THE ULTIMATE QUANTUM LIMITS ON THE ACCURACY OF MEASUREMENTS†

Horace P. Yuen
Department of Electrical Engineering and Computer Science
Department of Physics and Astronomy
Northwestern University
Evanston, IL 60208

ABSTRACT

A quantum generalization of rate-distortion theory from standard communication and information theory is developed for application to determining the ultimate performance limit of measurement systems in physics. For the estimation of a real or a phase parameter, it is shown that the root-mean-square error obtained in a measurement with a single-mode photon level N cannot do better than $\sim N^{-1}$, while $\sim exp\{-N\}$ may be obtained for multi-mode fields with the same photon level N. Possible ways to achieve the remarkable exponential performance are indicated.

INTRODUCTION

Given whatever physical constraints one has to operate with, what is the best possible system one can build for the measurement or estimation of a physical parameter of interest? It is evident that a systematic approach to the answer of this class of questions is of great interest in physics, which is so much concerned with the detection and accurate measurement of various quantities, from routine temperature gauging to the detection of very weak gravitational radiation. In this paper, I will describe a systematic theory for answering these questions. Conceptually, this theory is directly transplanted from ordinary (classical) information and communication theory, although technically the new quantum issues may greatly complicate the actual workout of a solution. As illustrations, I will provide the ultimate quantum limits on the accuracy of estimating a phase parameter, and also an arbitrary real parameter, when an optical field of a given power level is employed. Let N be the available number of photons of a narrowband optical field. For both the estimation of a phase parameter and a real amplitude parameter, the following results will be proved. For a single-mode field, the best root-mean-square error one may obtain is

$$\delta\phi \sim \frac{1}{N}, \quad \delta r \sim \frac{1}{N} \tag{1}$$

whereas for a multimode field with sufficiently many modes one may achieve

$$\delta\phi \sim \epsilon^{-N}, \quad \delta r \sim \epsilon^{-N}. \tag{2}$$

Moreover, the theory provides various indications on how one may actually approach the problem of realizing a multimode system that would yield the remarkable exponential performance given by (2). In the following, the underlying information theoretic results will first be explained before the quantum situation is discussed. Due to limitations in space-time, everything can only be briefly outlined. Nevertheless, I hope the discussion is self-contained and comprehensible.

RATE-DISTORTION LIMIT

The theory of information transmission pioneered by Shannon (refs. 1-3) can be immediately adapted to provide a systematic answer to the above class of questions. For a system described by classical physics, the solution goes as follows. First, we assign an a priori probability distribution $p(u)$ on the parameter u we are interested in estimating. This parameter would modulate a physical variable in whatever physical system we pick for extracting information about this parameter. For example, if u is the amplitude of a gravitational wave, the system may be a Michelson interferometer with the physical variable a certain optical phase of an electromagnetic field mode. Some measurement is to be made on the system, such as a determination of the field strengths of the mode, to extract information on the physical variable through which an estimate of u is to be obtained. Let x be the physical variable, and y the measurement variable which is in general random with conditional probability $p(y|x)$, with x itself a function of u depending on the specific scheme. Both the cases of discrete and continuous variables will be included throughout with proper interpretation of the probabilities as a distribution or a density function of the random quantities under discussion.

The condition probability $p(y|x)$ defines a *channel* in information theory, with x the channel *input* and y the channel *output*. For any input probability $p(x)$, the joint probability $p(y, x) = p(y|x)p(x)$ is specified from which one can evaluate the *average mutual information* between x and y,

$$I(\mathrm{x;y}) \equiv \int p(x, y) log \frac{p(x|y)}{p(x)} dx dy. \tag{3}$$

The entropy of a single random variable can be defined as average self-information

$$H(\mathrm{v}) \equiv I(\mathrm{v;v}) \tag{4}$$

in which $p(v|v)$ is to be interpreted as a Kronecker or Dirac delta. The units in (3) and (4) are given as bits per channel use (per channel input) and bits per source symbol if the log is taken to be of base 2, and as nats per use if the log is of base e. Shannon's channel coding theorem and its converse (ref. 1) state that successive independent samples of a random variable v can be transmitted over a channel $p(y|x)$ with zero probability of error if and only if $H(\mathrm{v}) \leq I(\mathrm{x;y})$. However, for a noisy channel, i.e., when y does not specify x uniquely, channel encoding and decoding are required to get zero error probability which is only obtained in the limit of arbitrarily long codes. Note that we have deviated somewhat from the standard notations in information theory to avoid conflict with later quantum notations. Also, the coding theorem is usually stated in terms of the *capacity* of the channel, which is defined to be the $I(\mathrm{x;y})$ obtained by maximizing over $p(x)$ under whatever further constraints one may impose on x. Typically, one assigns a cost function $\beta(x)$ and constraint the average cost to be under a given level B. The capacity $C(B)$ will then be an increasing function of B. *Roughly speaking, $C(B)$ is the maximum number of information bits one can transmit error-free over a channel with an average resource level B.*

