Adaptive single-shot phase measurements: The full quantum theory

H. M. Wiseman[1,*] and R. B. Killip[2]
[1]*Department of Physics, The University of Queensland, St. Lucia 4072, Australia*
[2]*Division of Physics, Mathematics, and Astronomy, California Institute of Technology, Pasadena, California 91125*
(Received 28 October 1997)

The phase of a single-mode field can be measured in a single-shot measurement by interfering the field with an effectively classical local oscillator of known phase. The standard technique is to have the local oscillator detuned from the system (heterodyne detection) so that it is sometimes in phase and sometimes in quadrature with the system over the course of the measurement. This enables both quadratures of the system to be measured, from which the phase can be estimated. One of us [H. M. Wiseman, Phys. Rev. Lett. **75**, 4587 (1995)] has shown recently that it is possible to make a much better estimate of the phase by using an adaptive technique in which a resonant local oscillator has its phase adjusted by a feedback loop during the single-shot measurement. In a previous work [H. M. Wiseman and R. B. Killip, Phys. Rev. A **56**, 944 (1997)] we presented a semiclassical analysis of a particular adaptive scheme, which yielded asymptotic results for the phase variance of strong fields. In this paper we present an exact quantum mechanical treatment. This is necessary for calculating the phase variance for fields with small photon numbers, and also for considering figures of merit other than the phase variance. Our results show that an adaptive scheme is always superior to heterodyne detection as far as the variance is concerned. However, the tails of the probability distribution are surprisingly high for this adaptive measurement, so that it does not always result in a smaller probability of error in phase-based optical communication. [S1050-2947(98)07303-X]

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

I. INTRODUCTION

In a typical textbook of quantum mechanics one might find a statement such as ''Every physical quantity $\mathcal{Z}$ has associated with it an Hermitian operator Z. A measurement of $\mathcal{Z}$ for a system with state matrix ρ will yield a result z, which is an eigenvalue of Z. The probability of getting the result z is equal to $\langle z|\rho|z\rangle$ where $Z|z\rangle = z|z\rangle$.'' Unfortunately the number of measurements of physical quantities for which this quantum measurement theory applies is very small. Nevertheless there are some in the context of quantum optics. It is only detector inefficiencies (now quite small) that limit the measurement of the photon number with operator $a^\dagger a$ and quadratures with operators such as $X = a + a^\dagger$ for single-mode optical fields. The former can be measured by direct photon counting and the latter by adding an essentially classical field of known phase (called the local oscillator) to the quantum field before counting photons (see, for example, Ref. [1]).

There is one obvious optical quantity of which we cannot make a quantum-limited measurement: the phase ϕ of the electromagnetic field. Despite the difficulties in defining a phase operator (which can be overcome [2]), the ''phase eigenstates'' $|\phi\rangle$ are independent of any phase operator (see Sec. II B) and have been recognized for a very long time [3]. The opinion is sometimes expressed that the reason one cannot measure phase is that the phase eigenstates do not have (even approximately) compact support on the number states, so that a measurement of phase would require infinite energy. This argument is specious, because the eigenstates of $a + a^\dagger$ also do not have compact support on the energy eigenstates, and yet in the limit of infinite local oscillator strength and perfect photodetection a homodyne measurement approaches a quadrature measurement. Nevertheless it is true that phase cannot be measured exactly, even in these ideal limits. The reason for this will be explored in the discussion section.

Although the quantum phase of a single mode field cannot be measured exactly, it can be measured approximately. As well as being interesting for theoretical reasons, there may be practical reasons for wishing to measure phase. For example, quantum-limited communication could be possible by encoding information in the phase of single-mode pulses of light. The first requirement for such a scheme would be to create states with very well-defined phase. This has been investigated by various authors (see Ref. [4] for some of these). The next step would be encoding the signal, which is easy to do using an electro-optic modulator. The third requirement is for the receiver to measure the encoded phase as accurately as possible. This is a problem that seems not to have received the amount of attention it deserves, given that it is as important to communication as the generation of states with well-defined phase. Another application for accurate phase measurements could be in inferring the properties of other quantum systems that can cause a phase shift, such as the presence of an atom at a particular point in a single-mode standing wave.

The standard way of measuring phase (approximately) is to use two simultaneous homodyne measurements of orthogonal quadratures (known as eight-port homodyne detection), or heterodyne detection, which are equivalent in an appropriate limit [5]. A way to improve upon this was first suggested by one of us [6]: single-shot adaptive measurements. By this we mean the use of measurement results from earlier stages of a *single measurement* to affect the condi-

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

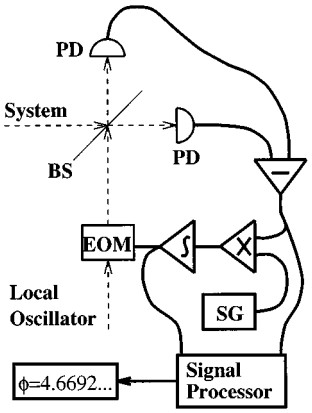

FIG. 1. Diagram for the experimental apparatus for making an adaptive phase measurement. Thin dashed lines indicate light rays and the thin continuous line labeled BS represents a 50/50 beam splitter. Medium lines represent electro-optic devices: photodetectors (PD) and an electro-optic phase modulator (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. III B.

tions of the measurement in its later stages. In this case it means using the photocurrent up to time t to control the local oscillator phase at time t by a feedback loop, during the detection of a single single-mode pulse. In Ref. [7] we investigated a particular feedback algorithm, illustrated in Fig. 1, using semiclassical theory. We showed that for large fields an adaptive measurement is a much closer approximation to a true phase measurement than is heterodyne detection.

In this paper we continue our analysis of the simple adaptive algorithm, but this time we present the full quantum theory of these adaptive phase measurements. The background theory required is presented in Sec. II. This introduces the theory of probability operator measures (POMs), which is required for approximate measurements. It also summarizes the theory of POMs for phase measurements and POMs for measurements using a large local oscillator. In Sec. III we derive expressions for the POMs for the two adaptive phase measurement schemes of Ref. [7]. In Sec. IV we use these POMs to calculate phase variances, for coherent states, and for phase-optimized states with an upper bound on the photon number. We compare our exact (quantum) numerical results to the asymptotic (semiclassical) analytical results obtained in Ref. [7]. One feature that can only be calculated using the full quantum theory is the overall shape of the probability distributions, including the tails. This is required for determining the probability of error in phase communication schemes. This aspect is investigated in Sec. V, again for coherent states and for phase-optimized states with an upper bound on the photon number. Section VI concludes with a discussion on the ultimate limits to phase measurements.

II. PROBABILITY-OPERATOR MEASURES

A. General theory of POMs

If (as in the present case) we are unconcerned about the fate of the system after it has been measured, then any measurement is completely described by the probability for each of the possible results to occur. Let the set of all possible measurement results λ be denoted Ω. Then the measurement is specified by a probability measure (PM) on Ω. If we denote the PM as P then for any subset $E \subseteq \Omega$, we can identify $P(E)$ as the probability to obtain a measurement result $\lambda \in E$. Of course this requires $P(\Omega)=1$.

For quantum mechanical systems, the most general way of generating a PM P is as the expectation value of an operator measure F on Ω. That is, for a quantum system with state matrix ρ,

$$P(E)=\mathrm{Tr}[\rho F(E)]. \tag{2.1}$$

Obviously $F(E)$ must be a positive operator, and by conservation of probability

$$F(\Omega)=1. \tag{2.2}$$

For this reason we call F a probability operator measure (POM), or sometimes an *effect*-valued measure [8,9]. Note that even for a subset E with a single element λ, $F(\lambda)$ is not necessarily a projector.

B. POMs for phase measurements

Now consider the case where the measured quantity is to be a phase ϕ of a single-mode photon field, so that F is a POM on $\Omega=[0,2\pi)$. Quantum mechanically this phase should in some sense be conjugate to the photon number operator $a^\dagger a$, but as long as we stick with POMs to describe the measurement there are none of the difficulties associated with defining a phase operator [2]. Since phase is a continuous variable, we will use $F(\phi)$ to denote the phase POM density. The completeness relation for a phase POM is therefore written as

$$\int_0^{2\pi} d\phi F(\phi)=1. \tag{2.3}$$

As explained in Ref. [7], for $F(\phi)$ to be invariant under phase shifts, and to be unbiased, implies that it can be written in the form

$$F(\phi)=\frac{1}{2\pi}\sum_{n,m=0}^{\infty} |m\rangle\langle n| e^{i\phi(m-n)} H_{mn}. \tag{2.4}$$

Here H is a positive-semidefinite Hermitian matrix with all entries real and positive, and $|m\rangle$ is the number state $a^\dagger a|m\rangle = m|m\rangle$.

The completeness condition (2.3) implies that

$$\forall m \geqslant 0 \quad H_{mm}=1. \tag{2.5}$$

The positivity condition on the matrix H 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 [5], although its special role was recognized very early in the history of quantum theory [3].

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 in Sec. IV. In fact, in all of the measurements we examine, we have

$$h(m)\equiv 1-H_{m,m+1}\leqslant O(m^{-1/2}). \tag{2.6}$$

For a canonical measurement $h(m)$ is identically zero. In this case we can write the POM (2.4) as

$$F^{\rm can}(\phi)=\frac{1}{2\pi}|\phi\rangle\langle\phi|, \tag{2.7}$$

where $|\phi\rangle$ is an unnormalized phase eigenstate

$$|\phi\rangle=\sum_{n=0}^{\infty} e^{in\phi}|n\rangle \tag{2.8}$$

as referred to in the Introduction.

C. POMs for dyne measurements

We now turn from the POMs for phase measurements of a single-mode field to the POMs for measurements on a single-mode photon field made by interfering the light from that field with another field that has a macroscopic coherent excitation. This can be done at a beam splitter, and the two output fields of the beam splitter can then be detected by normal photodetectors. The second field can be treated classically as a c number, and is known as a local oscillator. All practical phase-sensitive measurements require a local oscillator, to act as a phase reference. If the local oscillator is resonant with the system field then this type of measurement is known as *homodyne* detection. If the local oscillator is detuned (outside the bandwidth of the system field) then this is known as *heterodyne* detection. In considering phase measurements we will have to consider other sorts of measurements involving interference with a quasiclassical local oscillator. In ignorance of any received term for such measurements we will call them examples of *dyne* detection, so that homodyne and heterodyne are obviously special cases.

Let us assume that our single-mode signal field has a temporal pulse shape $u(t)$ that is positive and normalized as

$$\int_0^T u(t)=1. \tag{2.9}$$

Here we are obviously ignoring the phase variation at optical frequency ω; $u(t)$ is the envelope function. The total time T is necessarily much greater than ω^{-1}, so that the pulse can be considered monochromatic. This is essential in order for the dyne measurements (which are phase-sensitive measurements) to be quantum limited. That is, for quantum effects to provide the limit to the phase uncertainty in the measurement. If the characteristic spectral width of the pulse $\Gamma \gtrsim T^{-1}$ is too large then the phase uncertainty will be dominated by the term $\delta\phi\sim\Gamma/\omega$ coming from the uncertainty Γ in the frequency. In all that follows we assume this uncertainty to be negligible.

