

Fundamental limits upon the measurement of state vectors

K. R. W. Jones*

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

(Received 20 May 1994)

Using the Shannon information theory and the Bayesian methodology for inverting quantum data [K.R.W. Jones, *Ann. Phys. (N.Y.)* **207**, 140 (1991)] we prove a fundamental bound upon the measurability of finite-dimensional quantum states. To do so we imagine a thought experiment for the quantum communication of a pure state ψ , known to one experimenter, to his colleague via the transmission of N identical copies of it in the limit of zero temperature. Initial information available to the second experimenter is merely that of the allowed manifold of superpositions upon which the chosen ψ may lie. Her efforts to determine it, in an optimal way, subject to the fundamental constraints imposed by quantum noise, define a statistical uncertainty principle. This limits the accuracy with which ψ can be measured according to the number N of transmitted copies. The general result is illustrated in the physically realizable case of polarized photons.

PACS number(s): 03.65.Bz

I. QUANTUM UNCERTAINTY

In this paper we examine the intrinsic limits to the measurement of quantum states that are imposed by quantum uncertainty. To appreciate the open nature of this question, we must first examine the Heisenberg uncertainty principle [1,2]

$$\langle(\Delta\hat{A})^2\rangle\langle(\Delta\hat{B})^2\rangle \geq \frac{1}{4}|\langle[\hat{A}, \hat{B}]\rangle|^2 + \frac{1}{4}|\langle[\hat{A}, \hat{B}]^+\rangle|^2, \quad (1)$$

where $[\hat{A}, \hat{B}]^+$ denotes the anticommutator. The inequality fixes a fundamental limit upon the joint dispersion of noncommuting operators, as a consequence of the geometry of Hilbert space via the Schwarz inequality [3]. It bounds the joint dispersions in $\langle\hat{A}\rangle$ and $\langle\hat{B}\rangle$, but it places no restriction upon the measurement of ψ itself. Obviously, if that were *precisely known* then (1) is satisfied *automatically* [4].

The reason for this situation is that dispersion quantifies the statistical error for a *single measurement*. If we were given many *identically prepared* copies of a quantum state then we might (1) repeat measurements to improve the error in the mean; and (2) perform alternate measurements to obtain data for incompatible quantities. In practice, this is how methods like optical homodyne tomography (OHT) [5] can overcome the apparently insurmountable obstruction of quantum noise to the precise determination of ψ .

If we are to identify those restrictions which apply to the new class of experiments then we must seek the companion limit to (1). With one copy of a state the measurement should be poor, with many the statistics are improved. Our task must be to describe the accumulation and reduction of these data to account for the un-

certainty in ψ . Evidently, whatever answer we give is contingent upon three elements of choice: (1) the possible states; (2) the ensemble size N ; and (3) the kind of measurements we perform. Unlike the standard *algebraic* method of analysis [6–11], the challenge faced is to account for the quantum fluctuations, using a statistical theory [12] (an approach pioneered especially by Helstrom [13], Holevo [14], Wootters [15,16], and Wootters and Fields [17]). The object is to base all analysis upon the intrinsic, unavoidable, and classifiable nature of quantum noise, by exploiting the known, and constrained rule [18] for computing quantum probabilities [19].

Here we employ the Bayesian formalism of quantum inference [19–21], which unifies the methods of communication theory [22] with quantum measurement [23,24]. This approach treats the limits on state measurement within the paradigm of Shannon's analysis of physical communication channels, just as in the early quantum communication theory studies of Helstrom [13] and Holevo [14]. However, the question posed here lies closer to physics than engineering. It parallels those considered by Jaynes [25], Wootters [15], Jones [19–21], Braunstein and Caves [26], Peres and Wootters [27], Garrett [28], Dukes and Larson [29,30], and Larson [31]. These authors have all applied Bayesian methodology [32–35] to the analysis of quantum noise, and the information gained from quantum measurements.

The limit derived here is expressed as a channel capacity [36] for the transmission of information by N identical copies of a quantum state ψ . The use of this engineering measure is easily motivated by viewing quantum state determination from the standpoint of an experimenter equipped with a universal preparing device. His task is to communicate an arbitrary choice of ψ to a colleague equipped with a universal analyzing device which must measure the state starting from conditions of complete ignorance.

The first experimenter is thus using quantized information carriers to communicate the setting of his preparing

*Electronic address: jonesk@kelvin.physics.uq.oz.au.

device, which is a ψ . Unlike in the more usual classical problem, wherein in the limits upon the measurement of a microstate vanish in the limit of zero temperature, the quantum statistics remain *constrained* by the geometry of the physical Hilbert space of pure states. This distinguishing aspect of quantum measurement sets a fundamental geometrically defined limit analogous to the Heisenberg relation. Its expression is a bound upon the maximum information obtainable from N quantum measurements.

In outline: in Sec. II we formulate the general problem to be solved, and the basic tools required; in Sec. III we review the Bayesian theory of optimal measurement; in Sec. IV we prove a theorem, previously a conjecture, that is the fundamental limit, and our key result; in Sec. V we compute the upper limit imposed by quantum noise on various illustrative Hilbert spaces; and, finally, in Sec. VI we illustrate the limits for the simplest case of a two-level system.

II. GENERAL STATEMENT OF THE PROBLEM

To formulate fundamental limits we must account for the complete generality offered by the quantum formalism, while acknowledging the limitations of experimental practice.

The general perspective aids formulation of a tractable mathematical problem. Once we imagine the ideal situation we will discover that the quantum theory imposes a well-defined and universal fundamental limit. Its origins are geometrical, just as with the Heisenberg uncertainty principle. However, in pursuing a general limit one must ignore the practical obstructions to its realization. This aspect of the problem forces us to separate discussion of the finite-dimensional Hilbert space from its infinite-dimensional cousin.

The reason for this division is that the level of generality assumed here is *certainly not* attainable for infinite-dimensional systems. Analysis of this case must account for experimental limitations, which are closely tied to the physical system studied.

For example, the recent work in optical homodyne tomography [5] proves the practical measurability of states for infinite-dimensional systems. However, it is a rare workable scheme among many imaginable possibilities, where most are impracticable. A general bound is unlikely to be useful, and it seems best to devote a special treatment to this method. This is not true of the finite-dimensional case, where there are a variety of realizable examples of complete state preparability and analysis. Among them we mention the case of polarized light, and the numerous studies upon multipole spin measurements [7–9]. Thus the present limit is sought with the aim of emphasizing what may be expected in the best case scenario of finite-state quantized information carriers. The enterprise is made worthwhile since, as we will see, the ideal performance is approachable, in practice.

Further, in order to emphasize the key role played by the Hilbert space dimension we will study only the zero-temperature limit of *pure preparations*. In this way,

quantum noise effects are isolated from the thermodynamic fluctuations. Since the latter must always degrade performance, and since they form the *only* classical constraint upon measurement, we feel it is best to separate the one from the other [37].

Within the general scheme of an arbitrary finite-dimensional system, we illustrate the nature of physical restrictions upon state preparability and analysis via a simple model based upon symmetry considerations [38]. The object is to show how the general constraints are set by the dimensionality of the Hilbert space, but that these may be degraded due to an incompletely accessible manifold of superpositions.

For example, in the physical realization of state-preparation or analysis schemes there are oftentimes certain standard, or easy options. For instance, the vacuum, a ground state, linearly polarized light, or a coherent state are readily preparable. Similarly, pass filters such as a linear polarizer or Stern-Gerlach device are easy to implement [39].

In attempting to widen the options one might, and often does, exploit the freedom to impose a dynamical transformation \hat{U} to the ingoing or outgoing state ψ . Then the quantum mechanical statistical correlation

$$p(\phi_j|\psi) = |\langle\phi_j|\psi\rangle|^2,$$

describing complete measurement in the realizable eigenbasis $\{|\phi_j\rangle\}_{j=1}^d$, becomes

$$p(\phi_j|\hat{U}\psi) = |\langle\phi_j|\hat{U}\psi\rangle|^2 = |\langle\hat{U}^\dagger\phi_j|\psi\rangle|^2 = p(\hat{U}^\dagger\phi_j|\psi),$$

which amounts to using the oppositely transformed eigenbasis $\{\hat{U}^\dagger|\phi_j\rangle\}_{j=1}^d$ to measure ψ .

This links the constraint of realizable measurement schemes to the subject of realizable physical symmetries. To illustrate the general idea, and to explore methods for analysis, in the simplest possible general classification, we will study the symmetry of time-reversal invariance [40] as a model test case. Apart from its simplicity, this choice is further motivated by its obvious connection with random matrix theory [41–45] (which is based upon time-reversal symmetries). The methods developed for computations in state inference are also useful in studies of the eigenvector statistics in quantum chaos [46–48]. The example is both illuminating and relevant to other topical studies.

A. The state determination gedanken experiment

Consider two experimenters, Norman and Noeline, who are skilled in the preparation of states ψ and the conduct of arbitrary quantum measurements upon these, see Fig. 1. Since we deal with the zero-temperature limit we will assume ψ is pure.

Norman inhabits one cool chamber where he has constructed a universal preparing device, denoted \mathcal{P} . It has a dial upon it, wired through an isolating enclosure, to the instrument itself. Upon his dial he may select a ψ , labeling the state of a physical information carrier, such as a particle with internal states, drawn from a Hilbert

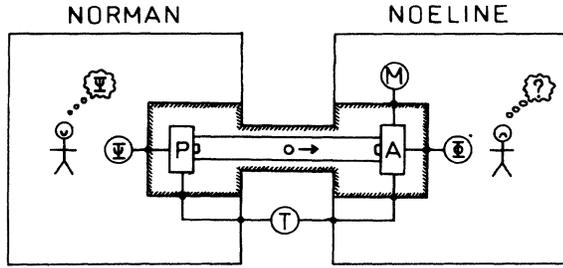


FIG. 1. Norman sets his preparing device \mathcal{P} to generate N identical copies of the state ψ . It could be the state vector for the internal degrees of freedom of some transmissible particle, such as a photon with two polarization states. Noeline, in receipt of minimal prior information about the kind of particle, and timing information T , must infer the selected ψ on the basis of N classical data items Φ using her analyzing device \mathcal{A} . She does so by setting various measurements \mathcal{M} , and then using the known statistical rule to infer the unknown state from her data. The state is assumed to be robustly transmissible, so that the noise temperature is zero.

space \mathcal{H} . Upon receipt of a GO clock pulse, Norman's preparator issues N identical copies of ψ through a portal.

In a neighboring chamber stands Noeline, who is equipped with a universal analyzing device, denoted \mathcal{A} . Her instrument is similarly isolated. It has one dial to set a quantum measurement, and another to read its result. Noeline is in receipt of clock pulses from Norman, so that they may synchronize their state preparation and analysis routine.

On receipt of GO Noeline implements N possibly different measurements upon the N copies of ψ that Norman has sent her. From her N items of classical data she must formulate an inference for the unknown ψ . Norman tells her nothing but the nature of the carrier and details of its Hilbert space. After N transfers, Norman issues a STOP command, and they meet to compare data, and judge Noeline's chosen scheme.

This N -trial experiment, repeated many times, tests the efficacy of some setting \mathcal{A}_N of Noeline's analyzer when faced with the preparation \mathcal{P}_N . Since Norman tells her nothing about the initial ψ , Noeline must account for this in her choice of \mathcal{A}_N . Her aim is to fix a choice that works best, on average, whatever \mathcal{P}_N it may please Norman to dial up. Our goal is to set a performance bound on the best that Noeline can do, based upon an analysis of the geometrical properties of the allowable quantum statistical rules [18].

This model formalizes a generic thought experiment that underlies previous studies in quantum state determination [6–11,15,17,19–21,27,29,30], the distinguishability of states [16], and limits on multiple measurement parameter estimation [13,49,50].

B. Description of the allowed states

Noeline knows the dimension d , and that the state ψ is pure, but she has no initial information about its orien-

tation. Her analyzer expresses its results with respect to some reference basis of complete states $\{|\phi_j\rangle\}_{j=1}^d$. However, in general she can do no more than assume the superposition

$$|\psi\rangle = |\phi_j\rangle\langle\phi_j|\psi\rangle, \quad (2)$$

with the d unknown complex quantities $\psi_j = \langle\phi_j|\psi\rangle$, that are normalized to unity. Removing the leading phase freedom and imposing normalization, Noeline must specify $2(d-1)$ real numbers with infinite precision to pin down ψ .

However, in order to illustrate the importance of symmetry restrictions, we will suppose that Norman and Noeline might be restricted to construct their states and measurements out of unitary transformations \hat{U} , acting upon standards. In view of our remarks, on the equivalence of state and basis changes, we need only consider these as applied to Norman.

Fix, therefore, a standard state ψ_0 , and suppose that Norman constructs all of his ψ via the rule $|\psi\rangle = \hat{U}|\psi_0\rangle$, for \hat{U} a unitary operator drawn from some subgroup \mathcal{G} of the full group $U(d)$, see Fig. 2. As is well known in the theory of generalized coherent states [51], this manifold is isomorphic to a coset space of the group \mathcal{G} [52].

