Quantum limits on bosonic communication rates

Carlton M. Caves

Center for Advanced Studies, Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131-1156

P. D. Drummond

Department of Physics, University of Queensland, St. Lucia, Queensland 4072, Australia

The capacity C of a communication channel is the maximum rate at which information can be transmitted without error from the channel's input to its output. The authors review quantum limits on the capacity that can be achieved with linear bosonic communication channels that have input power P. The limits arise ultimately from the Einstein relation that a field quantum at frequency f has energy $E = hf$. A *single* linear bosonic channel corresponds to a single transverse mode of the bosonic field—i.e., to a particular spatial dependence in the plane orthogonal to the propagation direction and to a particular spin state or polarization. For a single channel the maximum communication rate is $C_{\mathrm{WB}} = (\pi/\ln 2)\sqrt{2P/3\hbar}$ bits/s. This maximum rate can be achieved by a "number-state channel," in which information is encoded in the number of quanta in the bosonic field and in which this information is recovered at the output by counting quanta. Derivations of the optimum capacity C_{WB} are reviewed. Until quite recently all derivations assumed, explicitly or implicitly, a number-state channel. They thus left open the possibility that other techniques for encoding information on the bosonic field, together with other ways of detecting the field at the output, might lead to a greater communication rate. The authors present their own general derivation of the single-channel capacity upper bound, which applies to *any* physically realizable technique for encoding information on the bosonic field and to *any* physically realizable detection scheme at the output. They also review the capacities of coherent communication channels that encode information in coherent states and in quadrature-squeezed states. A three-dimensional bosonic channel can employ many transverse modes as parallel single channels. An upper bound on the information flux that can be transferred down parallel bosonic channels is derived.

CONTENTS

I. INTRODUCTION

A. Optimum capacity of a single linear bosonic channel

The capacity (Shannon, 1948) of a communication channel is the maximum rate, usually measured in bits per second, at which information can be transmitted reliably from the channel's input to its output. The capacity C is defined formally as the maximum mutual information per second that can flow from input to output,

 0034-6861/94/66(2)/481(57)/$10.70

where mutual information (see Gallager, 1968, for a general treatment of information theory and its application to communication) is a particular entropic measure of information content, of the sort introduced by Shannon. The suitability of this definition comes from Shannon's fundamental result (Shannon, 1948; Gallager, 1968) that, with suitable coding and decoding, information can be transmitted without error at any rate up to and including capacity, but that any attempt to transmit information at a rate beyond capacity inevitably introduces errors.

Although capacity is an important measure of a channel's performance, it is by no means the only measure and is often not the most important practical measure. To achieve an information transfer rate approaching capacity generally requires complicated coding and decoding algorithms, in order to match the input and output data to the channel's properties and to overcome errors introduced by channel noise. There is a trade-off between channel noise and coding complexity. Given two channels with the same capacity but with different bit-error rates, one prefers the channel with lower bit-error rate, because it requires less complex error-correcting codes to overcome channel noise.

Despite this caveat, an important *physical* question in communication theory concerns the channel capacity that can be obtained with a given input power P. This review focuses exclusively on a restricted, but still important, version of this question: *what is the maximum communication rate that can be achieved when information is carried by a linear bosonic field, such as the electromagnetic field propagating in vacuum or in a linear dielectric medium?*

To address the question of the capacity of a *single* linear bosonic channel, it is essential first to relate the physicist's description of a bosonic field to the communication theorist's concept of a channel. To do so, we must distinguish transverse and longitudinal "modes" of the bosonic field: a transverse mode is defined by a particular spatial dependence perpendicular to the propagation direction and by a particular spin state or polarization; a given transverse mode supports many longitudinal modes, each defined by its spatial dependence along the direction of propagation or, equivalently, by its temporal dependence.

Since orthogonal transverse modes can be separated unambiguously at the channel output, at least in principle, information can be transmitted independently down each transverse mode. Thus each transverse mode is an independent communication channel. This leads us to the following correspondence: *a single transverse mode of a linear bosonic field corresponds to a single communication channel.* A good example of a single channel is a transverse mode of an electromagnetic transmission line; throughout a good part of this article we couch our discussion in terms of this example. In analyzing *single-channel* capacities, the transverse properties of the channel are unimportant; what *is* important are the longitudinal properties, bandwidth and duration, which determine the number of distinguishable longitudinal modes.

There are many different kinds of channels that can be accommodated within a particular transverse mode and a bandwidth for that mode. Each kind of channel corresponds to a technique for encoding information on the field mode and to a scheme for detecting the information at the channel output. The encoding technique amounts to choosing a set of quantum states to be transmitted down the channel, and the detection scheme is some kind of measurement to be performed on the field at the output. The capacity of a particular kind of channel is the maximum communication rate that can be achieved with the given encoding technique and detection scheme, where the maximum is taken over the *probabilities* of the input quantum states. We reserve the term *optimum capacity* for the maximum communication rate that can be achieved by further maximizing over all possible encoding techniques and detection schemes.

Classical physics imposes no limitations on channel capacity. Quantum-mechanical capacity limitations arise ultimately from the Einstein relation that a quantum at frequency f carries energy $E = hf$, where h is Planck's constant. Indeed, the optimum capacity of a single channel can be obtained from what amounts to dimensional analysis. There are three rates, or frequencies, associated with a single channel: (i) f_0, the channel's "carrier" frequency; (ii) P/hf_0, the rate at which quanta are transmitted down the channel; and (iii) C, the information rate. Setting these three rates approximately equal yields the optimum capacity:

$$P/hf_0 \sim C_{\text{opt}} \sim f_0 \,. \tag{1.1}$$

The left-hand relation means that it is optimal to use about one quantum per bit of information, and the right-hand relation means that it is optimal to transmit about one bit per period. Since this requires that the field be modulated on the time scale of a period, achieving the optimum capacity requires wideband radiation—i.e., radiation whose bandwidth approaches its maximum frequency.

Eliminating f_0 yields a physically sensible relation between power and the optimum single-channel capacity,

$$P \sim hC_{\text{opt}}^2 \,. \tag{1.2}$$

Throughout the bulk of this article we rewrite this relation as an upper bound on single-channel capacity (Gordon, 1961; Lebedev and Levitin, 1963, 1966; Marko, 1965; Bowen, 1967; Pendry, 1983; Bekenstein, 1988; Bekenstein and Schiffer, 1990; Yuen and Ozawa, 1992):

$$C \le C_{\text{opt}} \sim \sqrt{P/h} \,. \tag{1.3}$$

This simple dimensional analysis gives approximately the correct quantum limit on single-channel capacity. Despite this success, the two relations in Eq. (1.1) should be treated with caution, because neither by itself gives an upper bound on capacity. The capacity can be greater than P/hf_0—i.e., there can be more than one bit per quantum—but only at a cost of having far less than one bit per period. The capacity can also be greater than

f_0—i.e., there can be more than one bit per period—but only at a cost of having far less than one bit per quantum. The optimal strategy, which leads to the optimum wideband capacity, is to satisfy both relations in Eq. (1.1).

This optimal strategy introduces a theme that permeates this article. One can modulate the field once a period by transmitting a sequence of longitudinal modes, each of which lasts roughly a period. The optimal strategy then corresponds to using about one quantum for each bit and to transmitting about one bit for each longitudinal mode. This optimal strategy transcends the example of a single wideband channel. Quite generally, it is true that one can transmit more than one bit per quantum, but only at a cost of having far less than one bit per mode, and that one can transmit more than one bit per mode, but only at a cost of having far less than one bit per quantum. The optimal strategy for transmitting information within a power constraint never departs very far from using about one quantum for each bit and transmitting about one bit per mode—a strategy that we capture in the slogan "one quantum–one bit–one mode."

Throughout this article we deal with lossless channels, on the grounds that losses introduce noise that degrades performance. We further assume that the channels have no gain, on the grounds that power arising from gain should be included in the input power P. In most cases we also ignore thermal or other nonquantum noise, which can only reduce communication rates relative to the ideal quantum-limited cases in which we are interested.

We emphasize at the outset that the single-channel upper bound (1.3) does not imply an unavoidable energy *cost* for transmitting information. First and foremost, P is only the power transmitted down the channel. There is no reason, in principle, why this power must be consumed (Landauer, 1987, 1988, 1989), and thus it does not impose an in-principle cost on communication. Second, the limit (1.3) applies only to a single channel, as is evident (Landauer and Woo, 1973; Levitin, 1982; Pendry, 1983; Bekenstein and Schiffer, 1990) from applying Eq. (1.3) separately to each of several parallel channels, which together have input power P. (We consider parallel bosonic channels—i.e., multiple transverse modes of a bosonic field—in Sec. VI.) Third, it is possible that a *nonlinear* bosonic channel, perhaps by making use of solitons, can beat the capacity maximum (1.3). Although we have no indication that this is so, Landauer's (1987, 1988, 1989) work on limits to communication suggests that linearity is the most important assumption in this article and is the assumption that should be lifted in further investigations of quantum limits on capacity.

Numerous workers (Stern, 1960; Gordon, 1961, 1962; Lebedev and Levitin, 1963, 1966; Marko, 1965; Takahasi, 1965; Bowen, 1967; She, 1968; Helstrom *et al.*, 1970; Helstrom, 1974; Yuen *et al.*, 1975; Kabanov, 1978; Pierce, 1978; Davis, 1980; Pierce *et al.*, 1981; Pendry, 1983; Yamamoto and Haus, 1986; Saleh and Teich, 1987, 1992; Bekenstein, 1988; Bekenstein and Schiffer, 1990; Slusher and Yurke, 1990; Yamamoto, 1990; Yamamoto *et al.*, 1990; Schiffer, 1991, 1992; Yuen and Ozawa, 1992; Hall, 1993; Hall and O'Rourke, 1993) over the past 30 years have considered quantum limitations on the capacity of linear bosonic channels. A part of this work (Gordon, 1961; Lebedev and Levitin, 1963, 1966; Marko, 1965; Bowen, 1967; Helstrom, 1974; Yuen *et al.*, 1974; Pierce *et al.*, 1981; Pendry, 1983; Bekenstein, 1988; Bekenstein and Schiffer, 1990; Schiffer, 1991, 1992; Yuen and Ozawa, 1992) has considered wideband communication rates.

Despite this large body of work, there had been, until recently (Yuen and Ozawa, 1992), no general proof of the capacity upper bound for a single linear bosonic channel. The reason is that previous investigators assumed, explicitly or implicitly, what we call a number-state channel, in which information is encoded in the number of quanta in the bosonic field and is read out by counting quanta. This work thus left open the possibility that other transmission/detection techniques might lead to greater capacity. Indeed, although most investigators supported the existence of a capacity upper bound, there were claims (Helstrom, 1974; Yuen *et al.*, 1975; Pierce *et al.*, 1981) that quantum mechanics imposes *no* limits on the wideband capacity of a bosonic channel.

In this review we draw attention to important contributions in the large body of previous work, and we comment critically on some of it. Our main aims, however, are threefold. First, we familiarize physicists with the information-theoretic description of a communication channel, thereby providing a physicist's formulation of quantum communication theory. In particular, we relate the physicist's description of a bosonic field to the communication theorist's description of a communication channel. Second, we introduce to physicists a theorem proved by Holevo (1973), which deserves to be called the *fundamental theorem of quantum communication theory.* Holevo's theorem establishes an upper bound on transmitted information in terms of quantum entropy, thereby making it possible to apply quantum statistical physics to the question of quantum capacity limits. Third, we present our own general proof, based on Holevo's theorem, of the single-channel capacity upper bound. Our proof aims to keep the relevant physical considerations prominently on display. Along the way we develop the theme encapsulated in our slogan, "one quantum–one bit–one mode," and use it to estimate optimum capacities.

A recent review (Bekenstein and Schiffer, 1990) of quantum limits on information rates differs from this review in two important respects: although it considers a more general set of questions, it does not provide the tools to *prove* capacity upper bounds.

To ensure that readers are not left behind as we pursue our three aims, we are deliberately pedagogical, at the risk of being tedious. For readers who may not want to wade through the entire result, however, we provide in the next subsection a set of guideposts for reading the paper. We indicate which parts are essential for the development of quantum communication theory, which parts provide heuristic and intuitive arguments, and which parts are mainly of specialized interest.

B. Guideposts: What can be found where

The core of this review is contained in Secs. II–IV, where we introduce the tools to derive and prove the optimum wideband single-channel capacity

$$C_{\mathrm{WB}} = \frac{\pi}{\ln 2}\sqrt{\frac{2P}{3h}} \text{ bits/s.} \tag{1.4}$$

The development is interrupted by heuristic arguments and special cases, which are intended to provide insight into the sterile general considerations necessary for a general proof. Section II begins by formulating the general information-theoretic description of a single bosonic channel and then proceeds to use the barest minimum of this description to develop a heuristic argument for the optimum single-channel capacity. Section III illustrates the general formulation by specializing to the most studied example, the number-state channel, and derives the wideband capacity of a number-state channel, which turns out to be the optimum wideband capacity (1.4). Section IV returns to the general formulation, introduces Holevo's theorem, and gives our general derivation of the optimum wideband single-channel capacity.

Section II.A reviews the information-theoretic formulation for the capacity of a single channel. This formulation specifies a channel by giving input and output alphabets and statistical descriptions for transmission of the input letters and reception of the output letters. We introduce the notion of mutual information as the appropriate statistical measure of information successfully transmitted from input to output and thus develop the definition of capacity in terms of the maximum value of the mutual information, where the maximum is taken over the probabilities of the input letters, subject to any channel constraints. The information-theoretic formulation is then mapped onto the physical description of a single transverse mode of a linear bosonic field. The input alphabet becomes a collection of quantum states, transmitted with various probabilities. The output alphabet becomes the possible results of a measurement on the bosonic field—a detection scheme—with the statistical description of the output given by the quantum-statistical description of the measurement. Section II.A is essential reading for anyone not familiar with quantum communication theory.

The general development is interrupted in Sec. II.B, where we use the bare minimum of information theory to develop heuristic arguments for the optimum wideband capacity. We first review Shannon's theorem (Shannon, 1948; Gallager, 1968) for the capacity of a channel with additive, white, Gaussian noise. Shannon's theorem is the fundamental theorem of *classical* communication theory and gives the capacity in terms of the signal-to-noise ratio and the channel bandwidth. By itself it cannot limit the capacity, because there is no bound on the signal-to-noise ratio of a classical signal. When Shannon's theorem is combined with a plausible estimate for the level of quantum noise, however, an approximate bound $P \gtrsim hC^2$ emerges. We use these considerations as a springboard to develop two heuristic arguments for a capacity upper bound, one of which applies to wavelike channels, where the wave aspect of the field predominates, and the other of which applies to particle-like channels, where the particle aspect predominates. These heuristic arguments clarify the relation of the capacity upper bound to the Heisenberg uncertainty principle and to the Einstein relation $E = hf$, and they reveal why the two relations in Eq. (1.1) lead approximately to the optimum capacity.

Valuable—indeed, essential—though these heuristic arguments are, they cannot be regarded as proofs of a capacity upper bound. Prior to the work of Yuen and Ozawa (1992), all derivations of the exact single-channel upper bound (1.4) (Lebedev and Levitin, 1963, 1966; Bowen, 1967; Pendry, 1983; Bekenstein, 1988; Bekenstein and Schiffer, 1990) fail as proofs, because they explicitly or implicitly assume a special case—what we call a frequency-multiplexed number-state channel. The input states to a number-state channel are particle-number eigenstates (Fock states), which we call number states for short. Information is transmitted as a sequence of numbers of quanta and is detected by counting quanta. Frequency multiplexing is a convenient, but by no means general, way of encoding information on a wideband channel by breaking it up into many narrowband channels.

In Sec. III we focus on number-state channels, because they illustrate the general description given in Sec. II.A and because they achieve the optimum capacity. Section III.A reviews previous work on the capacity of a narrowband number-state channel (Stern, 1960; Gordon, 1962; Marko, 1965; Takahasi, 1965; Lebedev and Levitin, 1966; Bowen, 1967; Yamamoto and Haus, 1986; Bekenstein and Schiffer, 1990; Slusher and Yurke, 1990; Yamamoto, 1990; Yamamoto *et al.*, 1990; Hall, 1993), and Sec. III.B generalizes to a wideband frequency-multiplexed channel. The result of the wideband considerations is the capacity (1.4) of a single wideband frequency-multiplexed number-state channel (Lebedev and Levitin, 1963, 1966; Bowen, 1967; Pendry, 1983; Bekenstein, 1988; Bekenstein and Schiffer, 1990). This capacity is often cited as the single-channel capacity upper bound; but as long as its derivation is tied to number states, quantum counting, and frequency multiplexing, there remains the *possibility* that other transmission/detection schemes could exceed this capacity. The need to deal with this possibility is particularly acute given the claims (Helstrom, 1974; Yuen *et al.*, 1975; Pierce *et al.*, 1981) of infinite wideband capacity in the literature.

The optimum single-channel capacity (1.4) is identical to the entropy flux of one-dimensional blackbody radiation. Previous derivations of this capacity, with the exception of the work of Yuen and Ozawa (1992), *assume* a direct connection between information transport and quantum entropy $-\mathrm{tr}(\hat{\rho}\log_2\hat{\rho})$. With this assumption, C_{WB} emerges as a general upper bound on the capacity

of any single linear bosonic channel. In Sec. IV, we prove this key result by appealing to Holevo's (1973) theorem, which *establishes* a connection between mutual information and quantum entropy. Specifically, we use Holevo's theorem to show that C_{WB} is a universal upper bound on the wideband capacity of *any* physically realizable single linear bosonic channel. The results in Sec. IV free the upper bound (1.4) from dependence on frequency multiplexing or the use of number states and show that it is a general entropy-based upper limit that applies to any input quantum states and to any detection scheme at the output.

Sections IV.A and IV.B are essential reading, because they lay the groundwork for proving the single-channel capacity upper bound. Section IV.A, continuing the development begun in Sec. II.A, formulates in more detail the general quantum-mechanical description of a single linear bosonic channel. Since any proof of the capacity upper bound must cover all possible detection schemes, we introduce in Sec. IV.A the formalism of generalized measurements, which encompasses *all* measurements that are consistent with the rules of quantum mechanics. Section IV.B states Holevo's theorem, which we view as the fundamental theorem of quantum communication theory, and draws attention to restrictive assumptions made in Holevo's proof of the theorem. These assumptions limit our proof to finite input and output alphabets and to a finite-dimensional Hilbert space.

The next three subsections (Secs. IV.C–IV.E) present proofs of the single-channel capacity upper bound for three different ways of implementing the power constraint on a long but finite transmission time $\mathcal{T}$: (i) a constraint that fixes the energy transmitted during $\mathcal{T}$ to be a particular value (Sec. IV.C); (ii) a constraint that places an upper bound on the energy transmitted during $\mathcal{T}$ (Sec. IV.D); and (iii) a constraint on the average energy transmitted during $\mathcal{T}$ (Sec. IV.E). The first two ways of implementing the power constraint lead to species of the microcanonical ensemble of statistical physics, whereas the third way leads to the canonical ensemble. Readers interested in connections to statistical physics will want to digest all three proofs, but readers interested in the shortest route to the optimum wideband capacity can safely consult just the first (and shortest) proof. Indeed, the reader who wants only to see a proof of the optimum wideband single-channel capacity (1.4) can read Secs. II.A and IV.A–IV.C and ignore everything else.

Section IV.F addresses the limitations on our proof that arise from the finiteness assumptions made in Holevo's proof of his theorem. We argue that physical realizability imposes these finiteness assumptions on *real* communication channels and thus that our proof applies to any *physically realizable* channel. We also mention the recent work of Yuen and Ozawa (1992), which points out that results already in the mathematical physics literature can be used to remove the finiteness assumptions made in Holevo's original proof, thus making appeals to physical realizability unnecessary. We consider briefly the claims of infinite wideband capacity (Helstrom, 1974; Yuen *et al.*, 1975; Pierce *et al.*, 1981), but defer a critical discussion of them to Appendix E.

Recent laboratory work on the generation of "nonclassical light" provides a reason to consider capacity upper bounds. For example, the direct observation of sub-shot-noise intensity fluctuations—sometimes called number squeezing—in light from a semiconductor laser [Machida *et al.* (1987); for discussion and references, see Yamamoto *et al.* (1991) or Saleh and Teich (1992)] demonstrates a potential for using number states in a communication system. Another possible source of number-squeezed light is photon twinning in a nondegenerate parametric oscillator, with feedback or feedforward from one twin [Heidmann *et al.* (1987); Reynaud *et al.* (1987); for discussion and references, see Saleh and Teich (1992)]. The parametric oscillator produces pairs of photons in two beams. By observing photocounts in one beam, one can control via feedback the rate of emission of the parametric oscillator or adjust directly via feedforward the intensity of the unobserved beam. In either case, one can stabilize the intensity fluctuations in the unobserved beam and thus produce a near number state in that beam.

Another example of nonclassical light that has been generated in the laboratory employs "quadrature-squeezed states." Advances in the production of quadrature squeezing stem from the pioneering work of Slusher *et al.* (1985), who first generated and detected quadrature-squeezed light with noise below the vacuum noise level. Quadrature-squeezed light has now been generated and detected in a number of laboratories around the world [for reviews see Kimble and Walls (1987), Loudon and Knight (1987), Teich and Saleh (1989, 1990), and Zaheer and Zubairy (1991)]. Motivated partly by these laboratory developments and partly by a desire to have examples to contrast with number-state channels, we consider, in Sec. V, two kinds of coherent communication channels, which we call coherent-state channels and quadrature-squeezed channels.

The input states to a coherent-state channel are coherent states of the bosonic field. Information is transmitted in the complex amplitude of the field—i.e., in both the (real) amplitude and the phase. At the output, information is retrieved by measuring both real and imaginary parts—often called quadrature components—of the complex amplitude, i.e., by measuring both the real amplitude and the phase. Although such a measurement involves noncommuting observables, it *can* be described in quantum mechanics. Its realization by optical heterodyne detection provides a concrete example of the generalized-measurement formalism introduced in Sec. IV.A. The example illustrates an important result: measurements of noncommuting observables are inevitably degraded by quantum noise over and above the conventional quantum uncertainties of the observables.

The input states to a quadrature-squeezed channel are quadrature-squeezed states. Relative to a coherent state, a quadrature-squeezed state has reduced quantum uncertainty in one quadrature component, called the squeezed

quadrature. Information is transmitted in the coherent excitation of the squeezed quadrature and is read out at the output by measuring the squeezed quadrature. At optical frequencies, measurement of one quadrature component can be realized by optical homodyne detection.

Section V.A gives a brief review of the definitions and pertinent properties of coherent states and quadrature-squeezed states. In Sec. V.B we review the capacities of a narrowband coherent-state channel (Gordon, 1962; She, 1968; Yamamoto and Haus, 1986) and a narrowband quadrature-squeezed channel (Yamamoto and Haus, 1986). In Sec. V.C we consider wideband channels, finding the capacity of a single wideband frequency-multiplexed coherent-state channel (Gordon, 1961; Marko, 1965),

$$C_{\rm CS} = \frac{1}{\ln 2}\sqrt{\frac{2P}{h}}\ \text{bits/s}, \tag{1.5}$$

and the capacity of a single wideband frequency-multiplexed quadrature-squeezed channel,

$$C_{\rm QS} = \frac{2}{\ln 2}\sqrt{\frac{P}{h}}\ \text{bits/s}. \tag{1.6}$$

The latter result shows that coherent quadrature communication can achieve $\sqrt{6}/\pi \simeq 78\%$ of the capacity maximum (1.4).

The recent advances in production of quadrature-squeezed states, together with the development of high-T_c superconductors, might make it feasible to approach the $\sqrt{P/h}$ limit. In Sec. V.C we mention briefly the possibility that a high-T_c superconducting waveguide, excited with microwave quadrature-squeezed states, could meet the *physical* conditions for approaching the wideband capacity (1.6). In this regard the report (Movshovich *et al.*, 1990) of microwave-frequency squeezing by a Josephson parametric amplifier is a promising start.

As has already been noted, parallel channels can give a higher communication rate for a given input power P than can a single channel (Landauer and Woo, 1973; Levitin, 1982 Pendry, 1983; Bekenstein and Schiffer, 1990). We give the first detailed analysis of this question in Sec. VI, which generalizes the single-channel results to multichannel bosonic communication. The multiple parallel channels are the many transverse modes of a three-dimensional linear bosonic field. The important quantum-mechanical limit is now a limit on information *flux*.

What survives into the multichannel results is the optimal strategy embodied in our slogan, "one quantum–one bit–one mode." To apply the slogan, let $\mathcal{A}$ be the area of the aperture through which the channel is transmitted; g_0, the number of spin states or polarizations for each spatially transverse mode; and c, the field's propagation speed. We can then introduce three fluxes: (i) $g_0 f_0/(c/f_0)^2$, the maximum flux of field "modes"—maximal because these "modes" last only a period (i.e., they are transmitted at rate f_0) and occupy a square wavelength $(c/f_0)^2$ in the transverse dimensions; (ii) $P/\mathcal{A}hf_0$, the flux of field quanta; and (iii) $C/\mathcal{A}$, the information flux. Setting these three fluxes approximately equal yields the optimum information flux:

$$\frac{P}{\mathcal{A}hf_0} \sim \frac{C_{\rm opt}}{\mathcal{A}} \sim \frac{g_0 f_0}{(c/f_0)^2}\,. \tag{1.7}$$

Eliminating f_0 yields the upper bound on information flux,

$$\frac{C}{\mathcal{A}} \le \frac{C_{\rm opt}}{\mathcal{A}} \sim \left(\frac{g_0}{c^2}\right)^{1/4}\left(\frac{P/\mathcal{A}}{h}\right)^{3/4}\,. \tag{1.8}$$

We show in Sec. VI that this is approximately the correct upper bound on information flux for a linear bosonic field. Just as is true for single channels, neither relation in Eq. (1.7) sets an upper bound on information flux by itself, but the two relations are approximately valid for an optimal channel.

The discussion of multiple parallel channels in Sec. VI proceeds in parallel with the previous discussions of single channels. Section VI.A considers narrowband parallel bosonic channels, using number-state, coherent-state, and quadrature-squeezed channels as specific examples. Section VI.B generalizes to wideband channels and replaces the heuristic bound (1.8) by an exact upper bound on the information flux that can be carried through a roughly circular aperture of area $\mathcal{A}$ by any linear bosonic field that has an isotropic dispersion relation:

$$\frac{C_{\rm BPC}}{\mathcal{A}} = \frac{4\pi}{3\ln 2}\left(\frac{\pi g_0}{15c^2}\right)^{1/4}\left(\frac{P/\mathcal{A}}{h}\right)^{3/4}\ \text{bits/s m}^2 \tag{1.9}$$

(BPC stands for "bosonic parallel channels"). In this upper bound c is the field's *minimum* phase velocity. The upper bound is attained by a dispersionless frequency-multiplexed number-state channel. We also give analogous capacities for frequency-multiplexed coherent-state and quadrature-squeezed channels. We indicate how the $P^{3/4}$ dependence of Eq. (1.9) goes over to a $P^{1/2}$ dependence when $\mathcal{A}$ is small enough that the channel accommodates only a single spatially transverse mode.

Five appendixes deal with matters that would interrupt the flow of the main text. Appendix A summarizes crucial properties of Shannon information and quantum entropy. Appendix B assesses a distinction between heralded and self-heralding signals, which was introduced by Bekenstein (1988) and by Bekenstein and Schiffer (1990). Appendix C argues that dispersion does not affect the description of a single bosonic channel developed in Sec. IV.A. Appendix D.1 outlines the formalism of generalized measurements, and Appendix D.2 analyzes a specific example of a generalized measurement, the simultaneous measurement of two quadrature components by optical heterodyne detection. Appendix E examines critically the claims (Helstrom, 1974; Yuen *et al.*, 1975; Pierce *et al.*, 1981) of infinite wideband capacity and describes how each of these claims arises from ignoring the

Einstein relation $E = hf$—specifically, from assuming that the energy of a quantum is a constant, independent of frequency.

II. SINGLE LINEAR BOSONIC CHANNEL

A. General description

To describe and analyze a single bosonic communication channel requires us to develop the information-theoretic formulation of channel capacity and to map this formulation onto the physical description of a bosonic field. This subsection begins the task, which is completed in Sec. IV.A. Our aim is to give a pedagogical, but necessarily brief, introduction to the relevant elements of information theory.

A communication theorist views a communication channel abstractly as having an input and an output. A message is transmitted down the channel as a stream of letters selected from an input alphabet A. We label the letters of the input alphabet by an index $a = 1, \ldots, \mathcal{A}$, where $\mathcal{A}$ is the number of input letters. From the channel output pours another stream of letters that come from an output alphabet B. We label the output letters by an index $b = 1, \ldots, \mathcal{B}$, where $\mathcal{B}$ is the number of output letters. Each transmission of an input letter and receipt of a corresponding output letter is called a "use" of the channel.

Perfect fidelity in transmitting messages from input to output—i.e., a noise-free channel—means that the input can be reconstructed unambiguously from the output. By grouping and relabeling output letters, perfect fidelity can always be reduced to the case of identical input and output alphabets, with transmission of an input letter leading to reception of precisely the same letter. More generally—for example, if there is noise—the properties of a channel are described by a set of conditional probabilities $p_{B|A}(b|a)$, where $p_{B|A}(b|a)$ is the probability to receive output letter b, given transmission of input letter a. A noise-free channel can always be reduced to the case $p_{B|A}(b|a) = \delta_{ba}$.

What remains is to provide a measure of the quantity of information carried by a letter or a message. Shannon's (1948) genius was to realize that the required measure has to do with the *statistics* of the letters. The primitive notion of information is that a string of N *bi*nary digi*ts* (0's and 1's) carries N *bits* of information. Shannon's statistical measure of information generalizes this primitive notion and can be constructed in two steps. We take the first step by noting that the number of N-digit binary strings is $\mathcal{N} = 2^N$; thus the information carried by a string can be written as $N = \log_2 \mathcal{N} = \log_2$ (number of N-digit binary strings). The first step generalizes the primitive notion of information by saying that the information carried by a message, measured in bits, is the base-2 logarithm of the number of possible messages. With this generalization, one already realizes that information content is not a property of a particular message, but rather a property of the set of possibilities from which a message is drawn.

Applying this first generalization to our input alphabet leads to the conclusion that the information carried by a letter is $\log_2 \mathcal{A}$. Suppose, however, that one of the input letters is very likely to be transmitted and that another letter is transmitted only rarely. One's gut feeling is that receipt of the rare letter provides much more information than receipt of the likely letter. Thus one realizes, first, that the generalization of the first step assumes implicitly that all possibilities are equally likely and, second, that information content is determined not just by the set of possibilities, but also by the probabilities of the various possibilities. The second step in generalizing the primitive notion of information allows for unequal probabilities of the input letters. Fortunately, the first step is sufficient to develop the proper generalization.

Suppose that input letter a is transmitted with probability $p_A(a)$. Consider a message consisting of N letters. In the limit of a very long message, the only messages with nonvanishing probability are those in which the frequency of occurrence of each letter is equal to its probability. A multinomial coefficient gives the number of such messages,

$$\mathcal{N} = \frac{N!}{\prod_a [N p_A(a)]!} , \tag{2.1}$$

all of which are equally likely, with probability

$$\mathcal{N}^{-1} = \prod_a [p_A(a)]^{N p_A(a)} , \tag{2.2}$$

where Stirling's formula relates Eqs. (2.1) and (2.2) for large N. Hence the information carried by each long message is

$$\log_2 \mathcal{N} = N H(A) , \tag{2.3}$$

where

$$H(A) = -\sum_a p_A(a) \log_2 p_A(a) , \tag{2.4}$$

Shannon's statistical measure of information (Shannon, 1948; Gallager, 1968), quantifies the *average* information per letter of input, measured in bits. Notice that $H(A)$ agrees with the above considerations. If the input letters are equally likely, i.e., $p_A(a) = 1/\mathcal{A}$, then $H(A) = \log_2 \mathcal{A}$ reduces to the base-2 logarithm of the number of possibilities. At the other extreme, if one letter is transmitted with unity probability, then $H(A) = 0$ reflects the fact that there is no information in a message whose content is known in advance. For probabilities between these extreme cases, $H(A)$ lies between zero and $\log_2 \mathcal{A}$. Important properties of Shannon information are gathered together in Appendix A.

At this point we can summarize the ingredients of a communication theorist's description of a communication

channel (Gallager, 1968): *(i) an input alphabet, the letters of which are labeled by an index* $a = 1, \ldots, \mathcal{A}$, *and a set of probabilities* $p_A(a)$ *for transmitting input letter* a *at each "use" of the channel; (ii) an output alphabet, labeled by* $b = 1, \ldots, \mathcal{B}$, *and a set of conditional probabilities* $p_{B|A}(b|a)$ *for receiving output letter* b, *given transmission of input letter* a.

These ingredients are sufficient to give a complete statistical description of the channel. The joint probability for transmitting input letter a and receiving output letter b is given by

$$p_{A,B}(a,b) = p_{B|A}(b|a)p_A(a) = p_{A|B}(a|b)p_B(b) , \quad (2.5)$$

where

$$p_B(b) = \sum_a p_{A,B}(a,b) = \sum_a p_{B|A}(b|a)p_A(a) \quad (2.6)$$

is the unconditioned probability for receiving output letter b, and where

$$p_{A|B}(a|b) = \frac{p_{B|A}(b|a)p_A(a)}{p_B(b)} \quad (2.7)$$

is the conditional probability that input letter a was transmitted, given receipt of output letter b. Equation (2.7), which inverts the order of conditioning, is sometimes called Bayes's theorem.

The Shannon information $H(A)$ of Eq. (2.4) is the information (in bits) transmitted down the channel at each use. It is worth emphasizing how to think of this information: if one knows only the transmission probabilities $p_A(a)$, $H(A)$ is the *average* amount of information one obtains when one determines which input letter was transmitted. Although $H(A)$ is the information transmitted down the channel, generally only a part of this information is successfully transmitted from input to output, because channel noise prevents one from determining the input letter from the output letter. Another important Shannon information is the information available at the output at each use,

$$H(B) = -\sum_b p_B(b)\log_2 p_B(b) . \quad (2.8)$$

Again it is worth emphasizing that $H(B)$ is the average amount of information one obtains when, knowing only the reception probabilities $p_B(b)$, one determines which output letter was received. Like $H(A)$, $H(B)$ cannot be regarded as the information successfully transmitted from input to output: $H(B)$ can be artificially large simply because the output alphabet is much bigger than the input alphabet, or more importantly, it can contain unwanted contributions from channel noise.

To find a measure of information transmitted from input to output, consider the Shannon information

$$H(B|a) = -\sum_b p_{B|A}(b|a)\log_2 p_{B|A}(b|a) , \quad (2.9)$$

which is the information available at the output at each use, given that *one knows that input letter* a *was transmitted.* Since one knows which input letter was transmitted, $H(B|a)$ is clearly not useful information concerning which input letter was transmitted; instead it is information about the channel properties, embodied in the conditional probabilities $p_{B|A}(b|a)$, that de-correlate the outputs from input letter a. To get a measure of information transmitted from input to output, the information $H(B|a)$ should be subtracted from the total information at the output; but since the inputs are actually unknown, $H(B|a)$ must first be averaged over the input letters.

The result of this argument is a conditional Shannon information

$$\begin{aligned} H(B|A) &= \sum_a p_A(a)H(B|a) \\ &= -\sum_{a,b} p_{B|A}(b|a)p_A(a)\log_2 p_{B|A}(b|a) , \end{aligned} \quad (2.10)$$

which is the *average* information available at the output at each use, given that one knows which input letter was transmitted, where the average is taken over both inputs and outputs. For the bosonic channels considered in this article, the conditional information $H(B|A)$ characterizes channel noise—either intrinsically quantum-mechanical noise or noise from some other source.

Subtracting the conditional information (2.10) from the total information at the output yields what is called the mutual information (Gallager, 1968) between output and input,

$$H(B;A) = H(B) - H(B|A) = \sum_{a,b} p_{B|A}(b|a)p_A(a)\log_2\left(\frac{p_{B|A}(b|a)}{p_B(b)}\right) \geq 0 . \quad (2.11)$$

Mutual information is the appropriate measure of information successfully transmitted from input to output at each use. That mutual information is non-negative, as implied by the inequality in Eq (2.11), follows immediately from the Gibbs inequality (A3). The mutual information is zero if and only if A and B are statistically independent, i.e., $p_{A,B}(a,b) = p_A(a)p_B(b)$.

Bayes's theorem (2.7) implies that mutual information is symmetric in A and B and thus can also be written as

$$H(B;A) = \sum_{a,b} p_{A|B}(a|b)p_B(b)\log_2\left(\frac{p_{A|B}(a|b)}{p_A(a)}\right) = H(A) - H(A|B) = H(A;B) , \quad (2.12)$$

where

$$H(A|B) = -\sum_{a,b} p_{A|B}(a|b)p_B(b)\log_2 p_{A|B}(a|b) \tag{2.13}$$

is a conditional input information. The mutual information can be thought of as the average information about the input that is obtained from determining the output or as the average information about the output that is obtained from determining the input.

The zoo of statistical informations is completed by defining the joint information

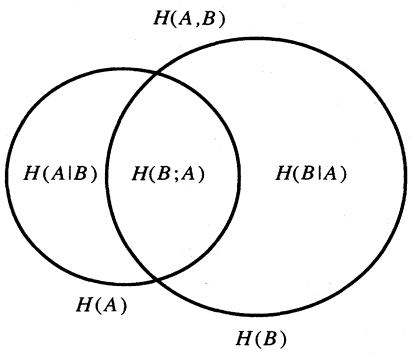
$$\begin{aligned} H(A,B) = H(B,A) &= -\sum_{a,b} p_{A,B}(a,b)\log_2 p_{A,B}(a,b) \\ &= H(A) + H(B|A) = H(B) + H(A|B)\ , \end{aligned} \tag{2.14}$$

which is the *average* information one obtains from determining both the input letter and the output letter. The relationships among the several information measures are summarized in Fig. 1.