For simplicity we will take the beam splitter at which the system and local oscillator fields are interfered to be balanced (50/50). Then, ignoring vacuum fluctuations, the two fields at the two output ports of the beam splitter are equal to

$$b_{\pm}(t)=\sqrt{u(t)/2}(a\pm\beta e^{i\Phi(t)})e^{-i\omega t}, \tag{2.10}$$

where a is the annihilation operator for the system and the real number β is the coherent amplitude of the local oscillator. This is normalized so that the instantaneous rate of photodetection at each detector is $\langle b_{\pm}^{\dagger}(t)b_{\pm}(t)\rangle$. We have assumed that the intensity profile of the local oscillator is the same as that of the system. However, we have included an arbitrary phase variation $\Phi(t)$ of the local oscillator relative to the system. The total number of photons in the local oscillator is β^2, so we are interested in the limit $\beta^2\gg 1,\langle a^{\dagger}a\rangle$. For homodyne detection $\Phi(t)=\Phi_0$, a constant. For heterodyne detection $\Phi(t)=\Phi_0+t\Delta$, where $\Delta\gg\Gamma$ is the detuning.

The signal of interest is simply the difference between the two photocurrents at the two detectors (labeled $\pm$). If we denote the number of photocounts at each of the detectors in the time interval $[t,t+\delta)$ by $\delta N_{\pm}(t)$ then we can define the signal photocurrent as

$$I(t)=\lim_{\delta t\to 0}\lim_{\beta\to\infty}\frac{\delta N_{+}(t)-\delta N_{-}(t)}{\beta\,\delta t}. \tag{2.11}$$

Note that the two limits here do not commute. The limit $\beta\to\infty$ implies that both photocounts will be dominated by the contribution from the local oscillator. The fact that the limit $\delta t\to 0$ is taken second indicates that we are only interested in the fluctuations in $I(t)$ on a time scale much greater than the mean time $\sim u(t)^{-1}\beta^{-2}$ between photodetections.

The general quantum theory of dyne measurements was derived by one of us in Ref. [10] for the case where the system mode is derived from an exponentially decaying cavity so that $u(t)=\gamma e^{-\gamma t}$ where γ is the cavity linewidth. This is easily generalized for arbitrary $u(t)$. First we define a scaled time variable

$$v=\int_0^t u(s)ds. \tag{2.12}$$

This is dimensionless, and increases monotonically with t from 0 to 1. For the case $u(t)=\gamma e^{-\gamma t}$ we have $v=1-e^{-\gamma t}$. The photocurrent in terms of v is scaled so that

$$I(v)dv=I(t)dt=dv\ I(t)/u(t). \tag{2.13}$$

Now the measurement result for a dyne measurement up to time t is the complete photocurrent record $I(t')$ from $t'=0$ to $t'=t$ [or equivalently, $I(v')$ from $v'=0$ to $v'=v$)]. This record is, in theory at least, a continuous infinity of real numbers, which is an impractically huge amount of data. Fortunately it turns out that there are just two sufficient statistics at scaled time v (henceforth called simply time), namely, the two complex numbers

$$A_v=\int_0^v I(u)e^{i\Phi(u)}du, \tag{2.14}$$

$$B_v = -\int_0^v e^{2i\Phi(u)}du. \tag{2.15}$$

We call these the sufficient statistics because, as shown in Ref. [10], the POM for the measurement at time $0\leqslant v<1$ is given by

$$G_v(A_v,B_v) = Q_v(A_v,B_v)\exp(\tfrac{1}{2}B_v a^{\dagger 2}+A_v a^\dagger) \times(1-v)^{a^\dagger a/2}\exp(\tfrac{1}{2}B_v^* a^2+A_v^* a), \tag{2.16}$$

where $Q_v(A_v,B_v)$ is a positive function, which will be defined shortly. This implies that the probability for obtaining any photocurrent $\{I(u):0\leqslant u<v\}$ is determined only by the two complex functionals of this current A_v and B_v. Any other features of $\{I(u):0\leqslant u<v\}$ are completely irrelevant.

It might be thought that the second integral B_v does not depend on $\{I(u):0\leqslant u<v\}$ at all because the photocurrent does not appear explicitly in Eq. (2.15). However, it may appear implicitly if the local oscillator phase $\Phi(v)$ depends upon $\{I(u):0\leqslant u<v\}$. This is precisely the situation we will consider later to construct a phase measurement. When we do so, the theory presented here shows that $\Phi(v)$ should be made to depend on $\{I(u):0\leqslant u<v\}$ only through the two integrals (2.14),(2.15). That is to say, we should have

$$\Phi(v) = f_v(A_v,B_v) \tag{2.17}$$

for some (possibly time-dependent) function f. This is an extremely powerful result, which is not at all intuitive.

In the limit $v\to 1$, $(1-v)^{a^\dagger a/2}\to|0\rangle\langle 0|$, where $|0\rangle$ is the vacuum state. So, dropping the subscript v when $v=1$, we can write the POM (2.16) as

$$G(A,B) = Q(A,B)|\tilde{\psi}(A,B)\rangle\langle\tilde{\psi}(A,B)|, \tag{2.18}$$

where $|\tilde{\psi}(A,B)\rangle$ is an unnormalized ket defined by

$$|\tilde{\psi}(A,B)\rangle = \exp(\tfrac{1}{2}Ba^{\dagger 2}+Aa^\dagger)|0\rangle. \tag{2.19}$$

With a little operator algebra it is easy to show that this is proportional to the squeezed state [11]

$$|\alpha,\epsilon\rangle = \exp(\alpha a^\dagger-\alpha^* a)\exp(\tfrac{1}{2}\epsilon^* a^2-\epsilon a^{\dagger 2})|0\rangle, \tag{2.20}$$

where

$$\alpha = \frac{A+BA^*}{1-|B|^2}, \tag{2.21}$$

$$\epsilon = \frac{-B\ \mathrm{atanh}|B|}{|B|}. \tag{2.22}$$

From Eq. (2.15), it is evident that $|B|\leqslant 1$. For the schemes we will consider $|B|<1$ with probability one, so that the two expressions (2.21), (2.22) are well defined.

If we rewrite the POM (2.16) in terms of α,ϵ instead of A,B, we have

$$G'(\alpha,\epsilon) = Q'(\alpha,\epsilon)|\alpha,\epsilon\rangle\langle\alpha,\epsilon|, \tag{2.23}$$

where Q' is some new positive function of α,ϵ. In this case the set of all measurement results is $\Omega=\mathbf{C}\otimes\mathbf{C}$, where $\mathbf{C}$ denotes the set of complex numbers. If we imagine varying the state of the system $|\psi\rangle$ (assumed pure), then the probability to obtain the result α,ϵ is

$$P(\alpha,\epsilon)\propto|\langle\alpha,\epsilon|\psi\rangle|^2. \tag{2.24}$$

Provided $\exp(|\epsilon|)\ll|\alpha|$, the squeezed state $|\alpha,\epsilon\rangle$ has a well-defined coherent amplitude α. Hence from Eq. (2.24) if the unknown system state $|\psi\rangle$ is also localized in the phase plane, it is highly likely that it must have a coherent amplitude close to α. This fact will be used later to good effect.

We must now address the issue of how $Q(A,B)$ is found. In Ref. [10] it is shown that $Q(A,B)$ is the joint probability distribution that A,B would have if the photocurrent $I(v)$ were given by

$$I(v)dv = dW(v), \tag{2.25}$$

where $dW(v)$ is the infinitesimal increment in a real Wiener process [12] satisfying

$$\langle dW(v)\rangle = 0, \tag{2.26}$$

$$dW(v)dW(v) = dt. \tag{2.27}$$

In Ref. [10], $Q(A,B)$ was called the *ostensible* probability distribution for A,B. It is the probability distribution that A,B would have if there were no signal whatsoever; that is, if the system were prepared in the vacuum state. The noise in Eq. (2.25) then represents the local oscillator shot noise (or vacuum fluctuations if a Heisenberg picture interpretation is preferred). The presence of a nonzero signal determines the *actual* probability distribution through the POM (2.18). That is to say, if the system state matrix is ρ then the true probability density is

$$P(A,B)d^2Ad^2B = Q(A,B)\langle\tilde{\psi}(A,B)|\rho|\tilde{\psi}(A,B)\rangle d^2Ad^2B. \tag{2.28}$$

Before moving onto specific examples in the following section, we will derive some general results regarding the ostensible distribution $Q(A,B)$. First, the ostensible mean of A is

$$\langle A\rangle_Q = \int_0^1\langle e^{i\Phi(v)}dW(v)\rangle = 0. \tag{2.29}$$

This holds true even if $\Phi(v)$ depends on the photocurrent record $\{I(u):0\leqslant u<v\}$ because $W(v)$ is a strictly Markovian process. Second,

$$\langle A^2\rangle_Q = \int_0^1\int_0^1\langle e^{i\Phi(v)+i\Phi(u)}dW(u)dW(v)\rangle = \int_0^1 dv\langle e^{2i\Phi(v)}\rangle = -\langle B\rangle_Q. \tag{2.30}$$

Third,

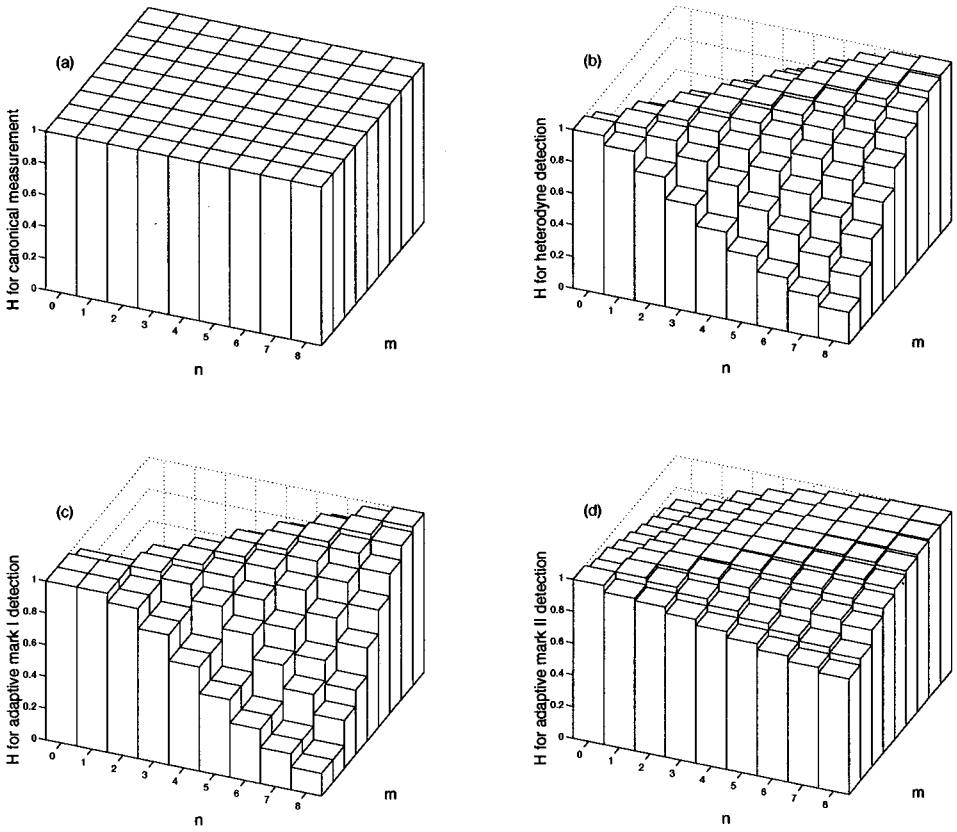

FIG. 2. Plot of the H matrix that defines the POM for phase measurements as in Eq. (2.4), for the four schemes (a) canonical, (b) heterodyne, (c) adaptive mark I, and (d) adaptive mark II.

$$\langle |A|^2 \rangle_Q = \int_0^1 \int_0^1 \langle dW(u) dW(v) \rangle = \int_0^1 dv = 1. \quad (2.31)$$

III. PHYSICALLY REALIZABLE PHASE MEASUREMENTS

A. Heterodyne measurements

As noted in Sec. II B the ideal form of phase measurement is a canonical phase measurement in which H_{mn} from Eq. (2.4) is equal to unity for all m,n. This is plotted in Fig. 2(a). All physically realizable phase measurements fall short of this ideal. The simplest method for making a phase measurement is via heterodyne detection. As explained above, this involves a local oscillator that is far detuned from the system. The linear variation of the phase is in fact not essential; all that is required is that all relative phases (of the system with respect to the local oscillator) be sampled equally and on a time scale much shorter than the reciprocal bandwidth of the system. As long as there is a record of the local oscillator phase as a function of time, the information in the photocurrent record can be recovered. For definiteness, however, we will take the local oscillator phase to simply change linearly with (scaled) time v. That is,

$$\Phi(v) = \Phi_0 + v\Delta, \quad (3.1)$$

where $\Delta \gg 1$.