In this finite-dimensional example there exists a subgroup \mathcal{K} , the maximal isotropy subgroup, defined as the largest subgroup of \mathcal{G} that leaves ψ_0 invariant:

$$\hat{U}|\psi_0\rangle = |\psi_0\rangle,$$

for all $\hat{U} \in \mathcal{K}$. The resulting manifold of accessible states ψ is then isomorphic to the coset \mathcal{G}/\mathcal{K} . In the case $\mathcal{G} = U(d)$, the maximal isotropy subgroup is $\mathcal{K} = U(d-1)$, and so the accessible manifold is

$$\mathcal{M} = U(1) \otimes \mathbf{PC}^{d-1} \approx U(d)/U(d-1),$$

the complex projective space of all rays, with a phase factor adjoined, which is just the full Hilbert space of normalizable states (our computations will average out

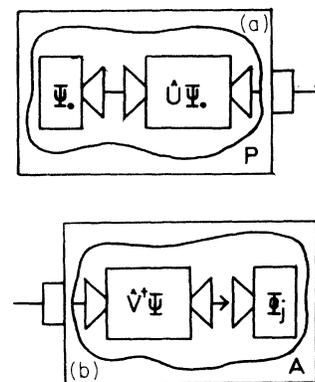


FIG. 2. Schematic cutaway diagram of Norman's universal preparing device (a); and Noeline's universal analyzing device (b). In both cases unitary mappings are applied to a standard state ψ_0 , or a standard measurement $\{\phi_j\}_{j=1}^d$. The choice of \hat{U} and \hat{V} is assumed open to control.

the phase).

More generally, a restriction upon the Hamiltonians \hat{H} must limit the accessible states. For instance, suppose that \hat{U} is generated by integrating

$$i\hbar \frac{d}{dt} |\psi\rangle = \hat{H} |\psi\rangle,$$

so that a class of \hat{H} acting over a finite time is the basis for Norman's universal preparator.

As an elementary and easily classifiable example of a physical symmetry that limits the chosen \hat{H} , and thus \hat{U} , we study time-reversal invariance.

The time-reversal operator \hat{T} , if it exists, is defined [40] as that which connects forward and backward evolving states, i.e., it must satisfy

$$\hat{T} |\psi(t)\rangle = |\psi(-t)\rangle.$$

Physical constraints dictate that $\hat{T}^2 = \pm 1$, or that it is absent entirely.

To such a symmetry there corresponds the constraint $[\hat{H}, \hat{T}] = 0$. Further, as familiar in the Kramers degeneracy [53], and at the root of random matrix theory [42–45], there lies the simple classification (Dyson called it the threefold way [41]):

$$\hat{T}^2 = +1 : \mathcal{G} = O(d), \quad (3)$$

$$\hat{T}^2 = -1 : \mathcal{G} = \text{USp}(2d); \quad (4)$$

or simply $U(d)$, if \hat{T} does not exist. Here $O(d) \subset U(d)$ is the orthogonal group, and $\text{USp}(2d) \subset U(2d)$ is the unitary symplectic group, which may be viewed as a complex representation of the d -dimensional unitary group on quaternionic numbers [54].

The resulting coset spaces are classified as (see the discussion of Gilmore [55])

$$O(d)/O(d-1) = \{\pm 1\} \otimes \mathbf{PR}^{d-1}, \quad (5)$$

$$U(d)/U(d-1) = U(1) \otimes \mathbf{PC}^{d-1}, \quad (6)$$

$$\text{USp}(2d)/U(2(d-1)) = \text{USp}(2) \otimes \mathbf{PH}^{d-1}, \quad (7)$$

which are, up to the relevant “phase factors,” the projective spaces corresponding to Hilbert spaces of dimension d over the real numbers \mathbf{R} , the complex numbers \mathbf{C} , or the quaternions \mathbf{H} .

Condensing these formal considerations to a simple statement, we are, when confined by time-reversal symmetry, restricted to the Hilbert spaces

$$\mathbf{F}^d = \mathbf{R}^d, \mathbf{C}^d, \text{ or } \mathbf{H}^d,$$

for which a general element may be written in the form

$$|\psi\rangle = \sum_{j=1}^d \sum_{k=1}^{\nu} |j\rangle x_{jk} \mathbf{e}_k, \quad (8)$$

where the x_{jk} are νd real numbers normalized to unity; the kets $|j\rangle$ are orthonormal; and the basis vectors \mathbf{e}_k are present to cater for the quaternionic case. Explicitly, we have $\mathbf{e}_1 = \mathbf{1}$, $\mathbf{e}_2 = \mathbf{i}$, $\mathbf{e}_3 = \mathbf{j}$, and $\mathbf{e}_4 = \mathbf{k}$, where

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1,$$

and $\mathbf{ij} = \mathbf{k}$ etc., as per the Pauli matrix algebra.

If we fix d , and choose $\nu = 1, 2$, or 4 , to select \mathbf{R}^d , \mathbf{C}^d , and \mathbf{H}^d , respectively, then one obtains a simple example of the restrictions imposed by a dynamical symmetry. In this comparison, one must consider \mathbf{H}^d to be a special case of \mathbf{C}^{2d} .

C. Description of the allowed measurements

The formulation of quantum limits, without need of thermodynamic considerations, would be impossible were it not for the fact that quantum statistical rules are generically inner products. The necessary restriction follows from Gleason's theorem [18], as developed in the theory of probability operator measures (POM's) due to Davies and Lewis [23], and applied by Helstrom [13] and Holevo [14] in the quantum decision and estimation theory.

One associates to the classical readings upon a physical measuring device a collection of M positive operators \hat{A}_k , which are subject to the closure constraint $\sum_{k=1}^M \hat{A}_k = \hat{1}$, but which need not be an orthonormal set of projectors. They may well be an overcomplete resolution of the identity such as in the theory of coherent states [56].

The important element, for our discussion, is that these give a mathematical handle upon the most general possible quantum measurement scheme. The probability of observing the classical datum which corresponds to the POM element \hat{A}_k given the state $\hat{\rho}$ is

$$p(\hat{A}_k | \hat{\rho}) = \text{Tr}[\hat{A}_k \hat{\rho}]. \quad (9)$$

Such generalized measurements are now commonly implemented in quantum optics using heterodyne and homodyne detection schemes [57].

Now we may describe Noeline's N -trial apparatus as the tensor product of N POM's $\mathcal{A}_{\text{POM}_k}$

$$\mathcal{A}_N \equiv \mathcal{A}_{\text{POM}_1} \otimes \cdots \otimes \mathcal{A}_{\text{POM}_N}, \quad (10)$$

describing the setting of her measuring instrument and the data as the tensor product

$$\Phi_N = \hat{A}_1 \otimes \cdots \otimes \hat{A}_N, \quad (11)$$

of N positive Hermitian operators $\hat{A}_k \in \mathcal{A}_{\text{POM}_k}$. Summation over the data Φ_N will involve use of the N closure relations: $\sum_{\hat{A}_k} \hat{A}_k = \hat{1}$, one for each separate POM.

D. Description of the prior knowledge

In classical physics the basic underlying measure of complete ignorance of a physical state is the invariant Liouville measure in classical phase space [58]. This is determined by the constraint of invariance under the canonical transformations realized by Hamilton's equations. It may be viewed as the natural invariant measure associated with the classical symmetry group of symplectic motions; and is the basis for statistical mechanics [59].

In direct analogy, we seek a measure upon the contin-

uous manifold of states ψ , denoted $d\hat{\Omega}_{\bar{\psi}}$, whose defining property is invariance under quantum canonical transformations. Since these are unitary symmetries, on the rays $\bar{\psi} \equiv |\psi\rangle\langle\psi|$, we demand that

$$d\hat{\Omega}_{\bar{\psi}} = d\hat{\Omega}_{\hat{U}\bar{\psi}\hat{U}^\dagger} \quad \text{and} \quad \int_{\mathcal{H}} d\hat{\Omega}_{\bar{\psi}} = 1, \quad (12)$$

for all $\hat{U} \in U(d)$.

The existence and uniqueness of such invariant measures has been discussed previously by Jones [19]. The theory of group invariant measures, called Haar or Hurwitz integration [60], establishes general results in this area. For the compact groups considered here [61], the measure on $U(d; \mathbf{F})$ is unique (up to a constant factor), and induces a unique invariant measure upon the coset space $U(d; \mathbf{F})/U(d-1; \mathbf{F})$, to fix $d\hat{\Omega}_{\bar{\psi}}$ [cf. the use of group invariant measures in classical statistical mechanics [59] (ergodic measure), random matrix theory [42] (unitary invariant matrix ensembles), the theory of geometric probability [62] (the measure of geometrical figures, and their intersections), and in the lattice gauge theory of Wilson [63] (the implementation of gauge invariance on a lattice via the gauge-group invariant sum over links)]. An explicit group theoretic construction (see Gilmore [64]) yields

$$d\hat{\Omega}_{\bar{\psi}} = 2^{-1}\Gamma(\nu d/2)\delta(1 - \sqrt{\langle\psi|\bar{\psi}\rangle}) d\psi d\bar{\psi},$$

where $d\psi d\bar{\psi} \equiv \prod_{j=1}^d \prod_{k=1}^{\nu} \pi^{-1/2} dx_{jk}$, $\langle\psi|\psi\rangle = \sum_{j=1}^d \sum_{k=1}^{\nu} x_{jk}^2$, and the x_{jk} range over $(-\infty, +\infty)$. In practice, we remove the square root using the identity

$$\begin{aligned} \delta(1 - \langle\psi|\bar{\psi}\rangle) &= \delta((1 - \sqrt{\langle\psi|\bar{\psi}\rangle})(1 + \sqrt{\langle\psi|\bar{\psi}\rangle})) \\ &= \delta(2(1 - \sqrt{\langle\psi|\bar{\psi}\rangle})) = 2^{-1}\delta(1 - \sqrt{\langle\psi|\bar{\psi}\rangle}), \end{aligned}$$

and take the expression

$$\int p(\psi, \bar{\psi}) d\hat{\Omega}_{\bar{\psi}} = \Gamma(\nu d/2) \int p(\psi, \bar{\psi}) \delta(1 - \langle\psi|\bar{\psi}\rangle) d\psi d\bar{\psi} \quad (13)$$

as the definition of Hilbert space integration. The final normalization factor $\pi^{\nu d/2}/\Gamma(\nu d/2)$ is then one-half of the surface area of a unit radius $\nu d/2$ -dimensional hypersphere.

Further simplifications are possible using the methods of Riemann-Liouville fractional integration, see Jones [65,66], and Weyl fractional integration, see Davies and Jones [67]. The result is a family of reduction formulas that enable complicated multidimensional integrals to be computed via a one-dimensional integral, e.g., the result [66]

$$\begin{aligned} \int f(\langle\psi|\hat{P}|\psi\rangle) d\hat{\Omega}_{\bar{\psi}} &= \frac{\Gamma(\nu d/2)}{\Gamma(\nu l/2)\Gamma(\nu(d-l)/2)} \\ &\times \int_0^1 f(w)(1-w)^{\nu(d-l)/2-1} \\ &\times w^{\nu l/2-1} dw, \end{aligned} \quad (14)$$

where f is an arbitrary function, subject to convergence of the integral, and \hat{P} is an arbitrary projector of rank l in a Hilbert space of dimension d , of type index ν .

To illustrate, consider the probability density

$$p_{\hat{P}}^n(\psi) = \mathcal{N}(\nu, d)^{-1} \left(|\langle\psi|\phi\rangle|^2 \right)^n, \quad (15)$$

which is uniform at $n = 0$, and approaches a δ function centered upon \hat{P} as $n \rightarrow \infty$. The normalization is computed from (14), and the beta function integral representation [68]

$$\int_0^1 (1-w)^{\mu-1} w^{\nu-1} dw = \frac{\Gamma(\mu)\Gamma(\nu)}{\Gamma(\mu+\nu)}. \quad (16)$$

Setting $f(w) = w^n$ and $l = 1$, we obtain

$$\mathcal{N}(\nu, d) = \frac{\Gamma(\nu d/2)\Gamma(\nu/2+n)}{\Gamma(\nu/2)\Gamma(\nu d/2+n)}, \quad (17)$$

which may be checked against the usual δ -function computations [45].

E. Description of the posterior knowledge

At the end of one N -trial experiment, Noeline has registered the data Φ_N , consisting of N positive Hermitian operators \hat{A}_k , expressed with respect to the reference basis $\{|\phi_j\rangle\}_{j=1}^d$. Her statistical analysis must proceed from the quantum likelihood function, defined as

$$p(\Phi_N = \{\hat{A}_1, \dots, \hat{A}_N\}|\psi) = \prod_{k=1}^N \langle\psi|\hat{A}_k|\psi\rangle. \quad (18)$$

Upon constructing this function she may compute the probability that any chosen ψ underlies the observed data Φ_N . Some ψ are more likely than others, in the sense of a greater value for $p(\Phi_N|\psi)$. The approach we will take is to invert this to obtain $p(\psi|\Phi_N)$.

Apart from its intuitive simplicity, the benefit of this method, when combined with information theory, is that the non-Euclidean nature of the underlying state manifold rarely intrudes upon analysis. The noncommuting aspects of the quantum statistics are relegated to the measure $d\hat{\Omega}_{\bar{\psi}}$. Although an explicit noncommuting statistical analysis can be done, such as with the quantum mechanical Cramér-Rao lower bound developed by Helstrom [69], it is easier to choose methods that are not tied to Euclidean assumptions [70].

