}$ can be treated as a small variable of order $\alpha^{-1/2}$. Keeping real terms up to order α^{-1} and imaginary terms up to order α^{-2} in Eq. (2.50), we find

$$\phi_{\rm II} = \operatorname{Im}\ln\left(1 - \frac{1}{2\alpha} + \frac{i\sigma}{2\alpha}\right). \qquad (2.52)$$

Here we have discarded the term $\hat{\varphi}\sigma/(2\alpha)$ because the real part of $\hat{\varphi}\sigma$ can be shown to be $O(\alpha^{-1})$, while its imaginary part is at most $O(\alpha^{-3/2})$. Expanding the logarithm finally yields

$$\phi_{\rm II} = \frac{\sigma + \sigma^*}{4\alpha}\left(1 + \frac{1}{2\alpha}\right) + o(\alpha^{-2}). \qquad (2.53)$$

From Eq. (2.48) we have

$$\langle (\sigma + \sigma^*)^2 \rangle = 2 + 2[1 - \alpha^{-1} + o(\alpha^{-1})], \qquad (2.54)$$

so that the mark II phase variance is

$$\langle \phi_{\mathrm{II}}^2 \rangle = \frac{1}{4\alpha^2} + \frac{1}{8\alpha^3} + o(\alpha^{-3}). \tag{2.55}$$

To leading order, this is one half the size of the variance of a heterodyne phase measurement of a coherent state of the same amplitude. A little extra calculation shows that the error term is, in fact, $O(\alpha^{-4})$.

E. Intrinsic and extrinsic phase noise

Given that the uncertainty in a mark II adaptive measurement of the phase of a coherent state is only a factor $1/\sqrt{2}$ smaller than the corresponding uncertainty from a standard phase measurement, it might be thought that it is an exaggeration to claim, as we have done, that the adaptive mark II scheme is far better than the standard scheme. In fact it is not an exaggeration, but to understand why it is necessary to consider intrinsic and extrinsic phase uncertainty. In the discussion so far we considered our system to be in a coherent state, and calculated the variance in the phase measured by three different detection schemes, without enquiring into the origin of that phase uncertainty. Because the phase variance is different in the three schemes, it is apparent that at least two of them (those with the greater variances) must put noise into the measurement result which is not inherent to the system. We call such introduced noise extrinsic phase noise, while that which is inherent to the system we call intrinsic phase noise.

In the semiclassical picture which we have been using, it appears that there is no intrinsic uncertainty in the phase $\arg\alpha$ of the state; all of the noise results from the shot noise in the measurement. One could imagine that a different measurement scheme, not involving photodetection, could determine α precisely in a single measurement. This is of course not correct, because of quantum mechanics. In the quantum-mechanical picture, at least some of the phase noise in the measurement result is due to the intrinsic phase noise of a coherent state, and that noise will turn up in any measurement of the phase, no matter how it is done. Thus the intrinsic phase noise can be defined to be the spread in the probability distribution $P(\theta)$ for obtaining the result θ from the best possible phase measurement. We will follow Ref. [1] in calling such a measurement the canonical phase measurement, so that θ is really shorthand for ϕ_{can}, and we will explain briefly in Sec. III how it is derived.

To estimate the intrinsic uncertainty in the phase, we make use of the following uncertainty relation between number and phase, proved by Holevo [11]

$$V(n)V^{\mathrm{H}}(\theta) \geqslant \frac{1}{4}. \tag{2.56}$$

This is a relation between the uncertainty in the number n of photons in a system, and the intrinsic uncertainty in its phase θ. Here $V(n)$ is simply the variance for the operator $n = a^\dagger a$, while $V^{\mathrm{H}}(\theta)$ is defined by

$$V^{\mathrm{H}}(\theta) = |\langle e^{i\theta} \rangle|^{-2} - 1, \tag{2.57}$$

where θ is the result of a canonical phase measurement with distribution $P(\theta)$. In ignorance of any common term for this $V^{\mathrm{H}}(\theta)$, we will call it the Holevo phase variance [12].

It is not possible to prove an uncertainty relation of the form (2.56) using the standard definition of variance $V(\theta) = \langle \theta^2 \rangle - \langle \theta \rangle^2$. This is easy to see, since $\langle \theta^2 \rangle$ is evidently bounded if θ is confined to the interval $[0,2\pi)$, yet $V(n) = 0$ if the state is a number state. Moreover, the standard phase variance is actually not well defined at all, because θ could be taken to be an element of $[\theta_0, \theta_0 + 2\pi)$ for any real θ_0. Different choices for θ_0 will yield (sometimes wildly) different results for the standard variance. For these reasons, the Holevo phase variance is a far superior measure of the spread in a distribution $P(\theta)$. For states having a distribution $P(\theta)$ which is narrow and symmetric about some $\overline{\theta}$, it is easy to verify that

$$V^{\mathrm{H}}(\theta) \simeq \langle (\theta - \overline{\theta})^2 \rangle, \tag{2.58}$$

so that the standard phase variance is a good approximation to the Holevo phase variance in this case. This justifies our use of the standard variance so far in this paper. In future, whenever we write $V(\phi)$ for any cyclic variable ϕ, we will mean $V^{\mathrm{H}}(\phi)$.