Having specified $\Phi(v)$ all that remains to completely describe this heterodyne measurement is to determine $Q(A,B)$, the ostensible probability distribution for the measurement results A,B. Because the above $\Phi(v)$ is independent of the photocurrent I, the ''result'' B is a constant (rather than a random variable) with value

$$B = -\int_0^1 dv \exp[2i(\Phi_0 + v\Delta)] \quad (3.2)$$

$$= \exp(2i\Phi_0)\frac{1-\exp(2i\Delta)}{2i\Delta} \to 0, \quad (3.3)$$

where the final limit results from taking $\Delta \to \infty$. The only variable in this case is therefore

$$A = \int_0^1 dv I(v) \exp[i(\Phi_0 + v\Delta)]. \quad (3.4)$$

To find the ostensible statistics for A we treat $I(v)dv$ as an independent Gaussian variable $dW(v)$ for each infinitesimal interval $[v, v+dv)$. Since A is just the sum of these Gaussian variables, it must ostensibly be a Gaussian variable itself. From Eqs. (2.29)–(2.31) with $B=0$ it follows that the ostensible distribution for A is the rotationally invariant Gaussian

$$Q^{\text{het}}(A) d^2A = \pi^{-1} \exp(-|A|^2) d^2A. \quad (3.5)$$

From these results and Eq. (2.18) we find the POM for heterodyne measurements to be

$$G^{\rm het}(A)=\pi^{-1}\exp(-|A|^2)|\tilde{\psi}(A,0)\rangle\langle\tilde{\psi}(A,0)|. \quad (3.6)$$

Now from Eqs. (2.20)–(2.22) it is easy to verify that $|\tilde{\psi}(A,0)\rangle$ is simply proportional to the coherent state $|A\rangle$ where A is the coherent amplitude usually denoted α. It turns out that the proportionality factor is just $\exp(|A|^2/2)$ so that we can rewrite Eq. (3.6) as

$$G^{\rm het}(A)=\pi^{-1}|A\rangle\langle A|. \quad (3.7)$$

This result has been obtained many times before by other means; for one example see Ref. [1]. The factor of π^{-1} remains because the coherent states are overcomplete.

In the context of this paper we are interested in heterodyne measurements only insofar as they enable us to make an estimate of the phase of the system. If there is no prior information about the system then Eq. (3.7) suggests a good estimate of the phase to be

$$\phi_{\rm het}=\arg A. \quad (3.8)$$

The POM for this phase estimate is found simply by marginalizing the modulus of A. That is,

$$F^{\rm het}(\phi)=\int_0^\infty |A|\ d(|A|)\ G^{\rm het}(|A|e^{i\phi}). \quad (3.9)$$

Evaluating this in the number state basis yields the matrix H of Eq. (2.4) to be

$$H^{\rm het}_{mn}=\frac{\Gamma[(n+m)/2+1]}{\sqrt{n!m!}}. \quad (3.10)$$

Clearly $H^{\rm het}_{nn}=1$, as required, while the off-diagonal elements decrease with distance away from the diagonal. These features can be seen in the matrix plot of $H^{\rm het}_{mn}$ in Fig. 2(b).

B. Adaptive measurements

A heterodyne phase measurement is not as good as a canonical measurement because it is actually a measurement of both phase and amplitude, with the latter information being thrown away. In order to make a better phase measurement one would like to concentrate on measuring the phase quadrature. This can be done by homodyne detection [7], but only if one already knows the phase of the system. A true phase measurement should work even if one has no information about the system phase. Nevertheless we can use this idea to construct a true phase measurement as follows. Rather than measuring a fixed quadrature, we control the local oscillator phase as a function of time in order to measure the *estimated* phase quadrature. That is, we set $\Phi(v)$ to be equal to

$$\Phi(v)=\hat{\varphi}(v)+\pi/2, \quad (3.11)$$

where $\hat{\varphi}(v)$ is the *estimated* phase of the system at time v.

Two questions remain to be decided. First, given our measurement record $\{I(u):0\leq u<v\}$ how do we decide $\hat{\varphi}(v)$? Second, what do we choose to be our best estimate of phase ϕ once the measurement is completed? We will postpone answering the second question. It was already noted above that the theory of dyne measurements implies that we should choose $\hat{\varphi}(v)=f_v(A_v,B_v)$ for some function f. For the remainder of this paper we choose

$$\hat{\varphi}(v)=\arg A_v, \quad (3.12)$$

as in Ref. [7]. As outlined in that reference, the motivations for this choice are as follows: (1) It is suggested by the above analysis for heterodyne detection. (2) As shown by one of us [6], it reproduces the canonical result if the system has at most one photon. (3) It gives a feedback algorithm that would be easy to implement experimentally. (4) It is mathematically tractable. When we say it can be exactly solved, we mean that we can determine the POM. To do this requires only the ostensible probability distribution $Q^{\rm ad}(A,B)$ given the feedback algorithm, Eqs. (3.11) and (3.12). To find this it is convenient to recast the ostensible integral equations (2.14),(2.15) as the ostensible Itô stochastic differential equations

$$dA_v=e^{i\Phi(v)}dW(v), \quad (3.13)$$

$$dB_v=e^{2i\Phi(v)}dv, \quad (3.14)$$

with the initial conditions

$$A_0=B_0=0. \quad (3.15)$$

With the above feedback algorithm we have $e^{i\Phi(v)}=iA_v/|A_v|$. This gives

$$dA_v=iA_v dW(v)/|A_v|. \quad (3.16)$$

This can be solved by transforming to polar coordinates $\hat{\varphi}(v)=\arg A_v$ and $|A_v|^2$. Using the Itô calculus we find

$$d|A_v|^2=dv, \quad (3.17)$$

$$d\hat{\varphi}(v)=dW(v)/|A_v|. \quad (3.18)$$

The first of these can be solved trivially to yield $|A_v|=\sqrt{v}$. That is, the modulus of A evolves deterministically and in particular $|A|=1$, as required by Eq. (2.31). Substituting this into the second gives

$$\hat{\varphi}(v)=\hat{\varphi}(0)+\int_0^1 dW(v)/\sqrt{v}. \quad (3.19)$$

Here $\hat{\varphi}(0)$ is an arbitrary initial phase. It is irrelevant to the problem because the divergence at $v=0$ of the integrand in this equation means that the initial phase will be randomized immediately:

$$\langle\hat{\varphi}^2\rangle_Q=\int_0^1 dv/v=\infty. \quad (3.20)$$

Thus the ostensible probability distribution for A is

$$Q_a^{\rm ad}(A)d^2A=\delta(|A|-1)|A|d(|A|)\frac{1}{2\pi}d(\arg A). \tag{3.21}$$

We require the joint ostensible probability distribution $Q^{\rm ad}(A,B)$. But rather than work with B_v it is more convenient to consider the variable

$$C_v=e^{-2i\hat{\varphi}(v)}\int_0^v e^{2i\hat{\varphi}(u)}du. \tag{3.22}$$

It is easy to prove that for $v=1$

$$C=BA^*/A, \tag{3.23}$$

so that A,C can replace A,B as the sufficient statistics. The advantage of the variable C_v is that, from Eq. (3.22) and Eq. (3.19), it obeys the stochastic Itô differential equation

$$dC_v=-\left[\frac{2idW(v)}{\sqrt{v}}+\frac{2dv}{v}\right]C_v+dv, \tag{3.24}$$

with the initial condition $C_0=0$. Since neither this initial condition nor the above differential equation involve the value of $\hat{\varphi}(0)$ (which is essentially random as noted above), the final value of C will be ostensibly independent of that of A. That is,

$$Q^{\rm ad}(A,C)=Q_a^{\rm ad}(A)Q_c^{\rm ad}(C). \tag{3.25}$$

In fact, given the above result Eq. (3.21) we need only $\hat{\varphi}=\arg A$ so that

$$Q^{\rm ad}(A,C)d^2A\ d^2C\rightarrow\frac{d\hat{\varphi}}{2\pi}Q_c^{\rm ad}(C)d^2C. \tag{3.26}$$

The problem remaining is thus to find $Q_c^{\rm ad}(C)$. It has not proven possible to find this analytically. However, we have been able to find the exact values of the moments

$$M_v^{n,m}=\langle C_v^n C_v^{*m}\rangle_Q \tag{3.27}$$

via a recurrence relation. This is done in Appendix A. For our purposes these moments are sufficient so we can assume the distribution $Q_c(C)$ known. From Eq. (2.18) The POM for the results $\hat{\varphi},C$ under the feedback algorithm (3.11) and (3.12) is thus

$$G^{\rm ad}(\hat{\varphi},C)\ d\hat{\varphi}\ d^2C=|\tilde{\psi}(e^{i\hat{\varphi}},e^{2i\hat{\varphi}}C)\rangle\langle\tilde{\psi}(e^{i\hat{\varphi}},e^{2i\hat{\varphi}}C)|\times\frac{d\hat{\varphi}}{2\pi}d^2CQ_c^{\rm ad}(C). \tag{3.28}$$

Since the point of this exercise is to construct a phase measurement, we want ultimately to calculate some phase $\phi_{\rm ad}(\hat{\varphi},C)$ from the sufficient statistics $\hat{\varphi},C$. We are not constrained to choose $\hat{\varphi}$ even though we have been using it as our estimated phase in the feedback loop. Therefore the general expression for the POM of our adaptive phase measurement is

$$F^{\rm ad}(\phi)=\int_0^{2\pi}d\hat{\varphi}\int\int d^2C\ G^{\rm ad}(\hat{\varphi},C)\,\delta(\phi-\phi_{\rm ad}(\hat{\varphi},C)). \tag{3.29}$$

There are constraints on the function $\phi_{\rm ad}(\hat{\varphi},C)$. Clearly if the phase of the state ρ is rotated by some angle θ, the probability distribution $P_{\rm ad}(\phi)={\rm Tr}[\rho F_{\rm ad}(\phi)]$ for ϕ should be shifted similarly. Now to rotate the phase of the state by θ is equivalent to rotating that of the POM by $-\theta$. This has the effect of replacing $|\tilde{\psi}(e^{i\hat{\varphi}},e^{2i\hat{\varphi}}C)\rangle$ by

$$e^{-i\theta a^\dagger a}|\tilde{\psi}(e^{i\hat{\varphi}},e^{2i\hat{\varphi}}C)\rangle=|\tilde{\psi}(e^{i(\hat{\varphi}-\theta)},e^{2i(\hat{\varphi}-\theta)}C)\rangle. \tag{3.30}$$

Thus the distribution $P_{\rm ad}(\phi)$ will shift by the desired amount if and only if $\phi_{\rm het}$ is given by

$$\phi_{\rm het}(\hat{\varphi},C)=\hat{\varphi}+g(C), \tag{3.31}$$

for some arbitrary real function g of C. Furthermore, it can be shown that for H_{mn} to be real and positive we need $g(C^*)=-g(C)$.