III. THE BAYESIAN THEORY OF OPTIMAL MEASUREMENT

Bayesian methods of statistical analysis are simple in concept, but difficult to apply in practice due to the problem of assigning prior distributions. As noted by Jones [19], quantum theory is special in that two outstanding historical difficulties are catered for; namely, the prob-

lem of prior selection, and the constraint of the statistical rule. Gleason's theorem [18] tells us the rule; while the principle of invariant prior selection, due to Jeffreys [71], Jaynes [72], and Hartigan [73] fixes the Laplacian notion of complete ignorance via the demand of a unitary invariant prior upon the quantum state manifold. This idea has been applied by Jones [19] to the general treatment of pure ensembles; and has been extended by Larson and Dukes [30] to enable the general treatment of impure ensembles also.

A. Quantum Bayes inversion

The theorem of Bayes asserts that [74]

$$p(A|B)p(B) = p(A, B) = p(B|A)p(A). \quad (19)$$

Mathematically, it is no more than a self-consistency constraint. However, in the theory of statistical inference it can be applied to the problem of data inversion [75].

Suppose $p(B|A)$ is the statistical correlation between data B and the unknown A . It is a characteristic of the experimental procedure employed. If we now specify a *prior* distribution $p_0(A)$, then one has the consequent *posterior* distribution [75]

$$p(A|B) = \frac{p(B|A)p_0(A)}{\sum_A p(B|A)p_0(A)} \quad (20)$$

as the basic rule of Bayesian inversion. It gives an inference for A , expressed as a probability density describing ignorance, fixed by the observed data B . As shown by Jones [19], these principles apply directly to quantum mechanics as soon as we have resolved the problem of how to integrate over the manifold of all ψ .

Using the results of Secs. IID and IIE, the joint distribution of quantum data and the unknown state reads as

$$p(\psi, \Phi_N) \equiv p(\Phi_N|\psi)p_0(\psi). \quad (21)$$

Here $p_0(\psi)$ is arbitrary, although we choose $p_0(\psi) = 1$ in the sequel. The quantum Bayes inversion procedure is now

$$p(\psi|\Phi_N) = \frac{p(\Phi_N|\psi)p_0(\psi)}{\int p(\Phi_N|\psi)p_0(\psi) d\hat{\Omega}_{\vec{\psi}}}, \quad (22)$$

yielding a normalized probability density $p(\psi|\Phi_N)$ parametrized by the observed data Φ_N .

In the quantum application $p(\Phi_N|\psi)$ is given by expression (18), and we choose the invariant prior $p_0(\psi) = 1$ to formulate, in a rigorous physical manner, the statistical idea of complete ignorance. This emphasizes the utility of quantum symmetry principles [19,30] in overcoming the ambiguities of the Bayesian paradigm [71–76]. It also shows how an arbitrary collection of N positive Hermitian operators defines a probability density over ψ . This, in turn, fixes a density matrix via the average

$$\hat{\rho}(\Phi_N) \equiv \int |\psi\rangle\langle\psi| p(\psi|\Phi_N) d\hat{\Omega}_{\vec{\psi}}, \quad (23)$$

which is the final Bayesian inversion of the data Φ_N consistent with $p_0(\psi)$.

The consistency is verified by considering a subsequent measurement. The probability of the result \hat{A} , given Φ_N , averaged over $p_0(\psi)$, may be computed directly from

$$p(\Phi_{N+1} = \Phi_N \otimes \hat{A}) \equiv \int \langle\psi|\hat{A}|\psi\rangle p(\Phi_N|\psi)p_0(\psi) d\hat{\Omega}_{\vec{\psi}},$$

or via the indirect route

$$\begin{aligned} p(\Phi_N \otimes \hat{A}) &= p(\hat{A}|\hat{\rho}(\Phi_N))p(\hat{\rho}(\Phi_N)) \\ &= \text{Tr}[\hat{A}\hat{\rho}(\Phi_N)]p(\hat{\rho}(\Phi_N)). \end{aligned}$$

The inversion procedure (23) ensures that both results are equal, i.e.,

$$\text{Tr}[\hat{A}\hat{\rho}(\Phi_N)]p(\hat{\rho}(\Phi_N)) = \int p(\Phi_N \otimes \hat{A}|\psi)p_0(\psi) d\hat{\Omega}_{\vec{\psi}}.$$

This property expresses a chainlike “rule of succession,” wherein new data update the density matrix in a self-consistent manner. It is *special* to the quantum statistical rule, and does not hold in the general setting of Bayesian inference. Finally, since $p(\psi|\Phi_N)$ is normalized, and positive, the matrix $\hat{\rho}(\Phi_N)$ will always be a physical density matrix.

B. The Shannon measure of mutual information

To formulate a criterion for optimal measurement we employ the Shannon information theory. From this viewpoint, Noeline's N -trial measuring device constitutes a possible design for an optimal receiver that decodes the N -state symbols issued from Norman's transmitter. Exactly as in the communication theory application, the optimality statement depends on the choice of $p_0(\psi)$, showing that the “optimal” measuring device is a concept defined relative to the available initial information.

Using the Shannon measure of information [22]

$$I(p_1, \dots, p_M) = \sum_{j=1}^M p_j \ln p_j,$$

we denote the prior information by I_0 , and the posterior information conditioned by the observed data Φ_N as $I_N(\Phi_N)$. Explicitly, these read

$$I_0 = + \int p_0(\psi) \ln p_0(\psi) d\hat{\Omega}_{\vec{\psi}}, \quad (24)$$

$$I_N(\Phi_N) = + \int p(\psi|\Phi_N) \ln p(\psi|\Phi_N) d\hat{\Omega}_{\vec{\psi}}, \quad (25)$$

where natural logarithms set the units as nats (base 2 for bits). Averaging over the possible data gives the expected information gain (mutual information)

$$\{\psi, \Phi_N\} = \sum_{\Phi_N} p(\Phi_N) I_N(\Phi_N) - I_0, \quad (26)$$

interpreted as the expected “decrease in uncertainty,”

due to the data Φ_N . As shown by Everett [77], this measures the statistical dependence between “data” and “states.” A good measuring device produces a strong “data-state” correlation.

The above quantity, called the discrete quantum channel capacity [36], measures the information that can be sent per N -state symbol in an encoding where the possible signal alphabet comprises all possible ψ . In measurement theory, the optimal \mathcal{A}_N is that which maximizes the expected information content of the data Φ_N . Both quantities are measured by the mutual information (26), showing the close connection between measurement and communication.

C. The von Neumann measure of purity

Another useful measure of performance is given by (23). Since density matrices comprise a convex set [91], with the pure states as extremal points, it is clear that $\hat{\rho}(\Phi_N)$ can only be pure if $p(\psi|\Phi_N)$ is concentrated upon a single point. This would be a *perfect* measurement.

Introducing the von Neumann entropy [78]

$$S[\hat{\rho}] = -\text{Tr}[\hat{\rho} \ln \hat{\rho}] \quad (27)$$

as a measure of purity, with $S[\hat{\rho}] \in [0, d]$, we define the quantity

$$\langle S[\hat{\rho}(\Phi_N)] \rangle = \sum_{\Phi_N} p(\Phi_N) S[\hat{\rho}(\Phi_N)] \quad (28)$$

as a comparative measure of performance. One is then asserting that the best instrument is that which minimizes the expected entropy of the inferred density matrix.

For the quantum invariant prior the initial density matrix is computable via symmetry considerations:

$$\begin{aligned} \hat{\rho}(0) &= \int |\psi\rangle\langle\psi| d\hat{\Omega}_{\hat{\psi}} = \int |\psi\rangle\langle\psi| d\hat{\Omega}_{\hat{U}\hat{\psi}\hat{U}^\dagger} \\ &= \int \hat{U}^\dagger |\psi\rangle\langle\psi| \hat{U} d\hat{\Omega}_{\hat{\psi}} = d^{-1} \hat{1}. \end{aligned}$$

Here we have exploited the unitary invariance of the measure to show that the final result must be invariant under all \hat{U} . Application of Schur’s lemma [79] shows the result must be a multiple of the identity, fixed by demanding a unit trace. Therefore, whereas the Shannon measure *increases* from zero at $N = 0$ and is unbounded, the von Neumann entropy *decreases* from $\ln d$ at $N = 0$, to vanish in the limit $N \rightarrow \infty$.

D. The fundamental variational problem

A formal statement of the quantum limits to uncertainty in states can now be given via the variational problem [80]:

$$\{\psi, \Phi_N\} \leq \max_{\mathcal{A}_N} \sum_{\Phi_N} p(\Phi_N) \int p(\psi|\Phi_N) \ln p(\psi|\Phi_N) d\hat{\Omega}_{\hat{\psi}}, \quad (29)$$

$$\langle S[\hat{\rho}] \rangle \geq \min_{\mathcal{A}_N} \sum_{\Phi_N} p(\Phi_N) S[\hat{\rho}(\Phi_N)], \quad (30)$$

where \mathcal{A}_N denotes the choice of N POM’s. Unlike with the usual uncertainty principle, we must now find an upper (lower) bound that is N dependent, via the optimization of \mathcal{A}_N .

To make the variational problems (29) and (29) more tractable, we convert them into the weaker bounding inequalities

$$\{\psi, \Phi_N\}[\mathcal{A}_N] \leq \max_{\Phi_N} \int p(\psi|\Phi_N) \ln p(\psi|\Phi_N) d\hat{\Omega}_{\hat{\psi}}, \quad (31)$$

$$\langle S[\hat{\rho}] \rangle \geq \min_{\Phi_N} S[\hat{\rho}(\Phi_N)]. \quad (32)$$

This amounts to the simple statement that the performance of any apparatus \mathcal{A}_N is limited by its best possible output. Since Φ_N is simply a choice of N positive Hermitian operators this is a much easier problem to solve.

IV. THEOREM OF THE ULTIMATE LIMIT

In this section we prove that the solution to the inequalities (31) and (32) is provided by the inferred density [cf. the computation at (15)]

$$p_{\max}^N(\psi|\phi \otimes \cdots \otimes \phi) = \frac{\Gamma(\nu/2)\Gamma(\nu d/2 + N)}{\Gamma(\nu d/2)\Gamma(\nu/2 + N)} (|\langle\psi|\phi\rangle|^2)^N, \quad (33)$$

where the choice of ϕ does not matter. This is the inferred density of maximal possible information. It represents an ultimate limit upon performance set by the best possible output of an N -trial instrument. The analysis proves an earlier conjecture due to Jones [80], and extends it to a general method of proof, in the style of earlier related results due to Holevo [81], Davies [82], and Bendjaballah and Charbit [83] on the optimal receiver problem in quantum communication theory.

A. Elements of convex analysis

Consider the general N -trial inferred density

$$p(\psi|\hat{A}_1 \otimes \cdots \otimes \hat{A}_N) = \frac{\prod_{k=1}^N \langle\psi|\hat{A}_k|\psi\rangle p_0(\psi)}{\int \prod_{k=1}^N \langle\psi|\hat{A}_k|\psi\rangle p_0(\psi) d\hat{\Omega}_{\hat{\psi}}}. \quad (34)$$

Whatever the chosen \mathcal{A}_N , even if it were an adaptive scheme where $\mathcal{A}_{\text{POM}_k}$ is determined by the previous data $\Phi_{k-1} = \hat{A}_1 \otimes \cdots \otimes \hat{A}_{k-1}$, the posterior density has this form. We will show that there is a maximum (minimum) for (31) [(32)] which is attained when all \hat{A}_k are equal to some rank-one projector \hat{P} .