Now a coherent state is a minimum uncertainty state for any pair of canonically conjugate quadrature operators. It is therefore not surprising that it is also, to a very good approximation, a minimum uncertainty state for number and phase, at least if it has a large coherent amplitude [15]. The number distribution for a coherent state is Poissonian, so $V(n) = \overline{n} = |\alpha|^2$. Substituting this into the Holevo relation (2.56) yields

$$\langle \theta^2 \rangle \simeq V(\theta) \simeq \frac{1}{4\alpha^2}, \tag{2.59}$$

where we have taken α to be real. This is the intrinsic phase variance of a coherent state. Subtracting it from the variances for the results of the various measurement schemes examined above thus gives the extrinsic noise introduced by those measurements. We find to leading order in α^{-1},

$$V_{\mathrm{coh}}(\phi_{\mathrm{het}}) - V_{\mathrm{coh}}(\theta) = \frac{1}{4\alpha^2} + O(\alpha^{-4}), \tag{2.60}$$

$$V_{\mathrm{coh}}(\phi_{\mathrm{I}}) - V_{\mathrm{coh}}(\theta) = \frac{1}{4\alpha} + O(\alpha^{-2}), \tag{2.61}$$

$$V_{\mathrm{coh}}(\phi_{\mathrm{II}}) - V_{\mathrm{coh}}(\theta) = \frac{1}{8\alpha^3} + O(\alpha^{-4}). \tag{2.62}$$

From these results we see that the variance $V(\phi_{\mathrm{II}})$ is due almost entirely to the intrinsic phase variance of the coherent state. The extrinsic noise in the mark II adaptive phase measurement is a factor $(2\alpha)^{-1}$ smaller than the extrinsic measurement of a heterodyne measurement (which is equal to the intrinsic phase variance of the coherent state). This is why it is correct to say that the adaptive mark II phase measurement is much better than a standard phase measurement such as by heterodyne detection. It also implies that by using nonclassical states, with a smaller phase uncertainty than coherent states of the same mean photon number, it would be possible for $V(\phi_{\mathrm{II}})$ to be much smaller than $V(\phi_{\mathrm{het}})$. Such nonclassical states cannot be described within the semiclassical theory we have used so far. For this reason, we turn in Sec. III to the quantum theory of phase estimation.

III. QUANTUM PHASE ESTIMATION

A. Quantum estimation theory

Before talking specifically of the quantum theory of phase estimation, we will summarize the general theory of quantum estimation as explained in Refs. [16,17]. Let the quantum system immediately before the measurement begins have the density operator ρ. Let the set of all possible measurement results λ be denoted Ω. Being a quantum-mechanical measurement, the result will in general be a random variable, so we can only talk about $P(E)=\mathrm{Pr}(\lambda\in E)$, the probability that the measurement result will be obtained in some subset $E\subseteq\Omega$. Then the most general possible formula for $P(E)$ is

$$P(E)=\mathrm{Tr}[\rho F(E)], \tag{3.1}$$

where F is a mapping from Ω onto the set of positive operators. That is, for any particular λ, $F(\lambda)$ is an operator with a positive semidefinite spectrum, and

$$F(E)=\int_{\lambda\in E}F(\lambda)d\lambda. \tag{3.2}$$

Note that $F(\lambda)$ is not necessarily a projector, or even proportional to a projector. To ensure that the normalization condition $P(\Omega)=1$ is satisfied for all states ρ, F must satisfy the completeness condition

$$F(\Omega)=1. \tag{3.3}$$

Since $P(\lambda)$ is a measure on the set Ω, $F(\lambda)$ is known as a positive-operator-valued measure (POVM) on the set Ω.

Now consider the case where the measured quantity is to be a phase ϕ, so that F is a POVM on $\Omega=[0,2\pi)$. The fact that ϕ is a cyclic variable implies that F should be invariant under a translation of the phase $\phi\to\phi+\theta$. Now a phase translation is effected by the unitary operator $R(\theta)=\exp(ia^\dagger a\theta)$, where $a^\dagger a$ is the number operator. Thus the invariance of F can be written as

$$R(\theta)F(\phi)R(-\theta)=F(\phi+\theta)\forall\theta,\phi\in\Omega. \tag{3.4}$$

It can be shown that this condition guarantees that $F(\phi)$ can be written as

$$F(\phi)=\frac{1}{2\pi}\sum_{n,m=0}^{\infty}|m\rangle\langle n|e^{i\phi(m-n)}H_{mn}, \tag{3.5}$$

where H is a positive Hermitian matrix and $|m\rangle$ is a number state. The completeness condition (3.3) implies that

$$\forall m\geq 0 H_{mm}=1. \tag{3.6}$$

The rotational invariance condition (3.4) does not capture all that we understand intuitively by saying that ϕ is a measurement of phase. For example, if $F(\phi)$ satisfies Eq. (3.4), then so will $F(\phi+\psi)$ for any $\psi\in\Omega$. To remove this and other degeneracies we simply impose the extra condition that H be a real matrix with all positive elements. This choice guarantees that in the semiclassical limit, the "mean" phase $\overline{\phi}$ is what one would expect from classical intuition. Consider first an arbitrary quantum state $|\psi\rangle$. Then

$$\langle e^{i\phi}\rangle=\int d\phi e^{i\phi}\langle\psi|F(\phi)|\psi\rangle$$

$$=\frac{1}{2\pi}\int d\phi\, e^{i\phi}\sum_{n,m=0}^{\infty}\langle\psi|m\rangle\langle n|\psi\rangle\, e^{i\phi(m-n)}H_{mn} \tag{3.7}$$

$$=\sum_{m=0}^{\infty}\langle\psi|m\rangle\langle m+1|\psi\rangle H_{m,m+1}. \tag{3.8}$$

Now for the semiclassical limit we want $|\psi\rangle=|\alpha\rangle$, a coherent state of amplitude α having the number state representation

$$\langle n|\alpha\rangle=e^{-|\alpha|^2/2}\frac{\alpha^n}{\sqrt{n!}}. \tag{3.9}$$

This gives

$$\langle e^{i\phi}\rangle=\sum_m P_m\frac{\alpha}{\sqrt{m+1}}H_{m,m+1}, \tag{3.10}$$

where $P_m=|\langle m|\alpha\rangle|^2$. Since all elements of H are assumed positive, we have the "mean" phase $\overline{\phi}=\arg\langle e^{i\phi}\rangle=\arg\alpha$, as desired.

