Estimating the spectrum of a density operator

M. Keyl* and R. F. Werner†
Institut für Mathematische Physik, TU Braunschweig, Mendelssohnstrasse 3, 38106 Braunschweig, Germany
(Received 7 March 2001; published 15 October 2001)

Given N quantum systems prepared according to the same density operator ρ, we propose a measurement on the N-fold system that approximately yields the spectrum of ρ. The projections of the proposed observable decompose the Hilbert space according to the irreducible representations of the permutations on N points, and are labeled by Young frames, whose relative row lengths estimate the eigenvalues of ρ in decreasing order. We show convergence of these estimates in the limit $N\rightarrow\infty$, and that the probability for errors decreases exponentially with a rate we compute explicitly.

DOI: 10.1103/PhysRevA.64.052311 PACS number(s): 03.67.Lx, 03.65.Ca

I. INTRODUCTION

The density operator of a quantum system describes the preparation of the system in all details relevant to statistical experiments. Like a classical probability distribution it cannot be measured on a single system, but can be estimated only on an ensemble sequence of identically prepared systems. In fact, if we could determine the density operator (or, in the pure case, the wave function) on a single quantum system, we could combine the measurement with a device repreparing several systems with the measured density operator, in contradiction to the well-known no-cloning theorem [1]. This points to a close connection between the problem of estimating the density operator and approximate cloning. In the case of inputs promised to be in a pure state the optimal solutions to both problems are known [2–4], and it turns out that in a sense the limit of the cloning problem for output number $M\rightarrow\infty$ is equivalent to the estimation problem. The ''optimal'' cloning transformation was shown in this case to be quite insensitive to the figure of merit defining optimality [3].

In the case of mixed input states much less is known about the cloning problem. It is likely that in this case there may be different natural figures of merit leading to inequivalent ''optimal'' solutions. Even the classical version of the problem is not trivial, and is related to the so-called bootstrap technique [5] in classical statistics.

The estimation problem certainly has many solutions. In fact, any procedure of determining the density matrix through the measurement of the expectations of a suitable ''quorum'' of observables [6], such as in quantum state tomography [7] is a solution. Other methods include adaptive schemes [8] where the result of one measurement is used to select the next one. In all these cases, the estimate amounts to the measurement of an observable on the full input state $\rho^{\otimes N}$, which factorizes into one-site observables. What we are concerned with here, as in the work of Vidal *et al.* [9], is the search for improved estimates, admitting arbitrary observables on the N-fold input system, including ''entangled'' ones. In contrast to [9], however, we are not interested in estimators that are optimal for a more or less general figure of merit, but in the asymptotic behavior if the number N of input systems goes to infinity (in this context see also the work of Gill and Massar [10]).

When $\mathcal{H}\cong\mathbf{C}^d$ is the Hilbert space of a single system, the overall input density operator of the estimation problem is $\rho^{\otimes N}$, which exists on the Nth tensor power $\mathcal{H}^{\otimes N}$. This space has a natural orthogonal decomposition according to the irreducible representations of the permutation group of N points, acting as the permutations of the tensor factors. Equivalently, this is the decomposition according to the irreducible representations of the unitary group on $\mathcal{H}$ (see below). It is well known that this orthogonal decomposition is labeled by *Young frames*, i.e., by the arrangements of N boxes into d rows of lengths $Y_1\geqslant Y_2\geqslant\cdots\geqslant Y_d\geqslant 0$ with $\Sigma_\alpha Y_\alpha=N$. There is a striking similarity here with the spectra we want to estimate, which are given by sequences of the eigenvalues of ρ, say, $r_1\geqslant r_2\geqslant\cdots\geqslant r_d\geqslant 0$, with $\Sigma_\alpha r_\alpha=1$. The basic idea of this paper is to show that this is not a superficial similarity: measuring the Young frame (by an observable whose eigenprojections are the projections in the orthogonal decomposition) is, in fact, a good estimate of the spectrum. More precisely, we show that the probability for the error $|Y_\alpha/N-r_\alpha|$ to be larger than a fixed ϵ for some α decreases exponentially as $N\rightarrow\infty$.

The group theoretic ideas just sketched are nothing new but go back to Weyl [11] and are in the meantime a standard tool within quantum mechanics. Examples of works where similar methods are used in quantum information are [12,9,13–15]. In particular, [15] is closely related to the present paper because similar techniques are used there. This concerns in particular the theory of large deviations [16], and a result by Duffield [17] on the large deviation properties of tensor powers of group representations. This will allow us to compute the rate of exponential convergence explicitly.

II. STATEMENT OF THE RESULT

In order to state our result, explicitly, we need to recall the decomposition theory for N-fold tensor products. Throughout, the one-particle space $\mathcal{H}$ will be the d-dimensional Hilbert space $\mathbf{C}^d$, with $d<\infty$. Two group representations play a crucial role: first, the representation $X\mapsto X^{\otimes N}$ of the general linear group $\mathrm{GL}(d,\mathbf{C})$ and, secondly, the representation

*Electronic address: m.keyl@tu-bs.de
†Electronic address: r.werner@tu-bs.de

$p \mapsto S_p$ of the permutations $p \in \mathbf{S}_N$ on N points, represented by permuting the tensor factors:

$$S_p \ \psi_1 \otimes \cdots \otimes \psi_N = \psi_{p^{-1}1} \otimes \cdots \otimes \psi_{p^{-1}N}. \tag{1}$$