The proof employs convex analysis [84], and thus covers a wide variety of optimality criteria [85]. First, we note that the set \mathcal{C} of all non-negative unit-trace Hermitian matrices comprise a convex set. If $\hat{A}, \hat{B} \in \mathcal{C}$ then $\lambda\hat{A} + (1 - \lambda)\hat{B} \in \mathcal{C}$, for all $\lambda \in [0, 1]$. Further, the ex-

tremal points of \mathcal{C} are the rank-one projectors, i.e., they must satisfy $\hat{P}^2 = \hat{P}$, as well as $\text{Tr}[\hat{P}] = 1$. Next, we assume a class of probability densities \mathcal{D} , parametrized by the points of \mathcal{C} , where

$$p(\lambda\hat{A} + (1-\lambda)\hat{B}) = \tilde{\lambda}(\lambda)p(\hat{A}) + [1 - \tilde{\lambda}(\lambda)]p(\hat{B}), \quad (35)$$

with $\tilde{\lambda}(\lambda) \in [0, 1]$ for all $\lambda \in [0, 1]$, and where $\tilde{\lambda}(1) = 1$, and $\tilde{\lambda}(0) = 0$. This technical requirement ensures that any convex functional defined on \mathcal{D} can be viewed as a convex function defined on \mathcal{C} . Upon the domain \mathcal{D} of probability densities p , we introduce a real-valued optimality functional $\mathcal{F}[\mathcal{D}]$ such that

$$\mathcal{F}[\lambda p + (1-\lambda)q] \leq \lambda\mathcal{F}[p] + (1-\lambda)\mathcal{F}[q], \quad (36)$$

which is called convex; or, with the opposing inequality, concave. Using (35) we have

$$\mathcal{F}[p(\lambda\hat{A} + (1-\lambda)\hat{B})]$$

$$\leq \tilde{\lambda}(\lambda)\mathcal{F}[p(\hat{A})] + [1 - \tilde{\lambda}(\lambda)]\mathcal{F}[p(\hat{B})]. \quad (37)$$

Obviously, any operator \hat{A} which can be decomposed as a convex combination leads to an improvement by choosing the maximum element of the right-hand side. Thus the maximum is attained on the boundary of \mathcal{C} , and so equals $\mathcal{F}[p(\hat{P})]$, for *some* rank-one projector \hat{P} .

$$p(\psi|\lambda\hat{A}'_1 + (1-\lambda)\hat{B}'_1 \otimes \dots \otimes \hat{A}'_N) = \tilde{\lambda}(\lambda)p(\psi|\hat{A}'_1 \otimes \dots \otimes \hat{A}'_N) + [1 - \tilde{\lambda}(\lambda)]p(\psi|\hat{B}'_1 \otimes \dots \otimes \hat{A}'_N), \quad (38)$$

where $\tilde{\lambda}$ reads

$$\tilde{\lambda}(\lambda) \equiv \frac{\lambda p(\hat{A}'_1 \otimes \dots \otimes \hat{A}'_N)}{\lambda p(\hat{A}'_1 \otimes \dots \otimes \hat{A}'_N) + (1-\lambda)p(\hat{B}'_1 \otimes \dots \otimes \hat{A}'_N)}, \quad (40)$$

which satisfies the requirements $\tilde{\lambda} \in [0, 1]$, with $\tilde{\lambda}(1) = 1$, and $\tilde{\lambda}(0) = 0$.

Step 4: Form the symmetrized tensor product $\mathcal{C}_N = \mathcal{C}_1 \otimes \dots \otimes \mathcal{C}_1$, with general elements

$$\hat{\mathcal{A}}_N = \frac{1}{N!} \sum_P \hat{A}'_{P(1)} \otimes \dots \otimes \hat{A}'_{P(N)}, \quad (41)$$

where $P(k)$ signifies the k th element of one of $N!$ permutations, which we sum over.

Step 5: Observe that \mathcal{C}_N is a convex domain, and that its elements $\hat{\mathcal{A}}_N$ are positive and of unit trace. Moreover, the formula (33) is permutation invariant. In particular,

$$p(\psi|\hat{\mathcal{A}}_N) = p(\psi|\hat{A}'_{P(1)} \otimes \dots \otimes \hat{A}'_{P(N)}), \quad (42)$$

for an arbitrary permutation P .

Step 6: Combining now step 3 and step 6 we see that for a convex \mathcal{F} one has

B. Proof of the general result

Theorem: For an arbitrary convex (concave) functional \mathcal{F} defined on the space of all N -trial inferred densities (33), the maximum (minimum) is attained when $\hat{A}_k = |\phi\rangle\langle\phi|$ for all $k \in [1, N]$, for *some* chosen ϕ [arbitrary when $p_0(\psi) = 1$].

Proof: It is a straightforward result dominated by the technical issue of introducing a convex domain \mathcal{C}_N defined as a symmetrized N -fold tensor product of the 1-trial \mathcal{C}_1 .

Step 1: Observe that (34) is invariant under the trace rescaling $\hat{A}_k \mapsto \hat{A}'_k = \hat{A}_k / \text{Tr}[\hat{A}_k]$, so that we may consider the inferred densities to be functionals defined on the convex domain \mathcal{C}_1 of non-negative unit-trace Hermitian matrices.

Step 2: Observe that the normalization

$$p(\hat{A}'_1 \otimes \dots \otimes \hat{A}'_N) = \int \prod_{k=1}^N \langle \psi | \hat{A}_k | \psi \rangle p_0(\psi) d\hat{\Omega}_{\psi} \quad (38)$$

is linear in each operatorial slot. For instance,

$$p(\lambda\hat{A}'_1 + (1-\lambda)\hat{B}'_1 \otimes \dots \otimes \hat{A}'_N)$$

$$= \lambda p(\hat{A}'_1 \otimes \dots \otimes \hat{A}'_N) + (1-\lambda)p(\hat{B}'_1 \otimes \dots \otimes \hat{A}'_N),$$

and similarly in the other slots.

Step 3: Observe, therefore, that

$$\mathcal{F}[p(\psi|\lambda\hat{\mathcal{A}}_N + (1-\lambda)\hat{\mathcal{B}}_N)]$$

$$\leq \tilde{\lambda}(\lambda)\mathcal{F}[p(\psi|\hat{\mathcal{A}}_N)] + [1 - \tilde{\lambda}(\lambda)]\mathcal{F}[p(\psi|\hat{\mathcal{B}}_N)], \quad (43)$$

where from (40) we find that

$$\tilde{\lambda}(\lambda) \equiv \frac{\lambda p(\hat{\mathcal{A}}_N)}{\lambda p(\hat{\mathcal{A}}_N) + (1-\lambda)p(\hat{\mathcal{B}}_N)}, \quad (44)$$

and similarly for \mathcal{F} concave.

Therefore the extremal values are attained for those $\hat{\mathcal{A}}_N$ which are rank-one projectors. In virtue of the symmetrization carried out at step 4, these are of the generic form

$$\hat{\mathcal{A}}_N = \hat{P} \otimes \dots \otimes \hat{P}, \quad (45)$$

i.e., all \hat{A}_k are equal to some rank-one projector, and the theorem is established. Obviously, the choice of \hat{P} is arbitrary when \mathcal{F} is independent of it, which follows if $p_0(\psi) = 1$.

C. Application to the Shannon information

The function $p \ln p$ is pointwise convex, so its integral, $I[p(\psi|\hat{\mathcal{A}}_N)]$, is a convex functional of the function p , de-

finned on the convex domain \mathcal{C}_N . Thus (31) is maximized by (33).

D. Application to the von Neumann entropy

The function $S[\hat{\rho}] = -\text{Tr}[\hat{\rho} \ln \hat{\rho}]$ is concave on all $\hat{\rho}[p(\psi|\hat{\mathcal{A}}_N)]$, and we have

$$\begin{aligned} \hat{\rho}[p(\psi|\lambda\hat{\mathcal{A}}_N + (1-\lambda)\hat{\mathcal{B}}_N)] \\ = \tilde{\lambda}(\lambda)\hat{\rho}[p(\psi|\hat{\mathcal{A}}_N)] + [1-\tilde{\lambda}(\lambda)]\hat{\rho}[p(\psi|\hat{\mathcal{B}}_N)], \end{aligned}$$

with $\tilde{\lambda}(\lambda)$ as in (44). Thus $S[p(\psi|\hat{\mathcal{A}}_N)]$ is concave on \mathcal{C}_N and (31) is minimized by (33).

V. THE INFERENCE UNCERTAINTY PRINCIPLE

To obtain an information theoretic statement of the inferential uncertainty principle we compute the Shannon information, and von Neumann entropy of the distribution (33).

A. Upper bound for the Shannon information

From the N -dependent normalization factor

$$\mathcal{N}(N) = \frac{\Gamma(\nu d/2)\Gamma(\nu/2 + N)}{\Gamma(\nu/2)\Gamma(\nu d/2 + N)},$$

as per (15), the information of (33) can be evaluated directly as

$$\begin{aligned} \Delta\mathcal{I}_{\max}(N) &= \int \frac{1}{\mathcal{N}(N)} (|\langle\psi|\phi\rangle|^2)^N \\ &\quad \times \ln \left[\frac{1}{\mathcal{N}(N)} (|\langle\psi|\phi\rangle|^2)^N \right] d\hat{\Omega}_{\tilde{\psi}} \\ &= \ln \frac{1}{\mathcal{N}(N)} + \frac{N}{\mathcal{N}(N)} \frac{d\mathcal{N}(N)}{dN}, \end{aligned} \tag{46}$$

where we have used the integral generating function

$$\begin{aligned} \int f(|\langle\psi|\omega\rangle|^2) \ln f(|\langle\psi|\omega\rangle|^2) d\hat{\Omega}_{\tilde{\omega}} \\ = \frac{d}{d\alpha} \int [f(|\langle\psi|\omega\rangle|^2)]^\alpha d\hat{\Omega}_{\tilde{\omega}} \Big|_{\alpha=1}. \end{aligned} \tag{47}$$

Using the fact that $\Psi(z) = \Gamma'(z)/\Gamma(z)$, where $\Psi(z)$ is the digamma function [68], we find

$$\begin{aligned} \{\psi, \Phi_N\} \leq \ln \left[\frac{\Gamma(\nu/2)\Gamma(\nu d/2 + N)}{\Gamma(\nu d/2)\Gamma(\nu/2 + N)} \right] \\ + N \{ \Psi(\nu/2 + N) - \Psi(\nu d/2 + N) \} \end{aligned} \tag{48}$$

as the exact upper bound for information obtainable from N identical copies of a d -state system in a Hilbert space of type index $\nu = 1, 2$, or 4.

This formula is made more transparent by using the asymptotic result [68]

$$\begin{aligned} \frac{\Gamma(z+a)}{\Gamma(z+b)} \sim z^{b-a} \{ 1 + (2z)^{-1}(a-b)(a+b-1) \\ + O(z^{-2}) \} \end{aligned} \tag{49}$$

for $|z| \rightarrow \infty$, to substitute

$$\begin{aligned} \frac{\Gamma(N + \nu d/2)}{\Gamma(N + \nu/2)} \stackrel{N \rightarrow \infty}{\sim} N^{-\nu(d-1)/2} \\ \times \left\{ 1 - (2N)^{-1}[\nu(d-1)/2] \right. \\ \left. \times [\nu(d+1)/2 - 1] + O(N^{-2}) \right\}. \end{aligned} \tag{50}$$

Dropping all those terms which vanish as $N \rightarrow \infty$ we find

$$\begin{aligned} \{\psi, \Phi_N\} \stackrel{N \rightarrow \infty}{\leq} \frac{\nu(d-1)}{2} \ln N - \frac{\nu(d-1)}{2} \\ - \ln \left[\frac{\Gamma(\nu d/2)}{\Gamma(\nu/2)} \right] + O(1/N). \end{aligned} \tag{51}$$

Although this is no longer a strict inequality, since the approach to the asymptote is from above, it does provide a bound upon the possible asymptotics.

As $N \rightarrow \infty$ the information diverges as one should expect, since precise determination of a quantum state requires an infinite precision in the specification of its components. As shown later, $\{\psi, \Phi_N\}/\ln 2$ estimates the total bit-string length required to specify the $\nu(d-1)$ numbers that fix a pure state.

To compare across the three classes of system with time-reversal invariance we introduce $\alpha(\nu, d) = \nu(d-1)$. Comparing \mathbf{R}^d with \mathbf{C}^d we compute $\alpha(1, d) = \alpha(2, d)/2$, so that, when $\ln N$ dominates, about half of the requisite information is missing. This is readily understood when we recall that a superposition with no phase information specified (i.e., real coefficients) leaves $(d-1)$ undetermined parameters, among the total $2(d-1)$ required.

Comparing \mathbf{C}^{2d} with \mathbf{H}^d we find $\alpha(4, d) = \alpha(2, 2d) - 2$. In this case the degradation of information is far less severe, as only two freedoms remain unspecified. This follows since the generic pure-state quaternionic vector is always an undetermined complex superposition of two kets $|1\rangle$ and $|2\rangle$. The possible complete measurements correspond to the eigenvectors of $2d \times 2d$ Hermitian matrices with d different eigenvalues, appearing in pairs. The above condition indicates that the relative phases and amplitudes among the $2d$ basis vectors can be determined except for one remaining pair, which leaves one phase, and one relative amplitude free, hence two unspecified parameters.

In all cases, the restrictions offered by incomplete measurement and preparability reduce the information gain, as one must expect. Further, given full preparability and measurement, it is clear that all systems having a common Hilbert space dimension d enjoy the same ba-

sic limit, although their physical preparability may differ greatly, and thus degrade performance. To know the dimension d constrains the ultimate performance.