A noise-free channel, in which the input can be reconstructed unambiguously from the output, is one for which the inverted conditional probabilities $p_{A|B}(a|b)$ are sharp: for each b, there is a unique a that is certain, i.e., has unity probability $p_{A|B}(a|b) = 1$. This is equivalent to the condition $H(A|B) = 0$ and thus to

$$H(B;A) = H(A) \quad \text{(noise-free channel)}\ ; \tag{2.15}$$

i.e., all the input information is successfully transmitted to the output. A noise-free channel does not require that the measurement conditional probabilities $p_{B|A}(b|a)$ be sharp—i.e., it does not require that $H(B|A) = 0$ or, equivalently, that $H(B;A) = H(B)$—because in a noise-free channel, there can be several output letters for each input letter. Given a noise-free channel, however, one can take the output letters that correspond to a particular input letter, group them into a single output letter with the same label as the corresponding input, and thus obtain a channel for which $p_{B|A}(b|a) = \delta_{ba}$.



$H(A) + H(B|A) = H(A,B) = H(B,A) = H(B) + H(A|B)$

$H(A) - H(A|B) = H(A;B) = H(B;A) = H(B) - H(B|A)$

FIG. 1. Relationships among information measures, conveniently summarized in a diagram where quantity of information is represented by area. The joint information $H(A,B) = H(B,A)$ [Eq. (2.14)] quantifies the combined information available at the input and the output and is represented by the area inside both circles. The input information $H(A)$ [Eq. (2.4)] is represented by the area of the left circle, and the output information $H(B)$ [Eq. (2.8)] is represented by the area of the right circle. The conditional informations $H(A|B)$ [Eq. (2.13)] and $H(B|A)$ [Eq. (2.10)] and the mutual information $H(A;B) = H(B;A)$ [Eqs. (2.11) and (2.12)] are represented by the three labeled areas, which together add up to the joint information. The diagram portrays correctly that mutual information is the information held in common by the input and the output.

The channel capacity per use, C (Gallager, 1968), is the maximum mutual information, where the maximum is taken over the input probabilities $p_A(a)$, subject to any constraints on the channel:

$$C \equiv \max_{\{p_A(a)\}} H(B;A)\ . \tag{2.16}$$

If one can identify the duration $\mathcal{T}$ of each use, then one can convert to capacity C as an information rate:

$$C = \frac{\mathcal{C}}{\mathcal{T}}\ . \tag{2.17}$$

The suitability of mutual information as a measure of information transmitted successfully down the channel is confirmed by Shannon's fundamental theorem (Shannon, 1948; Gallager, 1968), which states that the capacity, defined as the maximum mutual information rate, is the maximum rate at which information can be transmitted without error from input to output.

Having developed the information-theoretic description of a communication channel, we must now map it onto the physical description of a single bosonic channel—i.e, a single transverse mode of a bosonic field. This requires us to give physical realizations for the input and output letters and to specify how the channel statistics arise. This task is accomplished by the following correspondences: *(i) the input letters a correspond to quantum states (density operators) $\hat{\rho}_a$, one of which is transmitted during each use time $\mathcal{T}$, the probability for transmission of state $\hat{\rho}_a$ being $p_A(a)$; (ii) the output letters b correspond to the possible results of a measurement on the bosonic field at the channel output—i.e., the results of a detection scheme at the output—and the con-*

ditional probabilities $p_{B|A}(b|a)$ come from the quantum-statistical probabilities for the detection scheme and from the properties of any additional channel noise. It should be emphasized that a complete description of a channel requires specifying both the quantum states used—how is the information transmitted—and the measurement used at the output—how is the information detected.

Throughout most of this article, we restrict ourselves to quantum-limited channels; thus, throughout the bulk of the article, we consider only the intrinsically quantum-mechanical noise that arises from the quantum-statistical description of the output measurement acting on the input quantum states. The general physical description of a linear bosonic channel is depicted schematically in Fig. 2; we develop this description further and apply it to a general analysis of single bosonic channels in Sec. IV.

Practical techniques for encoding information on a transverse mode of the electromagnetic field all correspond to a particular way of specifying input quantum states, which, though not completely general, is important enough to warrant some attention. These practical techniques decompose the specification of input quantum states into two steps (see Fig. 2). In the first step, one chooses a set of orthogonal longitudinal field modes; and in the second step, one specifies quantum states for each mode. The overall quantum states $\hat{\rho}_a$ for time $\mathcal{T}$ are then *products* of states for the chosen longitudinal modes. These two steps neatly divorce the spatio-temporal properties of the field from the quantum-mechanical description of states. The spatio-temporal properties of the field along the propagation direction are contained in the choice of field modes. Each field mode can be regarded as a harmonic oscillator, and the relatively simple language and formalism of a quantum harmonic oscillator can be used to specify the input quantum states.

This two-step decomposition makes contact with practical techniques for encoding information on a transverse mode of the electromagnetic field. Specifically, a choice of

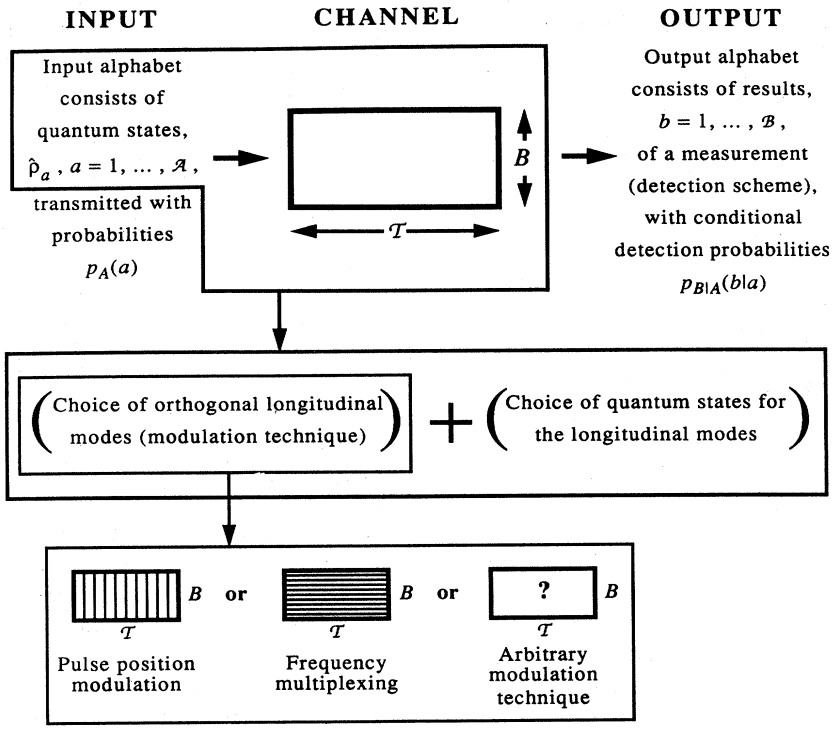


FIG. 2. Schematic physical description of a single linear bosonic channel. The input to the channel is a quantum state $\hat{\rho}_a$, selected with probability $p_A(a)$ from an alphabet of quantum states. The channel itself is described by a transmission time $\mathcal{T}$ and a bandwidth B and is depicted as a box in time-frequency space. The channel output is a result b of a measurement (a detection scheme); channel noise and the quantum mechanics of the detection scheme are described by conditional probabilities $p_{B|A}(b|a)$. The middle box illustrates a conventional, practical (but not completely general) way of decomposing the choice of input quantum states into two steps: (i) a choice of orthogonal longitudinal field modes within the allowed time-frequency box (this choice corresponds to a modulation technique for the bosonic field) and (ii) a choice of quantum states for the chosen modes, each of which is a harmonic oscillator. The total quantum state $\hat{\rho}_a$ is then a product of states for each of the longitudinal modes. The bottom box depicts modulation techniques schematically as ways of filling the time-frequency box with longitudinal modes: pulse position modulation uses short pulses, each of which fills the available bandwidth, whereas frequency multiplexing uses narrow frequency bins, each of which occupies the entire duration.

orthogonal longitudinal field modes corresponds to some technique for "modulating" the field to carry information (see Fig. 2). For example, a technique called pulse position modulation corresponds to choosing longitudinal field modes that are successive short pulses, each of which (in the simplest realization) can have one of two states, excited or unexcited. A contrasting modulation technique, called frequency multiplexing, divides the available bandwidth into narrow frequency bins, each of which corresponds to a longitudinal mode. Pulse position modulation and frequency multiplexing are extreme examples of allocating longitudinal modes between time and frequency. There is a potentially infinite number of modulation techniques, each corresponding to a choice for the longitudinal modes; a typical choice, of course, does not have such a simple description in terms of time and frequency.

General though this two-step decomposition is, it does not encompass all possibilities for specifying input quantum states, for two reasons. First, many quantum states cannot be written as product states for *any* choice of longitudinal modes, there being intrinsic entanglement among modes. Second, different input states might be product states for different choices of longitudinal modes; for example, one can easily imagine a channel—not necessarily a good one—in which one transmits sometimes pulse-position-modulated states and sometimes frequency-multiplexed states. We often use the two-step decomposition to describe particular kinds of channels, so it is worth emphasizing that our general analysis in Sec. IV is not in any way restricted to input quantum states specified by the two-step decomposition.

In the remainder of this article we introduce three kinds of channels as specific examples of quantum communication channels. The first of these, which we call a number-state channel, is considered in Sec. III. The inputs to a number-state channel are Fock states of the chosen orthogonal field modes—i.e., eigenstates of the number of quanta, which we call number states for short. Information is carried in the pattern of numbers of quanta among the various field modes. At the output the information is retrieved by counting the number of quanta in each field mode. For an optical-frequency electromagnetic channel, the counting could be done by a photodetector, a detection scheme called *direct detection.*[1]

In Sec. V we consider two kinds of coherent communication channels, which we call coherent-state channels and quadrature-squeezed channels. The inputs to a coherent-state channel are coherent states of the chosen field modes—i.e., eigenstates of the modal annihilation operators. Information is carried in the pattern of complex-amplitude excitations of the various field modes. At the output information is recovered by measuring both real and imaginary parts—often called quadrature components—of the complex amplitude of each field mode, i.e., by measuring both the real amplitude and the phase of each mode. Although such a measurement requires measuring two noncommuting observables, it *can* be described in quantum mechanics; and for an optical-frequency electromagnetic channel, it can be realized by a detection scheme called heterodyne detection (Yuen and Shapiro, 1978, 1980; Shapiro *et al.*, 1979).

The input states for a quadrature-squeezed channel are quadrature-squeezed states of the chosen field modes [for reviews of quadrature-squeezed states, see Kimble and Walls (1987), Loudon and Knight (1987), Teich and Saleh (1989, 1990), and Zaheer and Zubairy (1991)]. Relative to a coherent state, a quadrature-squeezed state has reduced quantum uncertainty in one quadrature component, called the squeezed quadrature. There is a corresponding increase in the uncertainty in the orthogonal quadrature component, called the amplified quadrature. One takes advantage of the reduced uncertainty by encoding information in the pattern of excitations of the squeezed quadratures of the various field modes. At the output information is retrieved by measuring the squeezed quadrature of each of the field modes. For an optical-frequency electromagnetic channel, measurement of one quadrature component can be realized by a detection scheme called homodyne detection (Yuen and Shapiro, 1978, 1980; Shapiro *et al.*, 1979; Yuen and Chan, 1983; Schumaker, 1984; Shapiro and Wagner, 1984; Shapiro, 1985; Yurke, 1985; Caves and Schumaker, 1986).

B. Heuristic arguments for single-channel capacity maximum

We now interrupt our development of the general formal description of a single bosonic channel, so that in this subsection we can apply physical intuition to the barest minimum of this description. The goal is to develop a *physical* understanding of the optimum single-channel capacities before proceeding with the formal development. The only information theory we need is that the Shannon information is the logarithm of the number of approximately equally likely possibilities.

We can gain a quick entry into thinking about quantum-limited capacities by recalling Shannon's theorem (Shannon, 1948; Gallager, 1968) for a single channel that has additive, white, Gaussian noise. Such a channel has capacity

$$C = B\log_2\left(1+\frac{P}{SB}\right) \text{ bits/s}, \tag{2.18}$$

[1]We assume that a photodetector is able to distinguish orthogonal longitudinal modes and thus to count separately the photons in each mode. For real photodetectors there is an obvious difficulty with this assumption, even for sequences of wave-packet modes, because the longitudinal modes generally overlap in time, even though they are orthogonal (the modes *must* overlap in time if they have a strictly finite bandwidth). Since orthogonal modes can be distinguished in principle, however, we ignore this difficulty throughout our analysis.

where P is the signal power, B is the channel bandwidth (in Hz), and S is the (white) power spectrum of channel noise (i.e., the noise power per Hz). Classical-wave channels typically have additive, white, Gaussian noise, so Shannon's theorem is a fundamental tool in classical communication theory. For a channel operating at frequency f, one might guess that quantum mechanics places a lower bound $S \gtrsim hf$ on the noise power spectrum. With this supposition Shannon's theorem yields a quantum-limited capacity

$$C = B\log_2\left(1+\frac{P}{hfB}\right) \text{ bits/s.} \tag{2.19}$$

A coherent-state channel is the closest thing to a classical-wave channel. As we discuss in Sec. V.B, the quantum-mechanical Shannon capacity (2.19) gives *exactly* the capacity of a single narrowband ($B \ll f$) coherent-state channel. Despite this success, Shannon's theorem is far too narrow a framework for a general analysis of quantum channel capacities, for two reasons: (i) quantum noise, even for narrowband channels, is generally neither additive nor Gaussian; (ii) quantum noise for wideband channels is never white—witness the nonwhite-noise power spectrum $S = hf$ for a coherent-state channel—because of the Einstein relation $E = hf$. The wider framework required for analyzing quantum-mechanical channels is provided by Holevo's (1973) theorem (see Sec. IV.B).

Nonetheless, we can gain considerable insight from the Shannon capacity (2.19) by putting it on a more general but admittedly heuristic footing. The dimensionless quantity P/hfB is the number of signal quanta transmitted down the channel per second per Hz. During a long transmission time $\mathcal{T}$, one can transmit at most $B\mathcal{T}$ orthogonal longitudinal modes (Gallager, 1968; Yamamoto and Haus, 1986). Thus P/hfB can also be regarded as the number of signal quanta per longitudinal mode. In this subsection we find it convenient to think of the longitudinal modes as successive pulses, which we call wave-packet modes, but the reader should keep in mind that the argument is not restricted to this particular choice. With this way of thinking, however, the bandwidth B becomes the transmission rate of wave-packet modes. In describing the bosonic field, one distinguishes a wave-like regime $P/hfB \gg 1$, where there are many signal quanta per mode, and a particle-like regime $P/hfB \ll 1$, where there are many modes per signal quantum (Yamamoto and Haus, 1986; Yamamoto, 1990; Yamamoto *et al.*, 1990). Recalling that Shannon's theorem arises from a classical-wave analysis, one might guess that the Shannon capacity (2.19) has general, but approximate, validity as the optimum capacity for narrowband channels in the wave-like regime.

This guess is supported by the following heuristic argument. For each wave-packet mode, construct a phase plane whose axes are scaled so that squared distance from the origin measures number of quanta. Then the power constraint means, crudely speaking, that the signals must be chosen from within a circle of radius $\simeq \sqrt{P/hfB}$, centered at the origin. Since each quantum state occupies a phase-plane area $\simeq \pi$, the maximum number of distinguishable quantum states that can be transmitted in each mode is $\gamma P/hfB$, where γ is a constant of order unity. The maximum information per wave-packet mode is $\log_2$(maximum number of distinguishable quantum states per mode) $= \log_2(\gamma P/hfB)$. The resulting optimum narrowband single-channel capacity in the wave-like regime is

$$C = \begin{pmatrix}\text{rate of}\\ \text{transmission}\\ \text{of modes}\end{pmatrix}\log_2\begin{pmatrix}\text{maximum number of}\\ \text{distinguishable states}\\ \text{per mode}\end{pmatrix}$$
$$= B\log_2\left(\frac{\gamma P}{hfB}\right) \text{ bits/s.} \tag{2.20}$$

The argument here is valid only in the wave-like regime, as is evident from the fact that the capacity (2.20) goes negative when $P/hfB < \gamma^{-1}$. The argument shows that quantum mechanics—i.e., the nonzero value of h—limits narrowband capacities in the wave-like regime in a manner reminiscent of the Shannon capacity (2.19). Furthermore, it reveals how to interpret the limitation: the Heisenberg uncertainty principle restricts the number of quantum-mechanical phase-space cells that can be crammed into a region of phase space whose size is limited by an energy constraint.

What about wideband capacities in the wave-like regime? One cannot just take the limit $B \to \infty$ because it takes one outside the wave-like regime. Indeed, taking the limit in the capacity (2.20) yields nonsense for just this reason. The Shannon capacity (2.19) does have a sensible limit $C \to P/hf\ln 2$ as $B \to \infty$, but this limit cannot be trusted for the same reason. Missing in taking the limit $B \to \infty$ is the frequency dependence of the energy of a quantum, which must be taken into account when the bandwidth becomes comparable to the channel frequency. For present purposes it is sufficient to estimate the wideband capacity by letting $B \sim f$—i.e., by letting the bandwidth be as large as possible consistent, at least crudely, with energy hf for a quantum. The resulting optimum wideband single-channel capacity in the wave-like regime is

$$C \sim f\log_2\left(\frac{\gamma P}{hf^2}\right) = \sqrt{\frac{\gamma P}{h}}\sqrt{\frac{hf^2}{\gamma P}}\log_2\left(\frac{\gamma P}{hf^2}\right)$$
$$\le \frac{2}{e\ln 2}\sqrt{\frac{\gamma P}{h}} \text{ bits/s,} \tag{2.21}$$

where the maximum on the right occurs for $f = e^{-1}\sqrt{\gamma P/h}$, just within the wave-like regime. This analysis indicates that in the wave-like regime the Einstein relation $E = hf$ is responsible for the $\sqrt{P/h}$ form of the capacity maximum (1.3).

A different heuristic argument yields an optimum narrowband capacity in the particle-like regime $P/hfB \ll 1$. Quanta are transmitted at a rate P/hf, and each

quantum can occupy one of at most $hfB/P \gg 1$ wave-packet modes. The information per quantum is $\log_2$(maximum number of distinguishable modes per quantum) $= \log_2(hfB/P)$. The resulting optimum narrowband single-channel capacity in the particle-like regime is

$$C = \begin{pmatrix} \text{rate of} \\ \text{transmission} \\ \text{of quanta} \end{pmatrix} \log_2 \begin{pmatrix} \text{maximum number of} \\ \text{distinguishable modes} \\ \text{per quantum} \end{pmatrix} = \frac{P}{hf} \log_2\left(\frac{hfB}{P}\right) \text{ bits/s.} \quad (2.22)$$

This argument is clearly valid only in the particle-like regime, as is evident from the fact that the capacity (2.22) goes negative when $P/hfB > 1$. Quantum mechanics limits narrowband capacities in the particle-like regime in the sense that the capacity (2.22) increases as h decreases (so long as $P/hfB < e^{-1}$). It is more to the point, however, to say that the particle-like regime is intrinsically quantum mechanical: if $h \to 0$, the particle-like regime disappears, leaving only a classical-wave regime that has no capacity restrictions.

What about wideband capacities in the particle-like regime? The limit $B \to \infty$ in Eq. (2.22) yields infinite capacity, but just as for the wave-like regime, this limit is unphysical because it neglects the Einstein relation $E = hf$. We can again take the Einstein relation into account crudely by letting $B \sim f$, thus finding an optimum wideband single-channel capacity in the particle-like regime,

$$C \sim \frac{P}{hf} \log_2\left(\frac{hf^2}{P}\right) = \sqrt{\frac{P}{h}} \sqrt{\frac{P}{hf^2}} \log_2\left(\frac{hf^2}{P}\right) \leq \frac{2}{e \ln 2} \sqrt{\frac{P}{h}} \text{ bits/s,} \quad (2.23)$$

where the maximum on the right occurs for $f = e\sqrt{P/h}$, just within the particle-like regime. This analysis indicates that in the particle-like regime the Einstein relation $E = hf$ is responsible for the $\sqrt{P/h}$ capacity maximum of Eq. (1.3).

These heuristic arguments capture the essence of the optimum single-channel capacities. It is useful to summarize the lessons they teach. For narrowband channels the dimensionless quantity P/hfB, the number of quanta per longitudinal mode, plays a crucial role: it delineates wave-like and particle-like regimes, and it determines the behavior of the capacity in both regimes [Eqs. (2.20) and (2.22)]. The nonzero value of Planck's constant limits directly the capacity in the wave-like regime and is responsible for the very existence of the particle-like regime.

For wideband channels, where $B \sim f$, the Einstein relation $E = hf$ leads, in both the wave-like and particle-like regimes, to an optimum capacity $C_{\text{opt}} \sim \sqrt{P/h}$. This optimum wideband capacity is achieved in the transition zone between the wave-like and particle-like regimes, by adjusting the frequency so that $P/hf^2 \sim 1$, which implies $1 \sim C/f$. The optimum wideband capacities thus reveal the origin of the two relations in Eq. (1.1), which we capture in our slogan, "one quantum–one bit–one mode." In the wideband wave-like capacity (2.21), the linear dependence on f outside the logarithm, which comes from the transmission rate of wave-packet modes, swamps the logarithmic dependence on f, which comes from the number of quanta per mode. This dictates choosing f nearly as large as is consistent with a wave-like channel, an optimal strategy described roughly by "one quantum–one bit–one mode." In the wideband particle-like capacity (2.23), the linear dependence on $1/f$ outside the logarithm, which comes from the transmission rate of quanta, swamps the logarithmic dependence on f, which comes from the number of modes per quantum. This dictates choosing f nearly as small as is consistent with a particle-like channel, again an optimal strategy described roughly by "one quantum–one bit–one mode."

III. CAPACITIES OF SINGLE NUMBER-STATE CHANNEL

We turn now to an analysis of a single number-state channel. This analysis is important for two reasons: it illustrates the general formalism begun in Sec. II.A, and it yields the optimum capacity proved in Sec. IV.

Consider then a single linear bosonic communication channel. A good example is a single transverse mode of an electromagnetic transmission line; throughout the remaining discussion of single channels (Secs. III–V), we use language appropriate to this example. We define a number-state channel as one in which information is transmitted as a pattern of precise numbers of photons in the various orthogonal field modes—i.e., the input states are number states of the field modes—and information is retrieved at the output using perfectly efficient direct photodetection of the number of photons in each orthogonal mode (see footnote 1).

We begin this section with yet another heuristic argument, which captures the essence of the quantum limit on wideband capacity. Suppose we transmit photons of maximum frequency f. The channel can be divided into N nonoverlapping bins of different frequency (frequency multiplexing), each with bandwidth $b \lesssim f/N$. Suppose further that we transmit binary (on-off) information down each frequency bin, the bits of information spaced in time by $1/b$. The capacity (in bits/s) of each frequency bin is b; so the capacity of the entire channel becomes $C = bN \lesssim f$. Several photons might be used for the "on" state in each bin; but to minimize the input power, it is clearly desirable to use only one photon for the "on" state, in which case each bin has average photon intensity $b/2$ (corresponding to 2 bits per photon). Since the average frequency is $\gtrsim f/2$, the average input power P to the entire channel satisfies

$$P \gtrsim h(f/2)(Nb/2) = \tfrac{1}{4}hfC \gtrsim \tfrac{1}{4}hC^2 \quad (3.1)$$

Gordon (1961) posited this inequality by counting the number of ways of populating energy-time cells whose size is determined by the energy-time uncertainty principle. The essence of the inequality (3.1) is that bigger capacity ultimately entails higher frequencies, where a photon requires more energy.

A. Narrowband number-state channel

1. Channel model and channel capacity

We proceed now to a rigorous analysis of a single number-state channel. We consider in this subsection a narrowband number-state channel that operates at frequency f within a bandwidth $B \ll f$, generalizing in Sec. III.B to the wideband case. It is convenient in the narrowband analysis to think of the channel as being excited by successive pulses of duration $\simeq B^{-1}$. The pulses, emitted at a rate B (Gallager, 1968; Yamamoto and Haus, 1986), make up a particular set of orthogonal longitudinal modes, which we call wave-packet modes. We can regard each wave packet as constituting a "use" of the channel (in the communication-theory jargon introduced in Sec. II.A); so the inverse bandwidth B^{-1} becomes the "use duration," and the bandwidth B becomes the "use rate" (number of uses per second). It should be emphasized that nothing in our analysis depends on the wave-packet choice of longitudinal modes. All that matters is that there are $B\mathcal{T}$ orthogonal longitudinal modes within a long transmission time $\mathcal{T}$. We indicate below how things work out for a general choice of longitudinal modes.

A description of the input is the first ingredient in the channel description. The inputs to the channel are photon-number eigenstates $|n\rangle$—number states—of the wave-packet modes; thus the input alphabet is the set of non-negative integers, and a transmitted message is a sequence of integers. The input is described statistically by probabilities $p_N(n)$ to transmit n photons down the channel at each use. We let

$$\bar{n} = \sum_{n=0}^{\infty} n p_N(n) \tag{3.2}$$

denote the average number of signal photons transmitted per use (or per mode)—i.e., the average number of photons per second per Hz. The average photon transmission rate is $B\bar{n}$, and hence the average input power to the channel is given by

$$P = Bhf\bar{n} \,. \tag{3.3}$$

The average number of photons per use, $\bar{n} = P/hfB$, is the crucial dimensionless quantity identified in Sec. II.B. For a narrowband channel with fixed frequency and bandwidth, we can implement the power constraint as the constraint that the probabilities $p_N(n)$ have the well-defined mean $\bar{n}$.

A description of the output is the second ingredient. At the channel output the photons are counted by an ideal photodetector (see footnote 1). Thus the output alphabet is also the set of non-negative integers, and a received message is a sequence of integers. Channel noise, should there be any, is characterized by a conditional probability $p_{M|N}(m|n)$, the probability to count m photons at the output when n photons are transmitted at the input. The unconditioned probability to count m photons at the output is

$$p_M(m) = \sum_n p_{M|N}(m|n) p_N(n) \,. \tag{3.4}$$

The information transmitted per use from input to output is quantified by the mutual information (Gallager, 1968),

$$H(M;N) = H(M) - H(M|N) = \sum_{n,m} p_{M|N}(m|n) p_N(n) \log_2 \left(\frac{p_{M|N}(m|n)}{p_M(m)} \right) \tag{3.5}$$

[cf. Eq. (2.11)], where

$$H(M) = -\sum_m p_M(m) \log_2 p_M(m) \tag{3.6}$$

[cf. Eq. (2.8)] is the total information available at the output and

$$H(M|N) = -\sum_{n,m} p_{M|N}(m|n) p_N(n) \log_2 p_{M|N}(m|n) \tag{3.7}$$

[cf. Eq. (2.10)] is a conditional output information that characterizes the channel noise (see Fig. 1).

Bekenstein (1988) and Bekenstein and Schiffer (1990) distinguish what they call heralded and self-heralding signals. A number-state channel provides an excellent forum for illustrating and assessing this distinction, but we relegate the discussion to Appendix B so as not to interrupt the capacity analysis.

The channel capacity per use (Gallager, 1968) or, more generally, per longitudinal mode, $\mathcal{C}$, is the maximum mutual information, where the maximum is taken over the possible input probabilities $p_N(n)$, subject to the constraint on average photon number $\bar{n}$ (i.e., subject to the constraint on average input power $P = Bhf\bar{n}$); in symbols, $\mathcal{C}$ is defined by

$$\mathcal{C} \equiv \max_{\{p_N(n)\}} H(M;N) \,, \tag{3.8}$$

where the maximum is taken subject to the normalization

constraint,

$$\sum_{n=0}^{\infty} p_N(n) = 1 , \tag{3.9}$$

and to the constraint (3.2) on the average number of photons per use (or per mode). Conversion from capacity per use to capacity in bits/s is accomplished by

$$C = B\mathcal{C} . \tag{3.10}$$

Notice that for an arbitrary choice of orthogonal longitudinal modes occupying a transmission time $\mathcal{T}$, we would write the information transmitted over time $\mathcal{T}$ as $B\mathcal{T}\mathcal{C}$ and thus find a capacity $C = B\mathcal{T}\mathcal{C}/\mathcal{T} = B\mathcal{C}$, just as in Eq. (3.10).

Suppose now that the channel is contaminated by "thermal noise" in the following way: before the signal photons are deposited in each wave-packet mode, the mode is already excited by a thermal radiator at temperature T. At each channel use the thermal radiator has probability

$$q(k) = \frac{1}{1+\bar{n}_T}\left(\frac{\bar{n}_T}{1+\bar{n}_T}\right)^k \tag{3.11}$$

to emit k photons, where

$$\bar{n}_T = \frac{1}{e^{hf/k_BT}-1} \tag{3.12}$$

is the mean number of thermal photons emitted at each use. The addition of n signal photons simply shifts the thermal distribution (3.11) so that $k = 0$ corresponds to n photons. Since the photons, signal or thermal, are counted with unit efficiency at the output, we model the channel by a conditional probability (Lebedev and Levitin, 1966; Bekenstein and Schiffer, 1990),

$$p_{M|N}(m|n) = \begin{cases} 0 , & m < n , \\ q(m-n) , & m \geq n , \end{cases} \tag{3.13}$$

to count m photons at the output, given transmission of n signal photons.

Hall (1993) has recently developed a quantum-mechanical description of Gaussian channel noise, which clarifies the physical interpretation of the above model of thermal noise. In particular, Hall's work prompted us to realize that the model corresponds to the scenario of initial thermal noise, to which the signal photons are added. Hall analyzes the capacity in a different scenario in which the ordering of noise and signal are reversed: the signal photons are first added to each mode, after which Gaussian noise is added, perhaps as the radiation propagates down the channel.

For the model of thermal noise embodied in Eq. (3.13), the conditional output information (3.7) reduces to the entropy of the thermal radiation (in bits),

$$\begin{aligned} H(M|N) &= -\sum_{k=0}^{\infty} q(k)\log_2 q(k) \\ &= \bar{n}_T \log_2(1+\bar{n}_T^{-1}) + \log_2(1+\bar{n}_T) , \end{aligned} \tag{3.14}$$

independent of the input probabilities. Thus, to maximize the mutual information $H(M;N)$, we need only maximize the output information $H(M)$. Indeed, *in this case*, we can find the maximum by maximizing $H(M)$ with respect to the *output* probabilities $p_M(m)$, subject to the constraints that the output distribution be normalized and that its mean be $\bar{n}+\bar{n}_T$. This problem is the same as finding the canonical distribution for a harmonic oscillator and has the standard result,

$$p_M(m) = \frac{1}{1+\bar{n}+\bar{n}_T}\left(\frac{\bar{n}+\bar{n}_T}{1+\bar{n}+\bar{n}_T}\right)^m , \tag{3.15}$$

which leads to maximum output information

$$\begin{aligned} H(M) = {} & (\bar{n}+\bar{n}_T)\log_2\left(1+\frac{1}{\bar{n}+\bar{n}_T}\right) \\ & + \log_2(1+\bar{n}+\bar{n}_T) . \end{aligned} \tag{3.16}$$

Generally, maximizing with respect to the output probabilities yields only an upper bound on capacity, because inversion of the linear equations (3.4) does not yield nonnegative input probabilities $p_N(n)$. In this case, however, it does, and the input probabilities were found by Lebedev and Levitin (1966; for a derivation, see Bekenstein and Schiffer, 1990):

$$p_N(n) = \begin{cases} \dfrac{p_M(0)}{q(0)} = \dfrac{1+\bar{n}_T}{1+\bar{n}+\bar{n}_T} , & n = 0 , \\ \dfrac{\bar{n}}{\bar{n}+\bar{n}_T} p_M(n) , & n > 0 . \end{cases} \tag{3.17}$$

This input distribution is not a thermal distribution (unless $\bar{n}_T = 0$), but it can be obtained from a thermal distribution with mean number of photons $\bar{n}+\bar{n}_T$ in the following way: remove a fraction $\bar{n}_T/(\bar{n}+\bar{n}_T)$ of the thermal probability from each possibility $n > 0$, and add the total probability removed, $\bar{n}_T/(1+\bar{n}+\bar{n}_T)$, to the thermal probability, $1/(1+\bar{n}+\bar{n}_T)$, for no photons.

Combining Eqs. (3.8), (3.5), (3.16), and (3.14) yields the capacity per use of a single narrowband number-state channel at temperature T (Gordon, 1962; Marko, 1965; Lebedev and Levitin, 1966; Yamamoto and Haus, 1986; Bekenstein and Schiffer, 1990; Yamamoto, 1990; Yamamoto *et al.*, 1990; Hall, 1993):

$$\begin{aligned} \mathcal{C}_T = {} & (\bar{n}+\bar{n}_T)\log_2\left(1+\frac{1}{\bar{n}+\bar{n}_T}\right) + \log_2\left(1+\frac{\bar{n}}{1+\bar{n}_T}\right) \\ & -\bar{n}_T\log_2\left(1+\frac{1}{\bar{n}_T}\right) . \end{aligned} \tag{3.18}$$

Hall (1993) analyzes the capacity of a number-state channel contaminated by his (different) model of Gaussian

noise, and Hall and O'Rourke (1993), using that model, have calculated bit-error rates for a number-state channel contaminated by Gaussian noise.

We review the nonzero-temperature capacity here to indicate the $T \neq 0$ corrections. Our main interest, however, is the $T = 0$ limit, where the capacity is largest. Specializing to $T = 0$, one finds the capacity per use (or per mode) of a single narrowband zero-temperature number-state channel:

$$\mathcal{C}_0 = \bar{n}\log_2(1+\bar{n}^{-1}) + \log_2(1+\bar{n}) \; . \tag{3.19}$$

Converting to bits/s [Eq. (3.10)] yields the corresponding capacity (Stern, 1960; Gordon, 1962; Marko, 1965; Takahasi, 1965; Lebedev and Levitin, 1966; Bowen, 1967; Yamamoto and Haus, 1986; Bekenstein and Schiffer, 1990; Slusher and Yurke, 1990):

$$C_0 = B\mathcal{C}_0 = B[\bar{n}\log_2(1+\bar{n}^{-1}) + \log_2(1+\bar{n})] \; . \tag{3.20}$$

This capacity is a combination of the wave-like and particle-like capacities (2.20) and (2.22). We emphasize that even though C_0 is derived here as the capacity of a number-state channel, the approach developed in Sec. IV shows that it is an upper bound on the capacity of *any* single narrowband linear bosonic channel.

Notice that at $T = 0$ capacity is achieved for thermal input probabilities

$$p_N(n) = p_M(n) = \frac{1}{1+\bar{n}}\left(\frac{\bar{n}}{1+\bar{n}}\right)^n \; . \tag{3.21}$$

At capacity the information *per photon* (in bits) is

$$\frac{\mathcal{C}_0}{\bar{n}} = \frac{C_0}{B\bar{n}} = \log_2(1+\bar{n}^{-1}) + \bar{n}^{-1}\log_2(1+\bar{n}) \equiv F_1(\bar{n}) \; . \tag{3.22}$$

Notice that the information per use (or per mode) is $\mathcal{C}_0 = C_0/B = F_1(\bar{n}^{-1})$.

2. Maximum channel capacity

Suppose now that we have additional freedom to vary the frequency and bandwidth of the channel in order to increase its capacity. For a real transmitting medium there are limits to how much the frequency f and the bandwidth B can be varied—limits set by absorption, dispersion, or other physical limitations of the medium. As a rough rule, however, the bandwidth can increase as the frequency increases. To model this relation between bandwidth and frequency, we assume that the channel has a fixed fractional bandwidth $\eta = B/f \ll 1$, and we maximize the capacity C_0 by varying f (or $\bar{n}$), subject to the constraints of fixed η and fixed input power

$$P = Bhf\bar{n} = \eta h f^2 \bar{n} \; . \tag{3.23}$$

Noting that

$$\sqrt{\frac{\eta P}{h}} = \eta f \bar{n}^{1/2} = B\bar{n}^{1/2} \tag{3.24}$$

is fixed by the constraints on fractional bandwidth η and input power P, we introduce a dimensionless capacity

$$\frac{C_0}{\sqrt{\eta P/h}} = \frac{\mathcal{C}_0}{\bar{n}^{1/2}} = \bar{n}^{1/2}\log_2(1+\bar{n}^{-1}) + \bar{n}^{-1/2}\log_2(1+\bar{n}) \equiv F_2(\bar{n}) \; , \tag{3.25}$$

which can be maximized easily as a function of $\bar{n}$. The dimensionless functions $F_1(\bar{n})$ and $F_2(\bar{n})$, defined in Eqs. (3.22) and (3.25), are plotted in Fig. 3. Since $F_2(\bar{n})$ is symmetric under the exchange $\bar{n} \leftrightarrow \bar{n}^{-1}$, it evidently has a maximum at exactly $\bar{n} = 1$, corresponding to frequency

$$f = \sqrt{\frac{P}{\eta h}} \tag{3.26}$$

and to maximum single-channel capacity

$$C_{\max} = 2B = 2\eta f = 2\sqrt{\frac{\eta P}{h}} \text{ bits/s.} \tag{3.27}$$

At maximum the information per photon is $C_{\max}/B\bar{n} = F_1(1) = 2$ bits, as in our heuristic argument, although this rigorous analysis shows that it is optimal to use all possible photon numbers, with an average of one photon per channel use, instead of the half photon per use that is optimal for binary signaling. The above results are summarized in Table I. If we regard frequency, rather than power, as the independent variable, we can write $C_{\max}$ and P in terms of ratios of fractional bandwidth and frequency to "typical" values for optical communication:

$$C_{\max} = 2\eta f \simeq 6.0\,\text{Gbits/s}\left(\frac{\eta}{10^{-5}}\right)\left(\frac{f}{3\times 10^{14}\,\text{Hz}}\right) \; , \tag{3.28a}$$

$$P = \eta h f^2 \simeq 0.60\,\text{nW}\left(\frac{\eta}{10^{-5}}\right)\left(\frac{f}{3\times 10^{14}\,\text{Hz}}\right)^2 \; . \tag{3.28b}$$

Yamamoto and Haus (1986) point out that the information per photon, $C_0/B\bar{n} = \mathcal{C}_0/\bar{n} = F_1(\bar{n})$ [Eq. (3.22)], goes as $\log_2 \bar{n}^{-1}$ for small $\bar{n}$ and thus diverges. To maximize the information per photon in a single narrowband channel, one lets $\bar{n}$ go to zero (as f goes to infinity) . For a given transmission period $\mathcal{T}$, however, there is a natural minimum value of $\bar{n}$ corresponding to making the frequency high enough that on average just one photon

TABLE I. Narrowband and wideband single-channel maximum capacities for three types of channels: number-state (NS) channels, quadrature-squeezed (QS) channels, and coherent-state (CS) channels. For each type of narrowband channel, the table lists the following quantities at maximum capacity: (i) the average number of photons per mode, $\bar{n}$; (ii) the dimensionless frequency, $f/\sqrt{P/\eta h} = \bar{n}^{-1/2}$; (iii) the information per mode, $C_{\max}/B = \mathcal{C}$; (iv) the dimensionless capacity, $C_{\max}/\sqrt{\eta P/h} = \mathcal{C}/\bar{n}^{1/2}$; and (v) the information per photon, $C_{\max}/B\bar{n} = \mathcal{C}/\bar{n}$. The first, third, and fifth of these quantities illustrate our slogan, "one photon–one bit–one mode." For each type of wideband channel, the table lists the dimensionless capacity, $C/\sqrt{P/h}$, and the information per photon, I/P. The information per photon is zero, because the low-frequency $1/f$ dependence of the photon-number spectra (3.35) and (5.48) leads to a logarithmic divergence in photon transmission rate; a low-frequency cutoff renders the information per photon finite, as in Eqs. (3.42) and (5.53).