According to the above arguments, a phase measurement is defined in terms of the POVM (3.5) with H a positive matrix with all elements real and positive and diagonal elements equal to unity. The positivity condition on the matrix obviously requires that the off-diagonal elements be less than or equal to unity. A unique phase measurement is defined by specifying that all of the off diagonal elements be equal to unity. This is what has recently been called a canonical phase measurement [1], although its uniqueness was recognized very early in the history of quantum theory [18]. In realistic phase measurements the off-diagonal elements $H_{m,n}$ will be less than unity, but for $|m-n|=1$ and $m\gg 1$ they should be close to unity if the measurement is to be a good phase measurement, as will be seen below. In fact, in all of the measurements we examine, we have

$$h(m)\equiv 1-H_{m,m+1}\leq O(m^{-1/2}). \tag{3.11}$$

For a canonical measurement $h(m)$ is identically zero.

B. Determining $h(m)$

Let us consider a coherent state $|\alpha\rangle$ with $\alpha\gg 1$ real, and an arbitrary phase measurement with the POVM $F(\phi)$. We have just shown that the mean phase $\arg\langle e^{i\phi}\rangle$ will be $\arg\alpha=0$, but we wish now to consider the spread in ϕ. As discussed in Sec. II E, this can also be measured from $\langle e^{i\phi}\rangle$. From Eq. (3.10) we have, for α real,

$$\langle e^{i\phi}\rangle=\sum_{m=0}^{\infty}P_m\frac{\alpha}{\sqrt{m+1}}[1-h(m)], \tag{3.12}$$

where P_m is the photon number distribution for a coherent state and $h(m)$ is as defined above Eq. (3.11). Now P_m is sharply peaked around $m=\alpha^2$ with a variance of α^2. Thus we can expand the sum about $m=\alpha^2$ to get

$$\langle e^{i\phi}\rangle \simeq \sum_{m=0}^{\infty} P_m\left[1-\frac{\delta m+1}{2\alpha^2}+\frac{3(\delta m+1)^2}{8\alpha^4}\right] \times[1-h(\alpha^2)-h'(\alpha^2)\delta m-\tfrac{1}{2}h''(\alpha^2)(\delta m)^2], \tag{3.13}$$

where $\delta m=m-\alpha^2$. Now if $h(m)\sim m^{-p}$ for large m and some positive power p (as will be shown), then $h'(\alpha^2)=O(h(\alpha^2)/\alpha^2)$ and $h''(\alpha^2)=O(h(\alpha^2)/\alpha^4)$. Thus from the moments of P_m we find

$$\langle e^{i\phi}\rangle = 1-(8\alpha^2)^{-1}-h(\alpha^2)+O(\alpha^{-4})+O(h(\alpha^2)\alpha^{-2}). \tag{3.14}$$

The (Holevo) phase variance of a distribution is defined above Eq. (2.57) as $|\langle e^{i\phi}\rangle|^{-2}-1$. In this case we thus have

$$V(\phi)=[(4\alpha^2)^{-1}+2h(\alpha^2)][1+O(\alpha^{-2})+O(h(\alpha^2))]. \tag{3.15}$$

Now for a canonical measurement $h(m)=0$, so only the first leading term is retained. This represents the intrinsic phase variance of a coherent state, as established above using the uncertainty relation (2.59). If $h(m)\neq 0$, we see that $2h(\alpha^2)$ can be interpreted as the extrinsic phase variance introduced by the measurement. Referring to Eqs. (2.60)–(2.62), we see that we can make the following identifications:

$$h_{\text{het}}(m)\simeq(8m)^{-1}+O(m^{-2}), \tag{3.16}$$

$$h_{\text{I}}(m)\simeq(8m^{1/2})^{-1}+O(m^{-1}), \tag{3.17}$$

$$h_{\text{II}}(m)\simeq(16m^{3/2})^{-1}+O(m^{-2}). \tag{3.18}$$

That is to say, from the semiclassical results for the measured phase variance of a large-amplitude coherent state we have been able to identify the important POVM matrix elements $H_{m,m+1}=1-h(m)$ for all three measurement schemes in the large m limit. These are the only elements we require for the analysis of the following section.

C. Optimized-state phase estimation

In Sec. II of this paper we derived from semiclassical photodetection theory the variance in the phase of a coherent state measured using three different schemes. As explained at the end of that section, because a coherent state has a relatively large intrinsic phase uncertainty, a better figure of merit is the variance in the phase of a state which has been optimized to have a low phase variance. The different detection schemes will in general have different optimized states. Of course, the optimization has to be constrained by something, because even the variance of the measured phase of a coherent state $|\alpha\rangle$ will go to zero as $\alpha\to\infty$. There are two obvious ways to constrain the states which are to be optimized: by putting an upper bound on the photon number states it is allowed to populate; and by fixing its mean photon number. In this paper we will consider the former of these, because it yields answers more simply.