B. Lower bound for the von Neumann entropy

To compute the density matrix

$$\hat{\rho}_{\min}(N) \equiv \frac{1}{\mathcal{N}(N)} \int |\psi\rangle\langle\psi| (|\langle\psi|\phi\rangle|^2)^N d\hat{\Omega}_{\vec{\psi}}, \quad (52)$$

we observe that unitary invariance of the measure $d\hat{\Omega}_{\vec{\psi}}$ implies that the matrix $\hat{\rho}_{\min}(N)$ is invariant under the maximal isotropy group of ϕ , so that

$$\hat{\rho}_{\min}(N) = \alpha(N)\hat{1} + \beta(N)|\phi\rangle\langle\phi|, \quad (53)$$

with $S[\hat{\rho}] = -(d-1)\alpha \ln \alpha - (\alpha + \beta) \ln(\alpha + \beta)$, where the unknown α and β are determined by taking two independent traces:

$$\text{Tr}[\hat{\rho}_{\min}(N)] = 1 = \alpha d + \beta, \quad (54)$$

$$\text{Tr}[\hat{\rho}_{\min}(N)|\phi\rangle\langle\phi|] = x = \alpha + \beta. \quad (55)$$

Here x is computable from (52) as

$$\begin{aligned} x &\equiv \frac{1}{\mathcal{N}(N)} \int (|\langle\psi|\phi\rangle|^2)^{N+1} d\hat{\Omega}_{\vec{\psi}} \\ &= \frac{\mathcal{N}(N+1)}{\mathcal{N}(N)} = \frac{\nu/2 + N}{\nu d/2 + N}, \end{aligned} \quad (56)$$

where we have used the identity $\Gamma(z+1) = z\Gamma(z)$. Solving now (54) and (55):

$$\alpha(N) = \frac{1}{d-1} \left\{ 1 - \frac{\mathcal{N}(N+1)}{\mathcal{N}(N)} \right\}, \quad (57)$$

$$\beta(N) = \frac{1}{d-1} \left\{ d \frac{\mathcal{N}(N+1)}{\mathcal{N}(N)} - 1 \right\}, \quad (58)$$

so that the inferred density matrix of minimum entropy reads

$$\hat{\rho}_{\min}(N)[\phi] = \frac{\nu/2}{\nu d/2 + N} \hat{1} + \frac{N}{\nu d/2 + N} |\phi\rangle\langle\phi|, \quad (59)$$

which yields the entropic bound

$$\begin{aligned} \langle S[\hat{\rho}(\Phi_N)] \rangle &\geq \left(\frac{\nu(d-1)/2}{\nu d/2 + N} \right) \ln \left(\frac{\nu d/2 + N}{\nu/2} \right) \\ &\quad + \left(\frac{\nu/2 + N}{\nu d/2 + N} \right) \ln \left(\frac{\nu d/2 + N}{\nu/2 + N} \right). \end{aligned} \quad (60)$$

Asymptotically, this reads

$$\begin{aligned} \langle S[\hat{\rho}(\Phi_N)] \rangle &\stackrel{N \rightarrow \infty}{\geq} \frac{\nu(d-1)}{2N} \ln N - \frac{\nu(d-1)}{2N} \ln(\nu/2) \\ &\quad + \frac{\nu(d-1)}{2N} + O\left(\frac{1}{N^2}\right), \end{aligned} \quad (61)$$

which vanishes as $\nu(d-1)(\ln N)/2N$ in the large N limit.

Comparing (51) and (61) we obtain the interesting relation

$$S_{\min}(N) \sim \Delta \mathcal{I}_{\max}(N)/N \quad (62)$$

showing that the von Neumann measure of purity is approximately the *information gained per state* in the limit of large N [86].

C. Universal scaling and a geometrical interpretation

We can write (51) in the general form [87]

$$\Delta \mathcal{I}(N) \sim \alpha \ln N + \beta, \quad (63)$$

where α and β are constants. To interpret them we consider classical measurement with the Gaussian instrument

$$p_1(x|y) = \frac{1}{\sigma_1 \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{y-x}{\sigma_1} \right)^2 \right\}, \quad (64)$$

where σ_1 is the standard error for a single trial. Repeated measurement, which classical physics allows to be non-disturbing [19], yields the N -trial rule

$$p_N(\bar{x}|y) = \frac{1}{\sigma_N \sqrt{2\pi}} \exp \left\{ -\frac{N}{2} \left(\frac{y-\bar{x}}{\sigma_1} \right)^2 \right\}, \quad (65)$$

where $\bar{x} = N^{-1} \sum_{k=1}^N x_k$, and $\sigma_N = \sigma_1/\sqrt{N}$. Since $\mathcal{I} = \ln(1/\sigma) - [\ln(2\pi) + 1]/2$, the relative gain on passing to N trials is

$$\mathcal{I}_N - \mathcal{I}_1 = \ln \left(\frac{\sigma_1}{\sigma_N} \right) = \frac{1}{2} \ln N. \quad (66)$$

Accuracy scales as \sqrt{N} , a result typical of the asymptotic behavior of measurements whose correlations obey the central limit theorem [88].

The prior information of an infinite region is infinite. However, one may adopt the strategy of box normalization and suppose that the measured variable is restricted to the interval $[-L/2, +L/2]$. Then we set $\mathcal{I}_0 = \ln(1/L)$, and so

$$\Delta \mathcal{I}(N) = \mathcal{I}_N - \mathcal{I}_0 = \ln(L/\sigma_N) - [\ln(2\pi) + 1]/2. \quad (67)$$

Comparing to (63) we have $\alpha = 1/2$, a scaling; and $\beta = \ln(L/\sigma_1 e^{1/2} \sqrt{2\pi})$, a volume factor which counts the number of nonoverlapping cells of size $\sigma_1 e^{1/2} \sqrt{2\pi}$ that will fit into the interval $[-L/2, -L/2]$. The accuracy measure based upon bits or nats is thus, up to an $O(1)$ constant, fixed by the desired confidence level, the bit-string length required to label each of the N -trial distinguishable cells (cf. Wootters and Fields [17], and Caves [89]).

Applying this to (51), the real dimension of the pure-state manifold is $\nu(d-1)$, which explains the value $\alpha = \nu(d-1)/2$. A mesh of cells, each of dimension $\nu(d-1)$, narrows by \sqrt{N} upon a side. The intercept

$\beta = \ln[\Gamma(\nu/2)/\Gamma(\nu d/2)] - \nu(d-1)/2$ counts the “number of distinguishable cells” for an effective one-trial instrument with the given asymptotic law. This term arbitrates optimality. For example, if two instruments have the same α but different β , we set $\Delta\mathcal{I}(N_1) = \Delta\mathcal{I}(N_2)$, and find that

$$\frac{N_1}{N_2} = \exp \left\{ \frac{\beta_2 - \beta_1}{\alpha} \right\},$$

so that instruments with the more positive β are better. They require fewer systems, which means their discretized outputs have more cells. The gain is simply a constant-ratio effect; e.g., 100 vs 105 systems, for equal information.

Thus one can see how an informational measure of bits gained is an arbiter of accuracy. The fineness of the mesh imposed upon state space yields an estimate of the number of bits one needs to specify which discretized cell the pure state lies in. The upper bound (48) obtained from the best possible N -trial quantum posterior distribution (33) sets the smallest possible quantum distinguishable cell for any instrument, and the optimal one must mimic this output as closely as possible [90].

VI. ILLUSTRATION FOR TWO-STATE SYSTEMS

The analysis of optimal measurement for two-state systems [15,19–21,29,30] provides a simple illustration of the basic principle at work. In this case there are numerous options for experiment, such as spin measurement, or observations upon a beam of perfectly polarized photons [15]. Thus it is feasible to verify the progress toward a precise measurement as $N \rightarrow \infty$. Here we illustrate the fundamental bound by showing how it is approached by an isotropic analyzer [19,20] that samples all possible states equally.

The analysis is simplest if we employ the Poincaré sphere representation of two-state systems [91]. An arbitrary pure state is then written

$$|\hat{\mathbf{a}}\rangle\langle\hat{\mathbf{a}}| = \frac{1}{2}(\hat{1} + a_i\sigma_i), \quad (68)$$

where the a_i are the three components of a unit vector $\hat{\mathbf{a}}$, and the σ_i are the Pauli matrices. The transition probability between two kets now becomes

$$p(\hat{\mathbf{a}}|\hat{\mathbf{b}}) = |\langle\hat{\mathbf{a}}|\hat{\mathbf{b}}\rangle|^2 = \frac{1}{2}(1 + \hat{\mathbf{a}} \cdot \hat{\mathbf{b}}). \quad (69)$$

Orthogonal states are represented by antipodal points $\pm\hat{\mathbf{a}}$, and the uniform integration is that over solid angle, written $d\hat{\Omega}_{\hat{\mathbf{r}}} \equiv 1/4\pi \sin\theta d\phi d\theta$.

In the case of polarized light, this is the Stokes sphere parametrization of a classical electromagnetic (e.m.) wave [92]. In quantum mechanics, the photons are spin 1, but massless, so we have a two-state helicity 1 system [93]. The north and south poles of the sphere may be taken as left and right helicity eigenstates, corresponding to left and right circularly polarized light. The sphere is covered upon taking all possible superpositions.

The analysis and preparation routine demand a po-

larization rotator, such as a quartz-wedge Babinet compensator plate [94], a Nicol prism to separate orthogonal polarization states, and a pair of photomultiplier detectors on the twin exit ports, see Fig. 3. Single photon detection is implemented by gating a cw laser ON and OFF through an attenuator, to ensure, with high probability, that there is only one photon received per pulse [95].

Using the polarization rotator an arbitrary basis change can be made to prepare, or measure, any given state $\hat{\mathbf{r}}$, along any chosen axis $\pm\hat{\mathbf{a}}$ upon the Poincaré sphere. Using the timing information, signal photons can be discriminated. Performing n trials, the data reduce to n_j counts of $+\hat{\mathbf{a}}_j$, and $n - n_j$ of $-\hat{\mathbf{a}}_j$ for the basis $\pm\hat{\mathbf{a}}_j$.

The entire procedure, including the data reduction to $\hat{\rho}(\Phi_N)$, is amenable to automation. After completing an N -trial run, the controlling automaton might present a printout of the result, and an estimate of confidence to the waiting experimenter.

Our statistical analysis of this scenario follows that of Jones [19]. For convenience, we envision a hierarchy of possible measurement schemes. They are characterized by the choice of m bases $\pm\hat{\mathbf{a}}_j$, used n times each, perhaps chosen at random, by the controlling automaton, with $N = nm$. The N -trial correlation is thus

$$p(\{n_1, \dots, n_m\}|\hat{\mathbf{r}}) = \prod_{j=1}^m \left(\frac{1 + \hat{\mathbf{a}}_j \cdot \hat{\mathbf{r}}}{2} \right)^{n_j} \left(\frac{1 - \hat{\mathbf{a}}_j \cdot \hat{\mathbf{r}}}{2} \right)^{n - n_j}, \quad (70)$$

which is inverted upon dividing by the normalization

$$p(\Phi_N = \{n_1, \dots, n_m\}) = \int p(\{n_1, \dots, n_m\}|\hat{\mathbf{r}}) d\hat{\Omega}_{\hat{\mathbf{r}}}, \quad (71)$$

for which a general combinatorial formula is known [19].

The mutual information of such a scheme is computed via the exact expression [19]

$$\{\hat{\mathbf{r}}, \Phi_N\} = -N/2 - \sum_{\{n_1, \dots, n_m\}} p(n_1, \dots, n_m) \times \ln p(n_1, \dots, n_m) \quad (72)$$

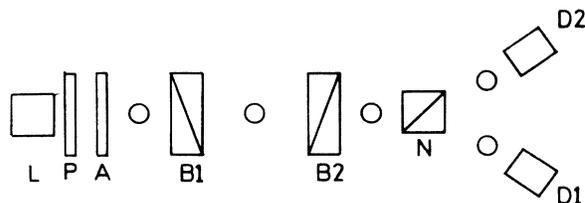


FIG. 3. Schematic of a photon polarization channel used as a testbed for studying the noise limits upon state measurement (cf. Bennett *et al.*, Ref. [95]). Light source L issues a beam that is polarized at P , and attenuated at A . Two quartz-wedge Babinet compensator plates $B1$ and $B2$ implement the \hat{U} and \hat{V} of Fig. 2, to carry out a polarization rotation at either end. The Nicol prism N performs a standard polarization measurement, sending photons of opposite linear polarization to the photomultipliers $D1$ and $D2$, where photon counts are accumulated.

(notice that the optimal instrument must maximize the entropy of its *outcome probabilities*). According to (48), this is strictly bounded above by

$$\{\hat{\mathbf{r}}, \Phi_N\} \leq \ln(N+1) - \frac{N}{1+N}. \quad (73)$$

Additionally, from (60), we have

$$\langle S[\hat{\rho}(\Phi_N)] \rangle \geq \frac{1}{N+2} \ln(N+2) + \frac{N+1}{N+2} \ln\left(\frac{N+2}{N+1}\right) \quad (74)$$

as the strict lower bound on entropy.

To analyze the asymptotic performance it is convenient to perform a reduction of the data Φ_N , as the complete combinatorial expansion is rather difficult to calculate. In the large N limit, the inference is near pure, so we replace the counting data Φ_N by its maximum likelihood estimator. This is defined as that $\hat{\mathbf{s}}(\Phi_N)$ where $p(\Phi_N|\hat{\mathbf{s}})$ has its peak. In the Appendix we show that the large N inferred density $p(\hat{\mathbf{r}}|\Phi_N)$ can be replaced by the local Gaussian

$$p(\hat{\mathbf{r}}|\hat{\mathbf{s}}) \sim 2N \{\det \mathbf{A}_\perp(\hat{\mathbf{s}})\}^{1/2} \times \exp\left\{-\frac{N}{2}(\hat{\mathbf{r}} - \hat{\mathbf{s}}) \cdot \mathbf{A}_\perp(\hat{\mathbf{s}}) \cdot (\hat{\mathbf{r}} - \hat{\mathbf{s}})\right\}, \quad (75)$$

which applies to the locally flat portion of the unit sphere around a small cap centered upon $\hat{\mathbf{s}}$. The anisotropy matrix \mathbf{A}_\perp depends upon the chosen bases

$$\begin{aligned} \mathbf{A}_\perp(\hat{\mathbf{s}}) &\equiv (1/m) \sum_{k=1}^m \frac{\hat{\mathbf{a}}_k \hat{\mathbf{a}}_k}{1 - (\hat{\mathbf{a}}_k \cdot \hat{\mathbf{s}})^2} \\ &= (1/m) \sum_{k=1}^m \hat{\mathbf{a}}_{k\perp}(\hat{\mathbf{s}}) \hat{\mathbf{a}}_{k\perp}(\hat{\mathbf{s}}). \end{aligned} \quad (76)$$

Here $\hat{\mathbf{a}}_{k\perp}(\hat{\mathbf{s}})$ denotes the normalized projection of the basis $\pm\hat{\mathbf{a}}_k$ in the tangent plane at $\hat{\mathbf{s}}$.