		Narrowband				Wideband	
Type	$\bar{n}$	$\frac{f}{\sqrt{P/\eta h}} = \bar{n}^{-1/2}$	$\frac{C_{\max}}{B} = \mathcal{C}$	$\frac{C_{\max}}{\sqrt{\eta P/h}} = \frac{\mathcal{C}}{\bar{n}^{1/2}}$	$\frac{C_{\max}}{B\bar{n}} = \frac{\mathcal{C}}{\bar{n}}$	$\frac{C}{\sqrt{P/h}}$	I/P
NS	1.000 0	1.000 0	2.000 0	2.000 0	2.000 0	3.700 7	0
QS	1.960 8	0.714 14	2.299 1	1.641 9	1.172 6	2.885 4	0
CS	3.921 6	0.504 98	2.299 1	1.161 0	0.586 28	2.040 3	0

is transmitted during $\mathcal{T}$; the information can be encoded, for example, in the photon's arrival time (pulse position modulation), which is defined with resolution B^{-1} Pierce (1978) and Pierce, Posner, and Rodemich (1981) have also drawn attention to the fact that a single photon can carry an arbitrarily large amount of information, encoded in its arrival time. *The above considerations show, however, that sending just one photon during the transmission period is not the best strategy for using the photon's energy.*

Suppose one transmits on average one photon at frequency f_1 within bandwidth B_1 during a long transmission time $\mathcal{T} \gg B_1^{-1}$ ($P = hf_1/\mathcal{T}$), with the information encoded in the photon's arrival time; this gives an average number of photons per use $\bar{n}_1 = 1/B_1\mathcal{T} \ll 1$ and thus yields, from Eq. (3.20), a capacity

$$C_1 \simeq B_1\bar{n}_1 \log_2 \bar{n}_1^{-1} = \frac{\log_2 \bar{n}_1^{-1}}{\mathcal{T}} = \frac{\log_2(B_1\mathcal{T})}{\mathcal{T}} = \frac{1}{\mathcal{T}} \log_2 \left(\begin{array}{c} \text{number of} \\ \text{distinguishable} \\ \text{arrival times} \end{array} \right) \tag{3.29}$$

One gets the same capacity if one encodes information in the photon's frequency (Yamamoto, 1990; Yamamoto *et al.*, 1990), because the frequency can be defined with resolution $\mathcal{T}^{-1}$, thus yielding $B_1\mathcal{T}$ distinguishable frequencies within bandwidth B_1. Indeed, one can transmit the information in the photon's occupation of any one of a set of orthogonal longitudinal modes, of which there are $B_1\mathcal{T}$ within a bandwidth B_1 and a duration $\mathcal{T}$. The information $\log_2(B_1\mathcal{T})$ is "placement information" (Landauer, 1989) arising from the $B_1\mathcal{T} = hf_1B_1/P$ possible modes that the photon can occupy (cf. the discussion of the particle-like regime in Sec. II.B).

Suppose now that one transmits the same energy hf_1 in the same time $\mathcal{T}$ at a lower frequency f_2 within a bandwidth B_2, where f_2 and B_2 are chosen to keep the fractional bandwidth fixed and to yield an average of one photon per channel use—i.e., $\bar{n}_2 = 1$. Since $B_1\bar{n}_1^{1/2} = \sqrt{\eta P/h} = B_2\bar{n}_2^{1/2} = B_2$ [Eq. (3.24)], these choices give a bandwidth $B_2 = B_1\bar{n}_1^{1/2} = B_1/\sqrt{B_1\mathcal{T}} \ll B_1$ and, from Eq. (3.27), a corresponding capacity

$$C_2 = 2B_2 = 2B_1\bar{n}_1^{1/2} = \frac{2\sqrt{B_1\mathcal{T}}}{\mathcal{T}} \gg C_1 . \tag{3.30}$$

The capacity at the lower frequency is larger even though the bandwidth is smaller.

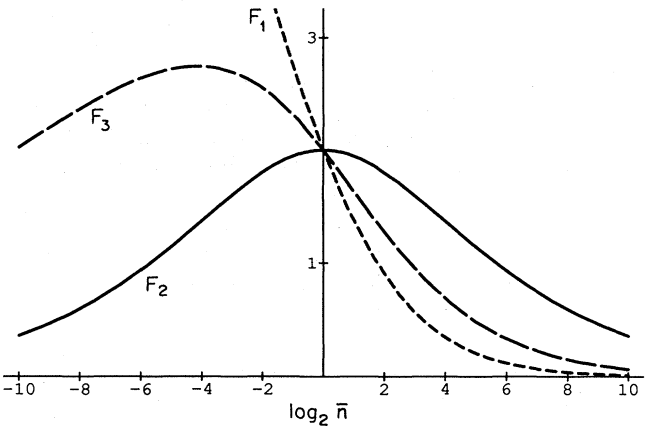


FIG. 3. Semilog plots of the dimensionless capacity functions $F_1(\bar{n})$ (short-dashed line), $F_2(\bar{n})$ (solid line), and $F_3(\bar{n})$ (long-dashed line) for number-state channels. The function $F_1(\bar{n})$ is the information per photon, defined for a single channel by Eq. (3.22) and for multiple parallel channels by Eq. (6.16). The function $F_2(\bar{n})$ is a dimensionless capacity, defined for a single channel by Eq. (3.25) and for multiple parallel channels by Eq. (6.25). The function $F_3(\bar{n})$ is a dimensionless information flux, defined for multiple parallel channels by Eq. (6.32). The functions $F_2(\bar{n})$ and $F_3(\bar{n})$ both tend to zero as $\log_2 \bar{n} \to \pm\infty$; $F_1(\bar{n})$ goes to zero as $\log_2 \bar{n} \to \infty$, but $F_1(\bar{n}) \to -\log_2 \bar{n} + 1/\ln 2$ as $\log_2 \bar{n} \to -\infty$.

This simple analysis illuminates the question of how much information a single photon can carry down a single narrowband channel. It is certainly possible to transmit an arbitrarily large amount of information on a single photon, the information transmitted as placement information, but doing so is not the most efficient way to use the photon's *energy*. One could send more information by using the *same* amount of energy to transmit many photons at a lower frequency. The optimal strategy for using the available energy, as Eq. (3.27) shows, is to use a frequency such that each photon carries on average 2 bits of information.

The flip side of this argument is that the information per mode, $\mathcal{C}_0$, can be made as large as desired by letting $\bar{n}$ go to infinity (as f goes to zero). Suppose, for example, that one works at frequency f_1 within bandwidth B_1, transmitting on average $\bar{n}_1 = P/hf_1B_1 \gg 1$ photons per mode. The corresponding capacity, from Eq. (3.20), is

$$C_1 \simeq B_1 \log_2 \bar{n}_1 \tag{3.31}$$

(cf. the discussion of the wave-like regime in Sec. II.B). If, instead, one has available the same power at a higher frequency f_2 within a bandwidth B_2, where f_2 and B_2 are chosen to keep the fractional bandwidth fixed and to yield an average of one photon per channel use, i.e., $\bar{n}_2 = 1$, then the new bandwidth is $B_2 = B_1\bar{n}_1^{1/2}$ and, from Eq. (3.27), the capacity at the higher frequency,

$$C_2 = 2B_2 = 2B_1\bar{n}_1^{1/2} \gg C_1 , \tag{3.32}$$

is larger. It is certainly possible to transmit an arbitrarily large amount of information in a single mode, but one could send more information by transmitting fewer photons at a higher frequency.

The lesson of both these arguments is summarized by our slogan, "one photon–one bit–one mode": the optimal strategy for using available power is to transmit about one photon in each mode, carrying about one bit of information.

The maximum capacity $C_{\max}$ [Eq. (3.27)] increases with increasing fractional bandwidth. Pushing the above derivation beyond its region of validity and setting $\eta = 1$ yields the capacity limit (3.1). To deal with such a wideband channel properly, however, it is mandatory to take into account the variation of photon energy with frequency.

B. Wideband frequency-multiplexed number-state channel

Consider a single zero-temperature *frequency-multiplexed* channel, in which each frequency bin is a narrowband number-state channel of the sort just considered; i.e., the inputs to each bin are number states which are detected at the output by an ideal photodetector. Let b be the bandwidth of each frequency bin; f_i, the frequency of the ith bin; and $b\bar{n}_i$, the average photon transmission rate in the ith bin. The input power and capacity of each frequency bin are given by Eqs. (3.3) and (3.20). The total input power

$$P = b\sum_i hf_i\bar{n}_i \tag{3.33}$$

and the overall information rate

$$C = \frac{b}{\ln 2}\sum_i [\bar{n}_i \ln(1+\bar{n}_i^{-1}) + \ln(1+\bar{n}_i)] \tag{3.34}$$

are evaluated by summing over the frequency bins.

For each frequency bin, the information rate is already maximized, subject to a constraint on the mean number of photons per mode, $\bar{n}_i$. What remains to find the wideband capacity is to maximize the overall information rate (3.34) by varying the distribution of power among frequencies—i.e., by varying $\bar{n}_i$ as a function of frequency, with P held constant. This process is precisely equivalent to maximizing the entropy in a channel of photons with positive momentum, and the resulting photon-number distribution is that of a one-dimensional blackbody,

$$\bar{n}_i = \frac{1}{e^{\beta hf_i} - 1} . \tag{3.35}$$

The Lagrange multiplier β has nothing to do with physical temperature T, which is here assumed to be zero; what β does do is to characterize, in a way identical to "inverse physical temperature," the optimal distribution of power and information among frequencies.

Replacing the sums (3.33) and (3.34) by integrals over all positive frequencies, one finds the standard results that

$$P = \frac{1}{h\beta^2}\int_0^\infty dx\,\frac{x}{e^x-1} = \frac{\zeta(2)}{h\beta^2} = \frac{\pi^2}{6h\beta^2} , \tag{3.36}$$

where $\zeta(2) = \pi^2/6$ is a value of the Riemann zeta function, and

$$\begin{aligned} C &= \frac{1}{h\beta\ln 2}\int_0^\infty dx\left(\frac{x}{e^x-1} - \ln(1-e^{-x})\right) \\ &= \frac{2\beta P}{\ln 2} = \frac{\pi^2}{3h\beta\ln 2} . \end{aligned} \tag{3.37}$$

Eliminating β between these two relations leads to the wideband capacity

$$C_{\mathrm{WB}} = \frac{\pi}{\ln 2}\sqrt{\frac{2P}{3h}}\ \text{bits/s} \tag{3.38}$$

of Eq. (1.4) (Lebedev and Levitin, 1963, 1966; Bowen, 1967; Pendry, 1983; Bekenstein, 1988; Bekenstein and Schiffer, 1990; see Table I). The generalization of C_{WB} to the capacity of a single wideband frequency-multiplexed *nonzero-temperature* number-state channel has been given by Lebedev and Levitin (1963, 1966) and is reviewed by Bekenstein and Schiffer (1990).

We emphasize that the wideband capacity C_{WB} is independent of propagation speed along the channel (Pendry,

1983; Bekenstein and Schiffer, 1990). Indeed, since the above derivation of $C_{\rm WB}$ is carried out in the frequency domain, it is valid even in the presence of dispersion; the propagation (group) velocity along the channel at frequency f simply scales all spatial dimensions along the channel at that frequency and thus does not appear in the result.[2] To achieve the capacity $C_{\rm WB}$, however, there must be no gaps in the frequency spectrum, since the presence of a gap means that information must be transmitted at higher frequencies, where the energy per photon is larger. Pendry (1983) gives a particularly neat derivation, which shows explicitly that $C_{\rm WB}$ is independent of dispersion and that $C_{\rm WB}$ decreases if there are gaps in the frequency spectrum. Pendry uses the same technique to derive an upper bound on the capacity of a single fermionic channel.

It is easy to estimate the effect of cutoffs in the frequency spectrum at very high and very low frequencies. There is a characteristic frequency

$$f_{\rm ch} = \frac{\pi^2}{3h\beta} = C_{\rm WB}\ln 2 = \pi\sqrt{\frac{2P}{3h}}\,, \tag{3.39}$$

near which most of the energy is carried. If there are low- and high-frequency cutoffs at frequencies $f_{\rm min}$ and $f_{\rm max}$, such that $f_{\rm min} \ll f_{\rm ch} \ll f_{\rm max}$, the capacity is little affected. The main effect is a fractional reduction in the capacity (3.38) by an amount

$$\frac{f_{\rm min}}{f_{\rm ch}}\ln\left(\frac{3ef_{\rm ch}}{\pi^2 f_{\rm min}}\right) \tag{3.40}$$

due to the elimination of frequencies below $f_{\rm min}$; the fractional reduction due to the high-frequency cutoff $f_{\rm max}$ is smaller than $(f_{\rm max}/f_{\rm ch})\exp(-\pi^2 f_{\rm max}/3f_{\rm ch})$.

Notice that in the absence of a low-frequency cutoff, the average photon intensity is infinite—hence the information per photon is formally zero—because of the $1/f$ divergence in $\bar{n}_i$ at low frequencies. Given a low-frequency cutoff at $f_{\rm min} \ll f_{\rm ch}$, however, the average photon transmission rate becomes

$$-\frac{3f_{\rm ch}}{\pi^2}\ln(1-e^{-\pi^2 f_{\rm min}/3f_{\rm ch}}) \simeq \frac{3f_{\rm ch}}{\pi^2}\ln\left(\frac{3f_{\rm ch}}{\pi^2 f_{\rm min}}\right)\,. \tag{3.41}$$

The corresponding information per photon,

$$\frac{\pi^2}{3\ln 2}\frac{1}{\ln(3f_{\rm ch}/\pi^2 f_{\rm min})}\,, \tag{3.42}$$

is of order unity for any reasonable value of $f_{\rm min}$.

Except for the work of Yuen and Ozawa (1992), previous derivations of the wideband capacity (1.4) apply only to a frequency-multiplexed number-state channel (Lebedev and Levitin, 1963, 1966; Bowen, 1967) or *assume* a connection between information and quantum entropy $-{\rm tr}(\hat{\rho}\log_2\hat{\rho})$ (Pendry, 1983; Bekenstein, 1988; Bekenstein and Schiffer, 1990)—a connection that is implemented in effect by assuming a frequency-multiplexed number-state channel. The upper bound $C_{\rm WB}$ is independent of the medium used and clearly has a general entropic origin. It should be possible to demonstrate that it is an upper bound on the capacity of any single linear bosonic channel, regardless of the technique used to modulate information onto the bosonic field, regardless of the input quantum states, and regardless of the detection scheme used at the output. To that demonstration we turn in Sec. IV.

IV. CAPACITY UPPER BOUNDS FOR SINGLE WIDEBAND LINEAR BOSONIC CHANNEL

In this section we consider again a single linear bosonic communication channel, which we think of as a single transverse mode of an electromagnetic transmission line. Our objective is to demonstrate that the capacity $C_{\rm WB}$ of a wideband frequency-multiplexed number-state channel is an upper bound on the capacity of any single linear bosonic channel. To that end, we return in Sec. IV.A to the general description of a single channel begun in Sec. II.A, extending that description and introducing the formalism of generalized measurements, which is required for analyzing general detection schemes at the channel output. To derive the optimum wideband capacity, we rely on a theorem due to Holevo (1973), introduced in Sec. IV.B, which *establishes* a connection between mutual information and quantum entropy. In Secs. IV.C–IV.E we derive, using methods from quantum statistical physics, the optimum wideband capacity for three different ways of imposing the constraint on input power P. For finite transmission times, the capacity limit depends on how the power constraint is imposed, but the capacity $C_{\rm WB}$ emerges in the limit of long transmission times. No restriction is placed on the communication method, except that the method be physically realizable in a sense that we discuss and justify in Sec. IV.F.

The reader interested in the minimal route to a proof of the optimum wideband capacity should consult Secs. IV.A and IV.B for essential background and then read just the first proof in Sec. IV.C, skipping the rest of the section. Readers interested in more detail will want to pick among the remaining subsections, choosing for careful reading those of most interest.

A. Quantum description of channel

In this section we assume a *finite* transmission time $\mathcal{T}$. We make this assumption because it is the first step

[2]For electromagnetic waves propagating in a linear dielectric medium, causality applied to the relation between polarization and electric field implies that dispersion must be accompanied by loss at some frequencies. Our wideband analysis applies strictly only if the loss occurs at frequencies much larger than the characteristic frequency (3.39). Since loss introduces noise, however, it always degrades performance and thus leads to a lower capacity.

toward a *finite-dimensional* channel Hilbert space, which is needed to apply Holevo's theorem as originally proved by Holevo. We are not interested here in small values of $\mathcal{T}$, and we assume explicitly that $\mathcal{T}$ is large compared to all other relevant times. For the fixed-energy channel considered in Sec. IV.C and the maximum-energy channel considered in Sec. IV.D, we derive *exact* expressions for the optimum capacity, we show that the $\mathcal{T} \to \infty$ limit of these expressions yields the optimum capacity $C_{\rm WB}$ of Eq. (1.4), and we derive the largest finite-$\mathcal{T}$ corrections to the $\mathcal{T} \to \infty$ limit.

For the average-energy channel considered in Sec. IV.E, we derive carefully the largest finite-$\mathcal{T}$ corrections to the $\mathcal{T} \to \infty$ capacity upper bound (1.4). Bekenstein and Schiffer (1990) have dealt with these same finite-$\mathcal{T}$ corrections. Unlike us, however, they are interested primarily in the regime where $\mathcal{T}$ becomes as small as $1/\sqrt{P/h}$, the inverse of the channel's characteristic frequency. For the average-energy channel, they derive the same leading-order corrections, plus one further term in the asymptotic expansion, which allows them to explore the regime $\sqrt{P/h}\mathcal{T} \sim 1$.

If the channel is dispersionless, then the assumption of a finite transmission time allows us to use the standard procedure of quantizing in terms of discrete frequency-domain longitudinal modes defined by periodic boundary conditions on a length $\mathcal{L} = c\mathcal{T}$ of the channel, where c is the propagation speed down the channel.[3] This procedure yields *one* frequency-domain longitudinal mode at each of a set of allowed frequencies

$$f_j = j/\mathcal{T} \ , \quad j = 1, 2, \ldots \ , \tag{4.1}$$

which are multiples of a minimum frequency $1/\mathcal{T}$. (There are actually two modes at each frequency, propagating in opposite directions; we use the mode that propagates in the direction of interest.)

It might seem that our work in this section is restricted to dispersionless channels, but we argue to the contrary in Appendix C. By considering the quantization of a channel with dispersion, that argument leads to the conclusion that the mode frequencies (4.1) apply even to a channel with dispersion. Here we stress the conclusion: *a single channel of duration $\mathcal{T}$ can be defined operationally by saying that there is one longitudinal frequency-domain mode at each of the frequencies (4.1).* The further results of this section use only this set of mode frequencies, without reference to dispersion or propagation speed along the channel, and thus are independent of dispersion and propagation speed.

The Hilbert space for the channel is spanned by photon-number eigenstates (Fock states) of the frequency-domain longitudinal modes,

$$|n_1, n_2, \ldots\rangle = |\{n_j\}\rangle \ , \tag{4.2}$$

where n_j is the number of photons in mode j. The Fock state $|\{n_j\}\rangle$ is an energy eigenstate with energy

$$E_N = \sum_j h f_j n_j = \frac{h}{\mathcal{T}} \sum_j j n_j = \frac{Nh}{\mathcal{T}} \ . \tag{4.3}$$

All the energy eigenvalues E_N are multiples of the quantum $h/\mathcal{T}$ for the lowest mode and can be labeled by the integer

$$N = \sum_{j=1}^{\infty} j n_j \ . \tag{4.4}$$

For N large the energy eigenstates are highly degenerate. If we let $\mathcal{N}_N$ denote the number of states with energy E_N, then $\mathcal{N}_N$ is the number of ways of writing the positive integer N as a sum (4.4). Determining $\mathcal{N}_N$ is precisely the number-theoretic problem of partitioning N—i.e., of determining the number of ways that up to N positive integers j can be added together to give N. Conventionally one chooses $\mathcal{N}_0 = 1$, which corresponds in our physical example to counting the vacuum state. The first few values of $\mathcal{N}_N$ are $\mathcal{N}_1 = 1$, $\mathcal{N}_2 = 2$, $\mathcal{N}_3 = 3$, $\mathcal{N}_4 = 5$, $\mathcal{N}_5 = 7$, and $\mathcal{N}_6 = 11$.

Our task now is to describe in more detail the two ingredients in a channel description: the input quantum states, together with their statistics, and the output detection scheme, together with its quantum statistics. The general description of a linear bosonic channel is depicted schematically in Fig. 2.

The input to the channel is *one* of the quantum states (density operators) $\hat{\rho}_a$, drawn from states labeled by $a = 1, \ldots, \mathcal{A}$. We emphasize that we are not talking about transmitting a sequence of the states $\hat{\rho}_a$ during time $\mathcal{T}$; each state $\hat{\rho}_a$ constitutes the *entire* message transmitted during time $\mathcal{T}$. We nonetheless refer to the set of input states as the input alphabet; a message consists of a single "letter" from the input alphabet. In communication-theory jargon the transmission time $\mathcal{T}$ becomes here a single "use" of the channel. If we applied the present description to the number-state channel analyzed in Sec. III.A, the input states $\hat{\rho}_a$ would be direct products of number states for successive wave-packet modes. The input is described statistically by probabilities $p_A(a)$ to transmit state $\hat{\rho}_a$. If one does not know which quantum state is transmitted, then one attributes to the channel an unconditioned density operator

$$\hat{\rho} = \sum_a p_A(a) \hat{\rho}_a \ . \tag{4.5}$$

We further emphasize that even though we build up the channel Hilbert space in terms of Fock states of frequency-domain modes, nothing restricts our analysis to frequency multiplexing or to number states. The mode

[3] Periodic boundary conditions on the length $\mathcal{L} = c\mathcal{T}$ provide a convenient way to count orthogonal longitudinal modes. The counting is equivalent to that obtained using physical modes that vanish outside of $\mathcal{L}$, except at the lowest frequencies near $\mathcal{T}^{-1}$.

frequencies (4.1) are just a way to count the number of orthogonal longitudinal modes per unit bandwidth, and the Fock states are just a fancy way to describe the channel Hilbert space as a direct product of harmonic-oscillator Hilbert spaces. If one thought in terms of the two-step procedure for specifying input quantum states (Sec. II.A)—first choose orthogonal longitudinal modes and then specify input quantum states for the chosen modes—then one would say that our analysis applies to arbitrary modulation techniques and to arbitrary choices of input quantum states for the modes. In fact, however, the present description of a linear bosonic channel is more general: not relying on the two-step procedure, it applies to completely general choices of input quantum states that cannot be described within the two-step procedure.

At the output one attempts to retrieve the transmitted information by using some kind of detection scheme, i.e., by making some sort of measurement. To prove general capacity upper bounds, we must somehow include in our analysis all possible measurements allowed by quantum mechanics. The most general quantum-mechanical measurement can be described in terms of a complete set of (bounded) non-negative Hermitian operators $\hat{F}_b$, where $b = 1, \ldots, \mathcal{B}$. These operators are called effects. [For a brief discussion of effects and generalized measurements, see Appendix D; for more extensive discussions, see Davies (1976), Holevo (1982), or Kraus (1983)]. The effects must be complete in the sense that they provide a resolution of the unit operator $\hat{1}$ as

$$\sum_b \hat{F}_b = \hat{1} \ . \tag{4.6}$$

Such a generalized measurement can be regarded as a measurement of a quantity B, a "generalized observable," whose possible values are labeled for convenience by $b = 1, \ldots, \mathcal{B}$. We call this set of values the output alphabet. Just as for the input, we emphasize that the output for time $\mathcal{T}$ consists of a single "letter" from the output alphabet. If we applied this description to the number-state channel analyzed in Sec. III.A, each letter in the output alphabet would be a sequence of detected photon numbers, and each effect would describe a sequence of measurements of photon number.

To describe the output statistically, we need the conditional probability to obtain output b, given that state $\hat{\rho}_a$ was transmitted (Davies, 1967; Holevo, 1982; Kraus, 1983):

$$p_{B|A}(b|a) = \mathrm{tr}(\hat{\rho}_a \hat{F}_b) \ . \tag{4.7}$$

These probabilities are real and non-negative because of the Hermiticity and non-negativity of the effects, and they are normalized to unity because of the completeness property (4.6). Indeed, the requirements that the probabilities (4.7) be real, non-negative, and normalized to unity for all quantum states $\hat{\rho}_a$ determine the required properties of the effects; thus it should not be surprising that effects can describe the most general quantum-mechanical measurement, because they are the most general way to generate probabilities from quantum states. If one does not know which quantum state was transmitted, then the probability for output b is the unconditioned probability

$$p_B(b) = \sum_a p_{B|A}(b|a) p_A(a) = \mathrm{tr}(\hat{\rho}\hat{F}_b) \ , \tag{4.8}$$

which depends only on the unconditioned channel density operator $\hat{\rho}$.

A generalized measurement reduces to a conventional quantum measurement if the effects are a complete set of *orthogonal projection operators* $\hat{\Pi}_b$, in which case the measurement probabilities (4.7) become

$$p_{B|A}(b|a) = \mathrm{tr}(\hat{\rho}_a \hat{\Pi}_b) \ . \tag{4.9}$$

We call a conventional quantum measurement ideal if, in addition to being orthogonal, the projection operators are *one-dimensional*. For ideal conventional measurements the projection operators have the form

$$\hat{\Pi}_b = \hat{\Pi}_b^{(1)} = |b\rangle\langle b| \ , \tag{4.10}$$

where the vectors $|b\rangle$ make up a complete, orthonormal basis, and the measurement probabilities reduce to the familiar formula

$$p_{B|A}(b|a) = \mathrm{tr}(\hat{\rho}_a \hat{\Pi}_b^{(1)}) = \langle b|\hat{\rho}_a|b\rangle \ . \tag{4.11}$$

The use of "ideal" conveys the notion that ideal conventional measurements have sufficient resolution to distinguish all the eigenvalues b—resolution that is sacrificed if the orthogonal projection operators are multidimensional.

Although important, conventional quantum measurements are by no means the most general. The effects can be multiples of *nonorthogonal* projection operators, which make up an (over)complete set; the coherent-state projectors of a single field mode provide an example, the measurement of which can be realized at optical frequencies by heterodyne detection (Yuen and Shapiro, 1978, 1980; Shapiro *et al.*, 1979; Shapiro and Wagner, 1984; Shapiro, 1985). Moreover, the effects need not be projection operators at all. We discuss effects and generalized measurements in Appendix D.1, where we also show, in Appendix D.2, that heterodyne detection is described by coherent-state projectors; readers unfamiliar with formalism of generalized measurements will find it useful to consult Appendix D.1 before proceeding.

Channel noise is characterized by the conditional probabilities $p_{B|A}(b|a)$. Again it is interesting to contrast the situation here with the analysis of a number-state channel in Sec. III.A. There, at zero temperature, there was no channel noise, because ideal photocounting of number states has perfect fidelity, i.e., $p_{M|N}(m|n) = \delta_{mn}$; channel noise was introduced by a finite-temperature thermal radiator. Here there is no source of noise such as a thermal radiator, yet the detection scheme generally has intrinsically quantum-mechanical noise, described by the

quantum-statistical conditional probabilities $p_{B|A}(b|a)$. A noise-free channel like a number-state channel can be obtained as follows: choose a complete set of orthogonal projection operators $\hat{\Pi}_a$, and then choose input states $\hat{\rho}_a = \hat{\Pi}_a/\mathrm{tr}(\hat{\Pi}_a)$ and effects $\hat{F}_b = \hat{\Pi}_b$, in which case $p_{B|A}(b|a) = \delta_{ba}$.

B. Holevo's theorem

Holevo (1973) has proved the fundamental theorem of quantum communication theory. Before stating and discussing Holevo's theorem, it is useful to recall the information measures introduced in Sec. II.A. The information successfully transmitted from input to output of a communication channel is quantified by the mutual information

$$H(B;A) = H(B) - H(B|A) = \sum_{a,b} p_{B|A}(b|a)p_A(a)\log_2\left(\frac{p_{B|A}(b|a)}{p_B(b)}\right) \geq 0 \tag{4.12}$$

of Eq. (2.11), where $H(B)$ [Eq. (2.8)] is the total information available at the output and $H(B|A)$ [Eq. (2.10)] is a conditional output information that characterizes channel noise. The mutual information is symmetric in A and B and thus can also be written as $H(B;A) = H(A) - H(A|B)$ [Eq. (2.12)], where $H(A)$ [Eq. (2.4)] is the total input information and $H(A|B)$ [Eq. (2.13)] is a conditional input information. The relationships among these information measures are summarized in Fig. 1. Since the conditional informations $H(A|B)$ and $H(B|A)$ are non-negative, one has immediately that the mutual information is bounded above by the input information—i.e., $H(B;A) \leq H(A)$—and by the output information—i.e., $H(B;A) \leq H(B)$.

Consider now a set of input quantum states $\hat{\rho}_a$ and probabilities $p_A(a)$. For any detection scheme at the output, described by a complete set of effects $\hat{F}_b$, one has available the channel statistics embodied in the conditional probabilities $p_{B|A}(b|a)$. These ingredients lead to the mutual information $H(B;A)$, which quantifies the amount of information successfully transmitted from input to output. Holevo's theorem establishes an upper bound on the mutual information over *all* possible generalized measurements at the output. The bound is in terms of the entropy of the unconditioned channel density operator and the entropies of the input states. In symbols the theorem reads

Holevo's theorem:

$$\max_{\{\hat{F}_b\}} H(B;A) \leq S(\hat{\rho}) - \sum_a p_A(a)S(\hat{\rho}_a)\,, \tag{4.13}$$

where the maximum of the mutual information $H(B;A)$ is taken over all complete sets of effects $\hat{F}_b$ and where

$$S(\hat{\rho}) = -\mathrm{tr}(\hat{\rho}\log_2\hat{\rho}) \tag{4.14}$$

is the quantum entropy (in bits) of a density operator $\hat{\rho}$. Important properties of quantum entropy are surveyed in Appendix A. Holevo's theorem was apparently first conjectured by Gordon (1964).

Equality holds in Eq. (4.13) if and only if all the input states commute, i.e., $[\hat{\rho}_a, \hat{\rho}_{a'}] = 0$ for all a and a'. One direction—commutation implies equality—of this if-and-only-if statement can be demonstrated easily. If all the input states commute, then they can all be made diagonal in the same orthonormal basis $|b\rangle$, i.e.,

$$\hat{\rho}_a = \sum_b q(b|a)|b\rangle\langle b|\,. \tag{4.15}$$

If we now choose effects $\hat{F}_b = |b\rangle\langle b|$, it follows that $p_{B|A}(b|a) = q(b|a)$ and, hence, that

$$\hat{\rho} = \sum_a p_A(a)\hat{\rho}_a = \sum_b p_B(b)|b\rangle\langle b|\,. \tag{4.16}$$

Since the density operators (4.16) and (4.15) are written in terms of a diagonal decomposition in an orthonormal basis, one has immediately that

$$S(\hat{\rho}) = -\sum_b p_B(b)\log_2 p_B(b) = H(B) \tag{4.17}$$

and that

$$\sum_a p_A(a)S(\hat{\rho}_a) = \sum_a p_A(a)\left(-\sum_b q(b|a)\log_2 q(b|a)\right) = H(B|A)\,, \tag{4.18}$$

which together give equality in Eq. (4.13). It is worth emphasizing the opposite direction—equality implies commutation—of the if-and-only-if statement: if the input quantum states do not commute, then *no* generalized measurement can achieve the upper bound in Eq. (4.13).

Holevo's theorem is of fundamental importance because it bounds the mutual information even in situations where one tries to transmit information $H(A)$ or to retrieve information $H(B)$ in excess of the quantum entropy $S(\hat{\rho})$. Physical intuition suggests that the quantum entropy $S(\hat{\rho})$ is the logarithm of the number of "distinguishable" quantum-mechanical states that can be accommodated within an unconditioned density operator $\hat{\rho}$. Should one try to transmit and detect information in excess of $S(\hat{\rho})$, this intuition holds that the channel must become noisy enough to prevent it. Holevo's theorem makes this physical intuition precise and establishes it for general inputs and outputs. For this reason it deserves to be called the *fundamental theorem of quantum communication theory.*

The same physical intuition allows one to interpret the second term, $-\sum_a p_A(a)S(\hat{\rho}_a)$, in the upper bound of Holevo's theorem (4.13). The quantum entropy $S(\hat{\rho})$ characterizes the number of distinguishable *pure* states that can be accommodated within density operator $\hat{\rho}$.

When input quantum state $\hat{\rho}_a$ is not pure and thus has nonzero quantum entropy $S(\hat{\rho}_a)$, it lumps together in a single input more than one distinguishable pure state. The resulting reduction in the availability of distinguishable states is taken care of by the second term in Holevo's upper bound.

It is instructive to consider an extended example that illustrates the significance of Holevo's theorem. Suppose that all the input states are *one-dimensional* projectors $\hat{\rho}_a = |a\rangle\langle a|$ (pure states), so that the second term in Holevo's upper bound (4.13) is zero. If, in addition, these input states are *orthogonal* one-dimensional projectors (orthogonal pure states), then

$$\hat{\rho} = \sum_a p_A(a)|a\rangle\langle a| \tag{4.19}$$

is a diagonal decomposition in an orthonormal basis, so one has immediately that

$$S(\hat{\rho}) = -\sum_a p_A(a)\log_2 p_A(a) = H(A)\ . \tag{4.20}$$

In this case, Holevo's upper bound is a trivial consequence of the information-theoretic inequality that the mutual information cannot exceed the input information: $H(B;A) \le H(A) = S(\hat{\rho})$.

What makes Holevo's theorem nontrivial and interesting is that it applies even when the input states are an (over)complete set of one-dimensional *nonorthogonal* projectors (nonorthogonal pure states, of which coherent states are an example that we consider in Sec. V), in which case one can attempt to transmit information $H(A) \gg S(\hat{\rho})$. To see this, consider, for example, $\mathcal{K}$ different complete, orthonormal bases within a $\mathcal{J}$-dimensional Hilbert space. Let the basis vectors be denoted by $|\bar{a},k\rangle$, where the index $k = 1,\ldots,\mathcal{K}$ specifies which basis and the index $\bar{a} = 1,\ldots,\mathcal{J}$ specifies which vector within a basis. Suppose that all these basis vectors are used as input quantum states

$$\hat{\rho}_a = \hat{\rho}_{(\bar{a},k)} = |\bar{a},k\rangle\langle\bar{a},k|\ , \tag{4.21}$$

where the index a, equivalent to the pair $(\bar{a},k)$, runs over all $\mathcal{A} = \mathcal{J}\mathcal{K}$ basis states; suppose further that these states are used with equal probabilities $p_A(a) = 1/\mathcal{A}$. The input information for such equally likely inputs is

$$H(A) = \log_2 \mathcal{A} = \log_2 \mathcal{J} + \log_2 \mathcal{K}\ . \tag{4.22}$$

Using the completeness of each basis, one finds that the unconditioned density operator,

$$\hat{\rho} = \sum_a p_A(a)\hat{\rho}_a = \frac{1}{\mathcal{J}}\left[\frac{1}{\mathcal{K}}\sum_{k=1}^{\mathcal{K}}\left(\sum_{\bar{a}}|\bar{a},k\rangle\langle\bar{a},k|\right)\right] = \frac{1}{\mathcal{J}}\hat{1}\ , \tag{4.23}$$

is a multiple of the unit operator $\hat{1}$ and thus is diagonal in any complete, orthonormal basis, with equal probabilities $1/\mathcal{J}$ on the diagonal. The corresponding quantum entropy

$$S(\hat{\rho}) = \log_2 \mathcal{J}\ , \tag{4.24}$$

although maximal for a $\mathcal{J}$-dimensional Hilbert space, is smaller than the input information by an amount, $\log_2 \mathcal{K}$, that can be made as large as desired. The input information can thus be made arbitrarily large compared to the quantum entropy $S(\hat{\rho})$, in which case the information-theoretic inequality $H(B;A) \le H(A)$ is of no help in relating mutual information to quantum entropy.