Let the maximum photon number allowed be denoted N. Then the general problem to be solved is to find the $N+1$ coefficients ψ_n defining the state

$$|\psi\rangle=\sum_{n=0}^{N}\psi_n|n\rangle, \tag{3.19}$$

subject to the normalization constraint $\Sigma|\psi_n|^2=1$, which minimizes the Holevo variance

$$V(\phi)=|\langle e^{i\phi}\rangle|^{-2}-1, \tag{3.20}$$

where from Eq. (3.8),

$$\langle e^{i\phi}\rangle=\langle\psi|\left[\sum_{m=0}^{N-1}|m\rangle\langle m+1|H_{m,m+1}\right]|\psi\rangle. \tag{3.21}$$

Minimizing the phase variance is equivalent to maximizing the modulus of $\langle e^{i\phi}\rangle$. Since the phase of this expectation value is arbitrary, we can choose it to be real. Then we can restate our aim to be to maximize the expectation value of the operator

$$\cos\phi=\sum_{m=0}^{N}[1-h(m)]\frac{|m\rangle\langle m+1|+|m+1\rangle\langle m|}{2}, \tag{3.22}$$

where $H_{m,m+1}=1-h(m)=H_{m+1,m}$ as before. Finally, since we are working in a finite subspace of the total Hilbert space, it is trivial that the state which maximizes the expectation value of $\cos\phi$ is the eigenstate of this operator with the largest eigenvalue. Thus the problem reduces to one of finding the eigenvalues λ_k of the operator (3.22).

1. Canonical measurement

For canonical measurements we have $h(m)=0$ and the problem becomes exactly soluble. The operator

$$2\cos\theta=\sum_{m=0}^{N-1}[|m\rangle\langle m+1|+|m+1\rangle\langle m|] \tag{3.23}$$

has eigenvalues

$$\lambda_k=2\cos\left(\frac{\pi k}{N+2}\right)k=1,\ldots,N+1 \tag{3.24}$$

corresponding to the eigenstates

$$|\psi\rangle_k\propto\sum_{m=0}^{N}\sin\left(\frac{(m+1)\pi k}{N+2}\right)|m\rangle. \tag{3.25}$$

Hence, the minimum Holevo variance achievable from a canonical measurement is

$$\left[\cos\left(\frac{\pi}{N+2}\right)\right]^{-2}-1=\frac{\pi^2}{(N+2)^2}+O(N^{-4}). \tag{3.26}$$

To leading order this agrees with the result obtained by Summy and Pegg [19], although they used the standard variance rather than the more natural Holevo variance.

2. Physically achievable measurements

There are three physically achievable phase measurements which we have analyzed, namely, heterodyne, adaptive mark I, and adaptive mark II. In all of these $h(m)$ is nonzero, and an analytical solution to the problem is not possible. Instead, we look for an approximate asymptotic solution for $N\gg 1$. In all three cases we can write to leading order [see Eqs. (3.16)–(3.18)]

$$h(m)=cm^{-p} \tag{3.27}$$

for some positive power $p\geqslant 1/2$ and positive coefficient c of order unity. For $N\gg 1$ we can treat the photon number m as a continuous variable and $\psi(m)=\psi_m$ as a twice-differentiable function. Then, noting that

$$-2+\sum_{m=0}^{N-1}[|m\rangle\langle m+1|+|m+1\rangle\langle m|] \tag{3.28}$$

is a finite-difference approximation to the second derivative operator with Dirichlet boundary conditions, we can use the approximation

$$2\cos\phi\approx 2+\frac{\partial^2}{\partial m^2}-2h(m). \tag{3.29}$$

This assumes that the phase variance is very small, as is the case in practice. From this we find that the eigenvalue equation we have to solve is

$$\left(-\frac{\partial^2}{\partial m^2}+2cm^{-p}\right)\psi(m)=(2-\lambda)\psi(m), \tag{3.30}$$

which is equivalent to a time-independent Schrödinger equation with the boundary conditions $\psi(0)=\psi(N)=0$. Note that the ''potential-energy'' term is lowest at $m=N$, which suggests that the solution of lowest ''energy'' will be localized in that region.

We are interested in the solution to Eq. (3.30) with the largest eigenvalue λ. For large N this eigenvalue will be very close to 2, as it is equal to $2\langle\cos\phi\rangle$. Also, since the solution will be localized at $m\approx N$, the potential-energy term can be linearized about that point. Changing variables to $y=1-N^{-1}m$ we thus transform Eq. (3.30) to

$$\left(-\frac{\partial^2}{\partial y^2}+by\right)\psi(y)=a_k\psi(y), \tag{3.31}$$

subject to the boundary conditions $\psi(1)=\psi(0)=0$, where

$$a_k=N^2(2-\lambda_k-2cN^{-p}), \tag{3.32}$$

$$b=2cpN^{2-p}. \tag{3.33}$$

This has the form of the time-independent Schrödinger equation for a bead on a frictionless vertical string attached at the floor and ceiling. Since we are interested in the solution of lowest energy (maximum λ), we can ignore the ceiling. That is to say, we can ignore the boundary condition at $y=1$ and let the string become semi-infinite. Then the normalizable solutions are the well-known Airy functions

$$\psi_k(y)\propto \mathrm{Ai}(z_k+b^{1/3}y), \tag{3.34}$$

for $y>0$, where z_k is the kth real zero of the Airy function satisfying$0>z_1>z_2>\cdots$. The corresponding eigenvalues are

$$a_k=-b^{2/3}z_k. \tag{3.35}$$

The smallest eigenvalue is a_1. In this case the solution (3.34) has a single zero, at $y=0$. The corresponding value for λ is

$$2-\lambda_1=2cN^{-p}+(-z_1)(2cp)^{2/3}N^{-2(1+p)/3}. \tag{3.36}$$

Since $\lambda=2|\langle e^{i\phi}\rangle|\approx 2$, the minimum Holevo variance is given by

$$V(\phi)=|\langle e^{i\phi}\rangle|^{-2}-1\simeq 2-\lambda_1+O((2-\lambda_1)^2). \tag{3.37}$$