1. Adaptive mark I measurements

The simplest choice is $g=0$. This corresponds to

$$\phi_{\rm I}=\hat{\varphi}=\arg A. \tag{3.32}$$

That is, the phase estimate $\hat{\varphi}$ used in the feedback loop is also used as the final phase estimate. We call this the adaptive mark I measurement. In this case the POM is

$$F^{\rm I}(\phi)=\int\int d^2C\ G^{\rm ad}(\phi,C).$$

$$=\int\int d^2CQ_c(C)|\tilde{\psi}(e^{i\phi},e^{2i\phi}C)\rangle\langle\tilde{\psi}(e^{i\phi},e^{2i\phi}C)|. \tag{3.33}$$

This POM can be easily evaluated in the number state basis using the definition (2.19). The result is in the form of Eq. (2.4) with the matrix H given by

$$H_{mn}^{\rm I}=\sum_{p=0}^{\lfloor m/2\rfloor}\sum_{q=0}^{\lfloor n/2\rfloor}\gamma_{mp}\gamma_{nq}\langle C^p(C^*)^q\rangle_Q, \tag{3.34}$$

$$=\sum_{p=0}^{\lfloor m/2\rfloor}\sum_{q=0}^{\lfloor n/2\rfloor}\gamma_{mp}\gamma_{nq}M^{p,q}. \tag{3.35}$$

Here $\lfloor m/2\rfloor$ is the integer part of $m/2$ and

$$\gamma_{mp}=\frac{\sqrt{m!}}{2^p(m-2p)!p!}. \tag{3.36}$$

This is an exact expression since the moments $M^{p,q}$ can be calculated exactly. It is not obvious from this definition $H_{nn}^{\rm I}=1$ for all n, but this can be verified computationally.

The matrix $H_{nn}^{\rm I}$ is plotted in Fig. 2(c). It appears not greatly different from that for the heterodyne measurement. One difference is that $H_{1,m}^{\rm I}=H_{0,m}^{\rm I}$ for all m, and in particular

that for $n,m \leqslant 1$, $H^{\mathrm{I}}_{n,m}=1$. This is identical to the canonical measurement and as good as possible, as first revealed in Ref. [6]. This result shows that for very weak fields the adaptive mark I measurement is significantly better than the standard heterodyne technique. For moderate fields it is not significantly better (as Fig. 2 shows). As we will show later, for large fields it is very much worse. Evidently the adaptive mark I scheme is not the scheme we would choose for most practical situations in which the photon number per pulse is very large.

2. Adaptive mark II measurements

A generally better result can be obtained by considering a final phase measurements $\phi_{\mathrm{ad}}=\hat{\varphi}+g(C)$ with $g(C)\neq 0$. Recall the result Eq. (2.24) obtained above, that the probability of obtaining a measurement result is proportional to the squared inner product of the system state with a squeezed state

$$P(\alpha,\epsilon)\propto|\langle\alpha,\epsilon|\psi\rangle|^2. \tag{3.37}$$

Here α,ϵ are defined in terms of A,B by Eqs. (2.21), (2.22). We are interested in the case when the state $|\psi\rangle$ has a well-defined (but unknown) phase. Since any physical state will have a finite mean photon number this means that it must have a large coherent amplitude. As argued in Sec. II C, it is most likely that this coherent amplitude will be close to α. Now in terms of the variables $\hat{\varphi},C$ we have

$$\alpha=\frac{e^{i\hat{\varphi}}(1+C)}{1-|C|^2}. \tag{3.38}$$

This suggests the mark II phase estimate

$$\phi_{\mathrm{II}}=\arg\alpha=\hat{\varphi}+\arg(1+C). \tag{3.39}$$

That is, we choose the function $g(C)$ so that

$$e^{ig(C)}=\sqrt{\frac{1+C}{1+C^*}}. \tag{3.40}$$

With this choice

$$F^{\mathrm{II}}(\phi)=\int\int d^2C\ G^{\mathrm{ad}}[\phi-\arg(1+C),C]. \tag{3.41}$$

The H matrix is therefore

$$H^{\mathrm{II}}_{mn}=\sum_{p=0}^{\lfloor m/2\rfloor}\sum_{q=0}^{\lfloor n/2\rfloor}\gamma_{mp}\gamma_{nq}\left\langle\left(\frac{1+C}{1+C^*}\right)^{(n-m)/2}C^p(C^*)^q\right\rangle_Q. \tag{3.42}$$

Unfortunately $[(1+C)/(1+C^*)]^{(n-m)/2}$ is not a polynomial in C and C^* so we cannot obtain an exact answer in terms of the known moments $M^{p,q}$. However, from the definition (3.22) it is apparent that the modulus of the random variable C is strictly bounded by unity. In fact $\langle C\rangle_Q=\langle C^*\rangle_Q=\langle C^*C\rangle_Q=1/3$, and all higher moments are smaller. Hence the MacLaurin series for $[(1+C)/(1+C^*)]^{(n-m)/2}$ will converge rapidly and so can be well approximated by a polynomial. Using an expansion to 100 terms, we have evaluated this POM matrix elements for n,m up to 100.

The matrix H^{II}_{mn} for n,m up to 8 is shown in Fig. 2(d). From this it is apparent that the adaptive mark II scheme is generally much closer to a canonical measurement in this range than are either the heterodyne or adaptive mark I scheme. Indeed, all the matrix elements are above 0.7, and all are greater than or equal to the heterodyne matrix elements. The only place where the adaptive mark II scheme is inferior to the adaptive mark I scheme is for very low photon numbers; $H^{\mathrm{II}}_{01}<1$ unlike H^{I}_{01}. We will show in the next section that the superiority of the mark II scheme over the other two schemes continues for large photon numbers, as quantified by the measured phase variance of various states.

IV. PHASE VARIANCE

A. Phase variance and H_{mn}

Because phase is a cyclic variable, the definitions of mean and variance that apply to the real line are not applicable. The sensible starting point for these two statistics for a cyclic variable with distribution $P(\phi)$ is

$$\mu=\int e^{i\phi}P(\phi)d\phi. \tag{4.1}$$

The mean phase can then be defined to be

$$\overline{\phi}=\arg\mu, \tag{4.2}$$

and the phase variance

$$V=|\mu|^{-2}-1. \tag{4.3}$$

It can easily be verified that these definitions go over to the usual ones appropriate for the real line when $P(\phi)$ is suitably localized (so that $1-|\mu|\ll 1$). There are of course other definitions of the variance in terms of $|\mu|$ that would also give the correct limit [13,14]. The advantage of the one presented here is that it can be used to derive an uncertainty relation

$$4V\geqslant(\langle a^\dagger a a^\dagger a\rangle-\langle a^\dagger a\rangle\langle a^\dagger a\rangle)^{-1}, \tag{4.4}$$

as shown by Holevo [15]. This inequality holds for the variance of any $P(\phi)$ arising from a phase measurement conforming to the definition in Sec. II B.

Without loss of generality we can consider a system state

$$|\psi\rangle=\sum_{n=0}^{\infty}\psi_n|n\rangle, \tag{4.5}$$

with real number state amplitudes ψ_n so that it is guaranteed to have a mean phase of zero. The probability distribution from a phase measurement described by a POM (2.4) with matrix H is

$$P(\phi)=\frac{1}{2\pi}\sum_{n,m=0}^{\infty}\psi_m\psi_n e^{i\phi(m-n)}H_{mn}. \tag{4.6}$$

For such a system we have

$$\mu=\sum_{n,m=0}^{\infty}\frac{1}{2\pi}\int d\phi e^{i\phi(m+1-n)}\psi_m\psi_n H_{mn} \tag{4.7}$$

$$=\sum_{n=0}^{\infty}\psi_{n+1}\psi_n H_{n+1,n}. \tag{4.8}$$

Thus the only part of H that contributes to the phase variance is the subdiagonal

$$H_{n+1,n}\equiv 1-h(n). \tag{4.9}$$

Although $h^{\mathrm{II}}(n)$ is not known exactly it was calculated to a very good approximation for n up to 100, as explained above. For heterodyne detection and adaptive mark I detection we have exact results and for a canonical phase measurement of course $h^{\mathrm{can}}(n)=0$. For large photon numbers it is more useful to have approximate asymptotic expressions for $h(n)$ for the three physically realizable schemes. These can be derived using semiclassical dyne detection theory [7]. The results are

$$h^{\mathrm{het}}(m)\simeq(8m)^{-1}+O(m^{-2}), \tag{4.10}$$

$$h^{\mathrm{I}}(m)\simeq(8m^{1/2})^{-1}+O(m^{-1}), \tag{4.11}$$

$$h^{\mathrm{II}}(m)\simeq(16m^{3/2})^{-1}+O(m^{-2}). \tag{4.12}$$

As will be shown in Secs. IV B and IV C this leads to a clear superiority of the adaptive mark II scheme over the heterodyne scheme, and of the latter over the mark I scheme, for measuring the phase of states with large photon numbers. Furthermore, it is shown at the end of Appendix B that the adaptive mark II scheme is the best scheme for measuring large fields given the feedback algorithm (3.12).

B. Coherent states

1. Canonical

A coherent state of mean phase equal to zero has coefficients

$$\psi_n=\exp(-\beta^2/2)\frac{\beta^n}{\sqrt{n!}}. \tag{4.13}$$

Thus for a canonical measurement we can use Eq. (4.8) with $H_{mn}=1$ to get

$$\mu=\exp(-\beta^2)\sum_{n=0}^{\infty}\frac{\sqrt{n}\beta^{2n}}{\beta n!}. \tag{4.14}$$

By expanding $\sqrt{n}$ in a Taylor series about $n=\beta^2$ while recognizing the moments of a Poisson distribution we obtain

$$\mu=1-\frac{1}{8\beta^2}-\frac{7}{128\beta^4}+O(\beta^{-6}). \tag{4.15}$$

Thus the variance from a canonical measurement of the phase of a coherent state is

$$V_{\mathrm{coh}}^{\mathrm{can}}=\frac{1}{4\beta^2}+\frac{5}{32\beta^4}+O(\beta^{-6}). \tag{4.16}$$

This can be regarded as the intrinsic phase variance of a coherent state. In Fig. 3 we have plotted the exact result obtained numerically from Eq. (4.14), and the asymptotic result Eq. (4.16) for β from 1 to 5. The latter corresponds to a mean photon number of 25, which is evidently large enough for the asymptotic results to hold quite well.