Similar considerations hold for the measurement at the output. One can measure an (over)complete set of effects [for example, single-mode coherent states measured by heterodyne detection (Yuen and Shapiro, 1978, 1980; Shapiro *et al.*, 1979; Shapiro and Wagner, 1984; Shapiro, 1985), an example that we consider in Sec. V] in an attempt to retrieve information $H(B) \gg S(\hat{\rho})$. To see this, consider $\mathcal{L}$ different complete, orthonormal bases, this time denoted by $|\bar{b},l\rangle$, where $l = 1,\ldots,\mathcal{L}$, and let the effects be

$$\hat{F}_b = \hat{F}_{(\bar{b},l)} = \frac{1}{\mathcal{L}}|\bar{b},l\rangle\langle\bar{b},l|\ , \tag{4.25}$$

where the index b, equivalent to the pair $(\bar{b},l)$, runs over all $\mathcal{B} = \mathcal{J}\mathcal{L}$ basis states. By virtue of the completeness of each orthonormal basis, these effects satisfy the completeness property (4.6):

$$\sum_b \hat{F}_b = \frac{1}{\mathcal{L}}\sum_{l=1}^{\mathcal{L}}\left(\sum_{\bar{b}}|\bar{b},l\rangle\langle\bar{b},l|\right) = \hat{1}\ . \tag{4.26}$$

Applying these effects to the nonorthogonal input pure states (4.21), one finds an unconditioned output probability that is uniform in b:

$$p_B(b) = \mathrm{tr}(\hat{\rho}\hat{F}_b) = \frac{1}{\mathcal{B}}\langle\bar{b},l|\hat{1}|\bar{b},l\rangle = \frac{1}{\mathcal{B}}\ . \tag{4.27}$$

This leads to an output information

$$H(B) = \log_2 \mathcal{B} = \log_2 \mathcal{J} + \log_2 \mathcal{L} \tag{4.28}$$

that exceeds the quantum entropy (4.24) by an amount, $\log_2 \mathcal{L}$, that can be made arbitrarily large. In this situation the information-theoretic inequality $H(B;A) \le H(B)$ is of no help in relating mutual information to quantum entropy.

Physically, what happens in this extended example is that the conditional detection probabilities

$$p_{B|A}(b|a) = \mathrm{tr}(\hat{\rho}_a\hat{F}_b) = \frac{1}{\mathcal{L}}|\langle\bar{b},l|\bar{a},k\rangle|^2 \tag{4.29}$$

cannot be sharp—i.e., there cannot be a uniquely determined b for each a (see Appendix D)—and, hence, the inverted conditional probabilities

$$p_{A|B}(a|b) = \frac{p_{B|A}(b|a)p_A(a)}{p_B(b)} = \frac{1}{\mathcal{K}}|\langle\bar{b},l|\bar{a},k\rangle|^2 \tag{4.30}$$

also cannot be sharp. The channel is inevitably noisy due to the quantum-mechanical indistinguishability of

nonorthogonal states. This example uses a particular way of decomposing the density operator $\hat{\rho} = \hat{1}/\mathcal{J}$ in terms of nonorthogonal one-dimensional projectors and a particular set of effects that are proportional to nonorthogonal one-dimensional projectors. Hughston, Jozsa, and Wootters (1993) have given the general decomposition of any density operator in terms of nonorthogonal one-dimensional projectors and the general decomposition of the unit operator into effects that are proportional to nonorthogonal one-dimensional projectors.

We can be both more precise and more general about the use of nonorthogonal input states by appealing to the inequality (Levitin, 1969; Balian, 1991; see Appendix A)

$$H(A) \geq S(\hat{\rho}) - \sum_a p_A(a) S(\hat{\rho}_a) , \tag{4.31}$$

in which equality holds if and only if all the input states $\hat{\rho}_a$ are orthogonal, i.e., $\mathrm{tr}(\hat{\rho}_a \hat{\rho}_{a'}) = 0$ for all $a \neq a'$ (which is equivalent to $\hat{\rho}_a \hat{\rho}_{a'} = 0$ for all $a \neq a'$). Putting this inequality together with Holevo's theorem, one can reach the following conclusions: (i) For noncommuting input states, which are necessarily nonorthogonal, the channel is always noisy, with $H(B;A) < S(\hat{\rho}) - \sum_a p_A(a) S(\hat{\rho}_a) < H(A)$ for all choices of detection schemes at the output; (ii) for nonorthogonal, but commuting, input states, which cannot be pure states, the channel is always noisy, but there is a detection scheme at the output (sketched above) for which $H(B;A) = S(\hat{\rho}) - \sum_a p_A(a) S(\hat{\rho}_a) < H(A)$; and (iii) for orthogonal input states, which are necessarily commuting, the channel can be noise-free, there being a detection scheme (sketched above) for which $H(B;A) = S(\hat{\rho}) - \sum_a p_A(a) S(\hat{\rho}_a) = H(A)$. Indeed, one may further assert that a channel can be made noise-free if and only if the input states are orthogonal.

Jozsa, Robb, and Wootters (1993) have recently found a *lower* bound on the maximum mutual information, in the case that the input states $\hat{\rho}_a$ are pure:

$$\max_{\hat{F}_b} H(B;A) \geq Q(\hat{\rho}) = -\sum_j \left(\prod_{k \neq j} \frac{p_j}{p_j - p_k} \right) p_j \log p_j . \tag{4.32}$$

Here the p_j are the eigenvalues of the unconditioned density operator $\hat{\rho}$. What the lower bound means is that no matter how $\hat{\rho}$ is made of pure input states, there is a detection scheme for which the mutual information is at least as big as $Q(\hat{\rho})$. Moreover, the lower bound is shown to be the tightest such bound that involves only properties of the unconditioned density operator.

For a given set of input states and a given detection scheme at the output, the channel capacity is obtained by maximizing the mutual information over the input probabilities $p_A(a)$, subject to any constraints on the channel, as in Eq. (2.16), and then dividing by $\mathcal{T}$ to get an information rate:

$$C \equiv \frac{1}{\mathcal{T}} \max_{\{p_A(a)\}} H(B;A) . \tag{4.33}$$

A further maximization over generalized measurements at the output and over input states yields the *optimum capacity* of the channel:

$$\begin{pmatrix} \text{optimum} \\ \text{capacity} \end{pmatrix} \equiv \max_{\{\hat{\rho}_a\}} \max_{\{\hat{F}_b\}} C = \frac{1}{\mathcal{T}} \max_{\{\hat{\rho}_a\}} \max_{\{p_A(a)\}} \max_{\{\hat{F}_b\}} H(B;A) \leq \frac{1}{\mathcal{T}} \max_{\{\hat{\rho}_a\}} \max_{\{p_A(a)\}} S(\hat{\rho}) = \frac{S_{\max}}{\mathcal{T}} . \tag{4.34}$$

Holevo's theorem (4.13) establishes the inequality here, which leads to a capacity upper bound in terms of the maximum entropy of the channel density operator,

$$S_{\max} = \max_{\{\hat{\rho}_a\}} \max_{\{p_A(a)\}} S(\hat{\rho}) = \max_{\hat{\rho}} S(\hat{\rho}) , \tag{4.35}$$

where the maximum is taken over all density operators, subject to any channel constraints.

Although the upper bound in Holevo's theorem cannot generally be achieved, the upper bound (4.34) on the optimum capacity can. To see this, find a complete, orthonormal set of basis states $|a\rangle$ in which the density operator $\hat{\rho}$ that maximizes $S(\hat{\rho}) = S_{\max}$ is diagonal, i.e.,

$$\hat{\rho} = \sum_a q(a) |a\rangle\langle a| , \tag{4.36}$$

and then choose $\hat{\rho}_a = |a\rangle\langle a| = \hat{F}_a$ and $p_A(a) = q(a)$. This choice uses an ideal conventional quantum measurement to achieve a noise-free channel with $H(A) = H(B) = H(B;A) = S_{\max}$. The equality condition for Holevo's theorem shows that such a choice of input states and input probabilities is the only way to achieve the optimum capacity. These considerations allow us to conclude that the maximum entropy, rather than just setting an upper bound, actually gives the optimum capacity:

$$\begin{pmatrix} \text{optimum} \\ \text{capacity} \end{pmatrix} = \frac{S_{\max}}{\mathcal{T}} . \tag{4.37}$$

Holevo (1973) proved his theorem for a finite-dimensional channel Hilbert space, a finite input alphabet ($\mathcal{A}$ finite), and a finite output alphabet ($\mathcal{B}$ finite). Yuen and Ozawa (1992) have recently appealed to results in the mathematical physics literature to remove these finiteness assumptions from the proof. In this article, however, we work within the restrictions of Holevo's original proof in considering the optimum capacity of a

linear bosonic channel. We do this partly because this paper was written before Yuen and Ozawa's work appeared, but more importantly, because we believe the physical origins of the optimum capacity for bosonic channels are more apparent when the problem is formulated within the finiteness assumptions of Holevo's original proof.

The first step toward a finite Hilbert space comes with our assumption of a large but finite transmission time, which leads to the discrete set (4.1) of frequency-domain modes. The second step is to use the average-power constraint to restrict the Hilbert space to finite dimension; we implement this second step in three different ways in the next three subsections. We emphasize that even though the alphabet sizes are finite, they can be arbitrarily large—in particular, arbitrarily large compared to the dimension of the Hilbert space. We return to the finiteness assumptions in Sec. IV.F and argue that they are satisfied by *physically realizable* channels.

C. Channel with fixed energy

One way to capture the power constraint is to say that over the very long transmission time $\mathcal{T}$, the energy transmitted is identically equal to

$$P\mathcal{T} = E_N = \frac{Nh}{\mathcal{T}} , \tag{4.38}$$

where the integer

$$N = \frac{P\mathcal{T}^2}{h} \tag{4.39}$$

is the number of photons, at the minimum channel frequency $1/\mathcal{T}$, that would be required to make up the entire energy $P\mathcal{T}$. Our assumption of a long transmission time means that $N \gg 1$. We stress that the fixed-energy assumption means that the input states must be energy eigenstates with energy eigenvalue E_N or incoherent mixtures of such eigenstates.

With the assumption of fixed energy there is a maximum photon frequency

$$f_{\max} = \frac{P\mathcal{T}}{h} = \frac{N}{\mathcal{T}} , \tag{4.40}$$

corresponding to a single photon carrying the entire energy $P\mathcal{T}$. The ratio of maximum to minimum frequencies is another way of writing N:

$$\frac{f_{\max}}{1/\mathcal{T}} = N . \tag{4.41}$$

The geometric mean of maximum and minimum frequencies,

$$\sqrt{f_{\max}/\mathcal{T}} = \sqrt{P/h} , \tag{4.42}$$

is the characteristic frequency at which most of the energy is carried. Our optimum capacity is an asymptotic expansion in the small parameter

$$\frac{\sqrt{P/h}}{f_{\max}} = \frac{1}{\sqrt{P/h}\,\mathcal{T}} = \frac{1}{\sqrt{N}} \ll 1 . \tag{4.43}$$

For long times $\mathcal{T}$ the fixed-energy assumption permits large fluctuations in *instantaneous* power, but insists that the *time-averaged* power be precisely equal to P. Indeed, the peak instantaneous power permitted by the fixed-energy assumption is, roughly speaking, $hf_{\max}^2$, corresponding to a single photon at the maximum frequency whose duration is a single period. The ratio of this peak instantaneous power to the time-averaged power P,

$$\frac{hf_{\max}^2}{P} = f_{\max}\mathcal{T} = N \gg 1 , \tag{4.44}$$

demonstrates the potential for large fluctuations in instantaneous power. Despite this potential, the fixed-energy assumption might be thought too restrictive; so we relax it progressively in the next two subsections, to allow fluctuations in time-averaged power. The relaxed assumptions yield the same optimum capacity in the limit $\mathcal{T} \to \infty$.

The fixed-energy assumption restricts the channel Hilbert space to the finite-dimensional fixed-energy subspace spanned by Fock states with energy eigenvalue E_N. We can apply Holevo's theorem directly to that subspace, which has dimension $\mathcal{N}_N$, as we discuss in Sec. IV.A. We denote the optimum capacity for this fixed-energy channel by

$$C_1 = \frac{S_{\max}}{\mathcal{T}} , \tag{4.45}$$

where $S_{\max}$ is the maximum entropy within the fixed-energy subspace. There being no other constraints within the subspace, the density operator that maximizes the entropy is a multiple of the unit operator $\hat{1}_N$ in the fixed-energy subspace,

$$\hat{\rho}_1 = \frac{1}{\mathcal{N}_N}\hat{1}_N , \tag{4.46}$$

and thus the maximum entropy is given exactly by

$$S_{\max} = \log_2 \mathcal{N}_N . \tag{4.47}$$

An exact expression for the maximum capacity follows from Eq. (4.45).

We are interested in large N, where there is an asymptotic expansion for $\mathcal{N}_N$ (Hardy and Ramanujan, 1918; Rademacher, 1937; Abramowitz and Stegun, 1964):

$$\mathcal{N}_N = \frac{1}{4\sqrt{3}N}\exp\left(\pi\sqrt{\frac{2N}{3}}\right) + O\left[\exp\left(\frac{\pi}{2}\sqrt{\frac{2N}{3}}\right)\right] . \tag{4.48}$$

The correction on the right yields exponentially small corrections in C_1 and in the capacity bounds derived below; so we ignore it in what follows. The result is an asymptotic expression for the optimum capacity of a fixed-energy channel,

$$C_1 = \frac{\pi}{\ln 2}\sqrt{\frac{2P}{3h}} - \frac{1}{\mathcal{T}}\log_2\left(\frac{4\sqrt{3}P\mathcal{T}^2}{h}\right) . \tag{4.49}$$

In the limit $\mathcal{T} \to \infty$, C_1 goes to the optimum capacity C_{WB} of Eq. (1.4); for finite $\mathcal{T}$, C_1 is smaller by a logarithmic correction that expresses a reduction, relative to the limit, in the available number of distinguishable input states.

To achieve the optimum capacity (4.49), one can use as input states any complete, orthonormal set of pure states within the fixed-energy subspace and transmit these states with equal probabilities. Though these states are, of course, energy eigenstates with energy E_N, they need not be Fock states of the frequency-domain modes.

D. Channel with energy maximum

It is perhaps more reasonable to implement the power constraint by saying that there is a *maximum* energy budget $E_N = P\mathcal{T} = Nh/\mathcal{T}$ (we introduce no restriction by assuming that N is an integer, since we can take N to be the integer part of $P\mathcal{T}^2/h$). During the transmission time one is willing to transmit any energy up to and including E_N, but no energy higher than E_N. Such a maximum-energy constraint allows fluctuations in time-averaged power and somewhat greater fluctuations in instantaneous power than does a fixed-energy constraint. We emphasize that the maximum-energy constraint is more restrictive than a constraint that the *average* energy be $P\mathcal{T}$, which allows transmission of energies larger than $P\mathcal{T}$. We consider the average-energy constraint in the next subsection.

The maximum-energy constraint restricts the Hilbert space to the finite-dimensional bounded-energy subspace spanned by all the states with energy eigenvalue $E_n \le E_N$, i.e., $n = 0, 1, \ldots, N$. We can apply Holevo's theorem directly to this subspace. We let

$$C_2 = \frac{S_{\max}}{\mathcal{T}} \tag{4.50}$$

denote the optimum capacity for this maximum-energy channel. The entropy within the bounded-energy subspace is maximized by the density operator

$$\hat{\rho}_2 = \left(\sum_{n=0}^{N}\mathcal{N}_n\right)^{-1}\sum_{n=0}^{N}\hat{1}_n , \tag{4.51}$$

which leads to an exact expression for the maximum entropy,

$$S_{\max} = \log_2\left(\sum_{n=0}^{N}\mathcal{N}_n\right) . \tag{4.52}$$

The sum over $\mathcal{N}_n$ is the dimension of the bounded-energy subspace. An exact expression for the maximum capacity follows from Eq. (4.50).

An asymptotic expansion of the sum in Eq. (4.52) is known, but we derive the leading correction here to illustrate how it arises. The number of states $\mathcal{N}_n$ rises so rapidly with n that the sum is dominated by the last few terms, and thus we write it as

$$\begin{aligned}\sum_{n=0}^{N}\mathcal{N}_n &= \sum_{k=0}^{N}\mathcal{N}_{N-k} \\ &= \frac{1}{4\sqrt{3}N}\sum_{k=0}^{N}\exp\left(\pi\sqrt{\frac{2N}{3}}\sqrt{1-k/N}\right)\left[1+O\left(\frac{k}{N}\right)\right] \\ &= \frac{1}{4\sqrt{3}N}\exp\left(\pi\sqrt{\frac{2N}{3}}\right)\sum_{k=0}^{N}\exp\left(-\frac{\pi}{\sqrt{6N}}k\right)\left[1+O\left(\frac{k}{N}\right)+O\left(\frac{k^2}{N^{3/2}}\right)\right] .\end{aligned} \tag{4.53}$$

The final sum in Eq. (4.53) gives

$$\frac{\sqrt{6N}}{\pi}\left[1+O\left(\frac{1}{\sqrt{N}}\right)\right] ; \tag{4.54}$$

the dimension of the bounded-energy subspace becomes

$$\sum_{n=0}^{N}\mathcal{N}_n = \frac{1}{\sqrt{8\pi^2 N}}\exp\left(\pi\sqrt{\frac{2N}{3}}\right)\left[1+O\left(\frac{1}{\sqrt{N}}\right)\right] . \tag{4.55}$$

The resulting asymptotic expression for the optimum capacity of a maximum-energy channel is

$$C_2 = \frac{\pi}{\ln 2}\sqrt{\frac{2P}{3h}} - \frac{1}{2\mathcal{T}}\log_2\left(\frac{8\pi^2 P\mathcal{T}^2}{h}\right) + \frac{1}{\mathcal{T}}O\left(\frac{1}{\sqrt{P/h\,\mathcal{T}}}\right) . \tag{4.56}$$

Again the $\mathcal{T} \to \infty$ limit gives the optimum capacity C_{WB} of Eq. (1.4).

The logarithmic correction to C_2 is roughly half as large as for a fixed-energy channel [Eq. (4.49)], reflecting the availability of more states. The additional contribution to $S_{\max}$, $\frac{1}{2}\log_2(6N/\pi^2)$, comes, crudely speaking, from states that lie within $\delta N \sim \sqrt{6N}/\pi$ below N. No-

tice that

$$\frac{\delta E}{E} = \frac{\delta N}{N} \sim \frac{1}{\sqrt{N}} = \frac{1}{\sqrt{P/h\mathcal{T}}} . \tag{4.57}$$

To achieve the optimum capacity (4.56), one can use as input states any complete, orthonormal set of pure states within the bounded energy subspace, transmitting these states with equal probabilities. This allows linear combinations of states with different energies, but Eq. (4.57) indicates that essentially all the states have mean energy very close to E_N.

E. Channel with average-energy constraint

To implement the power constraint in a way most closely related to the analysis in Sec. III.B, one should impose a constraint on the average energy $\bar{E}$ transmitted over time $\mathcal{T}$, i.e.,

$$\bar{E} = P\mathcal{T} . \tag{4.58}$$

To work within the original proof of Holevo's theorem, we need to make the Hilbert space finite-dimensional; so we assume that there is a maximum energy

$$E_{\max} = hf_{\max} = \frac{N_{\max}h}{\mathcal{T}} , \tag{4.59}$$

which is much larger than the average energy, i.e.,

$$E_{\max} \gg \bar{E} = P\mathcal{T} . \tag{4.60}$$

As noted in the previous subsection, the assumption of a maximum energy makes the Hilbert space finite-dimensional. We can get at the finite dimensionality in a different way: there are a finite number of frequency-domain modes $j = 1, \ldots, N_{\max}$, the jth of which can have a maximum number of photons given by the largest integer less than or equal to $N_{\max}/j$. As we show below, the energy maximum has negligible impact on the optimum capacity as long as it satisfies Eq. (4.60).

We denote the optimum capacity for this average-energy channel by

$$C_3 = \frac{S_{\max}}{\mathcal{T}} , \tag{4.61}$$

where $S_{\max}$ is the maximum entropy within the finite-dimensional subspace, given the average-energy constraint (4.58). This is a standard statistical physics problem: the entropy is maximized by the density operator

$$\hat{\rho}_3 = \frac{1}{Z}\sum_{n=0}^{N_{\max}} e^{-\beta E_n}\hat{1}_n \tag{4.62}$$

for the canonical ensemble, where

$$Z = \mathrm{tr}\left(\sum_{n=0}^{N_{\max}} e^{-\beta E_n}\hat{1}_n\right) = \sum_{n=0}^{N_{\max}} \mathcal{N}_n e^{-\beta E_n} \tag{4.63}$$

is the partition function, and β is a Lagrange multiplier—"inverse temperature"—chosen so that the average-energy constraint (4.58) is satisfied.

The analysis in Sec III.B suggests that $P \sim 1/h\beta^2$ and that most of the energy flows at a characteristic frequency $\sim \sqrt{P/h} \sim 1/h\beta$. Thus we seek an asymptotic expansion in the small parameter

$$\frac{1}{\sqrt{P/h\mathcal{T}}} \sim \frac{\beta h}{\mathcal{T}} = \alpha \ll 1 , \tag{4.64}$$

which is a dimensionless version of β. The maximum entropy is given by

$$S_{\max} = \frac{1}{\ln 2}(\ln Z + \beta\bar{E}) = \frac{1}{\ln 2}\left(\ln Z + \alpha\frac{P\mathcal{T}^2}{h}\right) , \tag{4.65}$$

where

$$\bar{E}\frac{\mathcal{T}}{h} = -\frac{\partial \ln Z}{\partial\beta}\frac{\mathcal{T}}{h} = -\frac{\partial \ln Z}{\partial\alpha} = \frac{P\mathcal{T}^2}{h} . \tag{4.66}$$

One method for evaluating the partition function Z of Eq. (4.63) is to rewrite it as a product of partition functions Z_j for the frequency-domain modes j—i.e., to write $\ln Z = \sum_j \ln Z_j$. One shows that the energy maximum $E_{\max}$ is irrelevant and can be taken to infinity, and one then derives the leading finite-$\mathcal{T}$ correction to $\ln Z$ by considering carefully the contributions from low-frequency modes. Bekenstein and Schiffer (1990) have used this method to evaluate $\ln Z$ when $E_{\max} \to \infty$; they use the Euler-Maclaurin summation formula to write the sum for $\ln Z$ as an integral plus correction terms that are an asymptotic expansion in α. We emphasize that this same method, applied to a narrow band of frequencies (where subtleties due to low-frequency modes do not arise), shows immediately that the capacity C_0 of Eq. (3.20) is, in fact, the optimum capacity of *any* single narrowband linear bosonic channel.

Here we use a different method to evaluate Z, more in the spirit of the preceding two subsections. The method is a standard statistical physics trick: write (with negligible error) the partition function as an integral,

$$Z = \int_0^{N_{\max}} dn\, \mathcal{N}_n e^{-\alpha n} ; \tag{4.67}$$

then, since the integrand $\mathcal{N}_n e^{-\alpha n}$ is highly peaked, evaluate the integral by the method of steepest descent. If we expand

$$\ln(\mathcal{N}_n e^{-\alpha n}) = \pi\sqrt{2n/3} - \ln(4\sqrt{3}n) - \alpha n \tag{4.68}$$

[Eq. (4.48)] to second order about its maximum at

$$n = \tilde{n} = \frac{\pi^2}{6\alpha^2} - \frac{2}{\alpha} + O(1) , \tag{4.69}$$

the integrand in Eq. (4.67) becomes a Gaussian. It is then clear that $N_{\max}$ can be taken to infinity with negligible error in evaluating Z and $S_{\max}$ (we are not making the

Hilbert space infinite-dimensional; we are only evaluating Z with negligible error). The resulting Gaussian integral for the partition function yields

$$Z = \mathcal{N}_{\tilde{n}} e^{-\alpha\tilde{n}} \sqrt{2\pi\sigma_{\tilde{n}}^2}\,[1 + O(\alpha)]\,, \tag{4.70}$$

where

$$\sigma_{\tilde{n}}^2 = -\left(\frac{\partial^2 \ln(\mathcal{N}_n e^{-\alpha n})}{\partial n^2}\bigg|_{n=\tilde{n}}\right)^{-1} = \frac{\pi^2}{3\alpha^3} + O\left(\frac{1}{\alpha^2}\right)\,. \tag{4.71}$$

To the required accuracy, we find

$$\begin{aligned}\ln Z &= \pi\sqrt{2\tilde{n}/3} - \ln(4\sqrt{3}\tilde{n}) - \alpha\tilde{n} + \tfrac{1}{2}\ln(2\pi\sigma_{\tilde{n}}^2) + O(\alpha) \\ &= \frac{\pi^2}{6\alpha} - \frac{1}{2}\ln\left(\frac{2\pi}{\alpha}\right) + O(\alpha)\,,\end{aligned} \tag{4.72}$$

which gives

$$\frac{P\mathcal{T}^2}{h} = -\frac{\partial \ln Z}{\partial\alpha} = \frac{\pi^2}{6\alpha^2} - \frac{1}{2\alpha} + O(1) \tag{4.73}$$

and which leads to maximum entropy

$$S_{\max} = \frac{1}{\ln 2}\left[\frac{\pi^2}{3\alpha} - \frac{1}{2}\ln\left(\frac{2\pi}{\alpha}\right) - \frac{1}{2} + O(\alpha)\right]\,. \tag{4.74}$$

Using the method described above, Bekenstein and Schiffer (1990) have derived Eqs. (4.72) and (4.74) for $\ln Z$ and $S_{\max}$ and, in addition, have evaluated the $O(\alpha)$ corrections, which are $-\frac{1}{24}\alpha$ for $\ln Z$ and 0 for $S_{\max}$. Inverting Eq. (4.73) gives

$$\frac{1}{\alpha} = \frac{1}{\pi}\sqrt{\frac{6P}{h}}\mathcal{T} + \frac{3}{2\pi^2} + O\left(\frac{1}{\sqrt{P/h}\,\mathcal{T}}\right)\,, \tag{4.75}$$

which allows us to solve for the optimum capacity of an average-energy channel in terms of P:

$$C_3 = \frac{\pi}{\ln 2}\sqrt{\frac{2P}{3h}} - \frac{1}{4\mathcal{T}}\log_2\left(\frac{24P\mathcal{T}^2}{h}\right) + \frac{1}{\mathcal{T}}O\left(\frac{1}{\sqrt{P/h}\,\mathcal{T}}\right)\,. \tag{4.76}$$

Again the $\mathcal{T}\to\infty$ limit gives the optimum capacity $C_{\rm WB}$ of Eq. (1.4).

The logarithmic correction to C_3 is roughly half as large as for a maximum-energy channel [Eq. (4.56)], reflecting the availability of even more states. Compared to a fixed-energy channel, the additional contribution to $S_{\max}$ is

$$\frac{1}{2}\log_2(2\pi\sigma_{\tilde{n}}^2) = \frac{1}{2}\log_2\left[\frac{2}{3}\left(\frac{6P\mathcal{T}^2}{h}\right)^{3/2}\right] \tag{4.77}$$

[cf. Eq. (4.72)], which comes, crudely speaking, from states that lie within $\sigma_{\tilde{n}} \sim (P\mathcal{T}^2/h)^{3/4}$ of $\tilde{n} \simeq P\mathcal{T}^2/h$. Notice that

$$\frac{\sigma_{\tilde{n}}}{P\mathcal{T}^2/h} \sim \frac{1}{(\sqrt{P/h}\,\mathcal{T})^{1/2}} \tag{4.78}$$

[cf. Eq. (4.57)]. To achieve the optimum capacity (4.76), one must use as input states a complete, orthonormal set of pure states formed by choosing separate orthonormal bases for each of the fixed-energy subspaces, transmitting the states that have energy E_n with probability $e^{-\beta E_n}/Z$.

F. Discussion of optimum capacity

The previous three subsections derive the optimum capacity of a linear bosonic channel for three different ways of implementing the constraint on mean power:

Sec. IV.C. Fixed-energy channel: $$C_1 = \frac{\pi}{\ln 2}\sqrt{\frac{2P}{3h}} - \frac{1}{\mathcal{T}}\log_2\left(\frac{4\sqrt{3}P\mathcal{T}^2}{h}\right)\,; \tag{4.79}$$

Sec. IV.D. Maximum-energy channel: $$C_2 = \frac{\pi}{\ln 2}\sqrt{\frac{2P}{3h}} - \frac{1}{2\mathcal{T}}\log_2\left(\frac{8\pi^2P\mathcal{T}^2}{h}\right) + \frac{1}{\mathcal{T}}O\left(\frac{1}{\sqrt{P/h}\,\mathcal{T}}\right)\,; \tag{4.80}$$

Sec. IV.E. Average-energy channel: $$C_3 = \frac{\pi}{\ln 2}\sqrt{\frac{2P}{3h}} - \frac{1}{4\mathcal{T}}\log_2\left(\frac{24P\mathcal{T}^2}{h}\right) + \frac{1}{\mathcal{T}}O\left(\frac{1}{\sqrt{P/h}\,\mathcal{T}}\right)\,. \tag{4.81}$$

All three of these optimum capacities limit to the optimum capacity

$$C_{\rm WB} = \frac{\pi}{\ln 2}\sqrt{\frac{2P}{3h}}\ \text{bits/s} \tag{4.82}$$

of Eq. (1.4) as $\mathcal{T}\to\infty$. The finite-$\mathcal{T}$ logarithmic differences between the three bounds can be readily understood on statistical physics grounds. The optimum capacities for the fixed-energy and maximum-energy channels arise from density operators [Eqs. (4.46) and (4.51)] that one recognizes as species of the microcanonical ensemble, and the optimum capacity for the average-energy channel comes from the density operator (4.62) for the canonical ensemble. The finite-$\mathcal{T}$ logarithmic corrections are an expression of the typical $\ln(E/\delta E)$ differences between the entropies of microcanonical and canonical ensembles.

That gaps in the frequency spectrum reduce the capacity can be understood immediately in terms of the approach used in this section. If there are gaps, the dimension $\mathcal{N}_N$ of the subspace with fixed energy E_N is inevitably reduced, because, in number-theoretic terms, some of the integers are unavailable for partitioning N. Reducing the dimensions of the fixed-energy subspaces obviously reduces the optimum capacities obtained in the preceding three subsections.

The key to our work on wideband capacities is, of course, Holevo's theorem. As we mention in Sec. IV.B, Holevo proved his theorem for a channel that has a finite-dimensional Hilbert space and that uses finite input and output alphabets. By assuming a finite transmission time

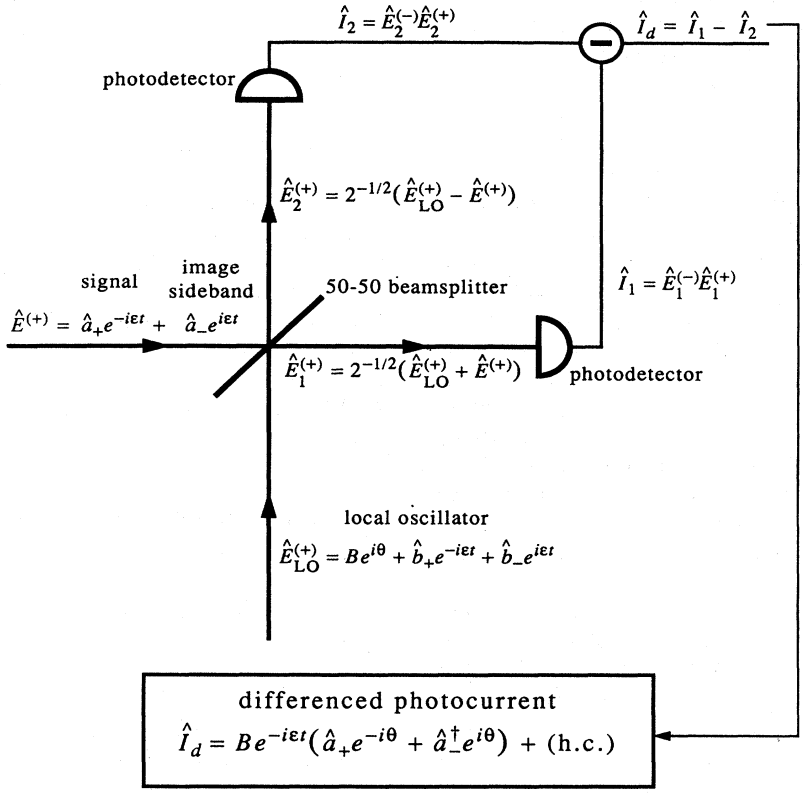


FIG. 4. Balanced heterodyne detection of an optical-frequency signal. The signal to be detected, at frequency $\Omega + \epsilon$, is combined at a 50-50 beam splitter with a much more powerful, coherent beam at frequency Ω. This powerful beam, produced by a laser, is called the local oscillator (LO). The two outputs of the beam splitter are directed onto photodetectors, whose output photocurrents are differenced. The differenced photocurrent $\hat{I}_d$ contains a contribution at the difference frequency ϵ, due to beats between the LO and the signal; it necessarily contains another contribution at the same frequency, due to beats between the LO and the "image sideband," which accompanies the signal into the detector, but which has frequency $\Omega - \epsilon$. The two inputs to the detector are labeled by positive-frequency field operators $\hat{E}^{(+)}$ (signal direction) and $\hat{E}_{\rm LO}^{(+)}$ (LO direction). These field operators have the common optical-frequency oscillation at frequency Ω removed. The two inputs are followed through the detector, and their contributions to the differenced photocurrent are calculated in the limit of a very powerful LO (LO amplitude $B \to \infty$; terms smaller than linear in B are neglected in $\hat{I}_d$). If the differenced photocurrent is filtered to pick out the beat terms at frequency ϵ, the result is detection of the quantity $\hat{a}_+ e^{-i\theta} + \hat{a}_-^\dagger e^{i\theta}$, where $\hat{a}_+$ is the modal annihilation operator for the signal mode, $\hat{a}_-^\dagger$ is the modal creation operator for the image sideband, and θ is the LO phase. Since the image sideband is in vacuum, heterodyne detection realizes a measurement of $\hat{a}_+$—i.e., a simultaneous measurement of both quadrature components of the signal (the real and imaginary parts of $\hat{a}_+$). The vacuum noise introduced by the image sideband represents the noise that must be added when both quadrature components are measured simultaneously. This version of heterodyne detection is called balanced because it uses the symmetry between the two outputs of the beam splitter to eliminate vacuum and excess noise that accompanies the LO at frequencies $\Omega \pm \epsilon$ (annihilation operators $\hat{b}_\pm$). Homodyne detection occurs if the differenced photocurrent is filtered to pick out the dc component ($\epsilon = 0$); then the signal and image sideband are the same, so homodyne detection realizes a measurement of the signal quadrature component $\frac{1}{2}(\hat{a}e^{-i\theta} + \hat{a}^\dagger e^{i\theta})$, which depends on the LO phase.

and by appropriate use of the power constraint, we are able to formulate the question of wideband capacities (in three different ways) in terms of a channel with a finite-dimensional Hilbert space. The input and output alphabets, though finite, can be arbitrarily large compared to the dimension of the Hilbert space.

The question of *infinite* input and output alphabets—either countably infinite or continuously infinite—is more difficult. The idealized coherent-state and quadrature-squeezed channels considered in Sec. V are examples of channels that require an infinite-dimensional Hilbert space and that use continuously infinite input and output alphabets. Nonetheless, we argue that our analysis is already sufficient to cover all *physically realizable* channels.

Our argument is based on the simple fact that real devices for transmission and detection have finite resolution and finite dynamic range. Thus, in practice, the number of possibilities for transmission or detection is always a finite set. We illustrate this argument by taking a closer look at the heterodyne and homodyne detection that are used in coherent-state and quadrature-squeezed channels. At optical frequencies both heterodyne and homodyne detection use powerful lasers as local oscillators—much more powerful than the signal to be detected—with the ultimate detection by photodetection (see Fig. 4). For finite local-oscillator power the dynamic range is limited by the local-oscillator power, and the resolution is set by single-photon events in the photodetectors. In the limit of *infinite* local-oscillator power, both the dynamic range and the resolution become infinite: heterodyne detection then realizes a measurement of both field quadratures (Yuen and Shapiro, 1978, 1980; Shapiro *et al.*, 1979; Shapiro and Wagner, 1984; Shapiro, 1985), and homodyne detection realizes a measurement of one quadrature (Yuen and Shapiro, 1978,1980; Shapiro *et al.*, 1979; Yuen and Chan, 1983; Schumaker, 1984; Shapiro and Wagner, 1984; Shapiro, 1985; Yurke, 1985; Caves and Schumaker, 1986). For finite local-oscillator power, however, heterodyne and homodyne detection measure, over a finite range, discrete approximations to the continuous quantities (Collett *et al.*, 1987; Braunstein, 1990). In both cases the physical output alphabet is finite.

The assumed finite resolution and finite dynamic range of the input and output do not appear in the optimum capacity (4.82). This is a tip-off that nothing goes wrong in the limit of infinite input and output alphabets. Nonetheless, it is of considerable mathematical interest that Yuen and Ozawa have recently used results from the mathematical physics literature to extend Holevo's theorem to apply to infinite-dimensional Hilbert spaces and to infinite input and output alphabets. This allows them to derive the optimum capacity (4.82) without any finiteness assumptions. Their work makes appeals to physical realizability unnecessary.

The capacity C_{WB} of Eq. (4.82) is thus a universal upper bound on the capacity of *any* single linear bosonic channel. It applies to any technique whatsoever for modulating a bosonic field—for example, pulse position modulation or any other kind of correlation between frequencies; moreover, it applies to any input quantum states and to any detection scheme—i.e., any kind of generalized measurement at the output. In particular, the optimum capacity C_{WB} is independent of the use of frequency multiplexing or of number states, although a frequency-multiplexed number-state channel offers one possibility for achieving the upper bound.

There are several claims (Helstrom, 1974; Yuen *et al.*, 1975; Pierce *et al.*, 1981) of infinite wideband capacity in the literature. Here it is sufficient to note that these claims arise from ignoring the Einstein relation $E = hf$, specifically, from assuming that the energy of a quantum is a constant, independent of frequency. There then being no increasing energy cost of a quantum with frequency, it is not surprising that infinite bandwidth yields infinite capacity, despite the power constraint. In Appendix E we examine the claims of infinite capacity carefully, showing how they can be repaired to be consistent with the optimum wideband capacity (4.82).

Because of their importance, both as examples to contrast with number-state channels and as potentially realizable channels, we consider separately, in the next section, wideband frequency-multiplexed coherent-state and quadrature-squeezed channels in the idealized limit of infinite local-oscillator power, and we show explicitly that such channels obey the capacity upper bound C_{WB} of Eq. (4.82).