Thus we have arrived at the expression we desire, the minimum phase variance for the states optimized for the various detection schemes, with a constrained maximum photon number N. Using $z_1\approx -2.338$ and substituting in the coefficients c and powers p for $h(m)$ from Eqs. (3.16)–(3.18), we obtain

$$V(\phi_{\mathrm{het}})\simeq \tfrac{1}{4}N^{-1}+0.9278N^{-4/3}, \tag{3.38}$$

$$V(\phi_{\mathrm{I}})\simeq \tfrac{1}{4}N^{-1/2}+O(N^{-1}), \tag{3.39}$$

$$V(\phi_{\mathrm{II}})\simeq \tfrac{1}{8}N^{-3/2}+0.7659N^{-5/3}. \tag{3.40}$$

We do not give an expression for the next-to-leading term in $V(\phi_{\mathrm{I}})$ because it is uncertain due to the uncertainty in $h(m)$ expressed in Eq. (3.17).

To leading order, we see the expected results due to the noise introduced by the measurements, and we see the great superiority of the adaptive mark II scheme over the standard (heterodyne) scheme. Our results for heterodyne detection disagree with the power law ''derived'' numerically by D'Ariano and Paris [20] for reasons to be explored in a future paper. The second term in each is due to the intrinsic phase uncertainty of the states, and becomes negligible compared to the leading term as $N\to\infty$. The width of the wave function $\psi(y)$ is of order $b^{-1/3}\sim N^{(p-2)/3}$, which also goes to zero as $N\to\infty$ since $p\leqslant 3/2$. This confirms that the solution ψ_m is concentrated at $m\approx N$. This argument also helps us to estimate the regime in which we expect the asymptotic results to be accurate. More than 0.995 of the area under the largest peak of the Airy function is confined to the interval $[-z_1,-z_1+5]$. Thus the width of $\psi(y)$ can be estimated as $5b^{-1/3}$. The assumption that $\psi(y)$ was concentrated at the lower end of the interval $[0,1]$ would then seem reasonable if $5b^{-1/3}\lesssim 1/2$. From Eq. (3.33), we can thus estimate that our asymptotic results will be valid if

$$N\gtrsim\left(\frac{10^3}{2cp}\right)^{1/(2-p)}. \tag{3.41}$$

Thus for an adaptive mark I measurement we require $N \gtrsim 400$; for heterodyne $N \gtrsim 4000$; and for adaptive mark II $N \gtrsim 3\times 10^7$. If these requirements are met then the estimates (3.38)–(3.40) should be good. However the converse is not necessarily true: the estimates (3.38)–(3.40) may be reasonable even for considerably smaller photon numbers N. This will be explored in a future paper.

IV. DISCUSSION

A. Summary

We have analyzed four different single-shot phase measurements schemes: canonical, heterodyne, adaptive mark I, and adaptive mark II. The first of these is the best possible phase measurement, but is not realizable physically. The second is one of the standard techniques (which are all equivalent) which is available to experimentalists at the present time. The last two are also experimentally realizable, and are based on the proposal in Ref. [5]. The essential feature of the adaptive measurements is that they use a feedback loop to change the detection system over the course of a measurement of a single pulse, using the results of the measurement up to that time. Both adaptive schemes use the same feedback algorithm. The difference between them is that the mark II adaptive scheme uses an improved formula for the final phase estimate of the system, using all of the recorded measurement data.

In this paper we have adopted an analysis based on semiclassical detection theory. It turns out that this is sufficient to derive asymptotic results for large photon number. A canonical phase measurement is the best measurement of phase allowed by quantum mechanics, so the minimum canonical phase variance for a state of maximum photon number N is a measure of the minimum intrinsic phase variance of such a state. This variance represents the ultimate quantum limit (UQL) to phase measurements. The minimum phase variances of the other three schemes is therefore a measure of the intrinsic phase variance plus the variance of the extrinsic phase noise introduced by the measurement. In the limit of asymptotically large N the extrinsic noise will always dominate.

We find that the four measurement schemes have minimum phase variances which scale in the following simple ways with maximum photon number N:

$$V_{\min}(\phi_{\rm can}) \simeq \pi^2 N^{-2}, \tag{4.1}$$

$$V_{\min}(\phi_{\rm II}) \simeq \tfrac{1}{8} N^{-3/2}, \tag{4.2}$$

$$V_{\min}(\phi_{\rm het}) \simeq \tfrac{1}{4} N^{-1}, \tag{4.3}$$

$$V_{\min}(\phi_{\rm I}) \simeq \tfrac{1}{4} N^{-1/2}. \tag{4.4}$$

The heterodyne measurement result represents the shot-noise limit or standard quantum limit (SQL), because a N^{-1} scaling is the minimum achievable from semiclassical states (that is, states which are mixtures of coherent states). The mark I adaptive scheme is thus much worse than the SQL for large photon numbers. The only attraction of this scheme is that it is the unique scheme which is as good as a canonical measurement for states with at most one photon, as shown in Ref. [5]. By contrast, the mark II adaptive scheme does much better than the SQL, with a variance lying intermediate to the SQL and the UQL.