The *rate-distortion function* $R(D)$ of a random variable u is defined to be the minimum $I(\mathrm{u;v})$ between u and any another random variable v such that the *average distortion* between u and v

$$E[d(u, v)] \equiv \int d(u, v)p(u)p(v|u)dudv \tag{5}$$

is at or below a given level D, where the distortion function $d(u, v)$ is a given measure of the difference between u and v. When u is a continuous real parameter, $d(u, v)$ is often chosen to be $|u - v|^2$ or $|u - v|$. The minimization of $I(\mathrm{u;v})$ is carried over $p(v|u)$ subject to the constraint

$$E[d(u, v)] \leq D. \tag{6}$$

One may think of v as a data-compressed version of the source variable u —— v represents u with an average distortion D but it requires less bits to represent v than u. Shannon's source coding theorem with a fidelity criterion and its converse (ref. 3) state that a source can be asymptotically represented with an average distortion D if and only if at least $R(D)$ bits per source symbol is provided. Again, source encoding and decoding are in general required to achieve such minimum distortion in the limit of long codes. Nevertheless, this result shows that *roughly speaking, $R(D)$ is the minimum number of information bits required to represent a source with an average distortion level D.*

What is the minimum average distortion D one can get for transmitting a source variable u over a channel with resource B? The answer is provided by Shannon's joint source-channel coding theorem (refs. 4-5) in what is often called the *rate-distortion limit* or *rate-distortion bound.* By combining the source and the channel coding theorems, the bound is

$$D \geq R^{-1}C(B) \tag{7}$$

where R^{-1} is the inverse of the monotone function $R(D)$. It is important to emphasize that the *Shannon theorem and its converse state that (7) is the ultimate limit and can be approached by an actual system that employs source coding and channel coding separately. It does not say that it can only be approached by separate source and channel coding.* In fact, the following example illustrating the power of the rate-distortion bound also shows that it is sometimes possible to achieve it (to get the actual minimum) without any coding or nonlinear modulation at all.

Let u be a zero-mean Gaussian random variable with variance σ_u^2, and distortion measure the squared error $d(u,v) = |u - v|^2$. The rate distortion function in this case is well known [refs. 3-5],

$$\begin{aligned} R(D) &= \tfrac{1}{2}log\tfrac{\sigma_u^2}{D} & 0 \leq D \leq \sigma^2 \\ &= 0 & D \geq \sigma^2 \end{aligned} \tag{8}$$

Consider an additive Gaussian noise channel

$$y = x + n \tag{9}$$

where n is a zero-mean Gaussian noise variable with variance $\mathcal{N}_0$ statistically independent of x. With a given power level, $E[x^2] \leq S$, it is wellknown [refs. 1-2, 4-5] that the capacity is

$$C(S) = \frac{1}{2}log(1 + \frac{S}{\mathcal{N}_0}). \tag{10}$$

The rate-distortion bound solves the following problem which cannot be solved in any other way, to my knowledge. Suppose each sample of u matches each use of x, i.e., the rate that u is generated is equal to the rate that x can be transmitted over the Gaussian channel. We are interested in finding the best signal processing scheme before transmission over the channel and after transmission in receiver processing that would yield an estimate of u with minimum mean-square error. From (7),(8) and (10) one gets

$$D \geq \sigma_u^2 \left(1 + \frac{S}{\mathcal{N}_0}\right)^{-1}. \tag{11}$$

It turns out that the right side of (11) can be obtained by simply letting $x = \sigma_u^{-1}S^{1/2}u$ and estimating $\hat{u}(y) = \sigma_u S^{-1/2}y$, i.e., direct linear transmission and estimation without coding or nonlinear modulation is already optimal as verified from the rate-distortion limit. On the other hand, a direct optimization approach to this or most other comunication problems is very difficult to just formulate, not to mention writing down the optimization conditions.