V. CAPACITIES OF SINGLE COHERENT-STATE AND QUADRATURE-SQUEEZED CHANNELS

In this section we develop the description of coherent-state and quadrature-squeezed channels, and we show explicitly that the capacities of single wideband frequency-multiplexed zero-temperature coherent-state and quadrature-squeezed channels lie below the capacity bound C_{WB} of Eq. (1.4). In addition, we point out that the development of high-T_c superconductors might make accessible the *physical* conditions for approaching the wideband quadrature-squeezed capacity C_{QS} of Eq. (1.6). Before turning to these tasks, however, we review briefly the properties of single-mode coherent and quadrature-squeezed states.

A. Coherent states and quadrature-squeezed states

Consider a single field mode that has frequency ω. The modal annihilation operator $\hat{a}$ and the modal creation operator $\hat{a}^\dagger$ obey the canonical commutation relation

$$[\hat{a}, \hat{a}^\dagger] = 1 \,. \tag{5.1}$$

If we remove from $\hat{a}$ the oscillation at frequency ω, then it is a constant non-Hermitian operator that represents the complex amplitude of the mode's oscillation (in units of number of quanta).

It is convenient to decompose $\hat{a}$ into Hermitian real

and imaginary parts, $\hat{x}_1$ and $\hat{x}_2$:

$$\hat{a} = \hat{x}_1 + i\hat{x}_2 \ . \tag{5.2}$$

The operators $\hat{x}_1$ and $\hat{x}_2$ are called quadrature components. Classically, x_1 and x_2 label the axes of a complex-amplitude diagram (or phase plane) in which the mode's oscillation is represented by a vector extending from the origin (see Fig. 5). The length of the vector is the real amplitude of oscillation (in units of number of quanta), and the angle the vector makes with the real axis is the phase of oscillation.

The canonical commutator (5.1) implies a commutator for the quadrature components,

$$[\hat{x}_1, \hat{x}_2] = \frac{i}{2}[\hat{a}, \hat{a}^\dagger] = \frac{i}{2} \ , \tag{5.3}$$

which, in turn, implies an uncertainty principle

$$\langle(\Delta\hat{x}_1)^2\rangle\langle(\Delta\hat{x}_2)^2\rangle \geq \frac{1}{16} \ . \tag{5.4}$$

The quadrature components are essentially identical to position and momentum operators, except that they are defined in natural units of number of quanta and they have the oscillation at ω removed.

Coherent states are defined in terms of a displacement operator (Glauber, 1963b, 1965)

$$\begin{aligned} D(\hat{a}, \alpha) = \exp(\alpha\hat{a}^\dagger - \alpha^*\hat{a}) &= e^{-|\alpha|^2/2} e^{\alpha\hat{a}^\dagger} e^{-\alpha^*\hat{a}} \\ &= e^{|\alpha|^2/2} e^{-\alpha^*\hat{a}} e^{\alpha\hat{a}^\dagger} \ . \end{aligned} \tag{5.5}$$

The first form is symmetrically ordered in $\hat{a}$ and $\hat{a}^\dagger$; the second and third forms are normally and antinormally ordered, respectively, and are obtained from the first by using the Baker-Campbell-Hausdorff identity. The most important property of the displacement operator is that it displaces the annihilation operator:

$$[D(\hat{a}, \alpha)]^\dagger \hat{a} D(\hat{a}, \alpha) = \hat{a} + \alpha \ . \tag{5.6}$$

The only other property we need is the product of two displacement operators,

$$\begin{aligned} [D(\hat{a}, \beta)]^\dagger D(\hat{a}, \alpha) &= D(\hat{a}, -\beta) D(\hat{a}, \alpha) \\ &= e^{(\alpha\beta^* - \alpha^*\beta)/2} D(\hat{a}, \alpha - \beta) \ , \end{aligned} \tag{5.7}$$

which follows from another application of the Baker-Campbell-Hausdorff identity.

A coherent state,

$$|\alpha\rangle = D(\hat{a}, \alpha)|0\rangle \ , \tag{5.8}$$

is obtained by applying the displacement operator to the vacuum (ground) state $|0\rangle$. The displacement property (5.6) implies that a coherent state is an eigenstate of the annihilation operator:

$$\hat{a}|\alpha\rangle = \alpha|\alpha\rangle \ . \tag{5.9}$$

A classical current source radiates a coherent state of the electromagnetic field (Glauber, 1963b, 1965). The phased dipoles in an ideal laser operated far above threshold approximate very closely a classical current source, except for a slow diffusion in phase; thus, for times short compared to the phase-diffusion time, a laser produces a close approximation to a coherent state.

Using the normally ordered form of the displacement operator (5.5) and the definition

$$|n\rangle = \frac{1}{\sqrt{n!}}(\hat{a}^\dagger)^n|0\rangle \tag{5.10}$$

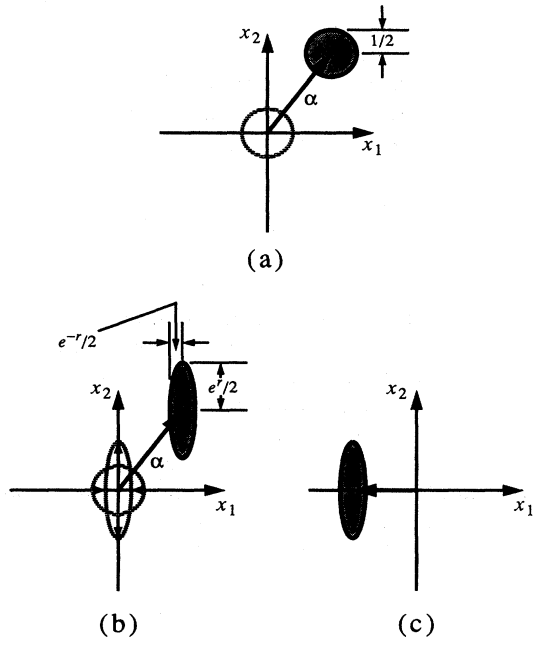


FIG. 5. Complex-amplitude diagrams. Axes are labeled by the quadrature components x_1 and x_2 [Eq. (5.2)]; the diagram is a phase plane with both axes scaled to be in units of number of quanta. (a) The coherent state $|\alpha\rangle$ [Eq. (5.8)] is represented by a quantum "error circle" that lies at the tip of an arrow that represents the mean complex amplitude α. The radius of the error circle is chosen to correspond to the quadrature uncertainties $\langle(\Delta\hat{x}_1)^2\rangle^{1/2} = \langle(\Delta\hat{x}_2)^2\rangle^{1/2} = \frac{1}{2}$. The diagram gives a pictorial representation of the operator equation (5.8) that generates a coherent state from vacuum: the vacuum-state error circle, centered at the origin, is displaced by α in the complex-amplitude plane. (b) The quadrature-squeezed state $|\alpha\rangle_{(r,0)}$ [Eq. (5.18) with $\phi = 0$] is represented by an "error ellipse" that lies at the tip of an arrow that represents the mean complex amplitude α. The principal radii of the error ellipse are chosen to correspond to the quadrature uncertainties $\langle(\Delta\hat{x}_1)^2\rangle^{1/2} = \frac{1}{2}e^{-r}$ and $\langle(\Delta\hat{x}_2)^2\rangle^{1/2} = \frac{1}{2}e^{r}$. The diagram gives a pictorial representation (for $\phi = 0$) of the operator equation (5.18) that generates a quadrature-squeezed state from vacuum: the vacuum-state error circle is first "squeezed" to form an ellipse, which is then displaced by α in the complex-amplitude plane. (c) A typical quadrature-squeezed state $|\alpha_1\rangle_{(r,0)}$ ($\alpha_2 = 0$) used in a quadrature-squeezed channel.

of a number state, one can derive the number-state expansion of a coherent state,

$$|\alpha\rangle = e^{-|\alpha|^2/2} e^{\alpha \hat{a}^\dagger} |0\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle \ , \quad (5.11)$$

which shows that the probability distribution of number of quanta is a Poisson distribution. For our analysis of coherent-state channels, we need the inner product of two coherent states,

$$\langle \beta | \alpha \rangle = e^{(\alpha\beta^* - \alpha^*\beta)/2} e^{-|\alpha-\beta|^2/2} \ , \quad (5.12)$$

which follows from Eqs. (5.8) and (5.7) and the normally ordered form of the displacement operator (5.5). Equation (5.12) shows that the coherent states are not an orthonormal set. Nonetheless, one can demonstrate that they satisfy an (over)completeness property (Glauber, 1963b, 1965)

$$\int \frac{d^2\alpha}{\pi} |\alpha\rangle\langle\alpha| \doteq \hat{1} \ , \quad (5.13)$$

where the integration measure is $d^2\alpha = d\alpha_1 d\alpha_2$ (α_1 and α_2 are the real and imaginary parts of α). The completeness property can be proved by considering the number-state matrix elements of the integral on the left.

The mean complex amplitude of the coherent state $|\alpha\rangle$ is given by

$$\langle \hat{a} \rangle = \alpha = \alpha_1 + i\alpha_2 = \langle \hat{x}_1 \rangle + i \langle \hat{x}_2 \rangle \ ; \quad (5.14)$$

the quadrature components are uncorrelated and have variances

$$\langle (\Delta \hat{x}_1)^2 \rangle = \langle (\Delta \hat{x}_2)^2 \rangle = \frac{1}{4} \ . \quad (5.15)$$

Coherent states are thus the quadrature minimum-uncertainty states that have equal quadrature uncertainties. The coherent-state variances (5.15) are often referred to as the "vacuum" or "zero-point" noise level; one can think of each quadrature component as carrying half of the half-quantum of zero-point noise. A coherent state can be represented in a complex-amplitude diagram by a quantum "error circle," representing the vacuum noise, which lies at the tip of a classical arrow that represents the mean complex amplitude α [see Fig. 5(a)].

To define quadrature-squeezed states, we need to introduce the squeeze operator (Stoler, 1970, 1971; Lu, 1971, 1972)

$$\hat{S}(r,\phi) = \exp\left(\frac{1}{2} r \left(\hat{a}^2 e^{-2i\phi} - \hat{a}^{\dagger 2} e^{2i\phi} \right) \right) \ , \quad (5.16)$$

where r is called the squeeze parameter and ϕ determines the phase of the squeezing. The squeeze operator transforms the annihilation operator according to

$$\hat{S}(r,\phi) \hat{a} [\hat{S}(r,\phi)]^\dagger = \hat{a} \cosh r + \hat{a}^\dagger e^{2i\phi} \sinh r \ . \quad (5.17)$$

A quadrature-squeezed state (Stoler, 1970, 1971; Lu, 1971, 1972; Yuen, 1976b; Hollenhorst, 1979; Caves, 1981; Caves and Schumaker, 1985; Schumaker and Caves, 1985),

$$|\alpha\rangle_{(r,\phi)} = D(\hat{a},\alpha) \hat{S}(r,\phi) |0\rangle \ , \quad (5.18)$$

is obtained by first squeezing the vacuum state and then displacing it. A squeezed state is generated in processes where several field modes, some pumped by an external laser, interact within a nonlinear medium [for reviews of quadrature-squeezed states, see Kimble and Walls (1987), Loudon and Knight (1987), Teich and Saleh (1989, 1990), and Zaheer and Zubairy (1991)]. By a rotation in the complex-amplitude plane ($\hat{a} \to \hat{a} e^{i\phi}$, $\alpha \to \alpha e^{i\phi}$), which for an optical-frequency field amounts to choosing an otherwise arbitrary phase, we can always arrange to set $\phi = 0$, which we do henceforth.

The mean complex amplitude of the quadrature-squeezed state $|\alpha\rangle_{(r,0)}$ is α, just as for a coherent state [Eq. (5.14)]; the quadrature components are uncorrelated and have variances

$$\langle (\Delta \hat{x}_1)^2 \rangle = \frac{1}{4} e^{-2r} \ , \ \ \langle (\Delta \hat{x}_2)^2 \rangle = \frac{1}{4} e^{2r} \ . \quad (5.19)$$

The quadrature-squeezed states with $\phi = 0$ thus constitute the entire class of quadrature minimum-uncertainty states. If $r > 0$, the first quadrature will have a variance reduced below the vacuum level (5.15) by a factor e^{-2r}, whereas the second quadrature will have a variance amplified above the vacuum level by e^{2r}; one usually says that $\hat{x}_1$ is the "squeezed quadrature" and that $\hat{x}_2$ is the "amplified quadrature." A quadrature-squeezed state can be represented in a complex-amplitude diagram in much the same way as a coherent state, except that the error circle is replaced by an "error ellipse" that depicts the reduction in $\hat{x}_1$ uncertainty and the increase in $\hat{x}_2$ uncertainty [see Fig. 5(b)].

To analyze quadrature-squeezed channels, we need the x_1 wave function of a quadrature-squeezed state (Schumaker, 1986):

$$\langle x_1 | \alpha \rangle_{(r,0)} = e^{i\delta} \left(\frac{1}{2\pi(\frac{1}{4}e^{-2r})} \right)^{1/4} e^{2i\alpha_2 x_1} \times \exp\left(-\frac{(x_1 - \alpha_1)^2}{4(\frac{1}{4}e^{-2r})} \right) \ . \quad (5.20)$$

Here δ is a phase factor which can be calculated, but which is irrelevant to our discussion.

B. Narrowband coherent-state and quadrature-squeezed channels

1. Channel models and channel capacities

As a preliminary, we consider narrowband channels, for which we use the language developed in Sec. III.A. We describe briefly both coherent-state and quadrature-

squeezed channels, partly to illustrate the formalism of generalized measurements introduced in Sec. IV.A.

a. Coherent-state channel

We use "coherent-state channel" as a shorthand for the following: at each use the transmitter emits a coherent state, which is read out at the output by heterodyne detection (see Fig. 4 and Appendix D.2)—i.e., by detection of both field quadratures. In other words, one is transmitting information in both quadrature components of each wave-packet mode—or, alternatively, in both the real amplitude and the phase of each mode—and one is attempting to read out the transmitted information by measuring both quadrature components at the output.

The input is described statistically by the probabilities $p_A(\alpha)d^2\alpha$ at each use to transmit a coherent state

$$\hat{\rho}_\alpha = |\alpha\rangle\langle\alpha| , \tag{5.21}$$

where α lies within $d^2\alpha$. The two-dimensional probability density $p_A(\alpha)$ turns out to be the Glauber-Sudarshan P function (Glauber, 1963a; Sudarshan, 1963), as is evident from writing the unconditioned channel density operator (per use),

$$\hat{\rho} = \int d^2\alpha\, p_A(\alpha)\hat{\rho}_\alpha . \tag{5.22}$$

The average-power constraint becomes a constraint on the mean number of photons per use,

$$\bar{n} = \mathrm{tr}(\hat{\rho}\hat{a}^\dagger\hat{a}) = \int d^2\alpha\, |\alpha|^2 p_A(\alpha) . \tag{5.23}$$

As we demonstrate in Appendix D.2, ideal heterodyne detection is described by an effect density (Yuen and Shapiro, 1978, 1980; Shapiro *et al.*, 1979; Shapiro and Wagner, 1984; Shapiro, 1985)

$$\hat{F}_\beta = \frac{1}{\pi}|\beta\rangle\langle\beta| . \tag{5.24}$$

The operators $\hat{F}_\beta$ are multiples of coherent-state projection operators; completeness of the effect density is simply completeness of the coherent states [Eq. (5.13)]. The measure $\hat{F}_\beta\, d^2\beta$ is sometimes called an effect-valued measure. The effect density gives the conditional probability density to read out $\beta = \beta_1 + i\beta_2$ at the output, given that α was transmitted:

$$p_{B|A}(\beta|\alpha) = \mathrm{tr}(\hat{\rho}_\alpha\hat{F}_\beta) = \frac{1}{\pi}|\langle\beta|\alpha\rangle|^2 = \frac{1}{\pi}\exp(-|\beta-\alpha|^2) \tag{5.25}$$

[cf. Eq. (5.12)].

Since the channel noise described by Eq. (5.25) is additive and Gaussian, we can invoke Shannon's theorem to find the capacity (Shannon, 1948; Gallager, 1968). The mutual information per use is maximized by a Gaussian input probability density

$$p_A(\alpha) = \frac{1}{\pi\bar{n}}\exp\left(-\frac{|\alpha|^2}{\bar{n}}\right) , \tag{5.26}$$

which makes the unconditioned channel density operator (5.22) a thermal state, and which, together with the conditional probability density (5.25), leads to an output probability density

$$p_B(\beta) = \int d^2\alpha\, p_{B|A}(\beta|\alpha)p_A(\alpha) = \frac{1}{\pi(1+\bar{n})}\exp\left(-\frac{|\beta|^2}{1+\bar{n}}\right) . \tag{5.27}$$

If one calculates an output information and a conditional output information with respect to the measures $d^2\alpha$ and $d^2\beta$, one finds

$$H(B) = -\int d^2\beta\, p_B(\beta)\log_2 p_B(\beta) = \log_2[e\pi(1+\bar{n})] , \tag{5.28}$$

$$H(B|A) = \int d^2\alpha\, p_A(\alpha)\left(-\int d^2\beta\, p_{B|A}(\beta|\alpha)\log_2 p_{B|A}(\beta|\alpha)\right) = \log_2(e\pi) \tag{5.29}$$

[Eqs. (2.8) and (2.10)]. Although neither $H(B)$ nor $H(B|A)$ is invariant under changes in integration variable—thus both are ambiguous—their difference, the mutual information (2.11), is invariant, as it must be, since it gives the capacity per use of a coherent-state channel (Gordon, 1962; She, 1968; Yamamoto and Haus, 1986):

$$\mathcal{C} = H(B;A) = H(B) - H(B|A) = \log_2(1+\bar{n}) . \tag{5.30}$$

Hall (1993) has generalized this capacity to a coherent-state channel degraded by additive Gaussian noise.

The mathematical description of a coherent-state channel has a compelling physical interpretation. Comparison of Eqs. (5.26) and (5.25) shows that the conditional probability density $p_{B|A}(\beta|\alpha)$ describes one quantum's worth of channel noise, half of which is the vacuum noise intrinsic to a coherent state [the noise depicted by the error circle in Fig. 5(a)] and the other half of which is additional noise that arises in *any* attempt to mea-

sure both field quadratures simultaneously (Arthurs and Kelly, 1965; Caves, 1982; Yamamoto and Haus, 1986; Arthurs and Goodman, 1988, 1991; Braunstein *et al.*, 1991; Ozawa, 1991; Stenholm, 1992).

The output information $H(B)$ is ambiguous because any continuous variable, viewed as a limit of a discrete variable, carries an infinite amount of information. How much of this infinity remains in the finite value of $H(B)$ depends on how the limit is taken, i.e., on which integration variable is introduced in taking the limit. The quantum-mechanical noise, however, sets a scale for distinguishing nearby coherent states, thus making the mutual information (5.30) finite and perfectly well defined. This scale is, in fact, the area of a quantum phase-space cell, which we invoked in our heuristic argument for wavelike channels in Sec. II.B.

Bekenstein and Schiffer (1990) have derived a coherent-state channel capacity $\log_2(e\bar{n})$. Their derivation is flawed because they ignore channel noise. Their capacity approximates the correct capacity for $\bar{n} \gg 1$, because they use a sensible, if *ad hoc*, integration measure; but their capacity goes badly wrong for $\bar{n} \lesssim 1$, becoming negative, as they note, for $\bar{n} \leq e^{-1}$.

b. Quadrature-squeezed channel

By a "quadrature-squeezed channel" we mean the following: at each use the channel is excited into a quadrature-squeezed state $|\alpha_1\rangle_{(r,0)}$, which is read out at the output by homodyne detection of the quadrature component $\hat{x}_1$. In other words, one is transmitting and reading out information on the quadrature component $\hat{x}_1$ and reducing its noise below the vacuum level by making it the squeezed quadrature [see Fig. 5(c)]. We choose the squeeze parameter to be the same for all the transmitted states,[4] and we choose the expectation value of the amplified quadrature $\hat{x}_2$ to be zero ($\alpha_2 = 0$), because unnecessary coherent excitation of the amplified quadrature wastes energy. By squeezing the first quadrature, however, one unavoidably excites the amplified quadrature incoherently [this is the amplified noise depicted in Fig. 5(b)], and this unavoidable excitation limits, through the power constraint, the amount of squeezing.

We denote the input density operators by

$$\hat{\rho}_{\alpha_1} = |\alpha_1\rangle_{(r,0)\,(r,0)}\langle\alpha_1| \tag{5.31}$$

and describe the input statistically by a probability density $p_{A_1}(\alpha_1)$ (defined with respect to the standard measure $d\alpha_1$) at each use to transmit the state $\hat{\rho}_{\alpha_1}$ (see footnote 4). The unconditioned channel density operator (per use) is

$$\hat{\rho} = \int d\alpha_1\, p_{A_1}(\alpha_1)\hat{\rho}_{\alpha_1}\ , \tag{5.32}$$

and the average-power constraint becomes

$$\bar{n} = \mathrm{tr}(\hat{\rho}\hat{a}^\dagger\hat{a}) = \sigma^2 + \sinh^2 r\ , \tag{5.33}$$

where

$$\sigma^2 = \int d\alpha_1\, \alpha_1^2 p_{A_1}(\alpha_1) \tag{5.34}$$

is the second moment of α_1 with respect to the input probability density $p_{A_1}(\alpha_1)$. The $\sinh^2 r$ contribution to $\bar{n}$ represents the unavoidable excitation of the amplified quadrature, which through the power constraint limits the degree of squeezing.

As we show in Fig. 4, ideal homodyne detection measures a quadrature component (Yuen and Shapiro, 1978, 1980; Shapiro *et al.*, 1979; Yuen and Chan, 1983; Schumaker, 1984; Shapiro and Wagner, 1984; Shapiro, 1985; Yurke, 1985; Caves and Schumaker, 1986) and thus is described by an effect density

$$\hat{F}_{x_1} = |x_1\rangle\langle x_1|\ . \tag{5.35}$$

The operators $\hat{F}_{x_1}$ are projection operators onto the δ-function-normalized eigenstates $|x_1\rangle$ of quadrature component $\hat{x}_1$; completeness of the effect density is the standard completeness property for δ-function-normalized eigenstates. The effect density gives the conditional probability density to read out value x_1, given that α_1 was transmitted,

$$p_{X_1|A_1}(x_1|\alpha_1) = \mathrm{tr}(\hat{\rho}_{\alpha_1}\hat{F}_{x_1}) = |\langle x_1|\alpha_1\rangle_{(r,0)}|^2 = \frac{1}{\sqrt{2\pi(\frac{1}{4}e^{-2r})}}\exp\left(-\frac{(x_1-\alpha_1)^2)}{2(\frac{1}{4}e^{-2r})}\right) \tag{5.36}$$

[cf. Eq. (5.20)].

Again the channel noise is additive and Gaussian, so Shannon's theorem (Shannon, 1948; Gallager, 1968), modified for a channel where only one quadrature is detected, tells us that the mutual information is maximized by a Gaussian input probability density,

$$p_{A_1}(\alpha_1) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{\alpha_1^2}{2\sigma^2}\right)\ , \tag{5.37}$$

and that the capacity per use is given by

$$C = \frac{1}{2}\log_2\left(1+\frac{\sigma^2}{\frac{1}{4}e^{-2r}}\right) = \frac{1}{2}\log_2[1+4e^{2r}(\bar{n}-\sinh^2 r)]\ . \tag{5.38}$$

[4] A more general quadrature-squeezed channel would not assume a constant squeeze parameter r, but rather would allow r to be a function of x_1. Such a channel might have a slightly larger capacity, but it would be difficult to analyze because, as Eq. (5.36) shows, it would not have Gaussian noise.

A further maximization with respect to the squeeze parameter r yields the capacity per use of a quadrature-squeezed channel (Yamamoto and Haus, 1986),

$$C = \log_2(1 + 2\bar{n}) , \tag{5.39}$$

which is achieved when (Yuen, 1976a; Yuen and Shapiro, 1978, 1980; Shapiro *et al.*, 1979)

$$e^{2r} = 2\bar{n} + 1 . \tag{5.40}$$

Hall (1993) has generalized the capacity (5.39) to a quadrature-squeezed channel degraded by additive Gaussian noise. Slusher and Yurke (1990) and Ohya and Suyari (1991) have gone beyond consideration of capacities and compared the bit-error rates attainable with coherent-state and quadrature-squeezed channels.

The coherent-state and quadrature-squeezed capacities (5.30) and (5.39) are examples of the narrowband wave-like capacity (2.20). Although a quadrature-squeezed channel does a bit better than a coherent-state channel (literally 1 bit better per use when $\bar{n} \gg 1$), it cannot do much better, because, as noted in Sec. II.B, quantum mechanics limits both these capacities by restricting the number of phase-space cells that are consistent with the power constraint. The 2 in Eq. (5.39) means that quadrature-squeezed states are twice as efficient as coherent states at tiling the phase plane, yet the logarithm converts this 2 to just a bit of increased capacity.

2. Maximum channel capacities

The capacities (in bits/s) of zero-temperature narrowband coherent-state and quadrature-squeezed channels can be summarized by

$$C = B\mathcal{C} = B\log_2(1 + \gamma\bar{n}) , \tag{5.41}$$

where $\gamma = 1$ applies to a coherent-state channel and $\gamma = 2$ applies to a quadrature-squeezed channel. At capacity the information per photon is

$$\frac{\mathcal{C}}{\bar{n}} = \frac{C}{B\bar{n}} = \bar{n}^{-1}\log_2(1 + \gamma\bar{n}) \equiv \gamma G_1(\gamma\bar{n}) . \tag{5.42}$$

If we now allow freedom to vary the frequency and bandwidth with fixed fractional bandwidth $\eta = B/f \ll 1$, as in Sec. III.A.2, we write

$$\frac{C}{\sqrt{\eta P/h}} = \frac{\mathcal{C}}{\bar{n}^{1/2}} = \bar{n}^{-1/2}\log_2(1 + \gamma\bar{n}) \equiv \gamma^{1/2}G_2(\gamma\bar{n}) . \tag{5.43}$$

The functions $G_1(\gamma\bar{n})$ and $G_2(\gamma\bar{n})$ are plotted in Fig. 6. The dimensionless capacity $G_2(\gamma\bar{n})$ has a maximum at $\gamma\bar{n} = 3.9216$, corresponding to frequency

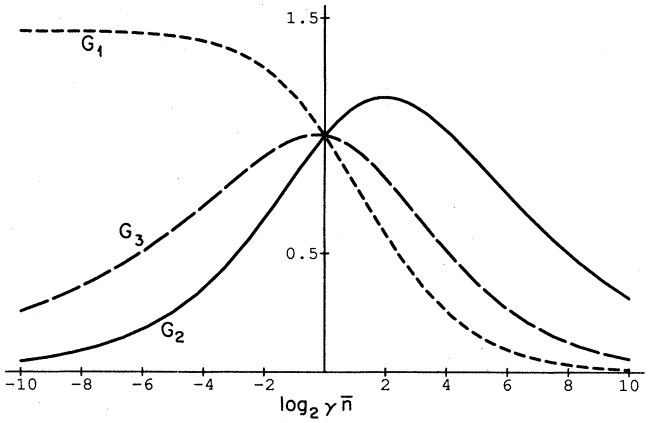


FIG. 6. Semi-log plots of the dimensionless capacity functions $G_1(\gamma\bar{n})$ (short-dashed line), $G_2(\gamma\bar{n})$ (solid line), and $G_3(\gamma\bar{n})$ (long-dashed line) for coherent-state ($\gamma = 1$) and quadrature-squeezed ($\gamma = 2$) channels. The function $\gamma G_1(\gamma\bar{n})$ is the information per photon, defined for a single channel by Eq. (5.42) and for multiple parallel channels by Eq. (6.16). The function $\gamma^{1/2}G_2(\gamma\bar{n})$ is a dimensionless capacity, defined for a single channel by Eq. (5.43) and for multiple parallel channels by Eq. (6.25). The function $\gamma^{3/4}G_3(\gamma\bar{n})$ is a dimensionless information flux, defined for multiple parallel channels by Eq. (6.32). The functions $G_2(\gamma\bar{n})$ and $G_3(\gamma\bar{n})$ both tend to zero as $\log_2(\gamma\bar{n}) \to \pm\infty$; $G_1(\gamma\bar{n})$ goes to zero as $\log_2(\gamma\bar{n}) \to \infty$, but $G_1(\gamma\bar{n}) \to 1/\ln 2$ as $\log_2(\gamma\bar{n}) \to -\infty$.

$$f = \sqrt{\frac{P}{\eta h\bar{n}}} = 0.50498\sqrt{\frac{\gamma P}{\eta h}} , \tag{5.44}$$

to maximum capacity

$$C_{\max} = 2.2991\,B = 1.1610\sqrt{\frac{\gamma\eta P}{h}}\ \text{bits/s}, \tag{5.45}$$

and to $C_{\max}/B\bar{n} = 0.58628\,\gamma$ bits per photon. These results are summarized in Table I.

Setting $\eta = 1$ in Eq. (5.45) gives an estimate of the capacity of wideband coherent-state and quadrature-squeezed channels, but a proper treatment of a wideband channel must include the frequency dependence of photon energy.

C. Wideband frequency-multiplexed coherent-state and quadrature-squeezed channels

Consider then a single zero-temperature frequency-multiplexed channel, in which each frequency bin is a narrowband coherent-state or quadrature-squeezed channel of the sort just considered. Just as in Sec. III.B, the total input power is given by Eq. (3.33) and the capacity by

$$C = \frac{b}{\ln 2}\sum_i \ln(1 + \gamma\bar{n}_i) . \tag{5.46}$$

The maximum capacity is obtained by varying $\bar{n}_i$ as a function of frequency, with P held constant. Carrying out this procedure and replacing sums by integrals, one finds that there is a natural upper cutoff frequency

$$f_c = \sqrt{\frac{2\gamma P}{h}} , \tag{5.47}$$

beyond which communication is too inefficient to be useful. The optimal photon-number distribution is nonthermal,

$$\gamma\bar{n} = \begin{cases} \dfrac{f_c}{f} - 1 , & 0 \le f \le f_c , \\ 0 , & f \ge f_c , \end{cases} \tag{5.48}$$

and the channel capacity for wideband frequency-multiplexed coherent-state and quadrature-squeezed channels is given by

$$C_{\mathrm{CQ}} = \frac{f_c}{\ln 2} = \frac{1}{\ln 2}\sqrt{\frac{2\gamma P}{h}} \text{ bits/s}, \tag{5.49}$$

where the subscript CQ reminds one that this capacity applies both to Coherent-state channels and to Quadrature-squeezed channels (see Table I). For a coherent-state channel ($\gamma = 1$) the results (5.47)–(5.49) were obtained by Gordon (1961) and by Marko (1965). For a quadrature-squeezed channel ($\gamma = 2$) the capacity (5.49) reduces to the capacity C_{QS} given in Eq. (1.6). Just as for a number-state channel, the photon intensity for the distribution (5.48) is infinite, and the information per photon is zero.

In this case we can evaluate explicitly the effect of a low-frequency cutoff at a frequency $f_{\min}$: the upper cutoff frequency is shifted upward to

$$f_c = \sqrt{\frac{2\gamma P}{h}} + f_{\min} , \tag{5.50}$$

and the capacity is reduced to

$$C_{\mathrm{CQ}} = \frac{1}{\ln 2}\left(f_c - f_{\min}\left[1 + \ln\left(\frac{f_c}{f_{\min}}\right)\right]\right)$$

$$= \frac{1}{\ln 2}\left(\sqrt{\frac{2\gamma P}{h}} - f_{\min}\ln\left(\frac{f_c}{f_{\min}}\right)\right) . \tag{5.51}$$

To keep the fractional reduction in capacity less than 10% requires $f_{\min} \le 0.028\, f_c$. The average photon transmission rate takes on the finite value

$$\frac{f_c}{\gamma}\left[\ln\left(\frac{f_c}{f_{\min}}\right) - 1 + \frac{f_{\min}}{f_c}\right]$$

$$= \sqrt{\frac{2P}{\gamma h}}\left[\ln\left(\frac{f_c}{f_{\min}}\right) - 1\right] + \frac{f_{\min}}{\gamma}\ln\left(\frac{f_c}{f_{\min}}\right) \tag{5.52}$$

[cf. Eq. (3.41)], which means that the dominant contribution to the information per photon,

$$\frac{\gamma}{\ln 2}\frac{1}{\ln(f_c/f_{\min})} , \tag{5.53}$$

has the same logarithmic behavior as for a number-state channel [cf. Eq. (3.42)].

Comparing the capacities of Eqs. (1.4) and (5.49), we find that the quadrature-squeezed channel achieves nearly 78% of the capacity of a number-state channel, whereas a coherent-state channel achieves just over 55%. Thus quadrature-squeezed inputs are an improvement over coherent inputs, especially when the practical advantages of higher signal-to-noise ratio—such as lower bit-error rate (Slusher and Yurke, 1990)—are taken into account. The quadrature-squeezed and coherent-state channels are ideally suited to a medium with an upper cutoff frequency, as all inputs have $f \le f_c$, while a number-state channel has no sharp upper cutoff frequency.

The required transmitting medium must be nearly lossless over a wide bandwidth from $f_{\min}$ to f_c. As we wish to neglect thermal effects, we must require $kT \ll hf_{\min}$, so that the thermal occupation number $\bar{n}_T \ll 1$ even at the minimum frequency. A superconducting waveguide [see, for example, Schrieffer (1983)] is an excellent example of this type of wideband medium, when it is operated at a temperature well below its transition temperature T_c. Such operation is necessary so that the cutoff frequency f_c remains below the absorption edge at $f = 2\Delta/h$, where 2Δ is the energy at which electrons are excited over the superconducting band gap and the waveguide is no longer lossless [$2\Delta = 3.5\, kT_c$ in a BCS superconductor; see, for example, Schrieffer (1983)].

The advantage of the recently developed high-T_c superconductors (Bednorz and Müller, 1986; Chu *et al.*, 1987) would be that much higher photon energies can be used than in low-T_c superconductors. The band gap is larger, mainly due to the larger T_c, but also because $2\Delta \simeq 8\, kT_c$ in these materials (Schlesinger, 1987). Suppose one chooses $f_{\min} = 0.028\, f_c$, as in the example mentioned above. With the readily obtainable transition temperature $T_c \simeq 90°$ K, an upper cutoff frequency $f_c \simeq 5$ THz is still a factor of 3 below the absorption edge. For $T \simeq 1°$ K the thermal occupation number at $f_{\min}$ is $\bar{n}_T \simeq 0.001$. The channel capacity with quadrature-squeezed inputs would then be $C \simeq 6$ Tbits/s, with input power $P \simeq 4$ nW. Thin-film experiments have shown very high reflectivities and low infrared absorption in Y-Ba-Cu-O films (Bozovic *et al.*, 1987). We suggest that microwave squeezed states (Movshovich *et al.*, 1990) propagating down cryogenic waveguides, with Josephson junctions (Bozovic *et al.*, 1987) or picosecond pulse sampling techniques (Sobolewski *et al.*, 1986), might meet the physical requirements for quantum-limited wideband communication.

VI. UPPER BOUNDS ON INFORMATION FLUX FOR MULTIPLE PARALLEL LINEAR BOSONIC CHANNELS

We turn now to upper bounds on the information flux—i.e., capacity per area—of channels that consist of many parallel single channels. That this is an important situation, where it is clearly possible to violate the $\sqrt{P/h}$ single-channel capacity bound, has been recognized by several authors (Landauer and Woo, 1973; Levitin, 1982; Pendry, 1983; Bekenstein, 1988; Bekenstein and Schiffer, 1990). To investigate this question, we adopt a specific channel model: a linear bosonic field with an *isotropic* (three-dimensional) dispersion relation,

$$f = f(k) , \tag{6.1}$$

where k is the magnitude of the wave vector $\vec{k}$. This dispersion relation corresponds to the isotropic group velocity

$$c_g(f) = 2\pi \frac{df}{dk} \tag{6.2}$$

and to an isotropic phase velocity

$$c_p(f) = 2\pi f/k = f\lambda , \tag{6.3}$$

where $\lambda = 2\pi/k$ is the wavelength at frequency f. The bosonic field is used to transmit information in a particular direction, which we call the $+z$ direction, through a roughly circular aperture of area $\mathcal{A}$ (regarded as fixed throughout this section) that is transverse to the propagation direction. (By roughly circular, we mean that the linear extent of the aperture is roughly the same in all directions.) The multiple parallel channels are the many transverse modes, both spatially transverse modes (characterized by orthogonal transverse spatial dependences) and polarization modes, that can be fitted into the area $\mathcal{A}$. As in previous sections, we refer to the field quanta as photons, even though the analysis applies equally well to other bosonic fields, such as phonons in a crystal.

We quantize the field in terms of periodic boundary conditions on a volume that has length $\mathcal{L}$ along the propagation direction and has square cross section of area $\mathcal{A}$ transverse to the propagation direction.[5] The periodic boundary conditions lead to a uniform density $\mathcal{L}\mathcal{A}/(2\pi)^3$ of allowed wave vectors. For each allowed wave vector there are g_0 distinct modes, corresponding to different spin states or to different polarizations. A particular mode is specified by giving its wave vector $\vec{k}$ and a polarization index j. To allow for the possibility that the channel may not make use of all the available polarization modes, we let $g \geq 1$ denote the number of polarization modes per wave vector that are actually used to transmit information. Thus the effective density of modes in wave-vector space is

$$\rho_{\vec{k}} = g\frac{\mathcal{L}\mathcal{A}}{(2\pi)^3} . \tag{6.4}$$

We let $\bar{n}_{\vec{k},j}$ be the mean number of photons in the mode specified by wave vector $\vec{k}$ and polarization index j. We can deal simultaneously with several different kinds of channels by writing the capacity of this mode as

$$\mathcal{C}_{\vec{k},j} = \delta\bar{n}_{\vec{k},j}\log_2(1+\bar{n}_{\vec{k},j}^{-1}) + \log_2(1+\gamma\bar{n}_{\vec{k},j}) , \tag{6.5}$$

where $\delta = \gamma = 1$ applies to a number-state channel [Eq. (3.20)], $\delta = 0$ and $\gamma = 1$ apply to a coherent-state channel [Eq. (5.30)], and $\delta = 0$ and $\gamma = 2$ apply to a quadrature-squeezed channel [Eq. (5.39)]. An extension of the arguments given in Sec. IV shows that the information fluxes derived in this section for number-state channels are actually upper bounds on the information flux of *any* linear bosonic channel with an isotropic dispersion relation.