B. Experimental practicalities

The asymptotic results presented above are very encouraging, in that they show that it is possible to make a phase measurement which is much closer to the ultimate quantum limit than previously thought possible. However, to achieve such a limit it would be necessary to create pulses of light with very large photon number and which are highly nonclassical. At least at first, an experimental attempt to realize the adaptive phase measurements proposed here would probably use a coherent light pulse, as this is much more readily available. This is exactly the scenario considered in Sec. II, and gave the following results:

$$V_{\rm coh}(\phi_{\rm II}) \simeq \frac{1}{4|\alpha|^2} + \frac{1}{8|\alpha|^3}, \tag{4.5}$$

$$V_{\rm coh}(\phi_{\rm het}) \simeq \frac{1}{2|\alpha|^2} + O(|\alpha|^{-4}), \tag{4.6}$$

$$V_{\rm coh}(\phi_{\rm I}) \simeq \frac{1}{4|\alpha|^1} + O(|\alpha|^{-2}). \tag{4.7}$$

Thus there would be an easily measurable difference between the three measurement schemes, although it would not show the dramatic difference in scaling between $V_{\min}(\phi_{\rm II})$ and $V_{\min}(\phi_{\rm het})$ presented earlier. However, these scalings must be taken with a grain of salt, because there are many other practical considerations which we have ignored which will tend to spoil these ideal results. Below we discuss two of these ''spoilers.''

1. Detector inefficiency

Detector inefficiency is well known as a destroyer of sub-SQL measurements. It might be thought that the adaptive measurements proposed here would be even more vulnerable to having a detector efficiency η less than one, because they rely on feeding back the measurement results. If the detectors are inefficient then the information being fed back is unreliable, and the performance of the device might be expected to suffer particularly badly. Fortunately this is not the case, as can be proven quite simply. The effect of a detector of efficiency η is completely equivalent to that of passing the pulse through a beam splitter of transmittance η. For a coherent state $|\alpha\rangle$ this has the simple effect of transforming it into the coherent state $|\sqrt{\eta}\alpha\rangle$. Thus the results (4.5)–(4.7) remain true, with α replaced by $\sqrt{\eta}\alpha$, and the difference between the measurement schemes will still be clear. However, for $\eta \leqslant 1/2$, the phase variance from the adaptive mark II scheme will be greater than that from a standard (heterodyne) measurement with $\eta = 1$. In this sense, we can say that it is necessary to have $\eta > 1/2$ in order to do better than the SQL.

Recall that for the adaptive mark II measurement with $\eta = 1$ the phase variance of a coherent state is almost entirely due to the intrinsic phase variance $\frac{1}{4}|\alpha|^{-2}$. With $\eta < 1$ this is no longer true, because the intrinsic phase noise should still be reckoned from the original state $|\alpha\rangle$, not from $|\sqrt{\eta}\alpha\rangle$.

Thus the noise introduced by the measurement is much larger. This follows through to the minimum phase variance of a state with constrained maximum number N. We find now

$$V_{\min}(\phi_{\mathrm{II}}) \simeq \frac{1-\eta}{4\eta N} + \frac{1}{8(\eta N)^{3/2}}, \tag{4.8}$$

$$V_{\min}(\phi_{\mathrm{het}}) \simeq \frac{2-\eta}{4\eta N}, \tag{4.9}$$

$$V_{\min}(\phi_{\mathrm{I}}) \simeq \frac{1}{4(\eta N)^{1/2}}. \tag{4.10}$$

Note that for η finitely less than one, the adaptive mark II result scales in the same way as the heterodyne result. That is, they both scale as N^{-1}, although the coefficient for the adaptive mark II case still puts it below the SQL provided $\eta > 1/2$.

2. Delay in the feedback loop

In contrast to inefficient detectors, a delay in the feedback loop is much harder to treat theoretically. Virtually all of the results of Secs. II C and II D rely on the assumption that the feedback is instantaneous. However, we can obtain a rough idea of the effect of such a delay by considering a toy mathematical problem which gives similar results if there is no delay, but which is simple enough to solve approximately when there is a delay. The details are lengthy and so are given in the Appendix, but the results are simple to state. Assuming a time delay τ and a pulse bandwidth Γ, we treat the product $\Gamma\tau$ as a small parameter. From this we find for a coherent state

$$V_{\mathrm{coh}}(\phi_{\mathrm{II}}) \simeq \frac{1}{4|\alpha|^2}[1+O(\Gamma\tau)] + \frac{1}{8|\alpha|^3}, \tag{4.11}$$

$$V_{\mathrm{coh}}(\phi_{\mathrm{I}}) \simeq \frac{1}{4|\alpha|} + O(\Gamma\tau) + O(|\alpha|^{-2}). \tag{4.12}$$

This implies that for a phase-optimized state of maximum photon number N we would find

$$V_{\min}(\phi_{\mathrm{II}}) \simeq \frac{1}{8N^{3/2}}[1+O(N^{1/2}\Gamma\tau)], \tag{4.13}$$

$$V_{\min}(\phi_{\mathrm{I}}) \simeq \frac{1}{4N^{1/2}}[1+O(N^{1/2}\Gamma\tau)]. \tag{4.14}$$

These results show that unless $\Gamma\tau \ll N^{-1/2}$ (which would be very hard to achieve for large photon numbers), the measured phase variance will be dominated by the effect of the delay. For the adaptive mark I measurement, this is a much worse effect than that arising from inefficient detectors. For the mark II scheme, the effect is much like that of an inefficient detector, with an inefficiency $1-\eta = O(\Gamma\tau) \ll 1$. Thus as long as the time delay is significantly less than the characteristic pulse length, the mark II scheme should still be superior to the SQL. Of course a real feedback loop will not suffer simply from a time delay; all of the electronic and electro-optic elements of the loop will have some characteristic response function. In that case, the total feedback loop will be characterized by a real positive response function $f(t)$, satisfying $\int_0^\infty f(t)dt = 1$, equal to the convolution of the response functions of the individual elements. The characteristic delay time τ could then be defined as $\tau = \int_0^\infty f(t)t\,dt$, provided this was suitably small.