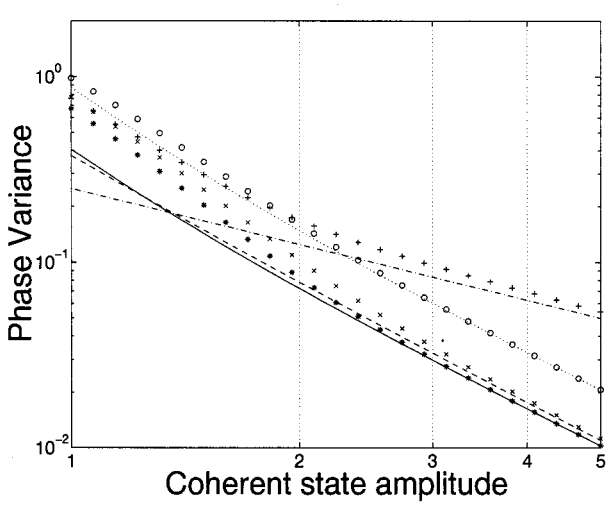

FIG. 3. Plot of the exact (points) and asymptotic (lines) expressions for the phase variance V_{coh} of a coherent state of amplitude β vs β under the four schemes: canonical (* and solid line), heterodyne (○ and dotted line), adaptive mark I (+ and dash-dotted line), and adaptive mark II (× and dashed line).

2. Heterodyne

For heterodyne detection we can use the exact expression Eq. (3.10) to get

$$\mu=\beta\exp(-\beta^2)\sum_{n=0}^{\infty}\frac{\Gamma\left(n+\frac{3}{2}\right)\beta^{2n}}{\Gamma(n+2)\Gamma(n+1)}. \tag{4.17}$$

In terms of confluent hypergeometric functions, this is

$$\mu=\beta\exp(-\beta^2)\frac{\Gamma\left(\frac{3}{2}\right)}{\Gamma(2)}{}_1F_1\left(\tfrac{3}{2};2;\beta^2\right). \tag{4.18}$$

Using the analog to Euler's formula, 4.2(1) of [16] asymptotic expansion

$$\mu=1-\frac{1}{4\beta^2}-\frac{3}{32\beta^4}+O(\beta^{-6}). \tag{4.19}$$

Thus the phase variance from a heterodyne measurement is

$$V_{\mathrm{coh}}^{\mathrm{het}}=\frac{1}{2\beta^2}+\frac{3}{8\beta^4}+O(\beta^{-6}). \tag{4.20}$$

To first (and almost to second) order this is twice that of the canonical phase variance. The reason for this is apparent from the expression Eq. (3.9) for the heterodyne POM. The probability distribution for a heterodyne phase measurement is

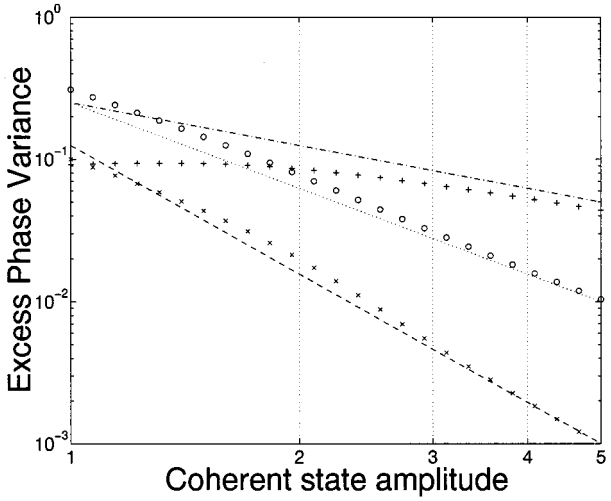

FIG. 4. Plot of the exact (points) and asymptotic (lines) expressions for the excess phase variance $V_{\rm coh}-V_{\rm coh}^{\rm can}$ of a coherent state of amplitude β vs β under the three dyne schemes: heterodyne ($\bigcirc$ and dotted line), adaptive mark I ($+$ and dash-dotted line), and adaptive mark II ($\times$ and dashed line).

$$P_{\rm het}^{\rm coh}(\phi)=\int_0^\infty |A|\ d(|A|)\ \langle\beta|F^{\rm het}(|A|e^{i\phi})|\beta\rangle \quad (4.21)$$

$$=\frac{1}{\pi}\int_0^\infty r\ dr|\langle\beta|re^{i\phi}\rangle|^2. \quad (4.22)$$

For ϕ close to the mean value of 0 the integrand will be strongly peaked at $r\simeq\beta\gg 1$. Thus

$$P_{\rm coh}^{\rm het}(\phi)\propto|\langle\beta|\beta e^{i\phi}\rangle|^2. \quad (4.23)$$

In other words, this distribution is approximately the convolution of the intrinsic phase distributions of two coherent states of amplitude β. Thus we expect the distribution to be approximately Gaussian, with a variance double that of a canonical measurement. The exact result from Eq. (4.18) and the asymptotic result Eq. (4.20) are plotted on Fig. 3. The excess phase noise in the heterodyne result is because the measurement is not as good as the canonical result. In fact, we have

$$V_{\rm coh}^{\rm het}-V_{\rm coh}^{\rm can}\simeq\frac{1}{4\beta^2}\simeq 2h^{\rm het}(\beta^2), \quad (4.24)$$

where $h(m)$ is the asymptotic expression for $H_{m,m+1}-1$ given in Eq. (4.10). The quantity in Eq. (4.24), which we will call the excess phase variance, is plotted in Fig. 4. From Eq. (4.8) it follows that, for states with a well-defined coherent amplitude, the excess phase variance for any scheme is approximately $2h(\beta^2)$.

3. Mark I adaptive

It was shown in Ref. [7] that for a coherent state of amplitude $\beta\gg 1$ the adaptive mark I phase $\hat{\varphi}$ can be approximated by a Gaussian random variable of mean zero and variance

$$V_{\rm coh}^{\rm I}=\frac{1}{4\beta}+O(\beta^{-2}). \quad (4.25)$$

This is plotted in Fig. 3 along with the exact result calculated from Eqs. (4.8) and (3.34) truncated at $n=100$. This result shows that the adaptive mark I is far worse than a heterodyne measurement for large β. Indeed, to the order calculated, the phase variance is entirely due to the excess phase variance

$$V_{\rm coh}^{\rm I}-V_{\rm coh}^{\rm can}=\frac{1}{4\beta}+O(\beta^{-2}). \quad (4.26)$$

This was the result used to obtain

$$h_{\rm I}(\beta^2)=\frac{1}{2}[V_{\rm coh}^{\rm I}-V_{\rm coh}^{\rm can}]=\frac{1}{8\beta}+O(\beta^{-2}), \quad (4.27)$$

as recorded above in Eq. (4.11). The asymptotic result (4.26) and its exact value are plotted in Fig. 4. This shows that for small coherent states, with amplitude less than about 2, the mark I measurement introduces less excess noise than the heterodyne measurement. For $\beta=5$ the asymptotic result is already a very good approximation.

4. Adaptive mark II

For our final scheme we again used semiclassical techniques in Ref. [7] to show that $P_{\rm coh}^{\rm II}(\phi)$ was approximately Gaussian with a variance

$$V_{\rm coh}^{\rm II}=\frac{1}{4\beta^2}+\frac{1}{8\beta^3}+O(\beta^{-4}). \quad (4.28)$$

Like the canonical result, this is dominated by the intrinsic phase noise of the coherent state. This asymptotic result, and the exact result from Eqs. (4.8) and (3.42), are plotted in Fig. 3. The excess phase noise in this case is

$$2h_{\rm II}(\beta^2)=V_{\rm coh}^{\rm II}-V_{\rm coh}^{\rm can}=\frac{1}{8\beta^3}+O(\beta^{-4}), \quad (4.29)$$

which is far below that of the other two dyne schemes. This asymptotic result, and the exact excess phase variance, are plotted in Fig. 4. Once again, the asymptotic behavior is evident for $\beta=5$.

C. Phase-optimized states

From the coherent state results, the marked superiority of the adaptive mark II measurement over the standard techniques is apparent only from considering the excess phase variance. A more direct measure is the *minimum* phase variance for each measurement scheme. In this measure, the state is optimized for each scheme, and is subject to the constraint of having a maximum photon number N. That is to say we have to optimize the unit-norm real vector $(\psi_0,\psi_1,\ldots,\psi_N)$ so as to *maximize*

$$\mu=\sum_{n=0}^{N}\psi_{n+1}\psi_n[1-h(n)]. \quad (4.30)$$

This can be rewritten as

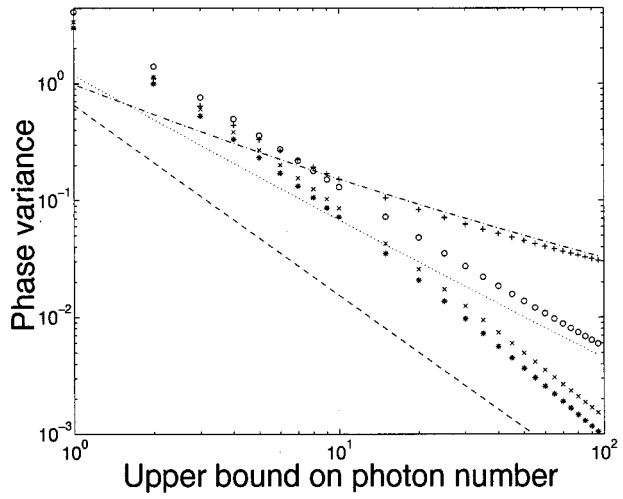

FIG. 5. Plot of the exact (points) and asymptotic (lines) expressions for the minimum phase variance $V_{\min}$ of the optimal state with at most N photons vs $N+1$ under the four schemes: canonical (* and solid line), heterodyne (○ and dotted line), adaptive mark I (+ and dash-dotted line), and adaptive mark II (× and dashed line).

$$\mu = \frac{1}{2}\sum_{m,n=0}^{N} \psi_m J_{mn} \psi_n , \tag{4.31}$$

where

$$J_{mn} = \frac{1}{2}[1-h(n)]\delta_{m,n+1} + \frac{1}{2}[1-h(m)]\delta_{m,n-1} . \tag{4.32}$$

The problem of maximizing μ thus reduces to that of finding the largest eigenvalue $\lambda_{\max}$ of the real symmetric matrix J. Since we have $h(n)$ for all schemes up to $n=100$ this can be done for a maximum photon number N up to 100.

For the canonical case with $h(m)=0$ the eigenvalue can be found exactly to be

$$\lambda_{\max} = \cos\left(\frac{\pi}{N+2}\right) \tag{4.33}$$

so that

$$V_{\min}^{\text{can}} = \tan^2\left(\frac{\pi}{N+2}\right) = \frac{\pi^2}{N^2} - 4\frac{\pi^2}{N^3} + O(N^{-4}). \tag{4.34}$$

For the dyne measurements there is no analytical solution but a numerical solution is easily obtained. The results are plotted in Fig. 5. This clearly shows the same order as established for coherent states with large photon numbers: the adaptive mark II measurement is best, followed by heterodyne, followed by adaptive mark I.