Despite the power of rate-distortion theory, we are faced with two complications in its application to measurement problems in physics. The first is derived from the fact that in a measurement

system, one may have very little or no room at all for source coding, basically because the parameter u in this case may be entirely out of one's control for further processing before modulating onto the physical variable x. Thus, while the bound (7) still remains a limit, one is *no longer sure that the limit can be achieved arbitrarily closely*. This problem can be overcome by replacing $R(D)$ by $\mathcal{R}(D)$, which is defined to be the number of bits required to represnet u to a distortion level D *given* a specific simple source coding scheme such as uniform quantization, or no coding at all. The second problem is similarly derived from the fact that no channel coding may be employed. In the same way, one can replace $C(B)$ by an average mutual information $\mathcal{I}(B)$ which incorporates whatever constraint one must face, including perhaps some modulation but no coding. In contrast to the source case $\mathcal{R}(D)$, in the evaluation of $\mathcal{I}(B)$ it may be difficult to actually take into account precisely the constraints one operates with. Anyhow, in a way exactly parallel to the Shannon joint source-channel coding theorem, the following generalized rate-distortion limit applies with whatever additional constraints in the present measurement situation,

$$D \geq \mathcal{R}^{-1}\mathcal{I}(B). \tag{12}$$

Depending on the specific case, the limit (12) may be much higher than (7).

In addition to providing an answer on how accurately one may actually perform a measurement through (7) or (12), this theory also indicates a way to approach the best performance, namely, through channel coding or modulation to achieve $C(B)$ or $\mathcal{I}(B)$, assuming source coding cannot be carried out. Illustrations will be given in the following quantum problems.

MEASUREMENT WITH A QUANTUM SYSTEM

The development of squeezed and nonclassical lights [ref. 6-7] has been strongly motivated by their possible applications to precision measurements. It is logical, in fact imperative, to ask for the ultimate limit of measurements in quantum physics; quantum fluctuations, being an intrinsic feature of nature as we understand it, would have to be taken into account in assessing such ultimate limits. Contrary to what one may first think, the uncertainty principle does not provide the answer either by itself or in conjuction with other additional considerations as discussed in the following two prime cases of continuous parameter estimation. Consider first the case of a real parameter λ defined over the whole real line $\Re$. For simplicity, let λ be a Gaussian random variable and N be the average available number of photons in a single-mode field that one can use to capture λ. That is, we wish to design the best measurement system, which is represented by the way λ modulates the quantum state ρ_λ of the mode and the quantum measurement one chooses to make on the mode, subject to the constraint

$$\int tr[\rho_\lambda a^\dagger a]p(\lambda)d\lambda \leq N \tag{13}$$

where $p(\lambda)$ is the probability of λ, a the model photon annihilation operator, and ρ_λ a density operator on the Hilbert space of quantum states $\mathcal{H}$. To include all possible quantum measurements such as heterodyning, a general quantum measurement on the system $\mathcal{H}$ is represented, as far as the measurement probability is concerned, by a *positive operator-valued measure* (POM) [ref. 8-10] generalizing the usual selfadjoint operator description. In a notation including possible operator-valued distributions, a POM X with measurement value $x \in \Re^n$ is a function $x \longmapsto X(x)$ such that each $X(x)$ is a bounded positive semidefinite selfadjoint operator and all the $X(x)$ sum to the identity operator, i.e.,

$$X(x) \geq 0, \tag{14}$$

$$\int X(x)dx = I. \tag{15}$$

|

When $X(x) = |x><x|$ are *orthogonal* projectors, the POM X can be desribed by a unique selfadjoint operator obeying the functional calculus

$$f(X) = \int f(x)X(x)dx. \tag{16}$$

For a general POM, (16) does *not* hold. When X is measured on a system in state ρ, the probability that x is obtained is given by $tr[\rho X(x)]$. Mathematically, the problem is to find a mapping $\lambda \mapsto \rho_\lambda, \lambda \in \Re$, a quantum measurement X, an estimate $\hat{\lambda}(x)$ of λ, such that the resulting mean-squared error between $\hat{\lambda}$ and λ is as small as possible subject to the constraint (13). It should be clear that the uncertainty principle is of little help in solving this problem, although some weaker conclusion may be obtained with its help [refs. 11-14]. Thus, if one assumes that X is to be a single field-quadrature operator, and the criterion is changed to average signal-to-noise ratio, then the uncertainty relation between conjugate quadratures

$$< \Delta a_1^2 >< \Delta a_2^2 > \geq \frac{1}{16} \tag{17}$$

can be used to show that the use of two-photon coherent states (TCS) or squeezed states in the narrow sense [refs. 6,15] is optimum. In fact, it yields a mean-square error given by, from (11),

$$D_0 = (\frac{\sigma_\lambda}{1+2N})^2. \tag{18}$$

As will be shown in the next section, this turns out to be very close to the best one can do.