To visualize \mathbf{A}_\perp , one recognizes that the $\hat{\mathbf{a}}_{k\perp}(\hat{\mathbf{s}})$ must populate the rim of a unit disk in the tangent plane at $\hat{\mathbf{s}}$. Choosing a coordinate system with z axis along $\hat{\mathbf{s}}$, one forms a unit vector from the x and y components of the $\hat{\mathbf{a}}_k$. The resulting 2×2 matrix is

$$\mathbf{A}_\perp(\hat{\mathbf{s}}) = (1/m) \sum_{k=1}^m \begin{pmatrix} a_{k\perp x} a_{k\perp x} & a_{k\perp x} a_{k\perp y} \\ a_{k\perp y} a_{k\perp x} & a_{k\perp y} a_{k\perp y} \end{pmatrix}(\hat{\mathbf{s}}). \quad (77)$$

It is obviously of unit trace, and must have positive eigenvalues because it is a sum of projectors. If the angle made between the i th and j th projected directions $\pm\hat{\mathbf{a}}_{i\perp}$ and $\pm\hat{\mathbf{a}}_{j\perp}$ is α_{ij} , then we find that

$$\det \mathbf{A}_\perp(\hat{\mathbf{s}}) = (1/m)^2 \sum_{i < j} \sin^2 \alpha_{ij}(\hat{\mathbf{s}}). \quad (78)$$

Only when the distribution of projected unit vectors is uniform will we have both eigenvalues being equal. Then the determinant assumes the value $1/4$. Otherwise it is smaller.

Thus $\mathbf{A}_\perp(\hat{\mathbf{s}})$ quantifies the unequal standard errors in $\hat{\mathbf{r}}$ along the family of great circle geodesics that radiate from the center $\hat{\mathbf{s}}$. Furthermore, it depends upon $\hat{\mathbf{s}}$, so that the accuracy of any given instrument depends upon which input state $\hat{\mathbf{r}}$ it measures.

Optimality of an m -basis instrument is now adjudged from the expression [19]

$$\{\hat{\mathbf{r}}, \Phi_N\} \sim \ln N - 1 + 1/2 \int \ln[4 \det \mathbf{A}_\perp(\hat{\mathbf{s}})] d\hat{\Omega}_{\hat{\mathbf{s}}}. \quad (79)$$

Obviously, the maximum value is attained only when $\det \mathbf{A}_\perp(\hat{\mathbf{s}}) = 1/4$, for all $\hat{\mathbf{s}}$. This corresponds to equal accuracy in all directions, for all possible input states $\hat{\mathbf{r}}$. Such an instrument has no prejudice, nor any preference.

For optimal performance we take an isotropic spread of measurement directions. An ideal single-trial POM with the same limiting statistical behavior is

$$p(\hat{\mathbf{a}}|\hat{\mathbf{r}}) = (1 + \hat{\mathbf{a}} \cdot \hat{\mathbf{r}}), \quad (80)$$

for which \mathbf{A}_\perp is one-half the identity. Thus

$$p(\hat{\mathbf{r}}|\hat{\mathbf{s}})_{\text{opt}} \sim N \exp\left\{-\frac{N}{4}(\hat{\mathbf{r}} - \hat{\mathbf{s}})_\perp \cdot (\hat{\mathbf{r}} - \hat{\mathbf{s}})_\perp\right\}. \quad (81)$$

The information gain is $\ln N - 1$, which attains the asymptotic value of (73).

Such an idealized POM can be closely approached in practice, using small numbers m of sample directions arranged in a symmetric array, such as that suggested by the face directions of the regular solids [19]. For instance, the simple choice $m = 3$, with axes $\pm\mathbf{x}$, $\pm\mathbf{y}$, and $\pm\mathbf{z}$, aligned on the six faces of a cube, requires only about 4% more systems than the optimal one. Passing to ten directions aligned with the 20 faces of an icosahedron this figure reduces to 0.37%. Using such a scheme, chosen according to the general principle of isotropic sampling, the ideal behavior (73) is effectively attained.

Thus it is generally expected that the upper bound given here can be closely approached in practice, provided that one can solve the basic technological problems of general state preparation and measurement. Recent advances in quantum optics for the preparation [96], and measurement [5] of arbitrary states of the radiation field encourage further work upon the statistical limitations that apply to infinite-dimensional systems.

VII. DISCUSSION

In assessing the role of the statistical approach to quantum state determination, it is helpful to compare the problem considered here to that which is more usual in quantum communication theory. The central problem there is to optimally discriminate among a set of M signal density matrices $\hat{\rho}_l$, $l \in [1, M]$, the so-called M -ary detection problem [97].

The prototypical theorem for this application is that of Holevo [98], which states that the channel capacity C is limited by

$$C \leq S[\hat{\rho}_C] - \sum_{l=1}^M p_l S[\hat{\rho}_l], \quad (82)$$

where \mathcal{A} is a single-trial POM, $\hat{\rho}_C$ is the channel density operator, defined as

$$\hat{\rho}_C = \sum_{l=1}^M p_l \hat{\rho}_l, \quad (83)$$

and p_l is the prior probability for transmission of the symbol $\hat{\rho}_l$. Our theorem differs since it is designed to handle multiple measurements upon the continuum of possible ψ .

The reason for this difference is that in a true communication system the information we wish to transmit will ordinarily have semantic content. The engineer must try and choose the $\hat{\rho}_l$, and the instrument \mathcal{A} , along with the p_l , to maximize channel performance. The receiver is not called upon to measure an arbitrary state. However, in the construction of an actual physical channel, or the verification of its specifications, one must check that the implemented $\hat{\rho}_l$ are indeed the desired set. To do this one would send each symbol $\hat{\rho}_l$ repeatedly, to check the channel fidelity, and to calibrate the distant receiver.

At the receiving end, whatever signal $\hat{\rho}_l$ was sent may well have suffered elements of a random, and/or systematic change. This may be due to a simple loss mechanism, such as a dissipative element. In a practical device this cannot be calculated. To characterize it, in the laboratory, one would send the identical test string, and measure the received state.

As a communication system, such an identical string is poor, there is no actual message. However, the ability to overcome the single measurement quantum noise by sending many copies is the key to “pump priming” a true quantum channel [21].

To best illustrate how the two problems are mutually entwined, consider again the photon example. If Norman and Noeline wanted to use polarization states of photons to converse they must first ensure that their respective polarizers are coaligned.

To check this Norman sends an identical string which Noeline measures so that she can align her receiver properly. If they had an infinite number of test photons to expend they could achieve perfect calibration (the classical limit of an intense light beam). Then an “up” state would be received as “up,” and “down” as “down,” with zero error. Using the optimal configuration of a binary symmetric channel [36], see Fig. 4, with equiprobable symbols, the channel capacity is one bit per photon.

However, given only N photons the perfect calibration is impossible. While Norman may really be sending a pure state, Noeline cannot find out its proper orientation. So far as she is concerned, Norman is signaling with an impure state characterized by her N -trial inferred density matrices

$$\rho(\hat{s}) \xrightarrow{N \rightarrow \infty} \frac{1}{2} (1 \pm \{1 - [4N \det \mathbf{A}_\perp(\hat{s})]^{-1}\} \hat{s} \cdot \sigma). \quad (84)$$

Although in our everyday theoretical calculations it is

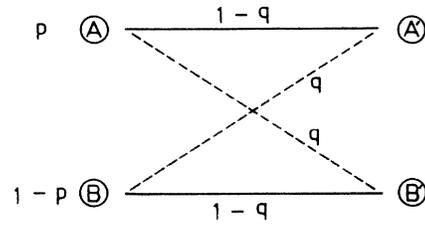


FIG. 4. A binary symmetric channel. Transmitted symbols A and B are sent with probabilities p and $1 - p$, respectively (a function of the chosen coding). In transit, errors cause a symmetric crossover to the received symbols A' and B' , with error probability q . The channel capacity is $C = q \log_2 q + (1 - q) \log_2 (1 - q) - w \log_2 w - (1 - w) \log_2 (1 - w)$, bits per symbol ($0 < C < 1$), where $w = p - 2pq + q$ (see Ref. [36]). At $p = 1/2$, $w = 1/2$, and the channel capacity is maximized for all q . At $q = 1/2$, for all p , noise obliterates the signal and $C = 0$.

customary to assume that states are “knowable” without measurement this is not the common situation. Usually that is the result of treating the problem with prior information. We take a system in equilibrium, in a ground state, or generally in some purpose-prepared state. Here it would not help simply to phone Norman, because the two must verify channel fidelity. In this, quantum mechanics *demand*s that Noeline be forever doubtful of *exactly* what basis Norman signals.

To find the resulting channel performance we substitute the matrices (84), for the optimal instrument, into Holevo’s bound (82) and get

$$C \sim 1 + (1/N) \log_2 (1/N) + (1 - 1/N) \log_2 (1 - 1/N) \quad (85)$$

[for a strict upper bound on C we substitute the expression (74)]. This is the estimated number of bits per photon that are transmissible using an optimized binary symmetric channel (see Fig. 4, and set $p = 1/2$ with the error rate $q = 1/N$), after N identical photons have been exchanged in the priming phase.

If the two *already* had a primed channel, the same N photons could transmit $N - \log_2 N - 1/\ln 2$ bits of intelligible human dialogue, rather than $\log_2 N - 1/\ln 2$ boring bits of quantum priming data. This example illustrates the importance of prior information when considering state measurement and inference. Only when the overall system has been fully calibrated does it become a useful communication device. Unlike in classical physics, this must entail a quantum cost, due to the fundamental theorem here established.

VIII. CONCLUSION

In summary, we have applied the Bayesian methodology for state inversion, and Shannon’s information theory to set the fundamental limits upon the determination of quantum states for arbitrary measurements on a finite-dimensional system.

The key result was the theorem proven in Sec. IV, which is the state-measurement analog of Holevo’s theo-

rem [98]. Since it reduces the general problem of finding bounds to an analysis of the N th powers of a one-trial correlation, this result greatly simplifies analysis of optimality questions. In spite of the sophistication of the methods required to prove this, we stress the extreme intuitive simplicity of this result. The limit follows in a straightforward way from the “most peaked” possible posterior distribution. This, in turn, is geometrically fixed, as to its functional form, by the celebrated theorem of Gleason [18], and our own. To stress the importance of symmetry considerations, we used the basic classification of time-reversal invariance. This illustrates how a suboptimal information gain arises when we can neither prepare nor analyze the full manifold of quantum superpositions.

Since the field of state measurement and its statistical analysis remains in its infancy it may be useful to peek forward to its future. First, it is desirable to analyze more fully the case of impure input states [30], due to a poor preparation, or the presence of thermal noise in the signal path. Secondly, one wants to know how these restrictions affect the empirical assignment of a scattering matrix. The main reason to be interested in these two theoretical problems lies with quantum communication and computation.

In both areas the quantum engineer needs “probe tools” with which to construct, analyze, and later calibrate modular quantum devices that effect particular state preparations, measurements, or unitary transformations. Such devices would form the basis for the construction and testing of quantum logic gates [99] in a quantum parallel computer [100] that implements the quantum logic of Birkhoff and von Neumann [101].

A second reason to pursue this possibility concerns the direct study of quantum dynamics at the quantum noise-limited level. Once we have seen how to overcome the quantum noise limitations to state measurement there is no obstacle, at least in principle, to the empirical study of quantum dynamical maps in simple systems. One would simply pass many copies of various probe states $|\psi_{\text{in}}\rangle$, or $\hat{\rho}_{\text{in}}$, through a device that applies an unknown unitary or nonunitary transformation and then measure the outgoing states $|\psi_{\text{out}}\rangle$, or $\hat{\rho}_{\text{out}}$, see Fig. 5. From the accumulated data we seek to assign an empirical scattering superoperator \mathcal{S} , generally a completely positive map [102] which can be compared with theoretical calculations [103]; or perhaps imprinted, with trademark, upon its epoxy casing.

The given example of a practically realizable two-state inference system demonstrates the feasibility of building devices that will perform the required state measurements at, or near, the quantum limit. Further, it shows how the familiar classical \sqrt{N} gain in accuracy applies to multiple measurements performed upon N members of a quantum ensemble.