Also plotted in Fig. 5 are the asymptotic results for the three dyne measurements. These were obtained in Ref. [7] using the asymptotic results for $h(n)$ of Eqs. (4.10)–(4.12). The results are most easily expressed by noting that these functions $h(n)$ can all be written as

$$h^{\text{dyne}}(n) = cn^{-p} \tag{4.35}$$

for some positive power $p \geq 1/2$ and positive coefficient c of order unity. From this we got

$$V_{\min}^{\text{dyne}} \approx 2cN^{-p} + (-z_1)(2cp)^{2/3} N^{-2(1+p)/3}, \tag{4.36}$$

were $z_1 \approx -2.338$ is the first zero of the Airy function. The leading term here is simply equal to $2h(N)$. This is essentially the excess noise introduced by the measurement, just as $2h^{\text{dyne}}(\beta^2)$ was for the coherent state. In this case the intrinsic noise (the second term) varies between the different schemes because the state is optimized for each measurement.

From Fig. 5 it is apparent that the exact numerical results are approaching this asymptotic result for the heterodyne and mark I measurements. However, the mark II exact results are a long way from the asymptotic results even with $N=100$. This is actually not surprising. A simple calculation carried out in Ref. [7] suggested that the asymptotic results would only become valid for

$$N \gtrsim N_{\text{as}} = \left(\frac{10^3}{2cp}\right)^{1/(2-p)} . \tag{4.37}$$

For an adaptive mark I measurement we have $N_{\text{as}}=400$; for heterodyne $N_{\text{as}}=4000$; and for adaptive mark II $N_{\text{as}} \approx 3 \times 10^7$. Evidently these requirements are overly conservative (as noted in our earlier paper). Nevertheless, it does explain why the minimum adaptive mark II phase variance is a long way from reaching its asymptote for $N=100$. This underlines the usefulness of the approximate asymptotic results. An exact numerical solution with $N=10^7$ would be severely impractical. It also points out the danger of trying to derive power laws such as Eq. (4.36) from numerical data for moderate photon numbers of a few hundred, as done by D'Ariano and Paris in Ref. [17]. A detailed comparison with their results for heterodyne detection for optimized states with a fixed mean photon number will appear in a future paper.

V. PHASE PROBABILITY DISTRIBUTIONS

A. $P(\phi)$ for coherent states

Although the semiclassical theory of Ref. [7] has proven invaluable for calculating the asymptotic phase variance for states of large photon number, it cannot readily yield the total phase distribution $P(\phi)$. This is the quantity that is needed for a proper analysis of optical communication based on encoding information in the phase of single-mode pulses. For a communication system there are certain phases that one would be expecting to receive, so what matters is not the mean-square error in the phase measurement, but the probability for mistaking one phase for another. This depends on the total $P(\phi)$, which requires knowledge of the full matrix H_{mn}:

$$P(\phi) = \frac{1}{2\pi}\sum_{n,m=0}^{\infty} \rho_{mn} e^{i\phi(m-n)} H_{mn} , \tag{5.1}$$

where ρ_{mn} is the density matrix for the system state in the photon number basis.

Before calculating probabilities of error it is informative simply to plot $P(\phi)$ for the various schemes with the system

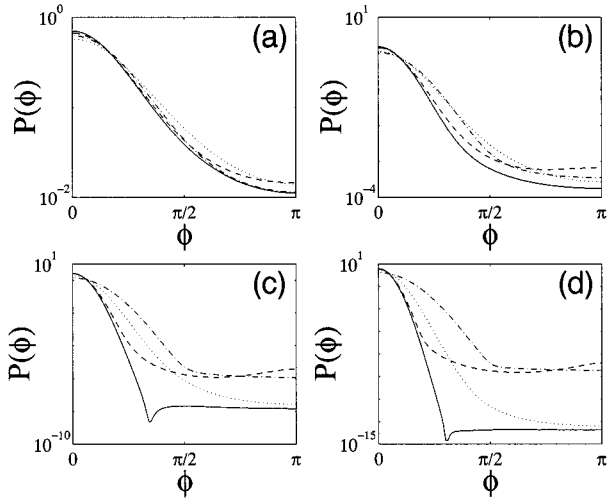

FIG. 6. Plot of the exact expressions for the log of the probability distribution $P_{\rm coh}(\phi)$ for coherent states under the four schemes: canonical (solid line), heterodyne (dotted line), adaptive mark I (dash-dotted line), and adaptive mark II (dashed line). The coherent amplitude is (a) $\beta=1$, (b) $\beta=2$, (c) $\beta=3.5$, (d) $\beta=5$.

in a coherent state. In Fig. 6 we plot $\log P_{\rm coh}(\phi)$ versus ϕ for various values of coherent amplitude β. One thing is clear: the canonical $P(\phi)$ is best by any definition. For small coherent amplitudes the adaptive mark I case is the best dyne measurement, and is almost indistinguishable from the canonical measurement. As β becomes larger the peak of $P^{\rm het}_{\rm coh}(\phi)$ becomes sharper and taller than that of $P^{\rm I}_{\rm coh}(\phi)$. The peak of $P^{\rm II}_{\rm coh}(\phi)$ becomes sharper and taller still, and for moderate β is indistinguishable from that of $P^{\rm can}_{\rm coh}(\phi)$. All of the curves are inverted parabolas for small ϕ, indicating that the distributions $P(\phi)$ are approximately Gaussian.

All of these features could be predicted from the above results. What is unexpected is the shape of the tails of the curves. First, as β increases, $P^{\rm can}_{\rm coh}(\phi)$ ceases to fall monotonically with distance from $\phi=0$, but suddenly reverses at $\phi\approx 1$ and has a broad local maximum at $\phi=\pi$. The heterodyne distribution has no such reversal, but nevertheless levels out and approaches the canonical value at $\phi=\pi$. The adaptive mark I case is also apparently smooth, but has much higher tails than the canonical heterodyne distributions. The big surprise is the adaptive mark II distribution. Like the canonical distribution it reverses (although smoothly) and has a broad local maximum at $\phi=\pi$. But the value of $P^{\rm II}_{\rm coh}(\pi)$ is actually the largest of all four schemes. In fact, for large β, $P^{\rm II}_{\rm coh}(\pi)$ closely follows $P^{\rm can}_{\rm coh}(\phi)$ until it reaches a floor, which is roughly the same as that of $P^{\rm I}_{\rm coh}(\phi)$.

These features are not easy to explain from the matrix elements H_{mn}. For example, the ratio of the probability density at $\phi=\pi$ to that at $\phi=0$ is given by

$$\frac{P(\pi)}{P(0)}=\frac{\sum_{mn} H_{mn}(-1)^{m-n}\beta^{m+n}/\sqrt{n!m!}}{\sum_{mn} H_{mn}\beta^{m+n}/\sqrt{n!m!}}. \tag{5.2}$$

Evidently this ratio depends crucially on the relative values of the matrix elements H_{mn} for $m,n\sim\beta^2$. In particular, just because $H^a_{mn}\geqslant H^b_{mn}\ \forall\ m,n$ it does not follow that $P^a(\pi)\leqslant P^b(\pi)$. That is, a measurement with a POM closer to the canonical POM, in the sense of having all elements of H_{mn} closer to unity, does not guarantee an unambiguously better phase probability distribution.

1. Heterodyne measurements

For heterodyne detection we can find an expression for $P(\pi)$ analytically. Recall that in this case the POM is

$$G'_{\rm coh}(\alpha)d^2\alpha=\frac{1}{\pi}|\alpha\rangle\langle\alpha|d^2\alpha, \tag{5.3}$$

where $|\alpha\rangle$ is a coherent state and the phase estimate is $\phi=\arg\alpha$. Clearly then the probability to obtain $\phi=\pi$ is

$$P^{\rm het}_{\rm coh}(\pi)=\frac{1}{\pi}\int_0^\infty r dr|\langle\beta|-r\rangle|^2 \tag{5.4}$$

$$=\frac{1}{\pi}\int_0^\infty r dr\exp[-(\beta+r)^2]. \tag{5.5}$$

This integral can be evaluated in terms of the error function, but for $\beta\gg 1$ it is well approximated by

$$P^{\rm het}_{\rm coh}(\pi)=\frac{1}{4\pi\beta^2}\exp(-\beta^2). \tag{5.6}$$

It can be verified from Fig. 6 that this is a very good approximation even for $\beta=5$. For very large β the most important contribution is the $\exp(-\beta^2)$ term. This scaling can be expressed as

$$\ln P^{\rm het}_{\rm coh}(\pi)\simeq-\beta^2. \tag{5.7}$$

2. Adaptive measurements

For the adaptive measurements we can also determine $P(\pi)$ by returning to the POM

$$G^{\rm ad}(\hat\varphi,C)\ d\hat\varphi\ d^2C=\frac{d\hat\varphi}{2\pi}d^2C Q_c(C)\times|\tilde\psi(e^{i\hat\varphi},e^{2i\hat\varphi}C)\rangle\langle\tilde\psi(e^{i\hat\varphi},e^{2i\hat\varphi}C)|, \tag{5.8}$$

where

$$|\tilde\psi(e^{i\hat\varphi},e^{2i\hat\varphi}C)\rangle=\exp(\tfrac{1}{2}e^{2i\hat\varphi}Ca^{\dagger 2}+e^{i\hat\varphi}a^\dagger)|0\rangle. \tag{5.9}$$

For a coherent state $|\beta\rangle$ with β real the probability density is

$$P_{\rm coh}^{\rm ad}(\hat{\varphi},C)=\frac{Q_c(C)}{2\pi}|\langle\beta|\tilde{\psi}(e^{i\hat{\varphi}},e^{2i\hat{\varphi}}C)\rangle|^2$$
$$=\frac{Q_c(C)}{2\pi}\exp(-\beta^2+\mathrm{Re}[e^{2i\hat{\varphi}}C\beta^2+2e^{i\hat{\varphi}}\beta]). \tag{5.10}$$

Consider first the adaptive mark I scheme for which $\phi=\hat{\varphi}$. The ratio of $P_{\rm coh}^{\rm I}(\pi)$ to $P_{\rm coh}^{\rm I}(0)$ is

$$\frac{P_{\rm coh}^{\rm I}(\pi)}{P_{\rm coh}^{\rm I}(0)}=\frac{\int\int d^2C P_{\rm coh}^{\rm ad}(\pi,C)}{\int\int d^2C P_{\rm coh}^{\rm ad}(0,C)} \tag{5.11}$$
$$=\frac{\int\int d^2C Q_c(C)\exp(-\beta^2+\mathrm{Re}[C\beta^2]-2\beta)}{\int\int d^2C Q_c(C)\exp(-\beta^2+\mathrm{Re}[C\beta^2]+2\beta)}$$
$$=\exp(-4\beta). \tag{5.12}$$

Now since $P_{\rm coh}^{\rm I}(\phi)$ is approximately Gaussian we have $P_{\rm coh}^{\rm I}(0)=(2\pi V_{\rm coh}^{\rm I})^{-1/2}=(\pi/4\beta)^{-1/2}$, so that

$$P_{\rm coh}^{\rm I}(\pi)\simeq\sqrt{4\beta\pi\exp(-4\beta)}. \tag{5.13}$$

This agrees excellently with the numerical result plotted in Fig. 6 for $\beta=5$. For very large β the dominant term is obviously the exponential, which we can express by the equation

$$\ln P_{\rm coh}^{\rm I}(\pi)\simeq-4\beta. \tag{5.14}$$

For the adaptive mark II scheme we expect the tail of the distribution to be at least as high as that for the adaptive mark I case, which is what is indeed seen. That is because

$$\phi=\hat{\varphi}+\arg(1+C), \tag{5.15}$$

and $\arg(1+C)$ lies between $-\pi/2$ and $\pi/2$. Thus irrespective of C, a result $\hat{\varphi}\approx\pi$ in the tail of the distribution of the mark I measurement must also give a result ϕ in the tail of the mark II measurement. By this crude argument we would also expect the log of the tail of the distribution of the mark II measurement to scale in the same way:

$$\ln P_{\rm coh}^{\rm II}(\pi)\simeq-4\beta. \tag{5.16}$$

Clearly the relative disparity between the height of tails of the adaptive measurements and those of the heterodyne or canonical measurements will continue to increase as β increases. A discussion about the reason for this disparity is to be found in Appendix B.