A. Multiple narrowband channels

1. Channel model and channel capacities

We consider first a narrowband channel that operates at frequency $f = c_p(f)/\lambda$ within bandwidth B, with fractional bandwidth $\eta = B/f \ll 1$. We let $\lambda_\perp \geq \lambda$ be the smallest transverse wavelength used for the *spatially* transverse modes. This minimum transverse wavelength corresponds to a maximum transverse wave number $k_\perp = 2\pi/\lambda_\perp$ or to a maximum polar angle $\theta_{\max}$ in wave-vector space, defined by

$$\sin\theta_{\max} = k_\perp/k = \lambda/\lambda_\perp \leq 1 . \tag{6.6}$$

In order to count modes accurately using the mode density (6.4), we must assume that (see footnote 5)

$$\lambda_\perp \ll \sqrt{\mathcal{A}} , \tag{6.7}$$

a condition equivalent to saying that the channel uses many *spatially* transverse modes.

It is important to note that there are other ways of getting many parallel channels, besides using the many spatially transverse modes that are available within a three-dimensional, isotropic medium. For instance, a bosonic field can have many "polarization" modes for each spatially transverse mode. In this case the parallel channels can be the polarization modes corresponding to a *single* spatially transverse mode; a trivial extension of our results for a single channel shows that the optimum capacity of such a channel increases as the square root of the number of "polarization" modes (Landauer and Woo, 1973; Levitin, 1982; Pendry 1983). A more important

[5] Periodic boundary conditions on a square transverse area provide a convenient way of counting orthogonal transverse modes. The counting is equivalent to that obtained using physical modes that vanish outside the actual roughly circular area, except at the longest transverse wavelengths near $\sqrt{\mathcal{A}}$.

practical possibility comes from the ability to confine a *single* (degenerate) spatially transverse mode within an area a smaller than a square wavelength. A good example is the TEM mode of a co-axial cable, where the field itself can be confined to an arbitrarily small area a between two conducting surfaces, which in principle can have negligible cross section. An extension of the single-channel results shows that by putting $n = \mathcal{A}/a$ such channels within an area $\mathcal{A}$, each carrying mean power P, one can achieve a capacity

$$C = n\sqrt{\frac{P/n}{h}} = \sqrt{\frac{nP}{h}} = \sqrt{\frac{P\mathcal{A}}{ha}} \,. \tag{6.8}$$

We return briefly to this "co-ax capacity" at the end of Sec. VI.B.

A mode whose wave vector has z component $k_z > 0$ propagates with speed $2\pi(df/dk_z) = c_g(f)(k_z/k) = c_g(f)\cos\theta$ in the $+z$ direction. Thus the power and information rate of the channel can be written as the sums

$$P = {\sum_{\vec{k},j}}' \frac{c_g(f)(k_z/k)}{\mathcal{L}} hf\bar{n}_{\vec{k},j} \,, \tag{6.9}$$

$$C = {\sum_{\vec{k},j}}' \frac{c_g(f)(k_z/k)}{\mathcal{L}} \mathcal{C}_{\vec{k},j} \,, \tag{6.10}$$

where the prime denotes a sum over allowed wave vectors $\vec{k}$ which have positive z component $k_z > 0$, whose magnitude lies within a range $\Delta k = (\Delta k/\Delta f)B = [2\pi/c_g(f)]B$ about $k = 2\pi f/c_p(f)$, and whose transverse magnitude satisfies $(k_x^2 + k_y^2)^{1/2} \le k_\perp$. The first task is to maximize C within the constraint of fixed power P by varying the mean photon numbers of the various modes. This maximization is trivial, the result being that all the modes should have the same mean number of photons, $\bar{n}$, which becomes the number of photons per mode. This allows us to write a capacity per mode,

$$\mathcal{C} = \delta\bar{n}\log_2(1+\bar{n}^{-1}) + \log_2(1+\gamma\bar{n}) \,. \tag{6.11}$$

Converting the power sum (6.9) to an integral over solid angle in wave-vector space yields

$$\begin{aligned} P &= \int d\Omega\, \rho_{\vec{k}} k^2 \Delta k \frac{c_g(f)\cos\theta}{\mathcal{L}} hf\bar{n} \\ &= \int d\Omega\, B\mathcal{A}\frac{g\cos\theta}{\lambda^2} hf\bar{n} \,, \end{aligned} \tag{6.12}$$

where the integral is restricted to polar angle $\theta \le \theta_{\max}$ [cf. Eq. (6.6)]. Equation (6.12) shows that $(g\cos\theta/\lambda^2)\bar{n}$ is the specific photon intensity—i.e., the mean number of photons per second per Hz per square meter per steradian. Equivalently, we can identify $g\cos\theta/\lambda^2$ as the number of modes per second per Hz per square meter per steradian. Doing the integral over solid angle leads to an energy flux

$$\frac{P}{\mathcal{A}} = B\frac{\pi g}{\lambda_\perp^2} hf\bar{n} \,, \tag{6.13}$$

where we can identify $\pi g/\lambda_\perp^2$ as the number of modes per second per Hz per square meter.

We can now write expressions for the information flux,

$$\frac{C}{\mathcal{A}} = B\frac{\pi g}{\lambda_\perp^2}\mathcal{C} \,, \tag{6.14}$$

and for the photon flux,

$$\frac{N}{\mathcal{A}} = B\frac{\pi g}{\lambda_\perp^2}\bar{n} \,. \tag{6.15}$$

These lead to familiar expressions for the information per photon:

$$\frac{C/\mathcal{A}}{N/\mathcal{A}} = \frac{\mathcal{C}}{\bar{n}} = \delta\log_2(1+\bar{n}^{-1}) + \bar{n}^{-1}\log_2(1+\gamma\bar{n}) \equiv \begin{cases} F_1(\bar{n}) \,, & \delta = \gamma = 1, \\ \gamma G_1(\gamma\bar{n}) \,, & \delta = 0 \end{cases} \tag{6.16}$$

[cf. Eqs. (3.22) and (5.42)].

It is useful to introduce a (dimensionless) multiplicity factor

$$\mu = \mathcal{A}\frac{\pi g}{\lambda_\perp^2} \,, \tag{6.17}$$

which, as the number of modes per second per Hz, quantifies the number of parallel channels. Since μ/g is the number of *spatially* transverse modes (per second per Hz), the requirement (6.7) for many spatially transverse modes can be written as

$$\frac{\mu}{g} = \frac{\pi\mathcal{A}}{\lambda_\perp^2} \gg 1 \,. \tag{6.18}$$

We can investigate approximately the case of a single spatially transverse mode by letting $\lambda_\perp^2 \sim \mathcal{A}$, or $\mu \sim g$, in which case μ reduces to the number of polarization states employed by the channel.

In terms of μ, the power (6.13) and the capacity (6.14) become

$$P = B\mu hf\bar{n} = \eta\mu hf^2\bar{n} \,, \tag{6.19}$$

$$C = B\mu\mathcal{C} = \eta\mu f\mathcal{C} \,. \tag{6.20}$$

These forms highlight the role of the dimensionless bandwidth-channel product $\eta\mu$. Levitin (1982) has given Eqs. (6.19) and (6.20) for a number-state channel in the limit $\bar{n} \ll 1$. It is also convenient to introduce a maximum multiplicity factor

$$\mu_0 = \mathcal{A}\frac{\pi g_0}{\lambda^2} = \mathcal{A}\frac{\pi g_0 f^2}{[c_p(f)]^2}\,, \tag{6.21}$$

which counts the number of available parallel channels, and a fractional multiplicity,

$$\xi = \frac{\mu}{\mu_0} = \frac{g}{g_0}\left(\frac{\lambda}{\lambda_\perp}\right)^2 = \frac{g}{g_0}\sin^2\theta_{\max} \le 1\,, \tag{6.22}$$

which gives the fraction of the available channels that are actually used.

2. Maximum channel capacities

We now investigate, in three different cases, how the capacity of Eq. (6.14) behaves when one or more properties of the channel are allowed to vary.

a. Case 1: varying multiplicity

As the first case, suppose that the number of parallel channels—i.e., the multiplicity factor μ of Eq. (6.17)—is allowed to vary, while the frequency and bandwidth are held constant. The power constraint implies that $\mu\bar{n}$ is fixed or, equivalently, that the photon flux is constant. Thus, in this case, maximizing the capacity means maximizing the information per photon (6.16). For all three kinds of channels we are considering, the information per photon is a decreasing function of $\bar{n}$ [see plots of $F_1(\bar{n})$ and $G_1(\gamma\bar{n})$ in Figs. 3 and 6]; hence the capacity is maximized by letting $\bar{n}$ be as small as possible—i.e., by letting the multiplicity factor go to its maximum value μ_0 of Eq. (6.21). This means choosing the number of photons per mode to be

$$\bar{n} = \frac{P/hf}{B\mu_0} \simeq 2.7\times10^{-3}\left(\frac{10^{-5}}{\eta}\right)\left(\frac{2}{g_0}\right)\left(\frac{P}{1\,\mathrm{mW}}\right)\left(\frac{10^{-4}\,\mathrm{m}^2}{\mathcal{A}}\right)\left(\frac{c_p}{3\times10^8\,\mathrm{m/s}}\right)^2\left(\frac{3\times10^{14}\,\mathrm{Hz}}{f}\right)^4. \tag{6.23}$$

The lesson of case 1 is that if you are stuck with a particular frequency, then it *always* pays to divide up the available photons among as many channels as possible. The payoff changes radically, however, at $\bar{n}\sim 1$, as is apparent from the plots of $F_1(\bar{n})$ and $G_1(\bar{n})$ in Figs. 3 and 6. If one starts with $\bar{n}\gg 1$ photons per mode and goes to one photon per mode by increasing the number of channels by a factor of $\bar{n}$, the capacity increases by a factor $\sim \bar{n}/\log_2\bar{n}$ for all three kinds of channels. In contrast, if one starts with one photon per mode and goes to $\bar{n}\ll 1$ photons per mode by increasing the number of channels by a factor of $\bar{n}^{-1}$, the capacity increases only by a logarithmic factor $\simeq \frac{1}{2}\log_2\bar{n}^{-1}$ for a number-state channel and at most by a factor $\gamma/\ln(1+\gamma)$ for coherent-state and quadrature-squeezed channels.

b. Case 2: varying frequency

The remaining two cases involve varying the channel frequency; they can be handled in general only by assuming that the channel is dispersionless over some range of frequencies. Thus, throughout the remainder of this subsection, we assume that the propagation speed $c = c_p = c_g$ is independent of frequency. We return to the question of dispersion in Sec. VI.B.

In our second case we keep the number of parallel channels, μ, constant, while allowing the frequency f and bandwidth B to vary, with the fractional bandwidth $\eta = B/f$ held constant. Equations (6.19) and (6.20) show, not surprisingly, that this case is precisely equivalent to the single-channel maximization problems considered in Sec. III.A.2 for a single number-state channel and in Sec. V.B.2 for single coherent-state and quadrature-squeezed channels, provided that one has made the replacement $\eta\to\eta\mu$.

In particular, noting that

$$\sqrt{\frac{\eta\mu P}{h}} = \eta\mu f\bar{n}^{1/2} = B\mu\bar{n}^{1/2}\,, \tag{6.24}$$

one can write a dimensionless capacity

$$\frac{C}{\sqrt{\eta\mu P/h}} = \frac{C}{\bar{n}^{1/2}} = \delta\bar{n}^{1/2}\log_2(1+\bar{n}^{-1}) + \bar{n}^{-1/2}\log_2(1+\gamma\bar{n}) \equiv \begin{cases} F_2(\bar{n})\,, & \delta=\gamma=1,\\ \gamma^{1/2}G_2(\gamma\bar{n})\,, & \delta=0\end{cases} \tag{6.25}$$

[cf. Eqs. (3.25) and (5.43)]. The dimensionless capacity functions $F_2(\bar{n})$ and $G_2(\gamma\bar{n})$ are plotted in Figs. 3 and 6. Notice that the maximum capacities for this case [Eqs. (3.27) and (5.45) with $\eta\to\eta\mu$] are proportional to the square root of μ, the number of parallel channels, just as for a channel that uses a single transverse mode and many polarization modes. There is one important difference between the single-channel and multiple-channel

results: whereas the fractional bandwidth η must satisfy $\eta \lesssim 1$, the bandwidth-channel product $\eta\mu$ can be much larger than 1.

c. Case 3: varying frequency and multiplicity

In case 1 we allow the multiplicity to vary, while holding the frequency constant, and find that it is optimal to take advantage of all the available channels. In case 2 we allow the frequency to vary, while holding the multiplicity fixed, and find results that are a simple extension of single-channel results. Our third case is intermediate between the first two: it allows both the frequency f and the number of channels μ to vary. A reasonable way to proceed, which makes the problem explicit, is to assume that both the fractional bandwidth $\eta = B/f$ and the fractional multiplicity $\xi = \mu/\mu_0$ of Eq. (6.22) are held constant. With these assumptions it is convenient to write the energy flux (6.13) and the information flux (6.14) in the forms

$$\frac{P}{\mathcal{A}} = Qhf^4\bar{n}\,, \tag{6.26}$$

$$\frac{C}{\mathcal{A}} = Qf^3\mathcal{C}\,, \tag{6.27}$$

where

$$Q = \eta\xi\frac{\pi g_0}{c^2} = \frac{\eta\pi g}{c^2}\left(\frac{\lambda}{\lambda_\perp}\right)^2 \tag{6.28}$$

is a characteristic quantity with dimensions of $\mathrm{s}^2/\mathrm{m}^2$, which is held fixed as the capacity is maximized. Notice that the bandwidth-channel product can be written as

$$\eta\mu = Q\mathcal{A}f^2 = \left(\frac{Q\mathcal{A}P}{h\bar{n}}\right)^{1/2}. \tag{6.29}$$

It is useful to introduce the maximum value of the characteristic quantity,

$$Q_0 = \frac{\pi g_0}{c^2}\,, \tag{6.30}$$

which applies to a wideband ($\eta = 1$) channel that uses all the available transverse modes ($\xi = 1$).

Noting that

$$Q^{1/4}(P/\mathcal{A}h)^{3/4} = Qf^3\bar{n}^{3/4} \tag{6.31}$$

is fixed by the constraints, we introduce a dimensionless information flux

$$\frac{C/\mathcal{A}}{Q^{1/4}(P/\mathcal{A}h)^{3/4}} = \frac{\mathcal{C}}{\bar{n}^{3/4}} = \delta\bar{n}^{1/4}\log_2(1+\bar{n}^{-1}) + \bar{n}^{-3/4}\log_2(1+\gamma\bar{n}) \equiv \begin{cases} F_3(\bar{n})\,, & \delta = \gamma = 1\,, \\ \gamma^{3/4}G_3(\gamma\bar{n})\,, & \delta = 0\,. \end{cases} \tag{6.32}$$

The dimensionless capacity functions $F_3(\bar{n})$ and $G_3(\gamma\bar{n})$ are plotted in Figs. 3 and 6.

Consider first a number-state channel ($\delta = \gamma = 1$). The dimensionless information flux $F_3(\bar{n})$ of Eq. (6.32) is maximized at $\bar{n} = 0.057166$, corresponding to frequency

$$f = \left(\frac{P/\mathcal{A}}{Qh\bar{n}}\right)^{1/4} = 2.0451\left(\frac{P/\mathcal{A}}{Qh}\right)^{1/4}, \tag{6.33}$$

to maximum information flux

$$\frac{C_{\max}}{\mathcal{A}} = 0.32081\,Qf^3 = 2.7440\,Q^{1/4}\left(\frac{P/\mathcal{A}}{h}\right)^{3/4}\ \mathrm{bits/s\,m^2}, \tag{6.34}$$

and to $C_{\max}/N = \mathcal{C}/\bar{n} = 5.6119$ bits per photon. These results are summarized in Table II. If we regard frequency, rather than energy flux, as the independent variable, we can write the information flux and energy flux in terms of ratios to "typical" values for optical communication:

$$\frac{C_{\max}}{\mathcal{A}} = 0.32\,\frac{\eta\xi\pi g_0}{c^2}f^3 \simeq 6.0\ \mathrm{Pbits/s\,m^2}\left(\frac{\eta}{10^{-5}}\right)\left(\frac{\xi}{10^{-6}}\right)\left(\frac{g_0}{2}\right)\left(\frac{3\times10^8\,\mathrm{m/s}}{c}\right)^2\left(\frac{f}{3\times10^{14}\,\mathrm{Hz}}\right)^3, \tag{6.35a}$$

$$\frac{P}{\mathcal{A}} = \frac{\eta\xi\pi g_0}{c^2}hf^4\bar{n} \simeq 0.21\,\mathrm{mW/m^2}\left(\frac{\eta}{10^{-5}}\right)\left(\frac{\xi}{10^{-6}}\right)\left(\frac{g_0}{2}\right)\left(\frac{3\times10^8\,\mathrm{m/s}}{c}\right)^2\left(\frac{f}{3\times10^{14}\,\mathrm{Hz}}\right)^4. \tag{6.35b}$$

TABLE II. Narrowband and wideband multiple-parallel-channel maximum capacities for three types of channels: number-state (NS) channels, quadrature-squeezed (QS) channels, and coherent-state (CS) channels. For each type of narrowband channel, the table lists the following quantities at maximum capacity: (i) the average number of photons per mode, $\bar{n}$; (ii) the dimensionless frequency, $f/(P/Q\mathcal{A}h)^{1/4} = \bar{n}^{-1/4}$; (iii) the information per mode, $C_{\max}/Q\mathcal{A}f^3 = \mathcal{C}$; (iv) the dimensionless capacity, $C_{\max}/(Q\mathcal{A})^{1/4}(P/h)^{3/4} = \mathcal{C}/\bar{n}^{3/4}$; and (v) the information per photon, $C_{\max}/N = \mathcal{C}/\bar{n}$. The first, third, and fifth of these quantities illustrate our slogan "one photon–one bit–one mode." For each type of wideband channel, the table lists the dimensionless capacity, $C/(Q_0\mathcal{A})^{1/4}(P/h)^{3/4}$, and the information per photon, C/N. The information per photon provides further illustration of our slogan. In the table $\mathcal{P} = P/\mathcal{A}h$.

Type	$\bar{n}$	Narrowband $\frac{f}{(\mathcal{P}/Q)^{1/4}} = \bar{n}^{-1/4}$	Narrowband $\frac{C_{\max}/\mathcal{A}}{Qf^3} = \mathcal{C}$	Narrowband $\frac{C_{\max}/\mathcal{A}}{Q^{1/4}\mathcal{P}^{3/4}} = \frac{\mathcal{C}}{\bar{n}^{3/4}}$	Narrowband $\frac{C_{\max}/\mathcal{A}}{N/\mathcal{A}} = \frac{\mathcal{C}}{\bar{n}}$	Wideband $\frac{C/\mathcal{A}}{Q_0^{1/4}\mathcal{P}^{3/4}}$	Wideband $\frac{C/\mathcal{A}}{N/\mathcal{A}}$
NS	0.057 166	2.045 1	0.320 81	2.744 0	5.611 9	3.070 7	5.196 0
QS	0.416 41	1.244 9	0.874 07	1.686 2	2.099 0	1.738 2	1.923 6
CS	0.832 83	1.046 8	0.874 07	1.002 6	1.049 5	1.033 5	0.961 80

It is instructive to write the maximum capacity of Eq. (6.34) approximately as

$$C_{\max} \simeq \left(\frac{Q\mathcal{A}P}{h}\right)^{1/4}\left(\frac{P}{h}\right)^{1/2} \simeq \sqrt{\frac{\eta\mu P}{h}}\,. \qquad (6.36)$$

The $P^{3/4}$ dependence of $C_{\max}$ for a channel with many spatially transverse modes arises in the following way: a $P^{1/2}$ contribution from the optimum capacity of a single transverse mode and a $P^{1/4}$ contribution, contained in the bandwidth-channel product (6.29), from the number of spatially transverse modes. In contrast, in the limit of only a single spatially transverse mode, where $\mu \sim g$ [Eq. (6.18)], $C_{\max}$ has the $P^{1/2}$ dependence of a single spatially transverse mode with g polarization channels.

Case 3 is a compromise between the first two cases—a compromise that is readily apparent in the plots in Figs. 3 and 6. Comparing case 3 to case 1, where the frequency and bandwidth are fixed, one sees that it is clearly favorable for a number-state channel to operate at a lower frequency with more photons per mode, as case 3 allows it to do. Comparing case 3 to case 2, where the number of channels is fixed, one sees that it is clearly favorable for a number-state channel to operate at a higher frequency with fewer photons per mode, provided that one can take advantage of the larger number of channels available at the higher frequency, as case 3 allows it to do.

In case 2 it is optimal to operate with $\bar{n}^{-1} = 1$ mode per photon, whereas in case 3 it is optimal to operate with $\bar{n}^{-1} \simeq 17$ modes per photon. Since case 2 already uses the longitudinal modes most efficiently, we can sensibly regard this ratio of 17 to 1 as measuring the optimum number of transverse modes per photon—i.e., the optimal degree of parallelism for a number-state channel. Though this ratio makes clear the substantial advantages of parallelism, the really important point is that the ratio is not *much* bigger than 1: it is not optimal to use, say, a thousand transverse modes per photon. These comparisons provide further illustrations of our slogan, "one photon–one bit–one mode."

This same prejudice in favor of parallelism is evident in the dependence of the maximum information flux (6.34) on the propagation speed c: as c gets smaller, the maximum information flux gets larger, even though individual bits of information transit the channel more slowly. This result can be readily understood. Our work on single channels and on case 2 shows that as far as the longitudinal properties of a channel are concerned, the propagation speed has no impact on the information rate. As was discussed in Sec. III, longitudinal wave-packet modes arrive at a rate B^{-1}, independent of propagation speed c; changing c scales lengths along the channel without affecting the arrival rate of modes. When one considers transverse properties of the channel, however, a smaller propagation speed leads to smaller wavelength, thus to more spatially transverse modes within a given area at a particular frequency, and hence to more capacity for the same power.

The advantages of parallelism, most apparent in case 1, where the frequency is held fixed, show up in case 3 as a prejudice in favor of higher frequencies and fewer photons per mode, relative to case 2. Still, case 3 shows that it is not an optimal use of *energy*, even for many parallel channels, to send an arbitrarily large amount of information on a single photon. Suppose, as we did in Sec. III.A.2, that one transmits on average a single photon at frequency f_1 within a bandwidth B_1 during a long transmission time $\mathcal{T} \gg B_1^{-1}$ ($P = hf_1/\mathcal{T}$), and suppose further that one uses a multiplicity factor $\mu_1 = \xi\mathcal{A}(\pi g_0 f_1^2/c^2) \gg 1$ [Eqs. (6.21) and (6.22)]. The information is encoded in the photon's arrival time *and* in which of the transverse modes it occupies—a more general version of the placement information considered in Sec. III.A.2. The number of modes available to the single photon is $\mu_1 B_1 \mathcal{T}$, thus giving a number of photons per mode $\bar{n}_1 = 1/\mu_1 B_1 \mathcal{T} \ll 1$, which yields, from Eqs. (6.11) and (6.20), a capacity

$$C_1 \simeq B_1\mu_1\bar{n}_1 \log_2 \bar{n}_1^{-1} = \frac{1}{\mathcal{T}}\log_2(\mu_1 B_1 \mathcal{T}) = \frac{1}{\mathcal{T}}\log_2\begin{pmatrix}\text{number}\\ \text{of modes}\end{pmatrix}\,. \qquad (6.37)$$

Suppose that instead one transmits the same energy hf_1 in the same time $\mathcal{T}$ at a lower frequency f_2 within a bandwidth B_2 and with a multiplicity factor $\mu_2 = \xi\mathcal{A}(\pi g_0 f_2^2/c^2)$, and suppose that f_2, B_2, and μ_2 are chosen, first, to keep the fractional bandwidth and the fractional multiplicity the same and, second, to give the optimum $\bar{n}_2 \simeq 0.057$ photons per mode. These choices give a frequency $f_2 = f_1(\bar{n}_1/\bar{n}_2)^{1/4}$ [Eq. (6.31)] and, from Eqs. (6.20) and (6.34), a capacity

$$C_2 \simeq 2.7\,(Q\mathcal{A})^{1/4}\left(\frac{P}{h}\right)^{3/4} = 2.7\frac{\bar{n}_1^{-1/4}}{\mathcal{T}} = 2.7\frac{(\mu_1 B_1 \mathcal{T})^{1/4}}{\mathcal{T}}\,. \tag{6.38}$$

The capacity at the lower frequency is larger even though the bandwidth is smaller *and* the number of transverse modes is smaller. The capacity gain in going to the lower frequency f_2 is not as great as in the comparable single-channel example in Sec. III.A.2—or as in case 2—because here going to a lower frequency entails a reduction in the number of transverse modes.

We can also consider case 3 for coherent-state and quadrature-squeezed channels ($\delta = 0$). The dimensionless information flux (6.32) is maximized at $\gamma\bar{n} = 0.83283$, corresponding to frequency

$$f = \left(\frac{P/\mathcal{A}}{Qh\bar{n}}\right)^{1/4} = 1.0468\left(\frac{\gamma P/\mathcal{A}}{Qh}\right)^{1/4}\,, \tag{6.39}$$

to maximum information flux

$$\frac{C_{\max}}{\mathcal{A}} = 0.87407\,Qf^3 = 1.0026\,Q^{1/4}\left(\frac{\gamma P/\mathcal{A}}{h}\right)^{3/4}\ \text{bits/s m}^2, \tag{6.40}$$

and to $C_{\max}/N = \mathcal{C}/\bar{n} = 1.0495\,\gamma$ bits per photon. These results are summarized in Table II. Parallelism is not as advantageous here as for a number-state channel. Just as for a number-state channel, we can quantify the optimal degree of parallelism by taking the ratio of the number of modes per photon in case 3, $\bar{n}^{-1} \simeq \gamma/0.83$, to the number of modes per photon in case 2, $\bar{n}^{-1} \simeq \gamma/3.9$; this ratio gives about five transverse modes per photon.

We can estimate an upper bound on wideband information flux by taking the maximum information flux (6.34) for a number-state channel and setting $\eta = 1$, to approximate a wideband channel, and $\xi = 1$, to describe a channel that uses all the available transverse modes. Under these circumstances the quantity Q of Eq. (6.28) takes on its maximum value $Q_0 = \pi g_0/c^2$ [Eq. (6.30)], and the resulting upper bound on wideband information flux is

$$\frac{C}{\mathcal{A}} \sim Q_0^{1/4}\left(\frac{P/\mathcal{A}}{h}\right)^{3/4}\,. \tag{6.41}$$

This estimate, which agrees with that obtained from our slogan, "one photon–one bit–one mode," in Eq. (1.8), is right on the mark; but a proper treatment requires us to confront two issues: first and obvious, frequency-dependent photon energy and, second, dispersion.

B. Multiple wideband channels

Consider then a frequency-multiplexed channel, in which each frequency bin is a narrowband channel consisting of many parallel channels, just like the narrowband channel in the preceding subsection. We now allow dispersion, so that the phase velocity varies from one frequency bin to another. Let b be the bandwidth of each frequency bin, and let f_i and $\lambda_i = c_i/f_i$ be the frequency and wavelength of the ith bin, where $c_i = c_p(f_i)$ is the phase velocity for the ith bin. We assume that each bin uses all available parallel channels, so the multiplicity factor of the ith bin is

$$\mu_i = \mathcal{A}(\pi g_0/\lambda_i^2) = \mathcal{A}(\pi g_0/c_i^2)f_i^2 \tag{6.42}$$

[i.e., $g_i = g_0$ and $(\lambda_\perp)_i = \lambda_i$; cf. Eq. (6.21)]. Using Eqs. (6.13) and (6.14), we can write the total-energy flux and total information flux as

$$\frac{P}{\mathcal{A}} = b\sum_i \frac{\pi g_0}{c_i^2} h f_i^3 \bar{n}_i\,, \tag{6.43}$$

$$\frac{C}{\mathcal{A}} = \frac{b}{\ln 2}\sum_i \frac{\pi g_0}{c_i^2} f_i^2[\delta\bar{n}_i \ln(1+\bar{n}_i^{-1}) + \ln(1+\gamma\bar{n}_i)]\,. \tag{6.44}$$

One must now maximize the information flux $C/\mathcal{A}$ by varying both $\bar{n}_i$ and c_i as functions of frequency, with $P/\mathcal{A}$ held constant. The result for the photon-number distribution is familiar: For a number-state channel, the optimal photon-number distribution is the thermal distribution

$$\bar{n}_i = \frac{1}{e^{\beta h f_i} - 1} \tag{6.45}$$

of Eq. (3.35), where β is a Lagrange multiplier that characterizes the optimal distribution of power and information among frequencies. For coherent-state and quadrature-squeezed channels, the optimal distribution is the nonthermal distribution

$$\gamma\bar{n}_i = \begin{cases} \dfrac{f_c}{f_i} - 1\,, & 0 \le f_i \le f_c\,, \\ 0\,, & f_i \ge f_c \end{cases} \tag{6.46}$$

of Eq. (5.48), where f_c is an upper cutoff frequency.

When one varies the phase velocity, one finds that the information flux increases monotonically as c_i decreases. This is the wideband version of our narrowband conclusion (in case 3 of Sec. VI.A.2) that the information flux

is larger for smaller phase velocities. What it means here is the following: if c is the *minimum* phase velocity for a particular bosonic field, we can establish upper bounds on that field's information flux by considering a dispersionless field whose phase velocity is c at all frequencies. Of course, for any particular dispersion relation, one could find a tighter upper bound, but the point is that we can proceed here without detailed knowledge of the dispersion relation and derive upper bounds assuming a dispersionless field.

One other issue, the question of gaps in the frequency spectrum, deserves mention. As we discuss in Secs. III.B and IV.F, gaps mean that information has to be transmitted at higher frequencies, where it requires more energy; thus we can find an upper bound on information flux by assuming, as we do here, that there are no gaps.

Focus now on a number-state channel. Converting the sums (6.43) and (6.44) to integrals over positive frequencies, one finds the standard results for the energy flux and information (entropy) flux of three-dimensional blackbody radiation,

$$\frac{P}{\mathcal{A}} = \frac{Q_0}{h^3\beta^4}\int_0^\infty dx\,\frac{x^3}{e^x-1} = \frac{Q_0}{h^3\beta^4}6\zeta(4) = \frac{\pi^4}{15}\frac{Q_0}{h^3\beta^4}\,, \tag{6.47}$$

$$\begin{aligned}\frac{C_{\rm BPC}}{\mathcal{A}} &= \frac{Q_0}{h^3\beta^3\ln 2}\int_0^\infty dx\,x^2\left(\frac{x}{e^x-1} - \ln(1-e^{-x})\right)\\ &= \frac{\frac{4}{3}\beta P/\mathcal{A}}{\ln 2} = \frac{4\pi^4}{45\ln 2}\frac{Q_0}{h^3\beta^3}\,.\end{aligned} \tag{6.48}$$

Here the subscript BPC stands for "bosonic parallel channels," $Q_0 = \pi g_0/c^2$ is the maximum characteristic factor of Eq. (6.30), and $\zeta(4) = \pi^4/90$ is a value of the Riemann zeta function. Combining Eqs. (6.47) and (6.48) yields an upper bound on information flux,

$$\frac{C_{\rm BPC}}{\mathcal{A}} = \frac{4\pi}{3\ln 2}\left(\frac{Q_0}{15}\right)^{1/4}\left(\frac{P/\mathcal{A}}{h}\right)^{3/4}\ \text{bits/s m}^2 \tag{6.49}$$

[cf. Eqs. (1.9) and (6.41); see Table II]. As derived here, the upper bound (6.49) applies to a wideband frequency-multiplexed number-state channel; but to repeat, the arguments given in Sec. IV can be extended to show that $C_{\rm BPC}/\mathcal{A}$ is an upper bound on the information flux of *any* linear bosonic channel that has an isotropic dispersion relation.

One can also use Eq. (6.15) to derive the standard result for the photon flux of blackbody radiation,

$$\begin{aligned}\frac{N}{\mathcal{A}} &= b\sum_i \frac{\pi g_0}{c^2}f_i^2\bar{n}_i = \frac{Q_0}{h^3\beta^3}\int_0^\infty dx\,\frac{x^2}{e^x-1}\\ &= 2\zeta(3)\frac{Q_0}{h^3\beta^3}\,,\end{aligned} \tag{6.50}$$

where $\zeta(3) = 1.2021$ is a value of the Riemann zeta function. Thus, in contrast to a *single* wideband number-state channel, the information per photon here is finite, with the value

$$\frac{C_{\rm BPC}/\mathcal{A}}{N/\mathcal{A}} = \frac{2\pi^4}{45\,\zeta(3)\ln 2} = 5.1960\ \text{bits/photon}, \tag{6.51}$$

which is very close to the information per photon for the narrowband number-state channel considered in case 3 of Sec. VI.A.2 (see Table II).

Turn now to coherent-state and quadrature-squeezed channels. Converting the sums (6.43) and (6.44) to integrals, one can first write the upper cutoff frequency in terms of the energy flux,

$$f_c = \left(\frac{12\gamma P/\mathcal{A}}{Q_0 h}\right)^{1/4}\,, \tag{6.52}$$

and then derive an upper bound on the information flux,

$$\begin{aligned}\frac{C_{\rm PCQ}}{\mathcal{A}} &= \frac{1}{9\ln 2}Q_0 f_c^3\\ &= \frac{2}{9\ln 2}(108Q_0)^{1/4}\left(\frac{\gamma P/\mathcal{A}}{h}\right)^{3/4}\ \text{bits/s-m}^2\end{aligned} \tag{6.53}$$

[cf. Eq. (6.41); see Table II], where the subscript PCQ reminds one that this upper bound applies to parallel coherent-state channels or quadrature-squeezed channels. A quadrature-squeezed channel can achieve nearly 57% of the parallel-channel upper bound (6.49), whereas a coherent-state channel can achieve nearly 34%. Calculating the photon flux for coherent-state and quadrature-squeezed channels,

$$\frac{N}{\mathcal{A}} = \frac{1}{6\gamma}Q_0 f_c^3\,, \tag{6.54}$$

one finds an information per photon,

$$\frac{C_{\rm PCQ}/\mathcal{A}}{N/\mathcal{A}} = \frac{2}{3\ln 2}\gamma = 0.96180\,\gamma\ \text{bits/photon}, \tag{6.55}$$

again very close to the information per photon found for comparable narrowband channels in case 3 of Sec. VI.A.2 (see Table II).

For a number-state channel, one can identify the peak of the blackbody spectrum as a characteristic frequency,

$$f_{\rm ch} \simeq \frac{\pi}{h\beta} \simeq \left(\frac{C_{\rm BPC}/\mathcal{A}}{Q_0}\right)^{1/3} \simeq \left(\frac{P/\mathcal{A}}{Q_0 h}\right)^{1/4}\,, \tag{6.56}$$

near which most of the energy flows. This characteristic frequency is approximately the same as the upper cutoff frequency (6.52) for coherent-state and quadrature-squeezed channels. For all three kinds of channels, satisfying the requirement (6.18) for many spatially transverse modes means satisfying it near the characteristic frequency. Thus we can write an explicit many-modes requirement for a wideband channel,

$$1 \ll \frac{\mu_{\rm ch}}{g_0} \simeq \frac{1}{g_0}\left(\frac{Q_0\mathcal{A}P}{h}\right)^{1/2}\,, \tag{6.57}$$

where

$$\mu_{\rm ch} = \mathcal{A}\frac{\pi g_0}{c^2} f_{\rm ch}^2 \simeq \left(\frac{Q_0 \mathcal{A} P}{h}\right)^{1/2} \tag{6.58}$$

is the maximum multiplicity factor (6.21) at the characteristic frequency [cf. Eq. (6.29) with $\eta = 1$].

It is instructive to write the maximum capacities (6.49) and (6.53) approximately as

$$C \simeq \left(\frac{Q_0 \mathcal{A} P}{h}\right)^{1/4} \left(\frac{P}{h}\right)^{1/2} \simeq \sqrt{\frac{\mu_{\rm ch} P}{h}}\,, \tag{6.59}$$

which shows, as in Eq. (6.36), that the $P^{3/4}$ dependence of the multichannel wideband capacity arises from the $P^{1/2}$ dependence for a single transverse mode combined with a $P^{1/4}$ dependence, contained in $\sqrt{\mu_{\rm ch}}$, from the number of spatially transverse modes.

The wideband many-modes requirement (6.57) tells us how big $\mu_{\rm ch}$, i.e., $\mathcal{A}$, must be so that we can count modes accurately using the mode density (6.4). For any finite value of $\mathcal{A}$, there are corrections to this counting, due to the discretization of the spatially transverse modes, particularly those with transverse wavelengths $\sim \sqrt{\mathcal{A}}$. We can argue, in a way familiar from our discussion of narrowband channels, that these corrections do not become terribly large, even in the limit $\mu_{\rm ch} \sim g_0$—i.e., $c/f_{\rm ch} \sim \sqrt{\mathcal{A}}$—when there is only about one spatially transverse mode at the characteristic frequency. In this limit the wideband capacity (6.59) becomes

$$C \sim \sqrt{\frac{g_0 P}{h}}\,, \tag{6.60}$$

which again shows how the $P^{3/4}$ dependence for many spatially transverse modes goes over to a $P^{1/2}$ dependence for a single spatially transverse mode, here with g_0 polarization channels. Of course, we assume in making this argument that the aperture remains roughly circular—i.e., remains two dimensional—as its area decreases to $\mathcal{A} \sim (c/f_{\rm ch})^2$. Should one dimension get small much faster than the other, so that the aperture becomes effectively one dimensional, then the capacity would assume a $P^{2/3}$ dependence.