To close, we may return now to (1) and observe that the multiple measurement strategy allows both $\langle \hat{A} \rangle$ and $\langle \hat{B} \rangle$ to be precisely determined, since the empirically accumulated averages must converge in probability to their predicted theoretical values. The importance of the Heisenberg uncertainty principle lies primarily in the pro-

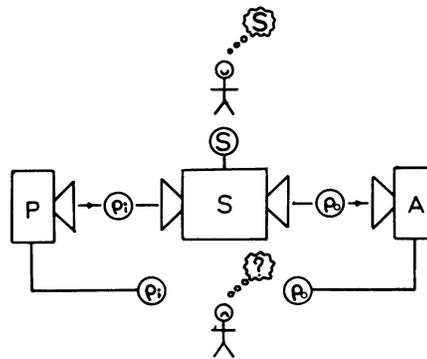


FIG. 5. The empirical assignment of evolution operators requires precalibrated preparation and analysis units. Norman now selects a superoperator, \mathcal{S} . Noeline attempts to determine this by sending a stream of known states $\hat{\rho}_{\text{in}}$ through the scattering region (perhaps an optical fiber in an optical communication link). On the other side she measures $\hat{\rho}_{\text{out}}$, and attempts to assign \mathcal{S} consistent with the data, i.e., $\hat{\rho}_{\text{out}} = \mathcal{S}[\hat{\rho}_{\text{in}}]$. In general, the problem requires many different $\hat{\rho}_{\text{in}}$ to be sent, and will have similar statistical limitations. It demands the technology of state measurement.

hibition it makes upon an exact determination of both, for a single system. This is very important in the context of Bell inequalities [104]. As shown by Wootters and Zurek [105], if it were not so, then correlated singlet photons would permit superluminal communication [28,106]. Fortunately, the quantum prohibits this, but allows, nevertheless, that we make precise measurements, provided only that a large supply of identically prepared quantum systems is available.

ACKNOWLEDGMENTS

K.R.W.J. thanks the Australian Research Council for their supporting Grant No. F69140440, and Andrew Davies for his constant encouragement, and assistance with some mathematical questions.

APPENDIX: TWO-STATE MAXIMUM LIKELIHOOD ANALYSIS

Elsewhere, I have presented a purely Bayesian analysis [19], here I obtain the same result obtained using Fisher’s method of maximum likelihood analysis [107,69,108], as applied by Braunstein *et al.* [49], and Lane *et al.* [50] to the problem of optimal quantum phase measurement. One begins with the N -trial likelihood for an m -basis instrument used n times,

$$\begin{aligned}
 p_N(\vec{X} = (X_1, \dots, X_m) | \mathbf{r}) &= \prod_{j=1}^m \binom{n}{n(1+X_j)/2} \\
 &\times \left(\frac{1 + \mathbf{a}_j \cdot \mathbf{r}}{2} \right)^{n(1+X_j)/2} \\
 &\times \left(\frac{1 - \mathbf{a}_j \cdot \mathbf{r}}{2} \right)^{n(1-X_j)/2},
 \end{aligned} \tag{A1}$$

where $X_j = 2k_j/n - 1$, with k_j the number of ‘‘up’’ counts on the basis $\pm \mathbf{a}_j$.

In the method of Fisher [107], a data reduction is made to estimate \mathbf{r} , and one replaces the above by the corresponding distribution of estimators.

Specifically, we recognize that as $n \rightarrow \infty$ the above distribution becomes strongly peaked near, but not necessarily on, the true value. From the data \vec{X} we obtain an estimate $\mathbf{s}(\vec{X})$, defined as the argument \mathbf{r} which maximizes the likelihood $p_N(\vec{X}|\mathbf{r})$. A statistical argument is then invoked to compute the distribution $p_N(\mathbf{s}|\mathbf{r})$, with the observed data eliminated.

The result is Fisher’s theorem [107,108,109], which states that the distribution of maximum likelihood estimators is asymptotically normal, here according to

$$p(\mathbf{s}|\mathbf{r}) \sim \mathcal{N} \exp \left\{ -\frac{1}{2} (\mathbf{r} - \mathbf{s})_\alpha F_{\alpha\beta}(\mathbf{s})(\mathbf{r} - \mathbf{s})_\beta \right\}, \quad (\text{A2})$$

where \mathcal{N} is a normalization, and summation over repeated Cartesian indices $\alpha \in [1, 3]$ is implicit. The Fisher information matrix corresponding to (A1) reads

$$F_{\alpha\beta}(\mathbf{r}) = \sum_{\vec{X}} p_N(\vec{X}|\mathbf{r}) [\partial_\alpha \ln p_N(\vec{X}|\mathbf{r})] [\partial_\beta \ln p_N(\vec{X}|\mathbf{r})], \quad (\text{A3})$$

which is an asymptotic result, so that we must require that n be large. Since

$$\begin{aligned} \partial_\alpha \ln p_N(\vec{X}|\mathbf{r}) &= \frac{1}{p_N} \frac{\partial p_N}{\partial \mathbf{r}_\alpha} \\ &= n \sum_{j=1}^m \frac{\mathbf{a}_{j\alpha}}{1 - (\mathbf{a}_j \cdot \mathbf{r})^2} (X_j - \mathbf{a}_j \cdot \mathbf{r}), \end{aligned}$$

we have

$$\begin{aligned} F_{\alpha\beta}(\mathbf{r}) &= n^2 \sum_{i,j} \frac{\mathbf{a}_{i\alpha} \mathbf{a}_{j\beta}}{[1 - (\mathbf{a}_i \cdot \mathbf{r})^2][1 - (\mathbf{a}_j \cdot \mathbf{r})^2]} \\ &\quad \times \langle (X_i - \mathbf{a}_i \cdot \mathbf{r})(X_j - \mathbf{a}_j \cdot \mathbf{r}) \rangle_{\vec{X}}. \end{aligned} \quad (\text{A4})$$

Using the exact binomial moment [110]

$$\begin{aligned} \sum_{\vec{X}} (X_i - \mathbf{a}_i \cdot \mathbf{r})(X_j - \mathbf{a}_j \cdot \mathbf{r}) p_N(\vec{X}|\mathbf{r}) \\ = \frac{\delta_{ij}}{n} [1 - (\mathbf{a}_j \cdot \mathbf{r})^2], \end{aligned} \quad (\text{A5})$$

we obtain the required result

$$F_{\alpha\beta}(\mathbf{r}) = N \sum_i \frac{1}{m} \frac{\mathbf{a}_{i\alpha} \mathbf{a}_{i\beta}}{[1 - (\mathbf{a}_i \cdot \mathbf{r})^2]}, \quad (\text{A6})$$

which is just (76) on specializing to unit vectors \mathbf{a}_i and \mathbf{r} . Here the given normalization is fixed by integration over the plane tangent to the sphere at \mathbf{s} , keeping the factor $1/4\pi$.

In the quantum context the useful feature of Bayesian methods lies in their applicability to *small data sets*. This is outside the domain of the central limit theorem, so that most mathematical statisticians would insist that more data be gathered, and they pay this case little attention. In quantum mechanics, and quantum technology, it is just this case that we want to know more about (i.e., where quantum noise dominates).

The useful feature of the asymptotic connection between maximum likelihood analysis and uniform prior Bayes methods [88] is that we may expect general agreement between optimizations performed using the Fisher criterion, see, e.g., Refs. [49,50], and the method here described. To further this comparison one would like to extend maximum likelihood performance analysis to small data sets. Recently, Braunstein [111] has broached the problem via a careful study of the *preasymptotic* convergence to Fisher’s asymptotic result [107] using corrections to the central limit theorem.

-
- [1] W. Heisenberg, *Z. Phys.* **43**, 172 (1927).
[2] H.P. Robertson, *Phys. Rev.* **34**, 163 (1929).
[3] J.J. Sakurai, *Modern Quantum Mechanics* (Addison-Wesley, New York, 1985), p. 34.
[4] As an example, we mention the recent test of the number-phase uncertainty relation using the method of optical homodyne tomography. See D.T. Smithey, M. Beck, J. Cooper, and M.G. Raymer, *Phys. Rev. A* **48**, 3159 (1993).
[5] K. Vogel and H. Risken, *Phys. Rev. A* **40**, 2847 (1989); D.T. Smithey, M. Beck, M.G. Raymer, and A. Fardani, *Phys. Rev. Lett.* **A 70**, 11 244 (1993); M. Beck, D.T. Smithey, and M.G. Raymer, *Phys. Rev. A* **48**, 890 (1993).
[6] The algebraic method has a long history going back to the celebrated Pauli problem on the (in)adequacy of the position and momentum distribution data for the empirical determination of ψ . Vogel and Risken, Ref. [5], solved it using a tomographic method. For the original statement see W. Pauli, in *Quantentheorie*, edited by H. Geiger and K. Scheel, *Handbüch der Physik* Vol. 24 (Springer-Verlag, Berlin, 1933), Pt. 1, p. 98. See also E. Feenberg, Ph.D. thesis, Harvard University, 1933, and the commentary in E.C. Kemble, *The Fundamental Principles of Quantum Mechanics* (McGraw-Hill, New York, 1937), p. 71.
[7] W. Band and J.L. Park, *Found. Phys.* **1**, 133 (1970); **1**, 211 (1971); **1**, 339 (1971).
[8] W. Gale, E. Guth, and G.T. Trammell, *Phys. Rev.* **165**, 1434 (1968).
[9] J. Bohn, *Phys. Rev. Lett.* **66**, 1547 (1991).
[10] S. Wiegert, *Phys. Rev. A* **45**, 1688 (1992).
[11] I.D. Ivanović, *J. Phys. A* **14**, 3241 (1991); *J. Math. Phys.* **24**, 1199 (1983).
[12] Increasingly, there is interest in the statistical analysis of inversion problems [see G.J. Daniell, in *Maximum En-*

- theory in Action*, edited by B. Bush and V.A. Macaulay (Clarendon, Oxford, 1991)]. Although algebraic methods of inversion appear “rigorous” and “elegant” they ignore the all-important imperfections of the data gathering process. Very often an initially well-posed algebraic inversion problem becomes ill posed when the realities of imperfect data are acknowledged. Nowhere is this more apparent than in the case where the data *underdetermine* the solution. This is the rule, rather than the exception, in the case of functional inference, i.e., the measurement of an unknown function. Of course, we must expect this situation in quantum mechanics because most operators are not easily measured.
- [13] C.W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
- [14] A.S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
- [15] W.K. Wootters, Ph.D. thesis, University of Texas at Austin, 1980.
- [16] W.K. Wootters, Phys. Rev. D **23**, 357 (1981).
- [17] W.K. Wootters and B.D. Fields, Ann. Phys. (N.Y.) **191**, 363 (1989).
- [18] A. Gleason, J. Math. Mech. **6**, 885 (1957). See also M. Readhead, *Incompleteness, Nonlocality and Realism* (Clarendon Press, Oxford, 1990), pp. 27–30.
- [19] K.R.W. Jones, Ann. Phys. (N.Y.) **207**, 140 (1991).
- [20] K.R.W. Jones, J. Phys. A **24**, 121 (1991).
- [21] K.R.W. Jones, Ph.D. thesis, University of Bristol, 1989.
- [22] C.E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948); **27**, 623 (1948); reprinted with commentary in C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, IL, 1949).
- [23] E.B. Davies and J.T. Lewis, Commun. Math. Phys. **17**, 239 (1970).
- [24] K. Kraus, *States, Effects and Operations* (Springer, Berlin, 1983).
- [25] E.T. Jaynes, in *E.T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, Synthese Library Vol. 158, edited by R.D. Rosenkrantz (Kluwer, Dordrecht, 1983).
- [26] S.L. Braunstein and C.M. Caves, Phys. Rev. Lett. **61**, 662 (1988).
- [27] A. Peres and W.K. Wootters, Phys. Rev. Lett. **66**, 1119 (1991).
- [28] A.J.M. Garrett, Found. Phys. **20**, 1475 (1990).
- [29] P.R. Dukes and E.G. Larson, in *Maximum Entropy and Bayesian Methods*, edited by W.T. Grandy, Jr. and L.H. Schick (Kluwer, Dordrecht, 1991), pp. 161–168.
- [30] E.G. Larson and P.R. Dukes, in *Maximum Entropy and Bayesian Methods*, edited by W.T. Grandy, Jr. and L.H. Schick (Kluwer, Dordrecht, 1991), pp. 181–189.
- [31] E.G. Larson, in *Maximum Entropy and Bayesian Methods*, edited by C.R. Smith *et al.* (Kluwer, Dordrecht, 1992), pp. 453–466.
- [32] H. Jeffreys, *Theory of Probability*, 3rd ed. (Clarendon, Oxford, 1961).
- [33] R.T. Cox, Am. J. Phys. **14**, 1 (1946).
- [34] E.T. Jaynes (unpublished).
- [35] J.A. Hartigan, *Bayes Theory* (Springer, New York, 1983).
- [36] A.B. Carlson, *Communication Systems—An Introduction to Signals and Noise in Electrical Communication* (McGraw-Hill, New York, 1986), Chap. 1 and Chap. 15, especially pp. 574–577 on mutual information, and the binary symmetric channel.
- [37] We mean to indicate that the problem at hand forms a mere component to the *physics of information* (see entries at Ref. [86] and Ref. [89]), i.e., the enterprise is larger than its parts. However, to split the question thus helps us to mentally isolate that which is special to quantum mechanics.
- [38] E.P. Wigner, *Group Theory and its Applications to the Quantum Mechanics of Atomic Spectra* (Academic Press, New York, 1959).
- [39] R.P. Feynmann, R.B. Leighton, and M. Sands, *The Feynmann Lectures on Physics* (Addison-Wesley, Reading, MA, 1965), Vol. III, Chap. 5.
- [40] See E. P. Wigner, Ref. [38], Chap. 26.
- [41] F.J. Dyson, J. Math. Phys. **3**, 1199 (1962).
- [42] F.J. Dyson, J. Math. Phys. **3**, 140 (1962); **3**, 157 (1962); **3**, 166 (1962).
- [43] M.L. Mehta, *Random Matrices and the Statistical Theory of Energy Levels* (Academic Press, New York, 1967).
- [44] *Statistical Theories of Spectra Fluctuations*, edited by C.E. Porter (Academic Press, New York, 1965).
- [45] T.A. Brody, J. Flores, J.B. French, P.A. Mello, A. Pandey, and S.S.M. Wong, Rev. Mod. Phys. **53**, 385 (1981).
- [46] F. Haake, *Quantum Signatures of Chaos* (Springer, Berlin, 1991).
- [47] K. Życzkowski, J. Phys. A **23**, 4427 (1990).
- [48] K.R.W. Jones, J. Phys. A **23**, L1247 (1990).
- [49] S.L. Braunstein, A.S. Lane, and C.M. Caves, Phys. Rev. Lett. **69**, 2153 (1992).
- [50] A.S. Lane, S.L. Braunstein, and C.M. Caves, Phys. Rev. A **47**, 1667 (1993).
- [51] A. Perelomov, *Generalized Coherent States and Their Applications* (Springer, Berlin, 1986).
- [52] R. Gilmore, *Lie Groups, Lie Algebras and Some of Their Applications* (Springer, Berlin, 1974); see especially Chap. 9, Sec. V on the analytical properties of cosets, and Chap. 3, Sec. V on the invariant Haar integral.
- [53] H.A. Kramers, Proc. Acad. Sci. Amsterdam **33**, 959 (1930).
- [54] The introduction of quaternions here has nothing to do with the suggestions for a wider quantum theory [J.M. Jauch, *Foundations of Quantum Mechanics* (Addison-Wesley, Reading, MA, 1968)]. As explained by Dyson [41], it is due to the classical result, of Froebenius, that there are only three kinds of associative division algebra which can form the basis for a continuous matrix algebra [L.S. Pontryagin, *Topological Groups* (Gordon and Breach, New York, 1966), p. 158]. They are an accident that arises for *even-dimensional* complex-valued quantum dynamical systems.
- [55] See R. Gilmore, Ref. [52], p. 371.
- [56] An early example of a POM, before the notion was formalized, is the homodyne method for the simultaneous measurement of noncommuting quantities. See E. Arthurs and J.L. Kelly, Bell Syst. Tech. J. **44**, 725 (1965); C.Y. She and H. Heffner, Phys. Rev. **152**, 11 103 (1966). See also J.H. Shapiro and S.R. Shepard, Phys. Rev. A **37**, 3795 (1991).
- [57] H.P. Yuen and V.W.S. Chan, Opt. Lett. **8**, 177 (1983), and references therein.
- [58] V.I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, Berlin, 1989).