In the second case, consider the estimation a real parameter defined over a finite interval, which for simplicity we take to be a phase parameter $\phi \in (-\pi, \pi]$. Mathematically, the problem is exactly the same as above except that $p(\lambda)$ is changed. The number-phase uncertainty relation in whatever form or interpretation,

$$\Delta N \Delta \Phi > \frac{1}{4} \tag{19}$$

is of no help at all here. In contrast to (17), (19) does not even place a limit on how small $\Delta \Phi$ may get under an average photon number constraint because ΔN may still be arbitrarily large. *More significantly*, in an actual measurement problem it is not the quantum fluctuation alone that is important in determining the limit. The total quantum state (the full statistics) and the way energy is distributed could be just as important. We now show how the rate-distortion theory can be generalized to provide the answers.

ULTIMATE QUANTUM LIMITS

To obtain the ultimate possible performance for the above two problems, we note that with the mean-square error criterion the rate distortion function $R(D)$ for a Gaussian random variable λ with variance σ^2 is given by (8) while that for a uniformly distributed $\phi \in (-\pi, \pi]$ is difficult to evaluate exactly. However, the wellknown Shannon upper and lower bounds [ref. 3] on $R(D)$ gives a very accurate estimate in this case: in nats per symbol

$$0.419 - log\sqrt{D} \leq R_\phi(D) \leq 0.595 - log\sqrt{D} \tag{20}$$

If the magnitude distortion function $d(u,v) = |u - v|$ is employed instead, the $R(D)$ for the uniform phase parameter is known exactly while that of a Gaussian random variable is known parametrically [ref. 16]. In both cases, they are quite close to that given by the Shannon lower bound, and are approximately the same as the mean-square case with the natural replacement of D by $\sqrt{D}$. Moreover, for the uniform phase variable the $\mathcal{R}(D)$ function obtained from a *uniform quantizer* (digitizer) can be easily evaluated. For the mean-square-criterion,

17

$$\mathcal{R}_\phi(D) \sim 0.595 - log\sqrt{D} \qquad (21)$$

which is exactly the upper bound part of (20)! Thus, uniform digitization without coding is quite close to optimum in this case. For the Gaussian case, uniform quantization also leads to a $\mathcal{R}(D)$ with a similar functional form to $R(D)$, but with a further fixed constant difference. In fact, it is well known that for a large class of memorylas sources and distortion measures simple quantization already leads to a performance close to the rate-distortion limit. The upshot of our discussion is that independently of the exact distortion criterion one chooses and without the need of coding, the $\mathcal{R}(D)$ functions for our two cases can be accurately estimated and they are close to the rate-distortion limit $R(D)$.

Given $\mathcal{R}(D)$ or $R(D)$, the quantum limitation on communication or measurement is determined by substituting the ultimate quantum information transmission capacity $\mathcal{C}$ into $C(B)$ in the bound (7). For a given system, the *ultimate quantum capacity* $\mathcal{C}$ is the maximum average mutual information $I(x;j)$ one may obtain by picking an input alphabet J, discrete or continuous, probability p_j on J, a map $j \mapsto \rho_j, j \in J, \rho_j$ density operators on the system state space $\mathcal{H}$, and a POM $X(x)$ subject to whatever constraints one may have. It is clear that the channel coding theorem and its converse hold for this capacity $\mathcal{C}$. The actual evaluation of $\mathcal{C}$ can be very complicated due to the added optimization over ρ_j and X which are entirely of quantum mechanical origin. However, for certain cases including the following ones, the evaluation can be carried out with the help of an entropy bound [ref. 10]. Thus, for a single-mode optical field with average photon number constraint N,

$$\sum_j p_j tr[\rho_j a^\dagger a] \le N, \qquad (22)$$

the ultimate quantum capacity is achieved by photon number eigenstates with the result [ref. 10]

$$C(N) = (N+1)log(N+1) - NlogN. \qquad (23)$$

For a narrowband optical field with m modes of approximately the same frequency and a constraint N on the *total* number of average photons in all m modes, the ultimate capacity is [ref. 10]