B. *M*-ary encoding with coherent states

As stated above, one reason for wishing to know the complete phase probability distributions, including the tails, is for calculating the effectiveness of the various schemes for digital communication using phase encoding. The canonical and heterodyne POMs have been examined before by Hall and Fuss [18]. Here we follow their approach, and consider M-ary encoding; that is, the transmission of data as the string of M-ary digits $\{0,1,\ldots,M-1\}$. Each digit is represented by a rotated version of some single quantum state $|\psi\rangle$ whose phase distribution is peaked about zero. The digit n is encoded as $\exp[(2in\pi/M)a^\dagger a]|\psi\rangle$. The receiver makes a phase measurement (as defined in Sec. II B) on this state and infers from the result which digit was sent. That is, a result ϕ in the interval $2\pi n/M\pm\pi/M$ is interpreted as the digit n.

The essential measure of any mode of digital communication is the probability that an error occurs. For each of the four measurement schemes we have calculated the minimal probability of error that may be achieved for each of two types of transmitted states. The first type is coherent states. These are important because, with the exception of squeezed states [11], they are perhaps the only pure single-mode quantum states that can be produced readily enough to be considered for communication applications.

Under the decoding scheme described above the probability of error is independent of the digit encoded. For the zero state it is

$$E=\int_{\pi/M}^{2\pi-\pi/M}P(\phi)d\phi. \tag{5.17}$$

It is easy to see that E is the expectation value of the positive operator $F_E=1-F_C$ where

$$F_C=\sum_{n,m=0}^{\infty}\frac{\sin[\pi(m-n)/M]}{\pi(m-n)}H_{m,n}|m\rangle\langle n|. \tag{5.18}$$

Using this operator, the expansion of a coherent state in terms of number states, and the values of $H_{m,n}$ for $0\leqslant m,n\leqslant 100$ computed earlier, one may easily determine the probability of error for coherent states with small β.

We can find approximate asymptotic analytic expressions for E by returning to Eq. (5.17). The logarithm of E will be well approximated by the logarithm of the largest value of the integrand in Eq. (5.17). Since $P(\phi)$ for coherent states is approximately monotonically decreasing from $\phi=0$ to $\phi=\pi$ for all schemes, we can thus say

$$\ln E_{\rm coh}\approx\ln P_{\rm coh}(\pi/M). \tag{5.19}$$

To proceed further we make the approximation that $P_{\rm coh}(\phi)$ is Gaussian until it hits the floor value $P(\pi)$. That is,

$$\ln P_{\rm coh}(\phi)\approx\max\{-\phi^2/2V_{\rm coh},\ln P_{\rm coh}(\pi)\}, \tag{5.20}$$

so that

$$\ln E_{\rm coh}\approx-\min\left\{\frac{\pi^2}{2M^2V_{\rm coh}},\ln P_{\rm coh}(\pi)\right\}. \tag{5.21}$$

From the results of Sec. IV B and Sec. V A we can evaluate this expression for the probability of error for the various schemes.

$$\ln E_{\rm coh}^{\rm can}\approx-\beta^2\min\{2(\pi/M)^2,1\}, \tag{5.22}$$

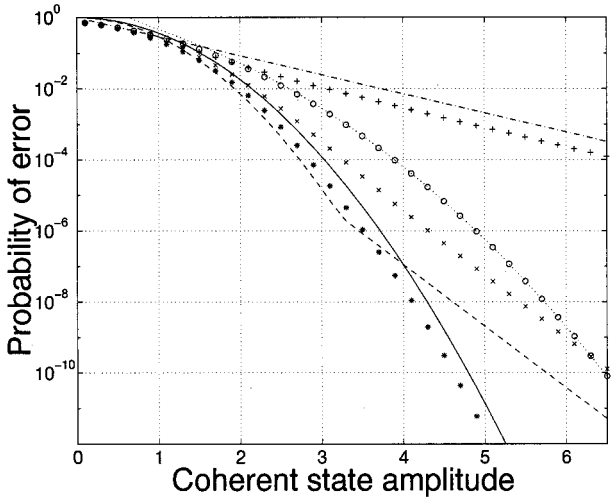

FIG. 7. Plot of the exact (points) and asymptotic (lines) expressions for the log of the probability of error $E_{\rm coh}$ for quaternary phase encoding using coherent states of amplitude β vs β under the four schemes: canonical (* and solid line), heterodyne ($\bigcirc$ and dotted line), adaptive mark I ($+$ and dash-dotted line), and adaptive mark II ($\times$ and dashed line).

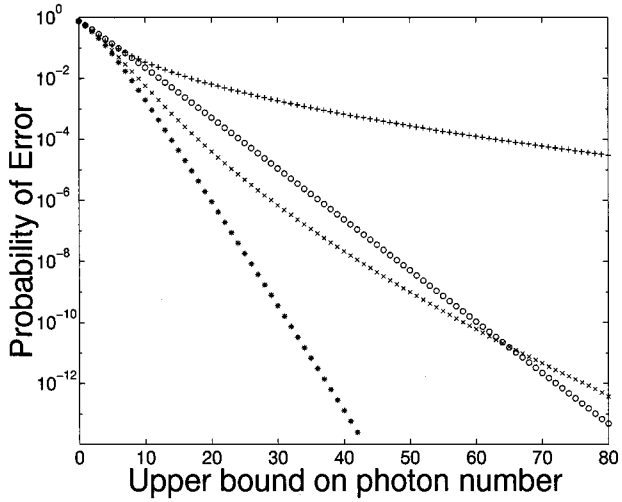

FIG. 8. Plot of the exact (points) expressions for the log of the minimum probability of error $E_{\rm coh}$ for quaternary phase encoding using the optimal state with at most N photons vs N under the four schemes: canonical (*), heterodyne ($\bigcirc$), adaptive mark I ($+$), and adaptive mark II ($\times$).

$$\ln E_{\rm coh}^{\rm het} \approx -\beta^2 \min\{(\pi/M)^2, 1\}, \tag{5.23}$$

$$\ln E_{\rm coh}^{\rm I} \approx -\beta \min\{2(\pi/M)^2, 4\}, \tag{5.24}$$

$$\ln E_{\rm coh}^{\rm II} \approx -\beta \min\{2\beta(\pi/M)^2, 4\}. \tag{5.25}$$

As long as $\beta > 2(M/\pi)^2$ we have the simple results that $-\ln E$ scales quadratically with β for canonical and heterodyne measurements, and linearly with β for the two adaptive measurements. For $\beta < 2(M/\pi)^2$ the adaptive mark II measurement scales quadratically.

From Fig. 6 it is evident that the approximation of $P(\phi)$ as a Gaussian plus a constant tail is poorest for the heterodyne measurement. Thus we would not expect the expression (5.23) to be particularly good. However, for this measurement scheme we can find the following expression for E:

$$1 - E_{\rm coh}^{\rm het} = \frac{1}{\pi} \int_0^\infty \int_0^{ay} e^{-(\beta - x)^2 - y^2} dx dy, \tag{5.26}$$

where $a = \cot(\pi/M)$. After quite some effort this yields the asymptotic expression

$$\ln(E_{\rm coh}^{\rm het}) \simeq -\beta^2/(1+a^2) + \ln\left(\frac{(1+a^2)^5 - a^{10}}{\sqrt{\pi}(1+a^2)^{9/2}}\right) + \ln(\beta) + O(\beta^{-1}). \tag{5.27}$$

The leading term of this differs from the above result (5.23) by at most 25% (for $M=3$) and approaches it for large M. The full expression (5.27), and the above approximate expressions (5.22), (5.24), and (5.25) are plotted as a function of β in Fig. 7 for $M=4$. Also plotted are the exact numerical calculations of the probability of error. The expression (5.27) is evidently a very good approximation. The other analytical expressions match quite well the slopes of the curves, but are displaced vertically. For large β the slope is of course the more important feature, and it is interesting that Eq. (5.25) does correctly predict the change from quadratic to linear behavior of $\ln E_{\rm coh}^{\rm II}$ at $\beta \approx 2(4/\pi)^2 \approx 3.24$.

From the asymptotic results it is clear that for large β the adaptive mark II measurement has a higher probability of error than heterodyne detection. Specifically, for $M > 3$ the crossover point is at

$$\beta \approx 4(M/\pi)^2. \tag{5.28}$$

For $M=4$ this is $\beta \approx 6.48$, which agrees well with the numerical data in Fig. 7. At this point the error is

$$\ln E_{\rm coh} \approx -16(M/\pi)^2. \tag{5.29}$$

Thus depending on whether the acceptable error level is less than or greater than this amount, the best dyne measurement scheme to use (in the sense of requiring the least energy $\hbar\omega\beta^2$ per pulse) will be heterodyne or adaptive mark II, respectively.

C. *M*-ary encoding with optimal states

In this section we consider the probability of error for optimized states subject to a maximum-photon-number constraint. Since the probability of error is

$$E = \langle \psi | 1 - F_C | \psi \rangle, \tag{5.30}$$

it is readily seen that the problem of finding the minimal probability of error for states of the form $\Sigma_{n=0}^{N} c_n |n\rangle$ is precisely that of finding the largest eigenvalue of the matrix formed by truncating the number-state matrix for F_C of Eq. (5.18). For small N this eigenvalue problem can be solved using MATLAB and the H_{mn} matrices computed earlier.

Figure 8 depicts the results for quaternary ($M=4$) encoding. It is clear from this graph that the log of the $E_{\rm opt}$ for optimized states has the same sort of dependence of the maximum photon number N as the log of $E_{\rm coh}$ has on the

mean photon number β^2. That is, for large N, the heterodyne and canonical measurements scale linearly with N (with the latter having the greater slope) while the adaptive measurements scale as the square root of N (with the adaptive mark II having the greater slope). Once again the adaptive mark II measurement is the best realizable measurement for moderate N, while the heterodyne measurement becomes superior for large N. We would expect the crossover point to scale as M^4, and for $M=4$ the numerical data show that it is at $N \approx 64 \approx 25(M/\pi)^4$.