C. Conclusion

By incorporating a real-time feedback loop into an optical detection scheme it is possible to create a single-shot measurement of phase which is far superior to standard single-shot measurements of phase. The device is based on balanced detection using a local oscillator, and it is the local oscillator phase which is controlled by the feedback. For the adaptive algorithm presented in this paper, the only elements required in the feedback loop are a signal generator, a variable amplifier, an integrator, and an electro-optic phase modulator. Thus the scheme should be experimentally practical. Under real experimental conditions, detector inefficiencies and the non-instantaneous response of the feedback loop will spoil the ideal results to some extent. However, as long as the detector inefficiency is not too large, and the feedback delay not too long compared to the pulse duration, the superiority of the adaptive scheme should still be evident. To be precise, if one had a sequence of pulses with randomly prepared phases, then the adaptive technique would give a mean-squared difference between measured phase and prepared phase which is smaller than that from any other technique known.

ACKNOWLEDGMENTS

H.M.W. thanks A. Holevo and Y. Li for valuable discussions. This work was supported by the Marsden Fund of the Royal Society of New Zealand and the Australian Research Council.

APPENDIX: THE EFFECT OF A TIME DELAY

Rather than attempting to treat the effect of a time delay in the feedback loop exactly (which is probably impossible to do analytically), we introduce a simplified model which seems to capture the essential features of the full system. First we present the toy model with no time delay.

1. Toy model with no delay

To obtain our toy model we simply take Eq. (2.36) for the phase estimate $\hat{\varphi}$ and replace the time-dependent diffusion coefficient with a constant diffusion coefficient equal to its value at the final time. Using γ in place of 2α (as a reminder that this is only a toy model) and t instead of x we have

$$d\hat{\varphi}_t = -\gamma\hat{\varphi}_t dt + dW(t). \tag{A1}$$

The solution to this is

$$\hat{\varphi}_t = e^{-\gamma t}\left[\hat{\varphi}_0 + \int_0^t e^{\gamma s} dW(s)\right]. \tag{A2}$$

Letting the maximum time be $t=1$ and equating the mark I phase estimate with $\hat{\varphi}$ at this time, we find to leading order

$$\langle \phi_{\rm I}^2\rangle = \frac{1}{2\gamma}. \tag{A3}$$

Identifying $\gamma=2\alpha$, this replicates the result of the full model (2.38).

Now consider the mark II phase estimate. From Eq. (2.44), this is given by

$$\phi_{\rm II}=\hat{\varphi}_1+\arg[1+\exp(-2i\hat{\varphi}_1)B], \tag{A4}$$

where, assuming a flat pulse $u(t)=1$,

$$B=\int_0^1 dt\ \exp(2i\hat{\varphi}_t). \tag{A5}$$

Now assuming that $|\hat{\varphi}_0|\ll 1$ (as seems reasonable given the argument in Sec. II C), we can expand the above exponentials to first order to obtain

$$\phi_{\rm II}\approx\hat{\varphi}_1+\arg\left[2-2i\hat{\varphi}_1+2i\int_0^1 dt\,\hat{\varphi}_t\right], \tag{A6}$$

$$\approx\int_0^1 dt\,\hat{\varphi}_t. \tag{A7}$$

While this is of course only a toy calculation, it gives some further insight into the formula (2.44) for $\phi_{\rm II}$ as some form of time average of the crude phase estimate $\hat{\varphi}_t$. In this case we find to leading order

$$\langle \phi_{\rm II}^2\rangle=\frac{1}{\gamma^2}+\frac{\langle\hat{\varphi}_0^2\rangle}{\gamma^2}. \tag{A8}$$

Ignoring the second term (as is justified since we assumed that $|\hat{\varphi}_0|\ll 1$), we again find agreement to leading order with the result of the full calculation (2.55).

2. Toy model with a delay τ

A time delay τ in the feedback loop would mean that the local oscillator phase at time t would be determined by the estimate for the system phase at time $t-\tau$. That is to say, Eq. (A1) is replaced by

$$d\hat{\varphi}_t=-\gamma\hat{\varphi}_{t-\tau}dt+dW(t). \tag{A9}$$

The essence of our approach is to treat the delay τ perturbatively. Thus we write the solution to the perturbed equation (A9) as

$$\hat{\varphi}_t=\hat{\varphi}_t^{(0)}+\gamma\tau\hat{\varphi}_t^{(1)}+O(\gamma^2\tau^2). \tag{A10}$$

The zeroth-order term $\hat{\varphi}_t^{(0)}$ obeys Eq. (A1) so the first-order correction obeys

$$\gamma\tau d\hat{\varphi}_t^{(1)}=\gamma(\hat{\varphi}_t^{(0)}-\hat{\varphi}_{t-\tau}^{(0)})dt-\gamma\tau\gamma\hat{\varphi}_{t-\tau}^{(1)}dt. \tag{A11}$$

Thus to first order in τ we have

$$\begin{aligned}d\hat{\varphi}_t^{(1)}&=-\gamma\hat{\varphi}_t^{(1)}dt+d\hat{\varphi}_t^{(0)}\\&=-\gamma\hat{\varphi}_t^{(1)}dt-\gamma\hat{\varphi}_t^{(0)}dt+dW(t)\\&=-\gamma dt\,\hat{\varphi}_t^{(1)}-\gamma dt\,e^{-\gamma t}\left[\hat{\varphi}_0+\int_0^t e^{\gamma s}dW(s)\right]\\&\quad+dW(t).\end{aligned} \tag{A12}$$