$$C(N) = mlog(\frac{N}{m} + 1) + \frac{N}{m}log(\frac{m}{N} + 1). \qquad (24)$$

We may also be interested in the capacity of TCS with homodyne detection [ref.17]

$$C^{HOM}_{TCS}(N) = mlog(2\frac{N}{m} + 1) \qquad (25)$$

and the capacity of coherent states with heterodyne detection

$$C^{HET}_{CS}(N) = mlog(\frac{N}{m} + 1). \qquad (26)$$

Going back to our single-mode optimal measurement problem, it follows from (8) and (23) that for a Gaussian parameter r with variance σ^2, the best root-mean-square error $\delta r \equiv \sqrt{D}$ one may obtain is

$$\delta r = \frac{\sigma}{N+1}(1 + \frac{1}{N})^{-N} \sim \frac{\sigma}{eN} \quad , N >> 1 \qquad (27)$$

The suboptimum TCS and CS performance are close to the optimum (27); from (25)-(26) with $m = 1$,

$$\delta r^{TCS} = \frac{\sigma}{2N+1}, \quad \delta r^{CS} = \frac{\sigma}{N+1} \quad . \tag{28}$$

Note that δr^{TCS} can be achieved *without* coding or nonlinear modulation from (18) as discussed in the previous section, while $\delta r^{CS} = \frac{\sigma}{\sqrt{N+1}}$ without coding or modulation. Thus, *the use of TCS can be viewed as an alternative to coding or nonlinear modulation* in at least the single-mode case. For the phase parameter ϕ with uniform distribution, it follows from (21) and (23) that the ultimate limit is

$$\delta\phi \sim \frac{1}{eN}, \quad N >> 1 \tag{29}$$

while with TCS and coherent states

$$\delta\phi^{TCS} \sim \frac{1}{2N}, \quad \delta\phi^{CS} \sim \frac{1}{N}, \quad N >> 1. \tag{30}$$

Again, it is known that the use of TCS or other phased-squeezed states would lead directly to $\delta\phi \sim \frac{1}{N}$ [ref. 18], while the use of coherant states without coding or modulation yields only $\delta\phi^{CS} \sim \frac{1}{\sqrt{N}}$.

Consider now the multimode limit under the constraint of the *same* number of photons N. From (8) and (24), we have

$$\delta r = \sigma \left(\frac{N}{m}\right)^{-m} \left(1 + \frac{m}{N}\right)^{-N} \tag{31}$$

which implies that δr would go to zero at least as quickly as e^{-N} for $m \geq 0.1N$. For TCS and coherent states,

$$\delta r^{TCS} = \sigma \left(1 + \frac{2N}{m}\right)^{-m}, \quad \delta r^{CS} = \sigma \left(1 + \frac{N}{m}\right)^{-m} \tag{32}$$

which implies that they would go to zero as e^{-N} for $m \geq N$. Similarly for the phase parameter ϕ,

$$\delta\phi \sim \left(\frac{N}{m}\right)^{-m} \left(1 + \frac{m}{N}\right)^{-N} \tag{33}$$

$$\delta\phi^{TCS} \sim \left(1 + \frac{2N}{m}\right)^{-m}, \quad \delta r^{CS} \sim \left(1 + \frac{N}{m}\right)^{-m} \quad . \tag{34}$$

This multimode behavior as indicated by equ (2) is not unexpected from communication theory, as a larger number of modes is equivalent to signal space of higher dimension which means that the different messages can be placed farther apart in signal space to combat the effect of noise [refs. 2,19]. This is familiar in what is called FM quieting in frequency modulation, and is commonly referred to as the exchange of bandwidth with signal-to-noise ratio. The remarkable feature is that a large number of modes moves the N^{-1} dependence of the ultimate limit to $\exp\{-N\}$ which is so much more accelerated!

There are several approaches one may consider for obtaining such exponential performance, although in a measurement rather than a communication situation one cannot be sure that the above capacities can be actually obtained. Since the number-states channel is noise-free, its capacity can be achieved *without* coding. Indeed it is achieved by a rather simple modulation scheme and the effect of a small nonideal residue noise is not expected to affect the resulting performance too much. The problem remains to find a scheme which, for a measurement system,

would naturally capture the parameter λ (either r or ϕ) of interest in such a modulation scheme or another one which is nearly as good. On the other hand, one may consider the use of nonlinear modulation on TCS or coherent states; different nonlinear modulation schemes are known to get quite close to the rate-distortion limit in many classical communication situations [ref. 20]. In particular, if nonlinear modulation or coding is to be employed, one may consider dispensing with the use of TCS and staying with coherent states, with the resulting loss of a factor of 2 in the exponent but a tremendous gain in practicality. Many different nonlinear modulation schemes may be employed. For example, it is wellknown that a simple pulse frequency modulation in which the modulated signal is given by