The maximum multiplicity factor (6.58) at the characteristic frequency also arises in the ratio of the "co-ax capacity" (6.8) to the wideband capacity (6.59). Indeed, the co-ax capacity is bigger if the area per co-ax, a, satisfies

$$a \lesssim \frac{\mathcal{A}}{\mu_{\rm ch}} = \frac{1}{\pi g_0}\left(\frac{c}{f_{\rm ch}}\right)^2 . \tag{6.61}$$

Co-axial cables are, of course, limited by Ohmic losses associated with field penetration into the confining conductors, especially at high frequencies; even if the co-ax uses superconductors, losses near or above the band gap limit the bandwidth as in the discussion at the end of Sec. V.C.

VII. CONCLUSION

In our conversations with other physicists, we often find an intense interest in information theory combined with a lack of familiarity with information-theoretic concepts. In this article, by concentrating on a particular question—how much information can be communicated via a linear bosonic field?—we aim to familiarize physicists with basic concepts and tools of information theory.

To this end we develop the communication theorist's description of a communication channel and relate it to the quantum-mechanical description of a linear bosonic field. The quantum-mechanical description requires two ingredients: a description of how information is encoded onto the field and a description of how information is "read off" the field by a quantum detection scheme. To address the question of optimum communication rates, one must be able to formulate both these ingredients in complete generality. Many physicists can benefit, in particular, from our discussion of the theory of generalized quantum measurements, which permits us to include all detection schemes that are consistent with the rules of quantum mechanics.

We highlight the role played by Holevo's theorem, the fundamental theorem of quantum communication theory, which establishes the connection between information, as measured by Shannon's statistical measure of information, and quantum entropy. Without Holevo's theorem the connection between information and quantum entropy, though tantalizing, is incomplete.

We give a general proof of the maximum communication rate of a single linear bosonic channel that has finite power P. Our proof uses techniques drawn from quantum statistical physics and thus emphasizes the connections between information theory and physics. We also draw attention to the recent proof of Yuen and Ozawa (1992), which removes the finiteness assumptions that are required for our proof.

The theme that permeates this article is that maximum communication rates for bosonic fields are encapsulated in the slogan "one quantum–one bit–one mode." Although it is possible to send far more than a bit on a single quantum or far more than a bit in a single field mode, doing either is not the most efficient use of energy. Efficient use of energy involves a compromise between the particle-like aspects of the field, embodied in the number of bits per photon, and the wave-like aspects, embodied in the number of bits per mode. The compromise is captured, at least approximately, in our slogan.

Efficient transmission of information at the wideband quantum limit is an extremely ambitious goal. It is characterized by a wideband spectrum, similar to blackbody radiation, making it necessary to detect quanta on time scales of their period. In fact, this can be regarded as defining efficient communication from a quantum perspective. In this wideband limit, virtual quantum effects are extremely important (Drummond, 1987, 1988). This article serves as an introduction to and a motivation for

further work on ultra-wideband quantum noise, a virtually unexplored area of quantum noise theory and quantum measurement.

ACKNOWLEDGMENTS

C.M.C. thanks Ben Schumacher for bringing Holevo's theorem to his attention, and P.D.D. thanks Anne Streat and Brendan McKay for discussions on summing partition functions. We thank J. D. Bekenstein, C. W. Helstrom, R. Landauer, M. Schiffer, and H. P. Yuen for reading an early draft and for suggestions that improved subsequent versions, and we thank R. W. Hellwarth for pointing out the importance of confining a TEM mode as in a co-axial cable. Special thanks go to G. J. Milburn and H. J. Kimble, both of whom read thoroughly a near-final version and made extensive recommendations that led to the final version. Finally, we thank H. N. Barnum, L. B. Crowell, D. A. Fisher, C. A. Fuchs, K. Khosro-Shahroudi, and R. J. Kunkle for reading carefully the final version as part of a statistical-mechanics course at the University of New Mexico. This work was supported in part by the Defense Advanced Research Projects Agency through the National Center for Integrated Photonic Technology at the University of Southern California, by the Office of Naval Research (Contract No. N00014-88-K-0042 and Grant Nos. N00014-91-J-1167 and N00014-93-1-0116), and by a University of Queensland Special Projects Grant.

APPENDIX A: PROPERTIES OF SHANNON INFORMATION AND QUANTUM ENTROPY

In this appendix we denote the Shannon information for probabilities p_j by

$$H(\mathbf{p}) = -\sum_j p_j \log_d p_j \ , \tag{A1}$$

where $\mathbf{p}$, standing for the entire probability distribution, can be thought of as a vector consisting of the probabilities p_j. Throughout this appendix we allow the base d of the logarithm to be arbitrary, since the main results hold in any base; except where necessary, we omit specification of the base, with the understanding that it is arbitrary. The Shannon information for probabilities on $\mathcal{J}$ alternatives is bounded by

$$0 \le H(\mathbf{p}) \le \log \mathcal{J} \ , \tag{A2}$$

where equality on the left-hand side is equivalent to probabilities for which one alternative is certain and equality on the right-hand side is equivalent to uniform probabilities $p_j = 1/\mathcal{J}$.

An important inequality satisfied by Shannon information, sometimes called the Gibbs inequality, involves the relative information of probability distribution $\mathbf{p}$ with respect to probability distribution $\mathbf{q}$:

$$H(\mathbf{p}/\mathbf{q}) = -\sum_j p_j \log \frac{p_j}{q_j} \le 0 \quad \text{(Gibbs inequality)}. \tag{A3}$$

This inequality can be demonstrated by using a property of the logarithm,

$$\log_d x = \frac{\ln x}{\ln d} \le \frac{x-1}{\ln d} \ , \quad \text{equality if and only if } x = 1, \tag{A4}$$

to show that

$$H(\mathbf{p}/\mathbf{q}) = \sum_j p_j \log \frac{q_j}{p_j} = \sum_j \left(p_j \log \frac{q_j}{p_j} - \frac{q_j - p_j}{\ln d} \right) \le 0 \ . \tag{A5}$$

Equality holds here if and only if every term in the second sum vanishes. Thus relative information is zero if and only if $\mathbf{p}$ and $\mathbf{q}$ are identical, i.e., $p_j = q_j$ for all alternatives j.

The Gibbs inequality shows immediately that the mutual information (2.11) is non-negative,

$$H(B;A) = H(B) - H(B|A) = \sum_a p_A(a) \left[\sum_b p_{B|A}(b|a) \log_2 \left(\frac{p_{B|A}(b|a)}{p_B(b)} \right) \right] \ge 0 \ . \tag{A6}$$

The mutual information is zero if and only if $p_{B|A}(b|a) = p_B(b)$ for all b and for all a for which $p_A(a) \neq 0$. This condition is equivalent to A and B being statistically independent, i.e., $p_{A,B}(a,b) = p_A(a)p_B(b)$.

Shannon information has a concavity property: if $\mathbf{p}_1$ and $\mathbf{p}_2$ are two probability distributions, then

$$H(\lambda \mathbf{p}_1 + (1-\lambda)\mathbf{p}_2) \ge \lambda H(\mathbf{p}_1) + (1-\lambda) H(\mathbf{p}_2) \ , \quad \text{for } 0 < \lambda < 1. \tag{A7}$$

Equality holds here if and only if $\mathbf{p}_1$ and $\mathbf{p}_2$ are identical. In words, this property means that averaging two probability distributions increases the Shannon information. Equation (A6) is an example of property (A7). The property is an immediate consequence of the Gibbs inequality, since, letting $\mathbf{p} = \lambda \mathbf{p}_1 + (1-\lambda)\mathbf{p}_2$,

$$H(\mathbf{p}) - \lambda H(\mathbf{p}_1) - (1-\lambda) H(\mathbf{p}_2) = -\lambda H(\mathbf{p}_1/\mathbf{p}) - (1-\lambda) H(\mathbf{p}_2/\mathbf{p}) \ge 0 \ . \tag{A8}$$

The quantum entropy of a density operator is defined by

$$S(\hat{\rho}) = -\mathrm{tr}(\hat{\rho}\log\hat{\rho}) \ . \tag{A9}$$

One can always write the density operator in terms of a diagonal decomposition in a complete, orthonormal basis of vectors $|j\rangle$:

$$\hat{\rho} = \sum_j p_j |j\rangle\langle j| \ . \tag{A10}$$

The quantum entropy is the Shannon information of the probability distribution $\mathbf{p}$ along the diagonal of this decomposition:

$$S(\hat{\rho}) = -\sum_j p_j \log p_j = H(\mathbf{p}) \ . \tag{A11}$$

Hence the quantum entropy is bounded by

$$0 \le S(\hat{\rho}) \le \log \mathcal{J} \ , \tag{A12}$$

where $\mathcal{J}$ is the dimension of Hilbert space. Equality on the left-hand side is equivalent to $\hat{\rho}$ being a pure state (a one-dimensional projection operator), and equality on the right-hand side is equivalent to $\hat{\rho}$ being a multiple of the unit operator $\hat{1}$, i.e., $\hat{\rho} = \hat{1}/\mathcal{J}$.

Given a different complete, orthonormal basis of vectors $|k\rangle$, one can generate from $\hat{\rho}$ a different set of probabilities

$$q_k = \mathrm{tr}(\hat{\rho}|k\rangle\langle k|) = \langle k|\hat{\rho}|k\rangle = \sum_j |\langle j|k\rangle|^2 p_j \ . \tag{A13}$$

The Shannon information of these new probabilities is not smaller than the quantum entropy:

$$H(\mathbf{q}) \ge H(\mathbf{p}) = S(\hat{\rho}) \ . \tag{A14}$$

This property is a consequence of a special property, called double stochasticity, of the quantum-mechanical conditional probabilities $|\langle j|k\rangle|^2$. Double stochasticity is the property that these conditional probabilities sum to unity not only on k, as any conditional probabilities must, but also on j:

$$\sum_k |\langle j|k\rangle|^2 = 1 = \sum_j |\langle j|k\rangle|^2 \ . \tag{A15}$$

Demonstration of the property (A14) goes as follows (Peres, 1993):

$$\begin{aligned} H(\mathbf{p}) - H(\mathbf{q}) &= \sum_{j,k} |\langle j|k\rangle|^2 p_j \log\frac{q_k}{p_j} \\ &= \sum_{j,k} |\langle j|k\rangle|^2 \left(p_j \log\frac{q_k}{p_j} - \frac{q_k - p_j}{\ln d} \right) \le 0 \ . \end{aligned} \tag{A16}$$

Double stochasticity is used to insert the sum over $q_k - p_j$, and the inequality follows from property (A4). Equality holds in Eq. (A16) if and only if every term in the second sum vanishes, i.e., $p_j = q_k$ or $\langle j|k\rangle = 0$ for all j and k. This necessary and sufficient condition for equality is equivalent to $(p_j - q_k)\langle j|k\rangle = 0$ for all j and k, which in turn is equivalent to

$$0 = \sum_j (p_j - q_k)|j\rangle\langle j|k\rangle = (\hat{\rho} - q_k)|k\rangle \ , \quad \text{for all } k. \tag{A17}$$

Thus equality holds in Eq. (A14) if and only if $\hat{\rho}$ is diagonal in the $|k\rangle$-basis, i.e., $\hat{\rho} = \sum_k q_k |k\rangle\langle k|$.

The quantum analogue of the relative information (A3) is a relative entropy of density operator $\hat{\rho}$ with respect to density operator $\hat{\rho}'$:

$$S(\hat{\rho}/\hat{\rho}') = -\mathrm{tr}(\hat{\rho}\log\hat{\rho}) + \mathrm{tr}(\hat{\rho}\log\hat{\rho}') \le 0 \ . \tag{A18}$$

The inequality asserts that, like relative information, relative entropy is not positive. To prove the inequality, write $\hat{\rho}$ and $\hat{\rho}'$ in terms of diagonal decompositions in complete, orthonormal bases $|j\rangle$ and $|k\rangle$,

$$\hat{\rho} = \sum_j p_j |j\rangle\langle j| \ , \tag{A19}$$

$$\hat{\rho}' = \sum_k r_k |k\rangle\langle k| \ . \tag{A20}$$

By introducing the probabilities q_k associated with $\hat{\rho}$ in the $|k\rangle$ basis [Eq. (A13)], the relative entropy can be written as

$$S(\hat{\rho}/\hat{\rho}') = \sum_{j,k} |\langle j|k\rangle|^2 p_j \log\frac{r_k}{p_j} = -\sum_k q_k \log\frac{q_k}{r_k} + \sum_{j,k} |\langle j|k\rangle|^2 p_j \log\frac{q_k}{p_j} = H(\mathbf{q}/\mathbf{r}) + H(\mathbf{p}) - H(\mathbf{q}) \le 0 \ . \tag{A21}$$

The relative entropy is zero if and only if $\hat{\rho} = \hat{\rho}'$.

Quantum entropy shares the concavity property of Shannon information: if $\hat{\rho}_1$ and $\hat{\rho}_2$ are two density operators, then

$$S(\lambda\hat{\rho}_1 + (1-\lambda)\hat{\rho}_2) \ge \lambda S(\hat{\rho}_1) + (1-\lambda) S(\hat{\rho}_2) \ , \quad \text{for } 0 < \lambda < 1. \tag{A22}$$

Equality holds here if and only if $\hat{\rho}_1 = \hat{\rho}_2$. The proof is the same as for Shannon information, with relative entropy replacing relative information. In words, concavity means that averaging two density operators increases the quantum entropy.

Our final property of quantum entropy concerns bounds on the entropy of an average density operator. Suppose that density operators $\hat{\rho}_i$, $i = 1, \ldots, \mathcal{I}$, occur with probabilities q_i, giving an average density operator

$$\hat{\rho} = \sum_{i=1}^{\mathcal{I}} q_i \hat{\rho}_i \ . \tag{A23}$$

One can show that

$$0 \le S(\hat{\rho}) - \sum_{i=1}^{\mathcal{I}} q_i S(\hat{\rho}_i) \le H(\mathbf{q}) \ . \tag{A24}$$

Notice that the quantity bounded here is the same as the one that appears in Holevo's theorem (4.13). The lower bound on the left-hand side is a simple extension of the concavity property (A22) of quantum entropy, but the upper bound on the right-hand side (Levitin, 1969;[6] Balian, 1991) requires some work. Equality holds in the upper bound if and only if all the density operators $\hat{\rho}_i$ are orthogonal.

To prove the upper bound, we generalize slightly a proof given by Hughston, Jozsa, and Wootters (1993). First write each density operator $\hat{\rho}_i$ in terms of its diagonal decomposition in a complete, orthonormal basis:

$$\hat{\rho}_i = \sum_{l=1}^{\mathcal{L}_i} r_{l|i} |l,i\rangle\langle l,i| \ . \tag{A25}$$

The eigenvalues $r_{l|i}$ are written deliberately as conditional probabilities for l given i; by omitting zero eigenvalues from the sum, we may assume that all these conditional probabilities are positive. The quantum entropy of $\hat{\rho}_i$ becomes

$$S(\hat{\rho}_i) = -\sum_{l=1}^{\mathcal{L}_i} r_{l|i} \log r_{l|i} \ . \tag{A26}$$

We can introduce a normalized "joint probability" for i and l,

$$Q_\mu = Q_{i,l} = r_{l|i} q_i \ , \tag{A27}$$

where the index μ stands for both i and l. By construction, all the probabilities Q_μ are nonzero. The average density operator can be decomposed in terms of the states $|l,i\rangle = |\mu\rangle$,

$$\hat{\rho} = \sum_{i=1}^{\mathcal{I}} q_i \left(\sum_{l=1}^{\mathcal{L}_i} r_{l|i} |l,i\rangle\langle l,i| \right) = \sum_{\mu=1}^{\mathcal{M}} Q_\mu |\mu\rangle\langle\mu| \ , \tag{A28}$$

where $\mathcal{M}$ is the number of states $|\mu\rangle$. The Shannon information of the joint distribution $\mathbf{Q}$ is

$$H(\mathbf{Q}) = -\sum_{\mu=1}^{\mathcal{M}} Q_\mu \log Q_\mu = -\sum_{i=1}^{\mathcal{I}} q_i \log q_i + \sum_{i=1}^{\mathcal{I}} q_i \left(-\sum_{l=1}^{\mathcal{L}_i} r_{l|i} \log r_{l|i} \right) = H(\mathbf{q}) + \sum_{i=1}^{\mathcal{I}} q_i S(\hat{\rho}_i) \tag{A29}$$

[cf. Eq. (2.14)].

The average density operator can also be written in terms of a diagonal decomposition in a complete, orthonormal basis $|j\rangle$,

$$\hat{\rho} = \sum_{j=1}^{\mathcal{J}} p_j |j\rangle\langle j| \ , \tag{A30}$$

where $\mathcal{J}$ is the dimension of Hilbert space. The quantum entropy of the average density operator is

$$S(\hat{\rho}) = -\sum_{j=1}^{\mathcal{J}} p_j \log p_j = H(\mathbf{p}) \ . \tag{A31}$$

If $\hat{\rho}$ has any zero eigenvalues, we can omit those terms from the sum. To formalize this omission, suppose that there are $\bar{\mathcal{J}}$ nonzero eigenvalues, labeled by $j = 1, \ldots, \bar{\mathcal{J}}$, so that $p_j = 0$ for $j > \bar{\mathcal{J}}$. Then we can write the quantum entropy of the average density operator as

$$S(\hat{\rho}) = H(\mathbf{p}) = -\sum_{j=1}^{\bar{\mathcal{J}}} p_j \log p_j \ . \tag{A32}$$

The upper bound we wish to prove can now be written as

$$H(\mathbf{Q}) \ge H(\mathbf{p}) \ . \tag{A33}$$

We proceed by showing that $\mathbf{Q}$ is related to $\mathbf{p}$ by a doubly stochastic matrix. First expand the states $|\mu\rangle$ in terms of the complete, orthonormal basis $|j\rangle$:

$$|\mu\rangle = \sum_{j=1}^{\mathcal{J}} |j\rangle\langle j|\mu\rangle \ . \tag{A34}$$

[6] This article was translated for the authors by A. Bezinger and S. L. Braunstein; it has been essentially reprinted as part of two longer articles (Levitin, 1991 and Levitin, 1993).

By noting that

$$p_j = \langle j|\hat{\rho}|j\rangle = \sum_{\mu=1}^{\mathcal{M}} Q_\mu |\langle j|\mu\rangle|^2 , \tag{A35}$$

we can conclude that if $j > \bar{\mathcal{J}}$ (in which case $p_j = 0$), $\langle j|\mu\rangle = 0$ for all μ. Thus we can rewrite the expansion (A34) as

$$\sqrt{Q_\mu}|\mu\rangle = \sum_{j=1}^{\bar{\mathcal{J}}} M_{\mu j}\sqrt{p_j}|j\rangle , \tag{A36}$$

where the $\mathcal{M} \times \bar{\mathcal{J}}$ matrix $M_{\mu j}$ is defined by

$$M_{\mu j} = \sqrt{\frac{Q_\mu}{p_j}}\langle j|\mu\rangle , \quad \mu = 1,\ldots,\mathcal{M} , \quad j = 1,\ldots,\bar{\mathcal{J}} . \tag{A37}$$

It is easy to demonstrate that the $\bar{\mathcal{J}}$ columns of $M_{\mu j}$ are orthogonal $\mathcal{M}$-dimensional vectors (which requires that $\mathcal{M} \geq \bar{\mathcal{J}}$):

$$\sum_{\mu=1}^{\mathcal{M}} M_{\mu j}M^*_{\mu j'} = \frac{\left\langle j\left|\left(\sum_{\mu=1}^{\mathcal{M}} Q_\mu|\mu\rangle\langle\mu|\right)\right|j'\right\rangle}{\sqrt{p_j p_{j'}}} = \frac{\langle j|\hat{\rho}|j'\rangle}{\sqrt{p_j p_{j'}}} = \delta_{jj'} . \tag{A38}$$

Now, by adding columns, extend $M_{\mu j}$ to be an $\mathcal{M} \times \mathcal{M}$ unitary matrix $U_{\mu j}$. Taking the magnitude of Eq. (A36) yields

$$Q_\mu = \sum_{j=1}^{\bar{\mathcal{J}}} |M_{\mu j}|^2 p_j = \sum_{j=1}^{\mathcal{M}} |U_{\mu j}|^2 p_j , \tag{A39}$$

where the probabilities p_j for $j > \bar{\mathcal{J}}$ are chosen to be zero.

Since $U_{\mu j}$ is unitary, the transition matrix $|U_{\mu j}|^2$ is doubly stochastic,

$$\sum_{\mu=1}^{\mathcal{M}} |U_{\mu j}|^2 = 1 = \sum_{j=1}^{\mathcal{M}} |U_{\mu j}|^2 ; \tag{A40}$$

so we can repeat the proof in Eq. (A16):

$$H(\mathbf{p}) - H(\mathbf{Q}) = \sum_{j=1}^{\mathcal{M}}\sum_{\mu=1}^{\mathcal{M}} |U_{\mu j}|^2 p_j \log\frac{Q_\mu}{p_j} = \sum_{j=1}^{\mathcal{M}}\sum_{\mu=1}^{\mathcal{M}} |U_{\mu j}|^2 \left(p_j \log\frac{Q_\mu}{p_j} - \frac{Q_\mu - p_j}{\ln d}\right) \leq 0 . \tag{A41}$$

Equality holds here if and only if $(p_j - Q_\mu)U_{\mu j} = 0$ for all j and μ. This condition requires that $U_{\mu j} = 0$ for $j > \bar{\mathcal{J}}$, which, since $U_{\mu j}$ is unitary, means that $\mathcal{M} = \bar{\mathcal{J}}$. Hence we can restate the necessary and sufficient conditions for equality as $\mathcal{M} = \bar{\mathcal{J}}$ and

$$0 = \sum_{j=1}^{\bar{\mathcal{J}}} (p_j - Q_\mu)|j\rangle M_{\mu j}\sqrt{\frac{p_j}{Q_\mu}} = \sum_{j=1}^{\mathcal{J}} (p_j - Q_\mu)|j\rangle\langle j|\mu\rangle = (\hat{\rho} - Q_\mu)|\mu\rangle , \quad \text{for all } \mu. \tag{A42}$$

Thus equality holds in the upper bound of Eq. (A24) if and only if the states $|\mu\rangle$ provide an orthogonal decomposition of $\hat{\rho}$, with no state appearing more than once in the decomposition. This condition, in turn, is equivalent to the above-stated requirement that the density operators $\hat{\rho}_i$ be orthogonal.

A more complete discussion of quantum entropy can be found in Balian (1991).

APPENDIX B: HERALDED AND SELF-HERALDING SIGNALS

Bekenstein (1988) and Bekenstein and Schiffer (1990) distinguish what they call heralded and self-heralding signals. To illustrate and assess this distinction, we return to the narrowband number-state channel introduced in Sec. III.A.1, which we now assume to be noise-free—$p_{M|N}(m|n) = \delta_{mn}$—and we focus on the information carried by a particular wave-packet mode—a particular channel use—which lasts a time $\simeq B^{-1}$. This information is quantified by the Shannon information

$$H(M;N) = H(M) = H(N) = -\sum_{n=0}^{\infty} p_N(n)\log_2 p_N(n) . \tag{B1}$$

Bekenstein and Schiffer argue that if the receiver knows in advance—"anticipates" in their language—that the mode of interest carries a signal—for example, knows that it has been manipulated to carry a signal or knows that it is part of an information-bearing sequence of wave packets—then the vacuum state should be included as a signaling state. They call such an anticipated signal a heralded signal because the signal's existence is known in advance. In contrast, they argue that if the receiver does not know in advance—does not "anticipate"—that the mode of interest carries a signal, then one should not include the vacuum state as a signaling state, because

there is no way to distinguish receipt of the vacuum from the persistent vacuum that arrives in the absence of signaling. Should an unanticipated signal—i.e., a nonzero number of photons—happen to be received in this mode, they call that signal a self-heralding signal, because it informs the receiver of its own existence.

The distinction can be displayed formally by writing the total information (B1) in the form

$$H(N) = H_{\rm bin} + qH_{n\neq 0} \ , \tag{B2}$$

where $q = 1 - p_N(0)$ is the probability that the mode is not in vacuum. In Eq. (B2)

$$H_{\rm bin} = -q\log_2 q - (1-q)\log_2(1-q) \tag{B3}$$

quantifies the binary information that the receiver gets from determining whether the mode is in vacuum [this is the only information if $p_N(1) = q = \bar{n}$ is the only nonzero probability for $n \geq 1$, as in the discussion of placement information for a single photon at the end of Sec. III.A.2], and

$$H_{n\neq 0} = -\sum_{n=1}^{\infty} \frac{p_N(n)}{q} \log_2\left(\frac{p_N(n)}{q}\right) \tag{B4}$$

quantifies the further information that the receiver gets from determining the photon number, once it knows that the mode is not in vacuum. In the total information, $H_{n\neq 0}$ is properly weighted by q. Notice that $H_{n\neq 0}$ does not include any contribution from the vacuum state. In accord with their argument, Bekenstein and Schiffer use the total information $H(N)$ for heralded signals, but toss out the binary information $H_{\rm bin}$ and use only $H_{n\neq 0}$ to quantify the information carried by a self-heralding signal.

There is nothing wrong with the decomposition (B2) or with its interpretation. If one is interested in the binary information $H_{\rm bin}$ or in the further information $H_{n\neq 0}$, then by all means one should deal with those quantities. We are unconvinced, however, that it is justified to neglect the binary information in calculating the *total* information. The difficulty, we believe, lies in the very concept of a signal and, specifically, in what it means to "anticipate" a signal.

A noise-free channel such as we are considering is specified by giving the input alphabet and the input probabilities for each use. Whatever the receiver knows or "anticipates" about the channel—about a particular use or a set of uses—must be incorporated in the input probabilities; there is no other place to express such knowledge. In contrast to the assumption made throughout this article, the input probabilities could change from use to use, reflecting a change in one's expectations based on results of previous uses. Given the input probabilities, however, the Shannon information (B2) quantifies the (average) information available at reception; there is no justification for neglecting a part of the Shannon information.

Bekenstein and Schiffer are really arguing that receipt of the vacuum provides *no* information when the receiver has an overwhelming expectation to receive the vacuum. Whatever the receiver's expectation, however, it must be incorporated in the probability q. If the receiver does not expect to receive a nonzero number of photons, then q is very small—exactly how small depending on the receiver's precise expectation. In accord with one's intuition, the binary information $H_{\rm bin} \sim -q\log_2 q$ does become very small, but not zero. There is no justification for neglecting it.

APPENDIX C: SINGLE-CHANNEL DISPERSION

In this appendix we argue that even in the presence of dispersion, a single bosonic channel of duration $\mathcal{T}$ has one longitudinal frequency-domain mode at each of the allowed frequencies

$$f_j = j/\mathcal{T} \ , \quad j = 1,2,\dots \ , \tag{C1}$$

of Eq. (4.1). The easy way to argue is that in the presence of dispersion, we use periodic boundary conditions on the time $\mathcal{T}$. This leads directly to the mode sequence of Eq. (C1). The mode at frequency f_j extends along the channel a length $c_g(f_j)\mathcal{T}$, where

$$c_g(f) = 2\pi\frac{df}{dk} \tag{C2}$$

is the group velocity at frequency f. In order to use the group-velocity approximation, the mode spreading during time $\mathcal{T}$, $\Delta c_g \mathcal{T} = dc_g/df$, should be negligible compared to the mode's length; this imposes the requirement

$$\mathcal{T} \gg \frac{1}{c_g}\frac{dc_g}{df} \ . \tag{C3}$$

Since periodic boundary conditions in time are not a standard quantization procedure, however, we also argue from a standard procedure.

In the presence of dispersion we can quantize by using periodic boundary conditions on a length L, chosen to be much bigger than $c_g(f)\mathcal{T}$ for all frequencies f. The result is a constant density of longitudinal modes in wave number, $\rho_k = L/2\pi$. We are only interested in modes with positive wave number, which propagate in the desired direction. From these modes we construct wave-packet modes that have duration $\mathcal{T}$. At frequency f there are $L/c_g(f)\mathcal{T} \gg 1$ successive wave-packet modes within the length L. These modes occupy a wave-number interval

$$\Delta k = \frac{L/c_g(f)\mathcal{T}}{\rho_k} = \frac{2\pi}{c_g(f)\mathcal{T}} \ , \tag{C4}$$

corresponding to a separation between frequencies of the wave-packet modes,

$$\Delta f = \frac{\Delta f}{\Delta k}\Delta k = \frac{c_g(f)}{2\pi}\Delta k = \frac{1}{\mathcal{T}} \ . \tag{C5}$$

The allowed wave-packet-mode frequencies are thus the

same as in Eq. (C1).

At each allowed frequency one should imagine the $L/c_g(f)\mathcal{T}$ wave-packet modes lined up along the channel from output to input, each occupying a length $c_g(f)\mathcal{T}$ and having duration $\mathcal{T}$. To construct *arbitrary* signals within the duration $\mathcal{T}$ nearest the channel output, one needs one wave-packet mode at each allowed frequency, the one lying nearest the channel output. The bottom line is that there is one longitudinal mode at each of the allowed frequencies (C1), just as for a dispersionless channel.

APPENDIX D: GENERALIZED MEASUREMENTS

In this appendix we discuss generalized measurements and their relation to conventional quantum measurements, and we demonstrate how the effect density of coherent-state projectors applies to heterodyne detection.

1. Conventional measurements versus generalized measurements

Consider a quantum system with Hilbert space $\mathcal{H}$. Conventional quantum measurements on this system are described by projection operators on $\mathcal{H}$. A projection operator $\hat{\Pi}$ is a Hermitian operator all of whose eigenvalues are 0 or 1, a requirement equivalent to

$$\hat{\Pi}^2 = \hat{\Pi} \,. \tag{D1}$$

The projection operator $\hat{\Pi}$ projects state vectors onto a subspace $\mathcal{S}(\hat{\Pi})$ spanned by the eigenvectors that have eigenvalue 1. The dimension $d = \mathrm{tr}(\hat{\Pi})$ of the projection operator is the dimension of this associated subspace.

Two projection operators $\hat{\Pi}$ and $\hat{\Pi}'$ are said to be orthogonal if

$$\mathrm{tr}(\hat{\Pi}\,\hat{\Pi}') = 0 \,, \tag{D2}$$

which is equivalent to the condition

$$\hat{\Pi}\,\hat{\Pi}' = \hat{\Pi}'\,\hat{\Pi} = 0 \tag{D3}$$

that the associated subspaces $\mathcal{S}(\hat{\Pi})$ and $\mathcal{S}(\hat{\Pi}')$ be orthogonal. A one-dimensional projector can be written as

$$\hat{\Pi}^{(1)} = |\psi\rangle\langle\psi| \,, \tag{D4}$$

where $|\psi\rangle$ is a normalized state vector (pure state) in $\mathcal{H}$. A d-dimensional projection operator can be written as a sum of d orthogonal one-dimensional projectors,

$$\hat{\Pi}^{(d)} = \sum_{i=1}^{d} \hat{\Pi}_i^{(1)} = \sum_{i=1}^{d} |\psi_i\rangle\langle\psi_i| \,, \tag{D5}$$

where the state vectors $|\psi_i\rangle$ are any complete, orthonormal basis within the subspace $\mathcal{S}(\hat{\Pi})$. Orthogonal projection operators obviously commute; moreover, commuting *one-dimensional* projection operators are either identical or orthogonal.

A conventional quantum measurement is described by a complete set of orthogonal projection operators $\hat{\Pi}_b$:

$$\hat{\Pi}_b\hat{\Pi}_{b'} = \hat{\Pi}_{b'}\hat{\Pi}_b = \delta_{bb'}\hat{\Pi}_b \quad \text{(orthogonality)}, \tag{D6}$$

$$\sum_b \hat{\Pi}_b = \hat{1} \quad \text{(completeness)}. \tag{D7}$$

The probability for outcome b is given by

$$p_{B|A}(b|a) = \mathrm{tr}(\hat{\rho}_a\hat{\Pi}_b) \,. \tag{D8}$$

We call a conventional quantum measurement "ideal" if the projectors are not only orthogonal, but also one-dimensional, i.e.,

$$\hat{\Pi}_b = \hat{\Pi}_b^{(1)} = |b\rangle\langle b| \,. \tag{D9}$$

The state vectors $|b\rangle$, which form a complete, orthonormal basis in $\mathcal{H}$, can be regarded as the eigenvectors of a nondegenerate observable (or complete set of observables). The ideal measurement is interpreted as a measurement of that observable, with the probability of outcome b given by the standard formula

$$p_{B|A}(b|a) = \mathrm{tr}(\hat{\rho}_a\hat{\Pi}_b^{(1)}) = \langle b|\hat{\rho}_a|b\rangle \,. \tag{D10}$$

If not all the orthogonal projectors $\hat{\Pi}_b$ are one-dimensional, they can still be diagonalized in a (nonunique) complete, orthonormal basis $|\bar{b}\rangle$,

$$\hat{\Pi}_b = \sum_{\bar{b}\in S_b} \hat{\Pi}_{\bar{b}}^{(1)} = \sum_{\bar{b}\in S_b} |\bar{b}\rangle\langle\bar{b}| \,, \tag{D11}$$

where the sets S_b are exhaustive and disjoint. The vectors $|\bar{b}\rangle$ such that $\bar{b} \in S_b$ span the subspace $\mathcal{S}(\hat{\Pi}_b)$. The measurement described by the projectors $\hat{\Pi}_b$ can be interpreted as a measurement of the nondegenerate observable that has eigenvectors $|\bar{b}\rangle$ (since the vectors $|\bar{b}\rangle$ are not unique, neither is this observable), but a measurement that is nonideal because it has insufficient resolution to distinguish all the eigenvalues. Alternatively, the measurement can be interpreted as a measurement of an observable that is degenerate in each of the subspaces $\mathcal{S}(\hat{\Pi}_b)$. The probability for outcome b is an obvious generalization of Eq. (D10):

$$p_{B|A}(b|a) = \mathrm{tr}(\hat{\rho}_a\hat{\Pi}_b) = \sum_{\bar{b}\in S_b} p_{\bar{B}|A}(\bar{b}|a) = \sum_{\bar{b}\in S_b} \langle\bar{b}|\hat{\rho}_a|\bar{b}\rangle \,. \tag{D12}$$

An effect $\hat{F}$ is a Hermitian operator all of whose eigenvalues lie between 0 and 1, inclusive (formally, $\hat{0} \le \hat{F} \le \hat{1}$). Thus a projection operator is a special effect. When written in diagonal form, $\hat{F}$ is a linear combination of orthogonal one-dimensional projectors,

$$\hat{F} = \sum_i F_i \hat{\Pi}_i^{(1)} = \sum_i F_i |\psi_i\rangle\langle\psi_i| \,, \tag{D13}$$

where the eigenvalues F_i satisfy

$$0 \le F_i \le 1 \,. \tag{D14}$$

A generalized quantum measurement is described by a complete set of effects $\hat{F}_b$:

$$\sum_b \hat{F}_b = \hat{1} \quad \text{(completeness)}. \tag{D15}$$

The probability for outcome b is given by

$$p_{B|A}(b|a) = \mathrm{tr}(\hat{\rho}_a \hat{F}_b) \,. \tag{D16}$$

What distinguishes conventional quantum measurements from generalized ones? The answer is provided by the following theorem: if there exists a complete, orthonormal set of pure states $\hat{\rho}_a = |a\rangle\langle a|$ such that all the measurement probabilities are sharp, i.e., for each a, there is a unique b such that $p_{B|A}(b|a) = \langle a|\hat{F}_b|a\rangle = 1$, then the effects are orthogonal projectors. The proof begins by noting that the set of all a values can be partitioned into sets

$$S_b = \{\, a \mid p_{B|A}(b|a) = 1 \,\} \tag{D17}$$

that are exhaustive and disjoint. Let $\mathcal{S}_b$ be the subspace spanned by the vectors $|a\rangle$ such that $a \in S_b$; the subspaces $\mathcal{S}_b$ are, of course, orthogonal. Since $p_{B|A}(b|a) = \langle a|\hat{F}_b|a\rangle = 0$ if $a \notin S_b$ and since any effect has a Hermitian square root, we can conclude that

$$\hat{F}_b|a\rangle = 0 \,, \quad \text{for } a \notin S_b; \tag{D18}$$

i.e., $\hat{F}_b$ operates only within the subspace $\mathcal{S}_b$. Thus the effects operate in mutually orthogonal subspaces, and completeness implies that they are projectors onto those subspaces, i.e.,

$$\hat{F}_b = \sum_{a\in S_b} |a\rangle\langle a| = \hat{\Pi}_b \,. \tag{D19}$$

The theorem says that if the effects are not orthogonal projectors, then the measurements are necessarily imprecise in the sense that it is impossible to find a complete set of states such that the measurement outcomes are completely predictable. An essentially trivial version of this imprecision occurs when all the effects commute—i.e.,

$$[\hat{F}_b, \hat{F}_{b'}] = 0, \text{ for all } b \text{ and } b' \tag{D20}$$

—because then one can diagonalize all the effects in a common basis $|\bar{b}\rangle$:

$$\hat{F}_b = \sum_{\bar{b}} p_{B|\bar{B}}(b|\bar{b}) \hat{\Pi}_{\bar{b}}^{(1)} = \sum_{\bar{b}} p_{B|\bar{B}}(b|\bar{b}) |\bar{b}\rangle\langle\bar{b}| \,. \tag{D21}$$

The probability for outcome b becomes

$$p_{B|A}(b|a) = \mathrm{tr}(\hat{\rho}_a \hat{F}_b) = \sum_{\bar{b}} p_{B|\bar{B}}(b|\bar{b}) p_{\bar{B}|A}(\bar{b}|a) = \sum_{\bar{b}} p_{B|\bar{B}}(b|\bar{b}) \langle\bar{b}|\hat{\rho}_a|\bar{b}\rangle \,. \tag{D22}$$

The eigenvalues in Eq. (D21), which satisfy

$$0 \le p_{B|\bar{B}}(b|\bar{b}) \le 1 \,, \tag{D23}$$

are clearly the conditional probabilities for outcome b, given that the system is in state $|\bar{b}\rangle$. These conditional probabilities are normalized to unity by the completeness condition (D15).