This has the solution

$$\hat{\varphi}_t^{(1)}=e^{-\gamma t}\int_0^t ds\left\{e^{\gamma s}\xi(s)-\gamma\left[\hat{\varphi}_0+\int_0^s e^{\gamma r}dW(r)\right]\right\}. \tag{A13}$$

The mark I phase estimate is, in this approximation, given by $\phi_{\rm I}=\hat{\varphi}_1^{(0)}+\gamma\tau\hat{\varphi}_1^{(1)}$. To leading order in τ and γ^{-1} this has a variance of

$$\langle\phi_{\rm I}^2\rangle=\frac{1}{2\gamma}+2\gamma\tau\langle\hat{\varphi}_1^{(0)}\hat{\varphi}_1^{(1)}\rangle, \tag{A14}$$

$$=\frac{1}{2\gamma}+\frac{\tau}{2}. \tag{A15}$$

In this scaled time the error due to a finite time delay in the feedback loop is thus of order τ. In real time, the error would be of order $\Gamma\tau$, where Γ is the characteristic bandwidth of the pulse.

Following the argument from the first section of this appendix, we take the mark II phase estimate to be

$$\phi_{\rm II}=\int_0^1[\hat{\varphi}_t^{(0)}+\gamma\tau\hat{\varphi}_t^{(1)}]dt. \tag{A16}$$

After considerable calculation we find that to leading order in τ and γ^{-1} the variance of this estimate is

$$\langle\phi_{\rm II}^2\rangle=\frac{1}{\gamma^2}+\frac{\langle\hat{\varphi}_0^2\rangle}{\gamma^2}+2\gamma\tau\left\langle\int_0^1 dt\int_0^1 dt'\,\hat{\varphi}_t^{(0)}\hat{\varphi}_{t'}^{(1)}\right\rangle, \tag{A17}$$

$$=\frac{1}{\gamma^2}+\frac{\langle\hat{\varphi}_0^2\rangle}{\gamma^2}+\frac{2\tau}{\gamma^2}. \tag{A18}$$

Thus in terms of real time the delay τ causes an error of order $\Gamma\tau/\alpha^2$.

[1]

[1] U. Leonhardt, J. A. Vaccaro, B. Böhmer, and H. Paul, Phys. Rev. A **51**, 84 (1995).

[2] It cannot be emphasized too much that there is an enormous difference between (i) a measurement which can be used to determine the canonical phase distribution $P(\phi_{\rm can})$ of a system by making repeated measurements of various observables on an ensemble comprising an arbitrarily large number of copies of the system, and (ii) a measurement on a single copy of the system which yields a single result ϕ drawn from some phase distribution $P(\phi)$ [which in practice will generally be broader than a canonical distribution $P(\phi_{\rm can})$]. Examples of the first sort of measurement include homodyne tomography [3] and projection synthesis [4]. In these measurements, the result of the measurement of any individual copy of the system in not a phase ϕ, but rather one piece of information (such as one result of measuring the quadrature operator X_Φ in the case of Ref. [3]) which, together with the results of all of the other measurements, determines the phase distribution. Of course, even for the second sort of measurement which yields a result ϕ from a single measurement, it would be necessary to make repeated measurements on a large ensemble of identically prepared systems in order to determine $P(\phi)$; to do otherwise would violate the foundations of quantum mechanics. But the purpose of a phase measurement is not to determine a phase distribution. Rather, it is to give a result which is *as good as estimate as possible* of the phase of the system. The goodness of the estimate will be limited by the intrinsic quantum phase uncertainty of the system, and the resolution of the measurement device.

[3] M. G. Raymer, M. Beck, and D. F. McAlister, Phys. Rev. Lett. **72**, 1137 (1994).

[4] S. M. Barnett and D. T. Pegg, Phys. Rev. Lett. **76**, 4148 (1996).

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

[6] S. J. Dolinar (unpublished). See also: Ref. [17], p. 163; and A.S. Holevo, Sov Math. **26**, 3 (1982).

[7] H. M. Wiseman, Mod. Phys. Lett. B **9**, 629 (1995).

[8] H. M. Wiseman, Quantum Semiclass. Opt. **8**, 205 (1996).

[9] In practice, it may be easier to use a constant local oscillator. This would then entail multiplying the difference photocurrent by the mode shape of the system $u(t)$. Such modulation would be incorporated with the high-pass filtering process which is necessary in any case to remove the inevitable dc component which arises from the lack of a perfect 50/50 beam splitter.

[10] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 1985).

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

[12] Alternative definitions of the phase variance have also been proposed [13,14] which are simple functions of the Holevo phase variance. These become identical to the Holevo variance in the limit $|\langle e^{i\theta}\rangle|\to 1$, and so all three are equivalent for the purposes of this paper.

[13] A. Bandilla, and H. Paul, Ann. Phys. (N.Y.) **23**, 323 (1969).

[14] T. Opatrný, J. Phys. A **27**, 7201 (1994).

[15] D. T. Pegg and S. M. Barnett, Phys. Rev. A **39**, 1665 (1989).

[16] E. B. Davies, *Quantum Theory of Open Systems* (Academic, London, 1976).

[17] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic, New York, 1976)

[18] F. London, Z. Phys. **40**, 193 (1927).

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

[20] G. M. D'Ariano and M. G. A. Paris, Phys. Rev. A **49**, 3022 (1994).