$$s(t,\lambda) = \sqrt{\frac{2E}{T}} sin(\omega_0 + \beta\lambda)t, \quad 0 \leq t \leq T \tag{35}$$

where β is a known constant and E the energy of the signal, could lead to an increase in the signal-to-noise ratio for the estimation of a phase parameter in the presence of additive Gaussian noise by a fact m^2, where $m = WT$ is the total number of modes in $s(t,\lambda)$ with W the frequency bandwidth of the signal. While such a simple scheme may not lead to exactly an exponential performance (2), it may still be a large improvement as the N^{-1} performance of (1) becomes $(mN)^{-1}$.

In conclusion, the quantum generalized rate-distortion theory and the possible actual systems it may suggest seem to hold much promise for greatly improved precision measurements in physics, as our two important examples discussed in this paper amply demonstrate.

REFERENCES

1. Shannon, C.E., 1948, "A Mathematical Theory of Communication", *Bell Sys. Tech. J.* 27, pp. 379-423, 623-656.

2. Shannon, C.E., 1949, "Communication in the Presence of Noise", *Proc. IRE* 37, pp. 10-21.

3. Shannon, C.E., 1959, "Coding Theorems for a Discrete Source with a Fidelity Criterion", *IRE Nat. Cov. Rec.*, Part 4, pp. 142-163.

4. Gallager, R.G., 1968, *Information Theory and Reliable Communication*, Wiley, New York.

5. Blahut, R.E., 1987, *Principles and Practice of Information Theory*, Addison-Wesley, Reading, Mass.

6. Loudon, R. and Knight, P.L. 1987, "Squeezed Light", *J. Modern Opt.* 34, pp. 709-759.

7. Teich, M.C. and Saleh, B.E. A., 1988, "Photon Bunching and Anti-Bunching", *Progress in Optics* XXVI, ed. by E. Wolf, Elsevier, New York, pp. 1-104.

8. Davies, E.B., 1976, *Quantum Theory of Open Systems*, Academic Press, New York.

9. Yuen, H.P. 1982, "Generalized Quantum Measurements and Approximate Simultaneous Measurements of Noncommuting Observables", *Phys. Lett.* 91A, pp. 101-104.

10. Yuen, H.P., and Ozawa, M., 1991, "The Ultimate Information Carrying Limit of Quantum Systems", submitted to *Phys. Rev. Letters*.

11. Yuen, H.P., 1976, "States that Give the Maximum Signal-to-Quantum Noise Ratio for a Fixed Energy", *Phys. Lett.* 56A, pp. 105-106.

12. Yuen, H.P., and Shapiro, J.H., 1978, "Optical Communication with Two-Photon Coherent States Part I: Quantum-State Propagation and Quantum-Noise Reduction", *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 657-668.

13. Shapiro, J.H., Yuen, H.P., and Machado Mata, J.A., 1979, "Optical Communication with Two-Photon Coherent States – Part II: Photoemissive Detection and Structured Receiver Performance", *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 179-192.

14. Yuen, H.P. and Shapiro, J.H., 1980, "Optical Communication with Two-Photon Coherent States Part III: Quantum Measurements Realizable with Photoemissive Detectors", *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 78-92.

15. Yuen, H.P., 1976, "Two-Photon Coherent States of the Radiation Field", *Phys. Rev. A*, vol. 13, pp. 2226-2232.

16. Viterbi, A.J., and Omura, J.K., 1979, *Principles of Digital Communication and Coding*, McGraw Hill, New York.

17. Yuen, H.P., 1975, "Generalized Coherent States and Optical Communications", *Proc. 1975 Conf. Information Sciences and Systems*, John Hopkins Univ., Baltimore, MD, pp. 171-177.

18. Shapiro, J.H., and Shepard, S.R., 1991, "Quantum Phase Measurement: A System-Theory Perspective", *Phys. Rev. A*, vol. 43, pp. 3795-3818.

19. Wozencraft, J.M., and Jacobs, I.M., 1965, *Principles of Communication Engineering*, Wiley, New York.

20. Van Trees, H.L., 1971, *Detection, Estimation and Modulation Theory, Part II, Nonlinear Modulation Theory*, Wiley, New York.