Measurements described by commuting effects are easily understood generalizations of orthogonal-projector measurements. An ideal conventional quantum measurement, as in Eq. (D10), is a measurement of a nondegenerate observable. The orthogonal-projector measurement of Eq. (D12) can be regarded as a measurement of a nondegenerate observable that has eigenvectors $|\bar{b}\rangle$, but a measurement with insufficient resolution to distinguish all the eigenvalues. The commuting-effects measurement of Eq. (D22) can also be regarded as a measurement of a nondegenerate observable that has eigenvectors $|\bar{b}\rangle$, but a measurement with reduced resolution *plus* imprecision that makes the outcome unpredictable even when the system state is an eigenstate $|\bar{b}\rangle$ of the measured observable.

The situation is more interesting, but less amenable to general interpretation, if the effects do not commute. There being then no complete, orthonormal basis that diagonalizes all the effects simultaneously, there is no way to interpret the measurement in terms of a single conventional observable. The intrinsic imprecision of the measurement arises because one is really measuring two or more noncommuting observables simultaneously. The noncommuting observables are, in general, not unique; more discouraging is that, given the effects, there is no general prescription for identifying candidates for the noncommuting observables. For this reason a measurement described by noncommuting effects is usually referred to as a measurement of a *generalized* observable. Nonetheless, it is useful to keep in mind the potential interpretation in terms of noncommuting conventional observables; in the next subsection we consider a measurement that can be naturally interpreted as a simultaneous measurement of two quadrature components (equivalent to a simultaneous—but necessarily imprecise!—measurement of position and momentum).

We can categorize measurements of generalized observables in the same way we categorized measurements of single conventional observables. An "ideal" measurement of a generalized observable—one with all the resolution allowed by quantum mechanics—is described by effects that are proportional to nonorthogonal *one-dimensional* projectors,

$$\hat{F}_b = \hat{F}_b^{(1)} = f_b |b\rangle\langle b| \tag{D24}$$

[cf. Eq. (D9)], where $0 \le f_b \le 1$. The corresponding

probability for outcome b is

$$p_{B|A}(b|a) = \mathrm{tr}(\hat{\rho}_a \hat{F}_b^{(1)}) = f_b \langle b|\hat{\rho}_a|b\rangle \ . \tag{D25}$$

Ideal measurements of generalized observables—and the associated multiples of nonorthogonal one-dimensional projectors—are the building blocks that one uses to construct other measurements of generalized observables. For example, we can introduce a straightforward reduction in resolution by adding ideal effects,

$$\hat{F}_b = \sum_{\bar{b}\in S_b} \hat{F}_{\bar{b}}^{(1)} = \sum_{\bar{b}\in S_b} f_{\bar{b}} |\bar{b}\rangle\langle\bar{b}| \ , \tag{D26}$$

where the sets S_b are exhaustive and disjoint [cf. Eq. (D11)]. The probability for outcome b becomes

$$p_{B|A}(b|a) = \mathrm{tr}(\hat{\rho}_a \hat{F}_b) = \sum_{\bar{b}\in S_b} p_{\bar{B}|A}(\bar{b}|a) = \sum_{\bar{b}\in S_b} f_{\bar{b}} \langle\bar{b}|\hat{\rho}_a|\bar{b}\rangle \ . \tag{D27}$$

Finally, we can introduce a reduction in resolution plus *additional* imprecision by taking appropriate linear combinations of the ideal effects,

$$\hat{F}_b = \sum_{\bar{b}} p_{B|\bar{B}}(b|\bar{b}) \hat{F}_{\bar{b}}^{(1)} = \sum_{\bar{b}} p_{B|\bar{B}}(b|\bar{b}) f_{\bar{b}} |\bar{b}\rangle\langle\bar{b}| \ , \tag{D28}$$

where $p_{B|\bar{B}}(b|\bar{b})$ is the conditional probability for outcome b, given outcome $\bar{b}$ in the corresponding ideal measurement [cf. Eq. (D21)]. The probability for outcome b is given by

$$\begin{aligned} p_{B|A}(b|a) = \mathrm{tr}(\hat{\rho}_a \hat{F}_b) &= \sum_{\bar{b}} p_{B|\bar{B}}(b|\bar{b}) p_{\bar{B}|A}(\bar{b}|a) \\ &= \sum_{\bar{b}} p_{B|\bar{B}}(b|\bar{b}) f_{\bar{b}} \langle\bar{b}|\hat{\rho}_a|\bar{b}\rangle \ . \end{aligned} \tag{D29}$$

An arbitrary complete set of effects can be written (nonuniquely) in the form (D28).

2. Heterodyne detection and coherent-state projectors

It is easy to show how effects arise out of very general quantum measurement models. Suppose, for example, that a quantum system, the object to be measured, interacts with a meter, the object that is to display the result of the measurement. In addition to the system and the meter, let there be a surrounding environment that interacts with both. Before the system and the meter interact, suppose that the total density operator can be written as

$$\hat{\rho}_{\mathrm{tot}} = \hat{\rho}_a \hat{\rho}_{\mathrm{ME}} \ , \tag{D30}$$

where $\hat{\rho}_a$ is a system density operator and $\hat{\rho}_{\mathrm{ME}}$ is the density operator of the meter and the environment. The system, meter, and environment interact for some time, their joint evolution described by a unitary operator $\hat{U}$.

Suppose now that after the interaction, one makes a conventional quantum measurement on the meter, which is described by a complete set of orthogonal meter projection operators $\hat{\Pi}_b^{\mathrm{M}}$. The probability for result b is given by the standard formula

$$p_{B|A}(b|a) = \mathrm{tr}_{\mathrm{tot}}(\hat{U}\hat{\rho}_{\mathrm{tot}}\hat{U}^\dagger \hat{\Pi}_b^{\mathrm{M}}) = \mathrm{tr}_{\mathrm{tot}}(\hat{\rho}_{\mathrm{tot}}\hat{U}^\dagger \hat{\Pi}_b^{\mathrm{M}} \hat{U}) \ , \tag{D31}$$

where $\mathrm{tr}_{\mathrm{tot}}$ denotes a trace over the system, the meter, and the environment. One can write the probability (D31) in terms of a system trace,

$$p_{B|A}(b|a) = \mathrm{tr}(\hat{\rho}_a \hat{F}_b) \ , \tag{D32}$$

where the system operator

$$\hat{F}_b = \mathrm{tr}_{\mathrm{ME}}(\hat{\rho}_{\mathrm{ME}} \hat{U}^\dagger \hat{\Pi}_b^{\mathrm{M}} \hat{U}) \tag{D33}$$

is defined in terms of a meter-environment trace. It should be obvious that $\hat{F}_b$ is an effect for the system; completeness follows from the completeness of the meter projection operators $\hat{\Pi}_b^{\mathrm{M}}$. It should also be clear that one would generate a complete set of system effects if the meter measurement were instead a generalized measurement, i.e., if the meter projection operators $\hat{\Pi}_b^{\mathrm{M}}$ were replaced by a complete set of meter effects.

More important than this simple demonstration that measurement models lead to effects is the converse (Holevo, 1982; Kraus, 1983; Ozawa, 1984; Peres, 1990): any complete set of system effects can be realized in an appropriate measurement model, i.e., by a conventional measurement on an extended Hilbert space.

A good example of how effects arise from measurement models is provided by heterodyne detection (see Fig. 4). This example has the added virtue of illustrating how a complete set of effects is replaced by a complete effect density in the case of continuous measurement outcomes. The analysis in the caption of Fig. 4 shows that heterodyne detection measures the complex quantity

$$\hat{a}_+ e^{-i\theta} + \hat{a}_-^\dagger e^{i\theta} \ , \tag{D34}$$

where $\hat{a}_+$ is the annihilation operator of the signal mode, $\hat{a}_-$ is the annihilation operator of the image sideband, and θ is the phase of the local oscillator. The state of the signal mode is the density operator $\hat{\rho}_a$, and the state of the image sideband is assumed to be the vacuum state $|0_-\rangle$.

By rephasing the quantity (D34) and absorbing an irrelevant phase into $\hat{a}_-$, we can always bring the measured quantity into the form

$$\hat{\beta} = \hat{\beta}_1 + i\hat{\beta}_2 = \hat{a}_+ + \hat{a}_-^\dagger \ . \tag{D35}$$

It is easy to verify that the Hermitian operators $\hat{\beta}_1$ and $\hat{\beta}_2$ commute:

$$[\hat{\beta}_1, \hat{\beta}_2] = \frac{i}{2}[\hat{\beta}, \hat{\beta}^\dagger] = 0 \ . \tag{D36}$$

Hence they have a genuine quantum-mechanical joint probability density $p(\beta_1, \beta_2) = p(\beta)$, which is determined by the states of the signal and the image sideband.

Perhaps the easiest way to get at the probability density $p(\beta)$ is through its characteristic function

$$\Phi(\gamma) = \left\langle e^{\gamma\hat{\beta}^\dagger - \gamma^*\hat{\beta}} \right\rangle = \int d^2\beta\, p(\beta) e^{\gamma\beta^* - \gamma^*\beta} . \quad \text{(D37)}$$

The second form here is a Fourier transform in disguise, with inverse

$$p(\beta) = \int \frac{d^2\gamma}{\pi^2} \Phi(\gamma) e^{\beta\gamma^* - \beta^*\gamma} . \quad \text{(D38)}$$

We can now manipulate the characteristic function through the following steps:

$$\begin{aligned}\Phi(\gamma) &= \mathrm{tr}[\hat{\rho}_a D(\hat{a}_+, \gamma)]\langle 0_- | D(\hat{a}_-, -\gamma^*) | 0_- \rangle \\ &= \mathrm{tr}(\hat{\rho}_a e^{-\gamma^*\hat{a}} e^{\gamma\hat{a}^\dagger}) = \mathrm{tr}(e^{\gamma\hat{a}^\dagger} \hat{\rho}_a e^{-\gamma^*\hat{a}}) .\end{aligned} \quad \text{(D39)}$$

The first equality follows from factoring the expectation value in Eq. (D37) into separate expectation values of displacement operators for the signal and the image sideband; the second equality follows from writing the signal displacement operator in antinormal order and the image-sideband displacement operator in normal order [Eq. (5.5)]; and the third equality follows from the cyclic property of the trace. There remaining no explicit reference to the image sideband, we dispense with the + that distinguishes the signal mode in the last two equalities.

By using the coherent-state completeness property (5.13) to perform the trace in the coherent-state basis, we can put the characteristic function in the form

$$\begin{aligned}\Phi(\gamma) &= \int \frac{d^2\beta}{\pi} \langle\beta| e^{\gamma\hat{a}^\dagger} \hat{\rho}_a e^{-\gamma^*\hat{a}} |\beta\rangle \\ &= \int d^2\beta \frac{1}{\pi} \langle\beta|\hat{\rho}_a|\beta\rangle e^{\gamma\beta^* - \gamma^*\beta} .\end{aligned} \quad \text{(D40)}$$

Comparison with Eq. (D37) shows that

$$p(\beta) = \frac{1}{\pi}\langle\beta|\hat{\rho}_a|\beta\rangle = \mathrm{tr}(\hat{\rho}_a \hat{F}_\beta) , \quad \text{(D41)}$$

where $\hat{F}_\beta$ is the coherent-state effect density of Eq. (5.24). The probability density (D41), formed from the diagonal matrix elements of the density operator in the coherent-state basis, is called the Q function in quantum optics (Husimi, 1940).

The Q function $\langle\beta|\hat{\rho}_a|\beta\rangle/\pi$ can never be as sharply peaked as a δ function. Moreover, it cannot be sharply peaked in either β_1 or β_2 separately; this imprecision can be thought of as a consequence of noise that is added during any attempt to measure simultaneously both quadrature components—or both position and momentum (Arthurs and Kelly, 1965; Caves, 1982; Yamamoto and Haus, 1986; Arthurs and Goodman, 1988, 1991; Braunstein *et al.*, 1991; Ozawa, 1991; Stenholm, 1992). The Heisenberg uncertainty principle states that position and momentum obey a fundamental uncertainty product, which prevents both from having zero variance. In a simultaneous measurement of x and p, the situation is even worse: both x and p have finite variance, even if the input quantum state is a position or momentum eigenstate.

Homodyne detection is considerably easier, because the analysis in the caption of Fig. 4 shows that it measures a quadrature component

$$\tfrac{1}{2}(\hat{a}e^{-i\theta} + \hat{a}^\dagger e^{i\theta}) , \quad \text{(D42)}$$

which is determined by the phase θ of the local oscillator. If $\theta = 0$, homodyne detection measures the quadrature component $\hat{x}_1$ of Eq. (5.2).

APPENDIX E: CLAIMS OF INFINITE WIDEBAND CAPACITY

There are several claims (Helstrom, 1974; Yuen *et al.*, 1975; Pierce *et al.*, 1981) of infinite wideband capacity in the literature. It is important to examine these claims critically. Before doing so, however, it is useful to note a difference between communication theorists and physicists—a difference that in our experience can make the channel capacity go to zero when the two groups try to communicate. A communication theorist defines a channel by specifying input and output alphabets and by giving conditional probabilities that characterize channel noise; he or she then puts primary emphasis on deriving rigorous mathematical consequences, which are both important and useful whether or not the channel has a physical realization. A physicist, while acknowledging the mathematical importance of such results, puts primary emphasis on a physical realization, because a physicist is interested in how physical law affects the performance of communication systems.

With this preface we turn to the claims (Helstrom, 1974; Yuen *et al.*, 1975; Pierce *et al.*, 1981) of infinite wideband capacity. From a communication theorist's point of view, these claims are rigorous consequences of assuming a channel where photon energy is independent of frequency. From a physicist's point of view, the neglect of the Einstein relation $E = hf$ means that these information-theoretic analyses are physically applicable only to narrowband channels, in which case it is widely accepted that quantum channels have a finite capacity. Technically, these analyses find infinite wideband capacity because they constrain the average photon intensity, whereas the relevant physical constraint involves energy—i.e., a power constraint—instead of photon number. Beyond just concluding that these infinite-capacity claims are physically irrelevant, however, we can clothe the channel models in physical garb and estimate wideband capacities by letting the bandwidth approach the frequency.

Helstrom (1974) considers a finite transmission time $\mathcal{T}$. He imagines constructing M orthogonal longitudinal modes on the time interval $\mathcal{T}$ and signaling by exciting *one* of these modes into a coherent state that has a mean number N_s of photons, while leaving the rest of the modes in the vacuum state. For $M \gg 1$ Helstrom

finds a capacity very close to $\mathcal{T}^{-1}\log_2 M$. He concludes that "the coherent pure-state channel must have infinite capacity," because M can be made arbitrarily large. He further remarks that "the quantum-mechanical nature of the signals themselves does not ... limit the information-carrying capacity of a coherent optical channel." These conclusions are apparently meant to apply even to narrowband channels.

To evaluate these claims, we give the channel some physical properties. Assume that the channel operates at frequency f within bandwidth B; the input power is then $P = N_s hf/\mathcal{T}$. Since there are $M = B\mathcal{T} = N_s hfB/P$ orthogonal longitudinal modes, Helstrom's capacity becomes

$$C_{\rm H} = \mathcal{T}^{-1}\log_2(B\mathcal{T}) = \frac{P}{N_s hf}\log_2\left(\frac{N_s hfB}{P}\right) . \quad \text{(E1)}$$

Helstrom's channel is just like the particle-like channel discussed in Sec. II.B, except that $N_s h$ replaces h [cf. Eq. (2.22)]. A study of Helstrom's Fig. 1 and the last equation of his paper shows that to approach the capacity (E1) requires $N_s \gg 1$. Thus Helstrom's channel is considerably suboptimal, because one could achieve the capacity (E1) with $N_s = 1$ by using one-photon number states instead of coherent states (cf. the discussion of placement information in Sec. III.A.2).

For narrow bandwidths Helstrom's capacity (E1) is clearly limited. To say that quantum mechanics does not limit the narrowband capacity is misleading, because the channel is intrinsically quantum mechanical: letting $h \to 0$ with B and P fixed violates the assumption that $1 \ll B\mathcal{T} = N_s hfB/P$. Moreover, if the channel is allowed to become wideband so that $M = B\mathcal{T}$ becomes large, then one must take into account the frequency dependence of photon energy. The argument given for the particle-like regime in Sec. II.B shows that the wideband capacity of Helstrom's channel is bounded by $C_H \lesssim \sqrt{P/N_s h}$ [cf. Eq. (2.23)].

Yuen, Kennedy, and Lax (1975) consider a channel similar to Helstrom's and conclude that "reliable communication is possible at an arbitrarily high rate for any given fixed power, if an arbitrarily large amount of bandwidth is available." They note, however, that "the narrowband or constant frequency (f white) condition breaks down for large bandwidth," thus acknowledging that their conclusion does not apply to physical channels.

The most prominent and explicit claim of infinite wideband capacity has been advanced by Pierce, Posner, and Rodemich (1981; referred to as PPR in the following). PPR give a detailed physical description of their channel; to relate their work to ours, we use a mixture of their notation and ours. PPR consider a binary (on-off) Poisson photon channel, which operates at frequency f (with photon energy hf) and which is contaminated by thermal radiation at temperature T. In the "on" state the transmitter emits a pulse of photons that lasts a time $t = B^{-1}$, where B is the bandwidth of the pulse (PPR assume that $Bt = \frac{1}{2}$, which contributes to an error that we correct below). The photon-number distribution in each signal pulse is assumed to be a Poisson distribution with mean $n_s = w_0^2 = u$. In the "off" state the transmitter emits no photons. The "on" state is transmitted with probability α and the "off" state with probability $1-\alpha$. Hence the mean number of signal photons per pulse duration, i.e., per channel use, is $\bar{n} = \alpha n_s$, and the average signal power is $P = hf\bar{p} = hfB\bar{n} = hf\alpha n_s/t$. The mean number of thermal photons per channel use, $\bar{n}_T$, is given by Eq. (3.12); PPR also introduce a dimensionless field quadrature variance $\sigma_0^2 = \frac{1}{2}\bar{n}_T$ [because they assume that $Bt = \frac{1}{2}$, PPR use $\sigma_0^2 = \frac{1}{4}\bar{n}_T$, but the $\frac{1}{2}$ is dictated by their Eq. (29)].

PPR consider both peak-power and average-power constraints. Our interest here is the Poisson channel with an average-power constraint, which PPR treat in their Sec. IV and for which they derive a capacity

$$C_{\rm PPR} = \bar{p}\log_2\left(1+\frac{1}{\bar{n}_T}\right) = \frac{P}{kT\ln 2} \quad \text{(E2)}$$

[PPR's Eqs. (100), (108), and (40)]. The rightmost form of this capacity is an exact consequence of PPR's work, but it agrees with the "approximate" form they give in Eq. (2) instead of with the "exact" form they give in Eq. (3), the reasons being, first, that their Eq. (3) uses the erroneous relation $\sigma_0^2 = \frac{1}{4}\bar{n}_T$ and, second, that it uses an approximate low-temperature form for $\bar{n}_T$ introduced in PPR's Eq. (13). The PPR result (E2) yields infinite capacity *in the zero-temperature limit.* In addition to their derivation, PPR give a heuristic argument to "explain why it is reasonable that the capacity with an average power constraint be asymptotic to [Eq. (E2)] as $T \to 0$." A careful examination of the derivation in PPR's Sec. IV shows, however, that the capacity (E2) emerges in the simultaneous limits $\alpha \to 0$, $n_s \to \infty$, and $t \to 0$ ($B \to \infty$), with $\bar{p} = P/hf = \alpha n_s/t$ held constant and hence with

$$\bar{n} = \alpha n_s = \bar{p}t \to 0 . \quad \text{(E3)}$$

Thus the PPR capacity (E2) must be viewed with considerable caution, because in these limits it is mandatory to take into account the frequency dependence of photon energy, which PPR did not do.

The procedure used by PPR to find the capacity is flawed, because it maximizes with respect to bandwidth B (or pulse duration t) without constraining the bandwidth to be consistent with assumption of frequency-independent photon energy. Technically, the PPR procedure amounts to varying the bandwidth and maximizing the information *per photon.* Another way to treat varying channel bandwidth would be to allow the frequency (photon energy) and the bandwidth to vary at the same time, perhaps with fixed fractional bandwidth, which would lead to a maximization problem like that in Sec. III.A.2.

Despite these difficulties, we can estimate the wideband capacity of PPR's channel, at least in the low-temperature case $\bar{n}_T \ll 1$, which is PPR's primary con-

cern and for which PPR provide sufficient data. Our estimate comes from applying the appropriate bandwidth constraint, $B \lesssim f$, to PPR's work. We are able to show that the capacity of PPR's channel is consistent with the capacity upper bound (1.4); moreover, we find a condition for validity of the PPR capacity (E2) and show that it prohibits taking the zero-temperature limit.

We proceed by noting that the data in PPR's Table II indicate that to approach the capacity (E2) requires that $\alpha \leq \bar{n}_T$. Furthermore, in their Eq. (109) PPR note that

$$C_{\mathrm{PPR}} \leq \frac{H(\alpha)}{t} = \frac{-\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha)}{t}, \tag{E4}$$

which is the sensible statement that the capacity per use, $C_{\mathrm{PPR}}t$, cannot exceed the statistical information $H(\alpha)$ corresponding to the on-off probabilities. In the case of interest, $\alpha \leq \bar{n}_T \ll 1$, Eqs. (E2) and (E4) imply that

$$\bar{n} = \bar{p}t \leq \frac{H(\alpha)}{\log_2(1+\bar{n}_T^{-1})} \simeq \frac{\alpha \log_2 \alpha^{-1}}{\log_2 \bar{n}_T^{-1}} \leq \bar{n}_T . \tag{E5}$$

Although it might seem odd that the PPR capacity requires $\bar{n} \ll 1$ [cf. Eq. (E3)], this is a consequence of using a procedure that maximizes the information *per photon* (see discussion in Sec. III.A.2). Equation (E5) is our condition for the validity of the PPR capacity. Rewriting this validity condition in a form that allows us to impose the bandwidth constraint, we find

$$\bar{n}_T \gtrsim \bar{n} = P/hfB \gtrsim P/hf^2 , \tag{E6}$$

which forbids letting $T \to 0$ unless $P \to 0$ at the same time.

We can use the validity condition $\bar{n} \lesssim \bar{n}_T$ to bound the PPR capacity in the case $\bar{n}_T \ll 1$,

$$\begin{aligned} C_{\mathrm{PPR}} &\simeq \bar{p} \log_2 \bar{n}_T^{-1} = \sqrt{\frac{P}{hf}} \sqrt{B\bar{n}} \log_2 \bar{n}_T^{-1} \\ &\lesssim \sqrt{\frac{P}{h}\frac{B}{f}} \sqrt{\bar{n}_T} \log_2 \bar{n}_T^{-1} \ll \sqrt{\frac{P}{h}\frac{B}{f}} , \end{aligned} \tag{E7}$$

which shows that the capacity is much smaller than the $\sqrt{P/h}$ optimum capacity because of thermal noise. Although it might seem paradoxical that Eq. (E7) gives zero capacity when $\bar{n}_T = 0$, this result follows naturally from our validity condition for the PPR capacity: if $\bar{n}_T = 0$, then $\bar{n} = 0$ as well. We should note that the PPR capacity per use, $C_{\mathrm{PPR}}t$, agrees with the capacity per use of a narrowband number-state channel at temperature T [Eq. (3.18)] in the case $\bar{n} \ll \bar{n}_T \ll 1$, a case that satisfies the validity condition (E6).

Kabanov (1978) and Davis (1980) have derived capacities for Poisson channels using a different, considerably more powerful approach. The signal is encoded in a time-dependent rate λ_t of photon emission, which controls a Poisson process for counting photons at the output. The rate λ_t is bounded above by a maximum rate B. (Davis also includes a background dark-noise contribution to the rate, which we ignore here.) With a constraint on the average power P, which must satisfy $P \leq Bhf$, Kabanov (1978) and Davis (1980) find the channel capacity and show that it is achieved when λ_t takes on only two values, 0 (for the "off" state) and B (for the "on" state). In such a Poisson channel the information is encoded in arrival times of individual photons (pulse position modulation); to distinguish successive photons, the channel bandwidth must be the maximum rate B, as is implied by our notation. For comparison with our work, it is convenient to introduce the number of photons per channel use, $\bar{n} = P/Bhf$, which must satisfy $\bar{n} \leq 1$ for this kind of intrinsically quantum-mechanical channel.

Written in terms of $\bar{n}$, the Poisson channel capacity found by Kabanov and Davis becomes

$$C = \begin{cases} B\bar{n}\log_2 \bar{n}^{-1} , & \bar{n} \leq e^{-1}, \\ B/e\ln 2 , & e^{-1} \leq \bar{n} \leq 1 \end{cases} \tag{E8}$$

(cf. the discussion of particle-like channels in Sec. II.B). The $\bar{n} \ll 1$ limit of this capacity has been given by Stern (1960), Gordon (1962), and Yamamoto and Haus (1986). One can maximize the capacity (E8) further by assuming a fixed fractional bandwidth $\eta = B/f \ll 1$, as in Sec. III.A.2. The resulting maximum,

$$C_{\max} = \frac{2}{e^2 \ln 2} B = \frac{2}{e \ln 2}\sqrt{\frac{\eta P}{h}} \text{ bits/s}, \tag{E9}$$

is achieved at $\bar{n} = e^{-2}$, corresponding to frequency $f = e\sqrt{P/\eta h}$ and to $C_{\max}/B\bar{n} = 2/\ln 2 = 2.8854$ bits per photon. Comparison with Eq. (3.27) shows that the capacity of a narrowband Poisson channel is smaller than the capacity of a narrowband number-state channel by the factor $1/e\ln 2 \simeq 0.53$. Setting $\eta = 1$ yields an estimate for the capacity of a wideband Poisson channel, which lies below the upper bound C_{WB} of Eq. (1.4).

REFERENCES

Abramowitz, M., and I. A. Stegun, 1964, Eds., *Handbook of Mathematical Functions* (U.S. GPO, Washington, DC), Sec. 24.2.1.

Arthurs, E., and M. S. Goodman, 1988, Phys. Rev. Lett. **60**, 2447.

Arthurs, E., and M. S. Goodman, 1991, in *Quantum Aspects of Optical Communications*, Lecture Notes in Physics Vol. 378, edited by C. Bendjaballah, O. Hirota, and S. Reynaud (Springer-Verlag, Berlin), p. 18.

Arthurs, E., and J. L. Kelly, Jr., 1965, Bell Syst. Tech. J. **44**, 725.

Balian, R., 1991, *From Microphysics to Macrophysics: Methods and Applications of Statistical Physics*, Vol. 1 (Springer-Verlag, Berlin), Chap. 3.

Bednorz, J. G., and K. A. Müller, 1986, Z. Phys. B **64**, 189.

Bekenstein, J. D., 1988, Phys. Rev. A **37**, 3437.

Bekenstein, J. D., and M. Schiffer, 1990, Int. J. Mod. Phys. C **1**, 355.
Bowen, J. I., 1967, IEEE Trans. Inf. Theory **IT–13**, 230.
Bozovic, I., D. Kirillov, A. Kapitulnik, K. Char, M. R. Hahn, M. R. Beasley, T. H. Geballe, Y. H. Kim, and A. J. Heeger, 1987, Phys. Rev. Lett. **59**, 2219.
Braunstein, S. L., 1990, Phys. Rev. A **42**, 474.
Braunstein, S. L., C. M. Caves, and G. J. Milburn, 1991, Phys. Rev. A **43**, 1153.
Caves, C. M., 1981, Phys. Rev. D **23**, 1693.
Caves, C. M., 1982, Phys. Rev. D **26**, 1817.
Caves, C. M., and B. L. Schumaker, 1985, Phys. Rev. A **31**, 3068.
Caves, C. M., and B. L. Schumaker, 1986, in *Quantum Optics IV,* edited by J. D. Harvey and D. F. Walls (Springer-Verlag, Berlin), p. 20.
Chu, C. W., P. H. Hor, R. L. Meng, L. Gao, Z. J. Huang, and Y. Q. Wang, 1987, Phys. Rev. Lett. **58**, 405.
Collett, M. J., R. Loudon, and C. W. Gardiner, 1987, J. Mod. Opt. **34**, 881.
Davies, E. B., 1976, *Quantum Theory of Open Systems* (Academic, London).
Davis, M. H. A., 1980, IEEE Trans. Inf. Theory **IT–26**, 710.
Drummond, P. D., 1987, Phys. Rev. A **35**, 4253.
Drummond, P. D., 1988, Phys. Scr. **T21**, 70.
Gallager, R. G., 1968, *Information Theory and Reliable Communication* (Wiley, New York).
Glauber, R. J., 1963a, Phys. Rev. Lett. **10**, 84.
Glauber, R. J., 1963b, Phys. Rev. **131**, 2766.
Glauber, R. J., 1965, in *Quantum Optics and Quantum Electronics,* edited by C. DeWitt, A. Blandin, and C. Cohen-Tannoudji (Gordon and Breach, New York), p. 63.
Gordon, J. P., 1961, in *Advances in Quantum Electronics,* edited by J. R. Singer (Columbia University, New York), p. 509.
Gordon, J. P., 1962, Proc. IRE **50**, 1898.
Gordon, J. P., 1964, in *Quantum Electronics and Coherent Light,* edited by P. A. Miles (Academic, New York), p. 156.
Hall, M. J. W., 1993, "Gaussian noise and quantum optical communication," Australian National University preprint.
Hall, M. J. W., and M. J. O'Rourke, 1993, Quantum Opt. **5**, 161.
Hardy, G. H., and S. Ramanujan, 1918, Proc. London Math. Soc. Ser. 2 **17**, 75.
Heidmann, A., R. J. Horowicz, S. Reynaud, E. Giacobino, C. Fabre, and G. Camy, 1987, Phys. Rev. Lett. **59**, 2555.
Helstrom, C. W., 1974, Proc. IEEE (Lett.) **62**, 139.
Helstrom, C. W., J. W. S. Liu, and J. P. Gordon, 1970, Proc. IEEE **58**, 1578.
Holevo (Kholevo), A. S., 1973, Probl. Pereda. Inf. **9**(3), 3 [Probl. Inf. Trans. **9**, 177 (1973)].
Holevo, A. S., 1982, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam).
Hollenhorst, J. N., 1979, Phys. Rev. D **19**, 1669.
Hughston, L. P., R. Jozsa, and W. K. Wootters, 1993, Phys. Lett. A **183**, 14.
Husimi, K., 1940, Proc. Phys. Math. Soc. Jpn. **22**, 264.
Jozsa, R., D. Robb, and W. K. Wootters, 1994, Phys. Rev. A **49**, 668.
Kabanov, Yu. M., 1978, Theory Probab. Appl. **23**, 143.
Kimble, H. J., and D. F. Walls, 1987, Eds., *Squeezed States of the Electromagnetic Field,* special issue of J. Opt. Soc. Am. B **4**, 1453.
Kraus, K., 1983, *States, Effects, and Operations: Fundamental Notions of Quantum Theory* (Springer-Verlag, Berlin).
Landauer, R., 1987, Appl. Phys. Lett. **51**, 2056.
Landauer, R., 1988, Nature **335**, 779.
Landauer, R., 1989, in *Selected Topics in Signal Processing,* edited by S. Haykin (Prentice-Hall, Englewood Cliffs, NJ), p. 18.
Landauer, R., and J. W. F. Woo, 1973, in *Synergetics,* edited by H. Haken (Tuebner, Stuttgart), p. 97.
Lebedev, D. S., and L. B. Levitin, 1963, Dokl. Akad. Nauk SSSR **149**, 1299 [Sov. Phys. Dokl. **8**, 377 (1963)].
Lebedev, D. S., and L. B. Levitin, 1966, Inf. Control **9**, 1.
Levitin, L. B., 1969, in *Proceedings of the Fourth All-Union Conference on Information and Coding Theory,* Sec. II (Tashkent).
Levitin, L. B., 1982, Int. J. Theor. Phys. **21**, 299.
Levitin, L. B., 1991, in *Quantum Aspects of Optical Communications,* Lecture Notes in Physics Vol. 378, edited by C. Bendjaballah, O. Hirota, and S. Reynaud (Springer-Verlag, Berlin), p. 101.
Levitin, L. B., 1993, in *Workshop on Physics and Computation: PhysComp '92,* edited by D. Matzke (IEEE Computer Society, Los Alamitos, CA), p. 215.
Loudon, R., and P. L. Knight, 1987, J. Mod. Opt. **34**, 709.
Lu, E. Y. C., 1971, Nuovo Cimento **2**, 1241.
Lu, E. Y. C., 1972, Nuovo Cimento **3**, 585.
Machida, S., Y. Yamamoto, and Y. Itaya, 1987, Phys. Rev. Lett. **58**, 1000.
Marko, H., 1965, Kybernetik **2**, 274.
Movshovich, R., B. Yurke, P. G. Kaminsky, A. D. Smith, A. H. Silver, R. W. Simon, and M. V. Schneider, 1990, Phys. Rev. Lett. **65**, 1419.
Ohya, M., and H. Suyari, 1991, in *Quantum Aspects of Optical Communications,* Lecture Notes in Physics Vol. 378, edited by C. Bendjaballah, O. Hirota, and S. Reynaud (Springer-Verlag, Berlin), p. 203.
Ozawa, M., 1984, J. Math. Phys. **25**, 79.
Ozawa, M., 1991, in *Quantum Aspects of Optical Communications,* Lecture Notes in Physics Vol. 378, edited by C. Bendjaballah, O. Hirota, and S. Reynaud (Springer-Verlag, Berlin), p. 3.
Pendry, J. B., 1983, J. Phys. A **16**, 2161.
Peres, A., 1990, Found. Phys. **20**, 1441.
Peres, A., 1993, *Quantum Theory: Concepts and Methods* (Kluwer, New York).
Pierce, J. R., 1978, IEEE Trans. Commun. **COM–26**, 1819.
Pierce, J. R., E. C. Posner, and E. R. Rodemich, 1981, IEEE Trans. Inf. Theory **IT–27**, 61.
Rademacher, H., 1937, Proc. London Math. Soc. Ser. 2 **43**, 241.
Reynaud, S., C. Fabre, and E. Giacabino, 1987, J. Opt. Soc. Am. B **4**, 1520.
Saleh, B. E. A., and M. C. Teich, 1987, Phys. Rev. Lett. **58**, 2656.
Saleh, B. E. A., and M. C. Teich, 1992, Proc. IEEE **80**, 451.
Schiffer, M., 1991, Phys. Rev. A **43**, 5337.
Schiffer, M., 1992, in *Computation: The Micro and Macro View,* edited by B. A. Huberman (World Scientific, Singapore), p. 1.
Schlesinger, Z., R. T. Collins, D. L. Kaiser, and F. Holtzberg, 1987, Phys. Rev. Lett. **59**, 1958.
Schrieffer, J. R., 1983, *Theory of Superconductivity* (Benjamin, Reading, PA).
Schumaker, B. L., 1984, Opt. Lett. **9**, 189.
Schumaker, B. L., 1986, Phys. Rep. **135**, 317.

Schumaker, B. L., and C. M. Caves, 1985, Phys. Rev. A **31**, 3093.
Shannon, C. E., 1948, Bell Syst. Tech. J. **27**, 379 (Part I); **27**, 623 (Part II). Reprinted in book form, with postscript by W. Weaver: Shannon, C. E., and W. Weaver, 1949, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, IL).
Shapiro, J. H., 1985, IEEE J. Quantum Electron. **QE–21**, 237.
Shapiro, J. H., and S. S. Wagner, 1984, IEEE J. Quantum Electron. **QE–20**, 803.
Shapiro, J. H., H. P. Yuen, and J. A. Machado Mata, 1979, IEEE Trans. Inf. Theory **IT–25**, 179.
She, C. Y., 1968, IEEE Trans. Inf. Theory **IT–14**, 32.
Slusher, R. E., L. W. Hollberg, B. Yurke, J. C. Mertz, and J. F. Valley, 1985, Phys. Rev. Lett. **55**, 2409.
Slusher, R. E., and B. Yurke, 1990, J. Lightwave Technol. **8**, 466.
Sobolewski, R., D. P. Butler, T. Y. Hsiang, C. V. Stancampiano, and G. A. Mourou, 1986, Phys. Rev. B **33**, 4604.
Stenholm, S., 1992, Ann. Phys. (N.Y.) **218**, 233.
Stern, T. E., 1960, IRE Trans. Inf. Theory **IT–6**, 435.
Stoler, D., 1970, Phys. Rev. D **1**, 3217.
Stoler, D., 1971, Phys. Rev. D **4**, 1925.
Sudarshan, E. C. G., 1963, Phys. Rev. Lett. **10**, 277.
Takahasi, H., 1965, in *Advances in Communications Systems, Vol. 1*, edited by A. V. Balakrishnan (Academic, New York), p. 227.
Teich, M. C., and B. E. A. Saleh, 1989, Quantum Opt. **1**, 153.
Teich, M. C., and B. E. A. Saleh, 1990, Phys. Today **43**(6), 26.
Yamamoto, Y., 1990, Trans. Inst. Electron. Inf. Commun. Eng. E **73**, 1598.
Yamamoto, Y., and H. A. Haus, 1986, Rev. Mod. Phys. **58**, 1001.
Yamamoto, Y., S. Machida, and O. Nilsson, 1991, in *Coherence, Amplification and Quantum Effects in Semiconductor Lasers*, edited by Y. Yamamoto (Wiley, New York), p. 461.
Yamamoto, Y., S. Machida, S. Saito, N. Imoto, T. Yanagawa, M. Kitagawa, and G. Björk, 1990, Prog. Opt. **28**, 87.
Yuen, H. P., 1976a, Phys. Lett. A **56**, 105.
Yuen, H. P., 1976b, Phys. Rev. A **13**, 2226.
Yuen, H. P., and V. W. S. Chan, 1983, Opt. Lett. **8**, 177.
Yuen, H. P., R. S. Kennedy, and M. Lax, 1975, IEEE Trans. Inf. Theory **IT–21**, 125.
Yuen, H. P., and M. Ozawa, 1992, Phys. Rev. Lett. **70**, 363.
Yuen, H. P., and J. H. Shapiro, 1978, IEEE Trans. Inf. Theory **IT–24**, 657.
Yuen, H. P., and J. H. Shapiro, 1980, IEEE Trans. Inf. Theory **IT–26**, 78.
Yurke, B., 1985, Phys. Rev. A **32**, 300, 311.
Zaheer, K., and M. S. Zubairy, 1991, Adv. At. Mol. Phys. **28**, 143.