VI. DISCUSSION

In this paper we have presented the exact quantum theory of two adaptive phase measurements. From this we have confirmed the semiclassical results obtained in Ref. [7]. In particular, the phase variance from our adaptive mark II phase measurement is always less than that from a standard phase measurement (such as heterodyne detection). We have also applied our theory to an area inaccessible to the semiclassical theory, that is the complete shape of the probability distribution for the measured results ϕ. We find that the adaptive measurement phase probability distributions have surprisingly high tails. This has the consequence that the adaptive measurement is not necessarily better than standard phase measurements when it comes to communication using M-ary encoding of data in the phase of states.

The fact that the adaptive phase measurement is not necessarily superior to the standard phase measurement for M-ary phase encoding does not mean that it is a poor phase measurement, or that adaptive measurements in general are not useful. After all the situation of M-ary encoding does not really call for a phase measurement; rather it calls for a measurement which can distinguish as well as possible between a finite number of known different (but not orthogonal) states. For the case of binary phase encoding using coherent states (with phases 0 and π), there is an adaptive measurement which has been known for some time [19] which distinguishes these possible states as well as quantum mechanics allows. It is only when $M \sim N$, where N is the mean photon number of the states, that the measurement required is really a phase measurement. In this limit the variance of the distribution is the important factor, and the adaptive mark II phase measurement always gives a lower error rate than standard detection.

Although the asymptotics for the phase variance of the adaptive schemes were already known from the semiclassical theory of Ref. [7] the quantum theory presented here sheds knew light on these results and allows us to probe new issues. For example, what is the ultimate limit on the phase noise introduced by an adaptive phase measurement? In other words, how closely is it possible to approximate a canonical phase measurement by using a measurement involving dyne measurements (that is measurements using photodetection and a local oscillator with arbitrary time-varying phase)? Although we cannot answer this question at this stage, we can show that there is a lower bound on the amount of excess noise. This lower bound is not due to imperfections such as a finite local oscillator or inefficient detectors, but is a fundamental limitation of the method of measurement via photodetection. We proceed by using the analysis in Appendix B.

It was shown in Appendix B that the probability for obtaining a particular phase ϕ is determined largely by the maximum overlap between the system state and any of the pure states which contribute to the probability operator $F(\phi)$ for that phase. For dyne measurements, these pure states are squeezed states. As a result of this, the variance of the measured phase probability distribution will be (to a good approximation) equal to the true (canonical) phase variance of the system plus the phase variance of the maximum-overlap pure state. Furthermore, it was shown in Appendix B that in order to obtain a large overlap, the maximum-overlap squeezed state must have a well-defined coherent amplitude roughly equal to the coherent amplitude of the system.

From these considerations we can conclude that if the system has roughly N photons, then the excess phase variance will be approximately that of a squeezed state with a mean photon number of N. Now the minimum (canonical) phase variance of a squeezed state with a mean photon number of N has been investigated by Collett [20], who found the asymptotic result

$$V_{\rm ss}^{\rm can} \geqslant \frac{\ln N}{4N^2}. \tag{6.1}$$

This represents a lower bound on the excess phase variance introduced by any dyne measurement. So, for example, if N is sufficiently large then the minimum measured phase variance for a state with at most N photons would be

$$V_{\rm min}^{\rm dyne} \geqslant \frac{\ln N}{4N^2}. \tag{6.2}$$

This lower bound should is a long way below the variance achieved by the adaptive mark II scheme presented here, for which

$$V_{\rm min}^{\rm II} \simeq \frac{1}{8N^{3/2}}, \tag{6.3}$$

which itself is a long way below the variance achieved by standard measurements, namely,

$$V_{\rm min}^{\rm het} \simeq \frac{1}{4N}. \tag{6.4}$$

In fact, the lower bound (6.2) is very close to the absolute lower limit set by canonical measurement [21]

$$V_{\rm min}^{\rm can} \simeq \frac{\pi^2}{N^2}. \tag{6.5}$$

Exactly how close one can come to the lower bound (6.1) by using a different feedback algorithm is a matter for future research.

ACKNOWLEDGMENTS

H.M.W. would like to thank the Australian Research Council and R.B.K. the W. H. Pickering Foundation for financial support.

APPENDIX A: THE OSTENSIBLE MOMENTS OF C

Following the text, we denote the ostensible moments of C as

$$M_v^{n,m}=\langle C_v^n C_v^{*m}\rangle_Q . \tag{A1}$$

Using the rules of Itô calculus to evaluate

$$dM_v^{n,m}=\langle (C_v+dC_v)^n(C_v^*+dC_v^*)^m-C_v^nC_v^{*m}\rangle \tag{A2}$$

we find from Eq. (3.24)

$$\frac{dM_v^{n,m}}{dv}=-\frac{2(n-m)^2}{v}M_v^{n,m}+nM_v^{n-1,m}+mM_v^{n,m-1}. \tag{A3}$$

Since $M_v^{0,0}\equiv 1$ these equations may be solved recursively to find

$$M^{n,m}=\frac{nM^{n-1,m}+mM^{n,m-1}}{2(n-m)^2+n+m}. \tag{A4}$$

Recall that by convention $M^{n,m}=M_1^{n,m}$. For n or m equal to zero this recurrence relation can be solved to get

$$M^{n,0}=M^{0,n}=\frac{1}{(2n+1)(2n-1)\cdots 1}=\frac{1}{(2n+1)!!}. \tag{A5}$$

These boundary values allow us to rapidly compute all the desired moments $M^{n,m}$.

APPENDIX B: THE TAILS OF THE DISTRIBUTIONS

The reason for the different scaling of the tails of the adaptive measurements compared to the heterodyne measurement can be understood as follows. For heterodyne detection the dominant term is the inner product of the system state β with the coherent state $|-r\rangle$ for $r=0^-$. This maximizes the overlap while still maintaining $\phi=\arg r=\pi$:

$$\ln P_{\rm coh}^{\rm het}(0)\simeq \ln|\langle\beta|0\rangle|^2=-\beta^2. \tag{B1}$$

For the adaptive mark I technique the overlap will be with a squeezed state $|\alpha,\epsilon\rangle$, where (using $\phi=\hat{\varphi}=\pi$)

$$\alpha=-\frac{1+C}{1-|C|^2}, \tag{B2}$$

$$\epsilon=-\frac{C\,\mathrm{atanh}|C|}{|C|}. \tag{B3}$$

The problem is to determine the value of C that maximizes this overlap.

It is not difficult to see that the value of C we seek will be real and positive. In this case

$$\alpha=-(1-C)^{-1}, \tag{B4}$$

$$\epsilon=-\,\mathrm{atanh}C . \tag{B5}$$

This describes a squeezed state centered at $x=-2/(1-C)$ with an x variance

$$\exp(-2\epsilon)=\frac{1+C}{1-C}. \tag{B6}$$

The overlap between $|\beta\rangle$ and $|\alpha,\epsilon\rangle$ is

$$|\langle\beta|\alpha,\epsilon\rangle|^2=\frac{\exp[-(1+\tanh\epsilon)(\beta+\alpha)^2]}{\cosh\epsilon} \tag{B7}$$

$$\simeq\frac{\exp\{-(1-C)[\beta+1/(1-C)]^2\}}{\sqrt{1-C^2}}. \tag{B8}$$

Ignoring the negligible $\sqrt{1-C^2}$, this expression is maximized for

$$1-C=\beta^{-1}. \tag{B9}$$

This implies $\alpha=-\beta$ and $\exp(-2\epsilon)\simeq 2\beta$. Substituting this in Eq. (B8) gives

$$\ln P_{\rm coh}^{\rm I}(\pi)\simeq\ln|\langle\beta|\alpha,\epsilon\rangle|^2\simeq-4\beta, \tag{B10}$$

as obtained in the body of the paper.

This derivation in this appendix shows that the reason for the high tails of the adaptive distributions is the large x variance of the squeezed state $|\alpha,\epsilon\rangle$, giving it a much larger overlap with $|\beta\rangle$ than has $|0\rangle$ (from the heterodyne measurement). Although this large squeezing is responsible for the high tails, it is also what allows the narrow peak of the adaptive mark II measurement. This can be seen as follows.

The most likely result for the adaptive mark II case is $\phi=\hat{\varphi}+\arg(1+C)=0$. This is obviously most likely to occur for $\hat{\varphi}=0$, in which case the only difference is that

$$\alpha=\frac{1+C}{1-|C|^2}. \tag{B11}$$

Once again it is easy to see that the maximum overlap will be for $C\approx 1$. The overlap in this case is

$$\ln|\langle\beta|\alpha,\epsilon\rangle|^2\simeq-(1-C)\left(\beta-\frac{1}{1-C}\right)^2. \tag{B12}$$

This is maximized (with a value of zero) at exactly the same $C=1-\beta^{-1}$. This gives $\alpha=\beta$ as expected, and the same x variance.

In this case what is of more interest is the y variance

$$\exp(2\epsilon)\simeq(2\beta)^{-1}. \tag{B13}$$

The *intrinsic* phase variance of this squeezed state is thus

$$V_{\rm ss}\simeq\frac{\langle y^2\rangle}{\langle x\rangle^2}=\frac{\exp(2\overline{\epsilon})}{(2\beta)^2}\simeq\frac{1}{8\beta^3}. \tag{B14}$$

This is precisely equal to the asymptotic expression for the excess variance

$$V_{\rm coh}^{\rm II} - V_{\rm coh}^{\rm can} \simeq \frac{1}{8\beta^3}. \tag{B15}$$

The reason for this is that the measured phase distribution is at least as wide as a convolution of the true (canonical) phase distribution of the state with the true phase distribution of the most likely POM. This is completely analogous to the argument centered around Eq. (4.23) for the heterodyne case. For the adaptive mark I measurement the measured distribution is actually much wider, but the above calculation shows that for the adaptive mark II measurement all of the introduced noise is due to the quantum uncertainty in the states making up the POM. Thus the mark II phase estimate is, for large fields, the best possible estimate given the feedback algorithm (3.12).

[1] H. M. Wiseman, Quantum Semiclassic. Opt. **7**, 569 (1995).

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

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

[4] *Quantum Phase and Phase Dependent Measurements*, edited by W. P. Schleich and S. M. Barnett, special issue of Phys. Scr. **T48** (1993).

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

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

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

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

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

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

[11] D. F. Walls and G. J. Milburn, *Quantum Optics* (Springer, Berlin, 1994).

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

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

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

[15] 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 Mathematics Vol. 1055 (Springer, Berlin, 1984), p. 153.

[16] Y. L. Luke, *The Special Functions and Their Approximations* (Academic Press, New York, 1969), Vol. 1.

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

[18] M. J. Hall and I. G. Fuss, Quantum Opt. **3**, 147 (1991).

[19] S. J. Dolinar, Research Laboratory of Electronics, MIT, Quarterly Progress Report No. 111 (1973), p. 115; see also *Quantum Detection and Estimation Theory* (Ref. [9]), p. 163; and A. S. Holevo, Sov. Math. **26**, 3 (1982).

[20] M. J. Collett, *Quantum Phase and Phase Dependent Measurements* (Ref. [4]), p. 124.

[21] This result shows that we would only expect the asymptotic lower bound (6.2) to be relevant for $\ln N$ greater than $4\pi^2$; that is, for $N \gtrsim 10^{18}$. Thus the absolute lower bound due to the intrinsic phase variance of the squeezed states which comprise the POM is irrelevant unless we have a photon detection scheme that is perfect to 1 part in 10^{18}, which is practically impossible.