- [59] A.I. Khinchin, *Mathematical Foundations of Statistical Mechanics* (Dover, New York, 1949), pp. 13–19.
- [60] L. Nachbin, *The Haar Integral* (Van Nostrand Reinhold, New York, 1965).
- [61] R. Gilmore, Ref. [52], p. 78.
- [62] L.A. Santalo, *Integral Geometry and Geometric Probability* (Addison-Wesley, Reading, MA, 1976).
- [63] M. Creutz, *Quarks, Gluons and Lattices* (Cambridge University Press, Cambridge, England, 1983).
- [64] R. Gilmore, Ref. [52], pp. 391–395.
- [65] K.R.W. Jones, in *Recent Advances in Fractional Calculus*, edited by R.N. Kalia (Global Publishing, Sauk Rapids, MN, 1993).
- [66] K.R.W. Jones, *J. Phys. A* **24**, 1237 (1991).
- [67] A.J. Davies and K.R.W. Jones, *J. Math. Anal. Appl.* (to be published).
- [68] *Handbook of Mathematical Functions*, edited by A. Abramowitz and I.A. Stegun (Dover, New York, 1965), p. 930.
- [69] C. W. Helstrom, *Int. J. Theor. Phys.* **8**, 361 (1973).
- [70] Fisher has pointed out [R.A. Fisher, *Proc. R. Soc. London, Ser. A* **217**, 295 (1953)] that Gauss invented modern statistics to analyze errors in astronomical data. Well aware of curvature, he understood that the typical observational error was small enough that the celestial sphere could be approximated by its Euclidean tangent plane at the star in question. Thereafter, the Euclidean statistics has dominated our thinking. Recently, some mathematical statisticians have become interested in non-Euclidean statistics [*Differential Geometry in Statistical Inference*, edited by S.-I. Amari *et al.*, Institute of Mathematical Statistics Lecture Notes Vol. 10, edited by S.S. Gupta (IMS, Hayward, CA, 1987)]. Presently, the applications are few in number [but, see R.A. Fisher, *Proc. R. Soc. London, Ser. A* **217**, 295 (1953) with his early example of Icelandic paleomagnetism data]. However, quantum theory has this feature universally.
- [71] H. Jeffreys, *Proc. R. Soc. London, Ser. A* **186**, 453 (1946).
- [72] E.T. Jaynes, *IEEE Trans. Syst. Sci. Cybern.* **SSC4**, 227 (1968); reprinted in E.T. Jaynes, Ref. [25].
- [73] J. Hartigan, *Ann. Math. Stat.* **35**, 386 (1964).
- [74] M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, 4th ed. (Griffin, London, 1977) Vol. I, Chap. 8.
- [75] T. Bayes, *Philos. Trans. R. Soc. London* **53**, 370 (1763); reprinted in *Biometrika* **45**, 293 (1958).
- [76] H. Reichenbach, in *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, edited by J. Neyman (University of California Press, Berkeley, 1949), p. 14. Reichenbach was an ardent critic of Bayesian methods, chiefly because he recognized that symmetry principles are unavailable in many applications.
- [77] H. Everett, Ph.D. thesis, Princeton University, 1957; reprinted in *The Many-Worlds Interpretation of Quantum Mechanics*, edited by B.S. De Witt and N. Graham (Princeton University Press, Princeton, NJ, 1973).
- [78] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, NJ, 1955).
- [79] H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950), pp. 152–156.
- [80] K.R.W. Jones, *J. Phys. A* **24**, L415 (1991).
- [81] A.S. Holevo, *J. Multivar. Anal.* **3**, 337 (1973).
- [82] E.B. Davies, *IEEE Trans. Inf. Theory* **24**, 596 (1978).
- [83] C. Bendjaballah and M. Charbit, *IEEE Trans. Inf. Theory* **35**, 1114 (1989).
- [84] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, NJ, 1970).
- [85] Convex analysis is extremely useful in statistical optimization problems because the space of all probability densities comprises a convex set. Given a function $p(\mathcal{M})$ that is a probability density on the manifold \mathcal{M} , and another q , we may form a new one $r(\mathcal{M}) = \lambda p(\mathcal{M}) + (1 - \lambda)q(\mathcal{M})$, for any $\lambda \in [0, 1]$, where the addition is made pointwise. If we have to measure the concentration of p , then it is natural to choose a convex function, since this will take its maximum value upon the δ -function distributions, i.e., those which assign unit measure to a single point. If now the densities $p(\mathcal{M})$ are themselves constrained, as functionals defined upon some parameter space which is convex, then the methods of convex analysis are applicable. In quantum theory we are fortunate that the space of density operators is convex, and the observable probability densities are thus generally constrained to be functionals defined on a convex domain.
- [86] For discussion of the connection between entropy, information, and thermodynamics see L. Brillouin, *Science and Information Theory* (Academic Press, New York, 1956); E.T. Jaynes, *Phys. Rev.* **106**, 602 (1957); **108**, 171 (1957); C.H. Bennett, *Int. J. Theor. Phys.* **21**, 905 (1982); W.H. Zurek, *Phys. Rev. A* **40**, 4731 (1989); *Maxwell's Demon Entropy, Information, Computing*, edited by A.S. Leff and A.F. Rex (Princeton University Press, Princeton, NJ, 1990). The given relation is not thermodynamic. Rather it stems from the quantum noise, present even in the limit of zero temperature (at least that is so on our present understanding). It expresses the fact that there is a minimal quantum data set required to be confident of a given level of purity for the assigned density matrix. As the size of the data set N grows unbounded, a pure state is approached.
- [87] K.R.W. Jones, *Phys. Scr.* **T48**, 100 (1993).
- [88] B.S. Clarke and A.R. Barron, *IEEE Trans. Inf. Theory* **36**, 453 (1990), and cited articles.
- [89] C.M. Caves, *Phys. Rev. E* **47**, 4010 (1993).
- [90] Obviously, this is *always* possible asymptotically. Once we have some information about the state under study one may attempt to align the chosen measurement basis so that one of the eigenvector outcomes is as close as possible to the actual pure state. When this condition is met, and this is only possible in the limit as $N \rightarrow \infty$, then this single outcome will occur repeatedly with unit probability. The degradation below optimal performance lies always in the initial expenditure of states required to make this choice. Of course, the performance of such an adaptive analyzer remains limited by Eq. (48).
- [91] F.J. Bloore, *J. Phys. A* **9**, 2059 (1976).
- [92] M. Born and E. Wolf, *Principles of Optics*, 6th ed. (Pergamon Press, London, 1980), p. 31.
- [93] W. Heitler, *The Quantum Theory of Radiation* (Dover, New York, 1984), Appendix 1.
- [94] S.G. Lipson and H. Lipson, *Optical Physics*, 2nd ed. (Cambridge University Press, Cambridge, England, 1981), p. 121.

- [95] A prototype device similar to this has already been constructed by Bennett *et al.* to test their ideas on quantum public key distribution, as the basis for an unbreakable quantum cryptogram. C.H. Bennett, *Science* **257**, 752 (1992); C.H. Bennett and G. Brassard, *SIGACT News* **20**, 78 (1989); C.H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin, *J. Cryptology* **5**, 3 (1992); and A.K. Ekert, *Phys. Rev. Lett.* **67**, 661 (1991). For a description, and picture, of the first quantum communication channel using polarized photons operating at the single quantum level, see C.H. Bennett, G. Brassard, and A.K. Ekert, *Sci. Am.* **267**(4), 26 (1992).
- [96] K. Vogel, V.M. Akulin, and W.P. Schleich, *Phys. Rev. Lett.* **71**, 1816 (1993).
- [97] H.P. Yuen, R.S. Kennedy, and M. Lax, *Proc. IEEE* **58**, 1770 (1970).
- [98] A.S. Holevo, *Probl. Inf. Transm. (USSR)* **9**, 177 (1973).
- [99] R. P. Feynmann, *Found. Phys.* **16**, 507 (1986); P. Benioff, *J. Stat. Phys.* **22**, 563 (1980); K.K. Likharev, *Int. J. Theor. Phys.* **21**, 311 (1982); K. Obermayer, G. Mahler, and H. Haken, *Phys. Rev. Lett.* **58**, 1792 (1987); G.J. Milburn, *ibid.* **62**, 2124 (1989).
- [100] D. Deutsch, *Proc. R. Soc. London Ser. A* **400**, 97 (1985).
- [101] G. Birkhoff and J. Von Neumann, *Ann. Math.* **37**, 823 (1936).
- [102] G. Lindblad, *Commun. Math. Phys.* **48**, 119 (1975).
- [103] One obvious reason to be interested in this is to study the off-diagonal decay of density matrix elements due to the presence of a crossed decohering field in the transmission pathway. H.D. Zeh, *Found. Phys* **1**, 69 (1970); E. Joos and H.D. Zeh, *Z. Phys. B* **59**, 223 (1985); W. Zurek, *Phys. Rev. D* **24**, 1516 (1981); **26**, 1862 (1982); M. Gell-Mann and J.B. Hartle, *ibid.* **47**, 3345 (1993).
- [104] J.S. Bell, *Rev. Mod. Phys.* **38**, 448 (1966).
- [105] W.K. Wootters and W.H. Zurek, *Nature (London)* **299**, 802 (1982).
- [106] R.J. Glauber, *Ann. (N.Y.) Acad. Sci.* **480**, 336 (1986).
- [107] R.A. Fisher, *Proc. Cambridge Philos. Soc.* **22**, 700 (1925).
- [108] M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, 4th ed. (Griffin, London, 1977), Vol. II, Chap. 18.
- [109] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton NJ, 1946), pp. 500–504.
- [110] G.A. Korn and T.M. Korn, *Mathematical Handbook for Scientists and Engineers*, 2nd ed. (McGraw-Hill, New York, 1968).
- [111] S.L. Braunstein, *J. Phys. A* **25**, 3813 (1992).