The basic result [18] is that these two representations are "commutants" of each other, i.e., any operator on $\mathcal{H}^{\otimes N}$ commuting with all $X^{\otimes N}$ is a linear combination of the S_p, and conversely. This leads to the decomposition

$$\mathcal{H}^{\otimes N} \cong \bigoplus_Y \mathcal{R}_Y \otimes \mathcal{S}_Y, \tag{2}$$

$$X^{\otimes N} \cong \bigoplus_Y \pi_Y(X) \otimes \mathbf{1}, \tag{3}$$

$$S_p \cong \bigoplus_Y \mathbf{1} \otimes \hat{\pi}_Y(p), \tag{4}$$

where $\pi_Y : \mathrm{GL}(d,\mathbf{C}) \rightarrow \mathcal{B}(\mathcal{R}_Y)$ and $\hat{\pi}_Y : \mathbf{S}_N \rightarrow \mathcal{B}(\mathcal{S}_Y)$ are irreducible representations, and the restriction of π_Y to unitary operators is unitary. The summation index Y runs over all Young frames with d rows and N boxes, as described in the Introduction. We denote by P_Y the projection onto the corresponding summand in the above decomposition.

Let us consider now the estimation problem. As already discussed in the Introduction, we are searching for an observable E_N describing a measurement on N d-level systems, whose readouts are possible spectra of d-level density operators. The set of possible spectra will be denoted by

$$\Sigma = \left\{ s \in \mathbf{R}^d \middle| x \triangleright 0, \ \sum_{j=1}^{d} x_j = 1 \right\}, \tag{5}$$

where $x \triangleright 0$ denotes the ordering relation on $\mathbf{R}^d$ given by

$$s \triangleright 0 : \Leftrightarrow s_j > s_{j+1} \quad \text{for all} \quad j = 1, \ldots, d-1. \tag{6}$$

Technically, E_N must be a positive operator valued measure on this set, assigning to each measurable subset $\Delta \subset \Sigma$ a positive operator $E_N(\Delta) \in \mathcal{B}(\mathcal{H}^{\otimes N})$, whose expectation in any given state is interpreted as the probability for the measurement to yield a result $s \in \Delta$.

The criterion for a good estimator E_N is that, for any one-particle density operator ρ, the value measured on a state $\rho^{\otimes N}$ is likely to be close to the true spectrum $r \in \Sigma$ of ρ, i.e., that the probability

$$K_N(\Delta) := \mathrm{Tr}[E_N(\Delta) \rho^{\otimes N}] \tag{7}$$

is small when Δ is the complement of a small ball around r. Of course, we will look at this problem for large N. So our task is to find a whole sequence of observables E_N, $N = 1,2, \ldots$, making error probabilities like Eq. (7) go to zero as $N \rightarrow \infty$.

The search for efficient estimation strategies E_N can be simplified greatly by symmetry arguments. To see this, consider a permutation $p \in \mathbf{S}_N$. If we insert the transformed estimator $S_p E_N(\Delta) S_p^*$ into Eq. (7) we see immediately that $K_N(\Delta)$ remains unchanged. Replacing $E_N(\Delta)$ by the average $N!^{-1} \sum_{p \in \mathbf{S}_N} S_p E_N(\Delta) S_p^*$ shows that we may assume $[E_N(\Delta), S_p] = 0$ for all permutations p, without loss of estimation quality. A similar argument together with the fact that the quality of the estimate is judged by some criterion not depending on the choice of a basis in $\mathcal{H}$ shows that we may assume in addition that $E_N(\Delta)$ commutes with all unitaries $U^{\otimes N}$. But this implies according to Eqs. (3) and (4) that $E_N(\Delta)$ must be a function of the projection operators $P_Y : \mathcal{H}^{\otimes N} \rightarrow \mathcal{R}_Y \otimes \mathcal{S}_Y$ defined at the beginning of this section. If we require in addition that each $E_N(\Delta)$ be a projection, which is suggestive for ruling out unnecessary fuzziness, E_N must be of the form

$$E_N(\Delta) = \sum_{Y : s_N(Y) \in \Delta} P_Y, \tag{8}$$

where s_N is an arbitrary mapping assigning to each Young frame Y (with d rows and N boxes) an estimate $s_N(Y) \in \Sigma$. In other words, the estimation proceeds by first measuring the Young frame projections P_Y and then computing an estimate $s_N(Y)$ on the basis of the result Y.

The simplest choice is clearly to take the normalized Young frames themselves as the estimate, i.e.,

$$s_N(Y) = Y/N. \tag{9}$$

It turns out somewhat surprisingly that with this choice the $E_N(\Delta)$ from Eq. (8) form an asymptotically exact estimator. By this we mean that, for every ρ, the probability measures K_N from Eq. (7) converge weakly to the point measure at the spectrum r of ρ. Explicitly, for each continuous function f on Σ we have

$$\lim_{N \rightarrow \infty} \int_\Sigma f(s) K_N(ds) = \lim_{N \rightarrow \infty} \sum_Y f\left(\frac{Y}{N}\right) \mathrm{Tr}(\rho^{\otimes N} P_Y) = f(r). \tag{10}$$

We illustrate this in Fig. 1, for $d = 3$ and ρ a density operator with spectrum $r = (0.6, 0.3, 0.1)$. Then Σ is a triangle with corners $A = (1,0,0)$, $B = (1/2,1/2,0)$, and $C = (1/3,1/3,1/3)$, and we plot the probabilities $\mathrm{Tr}(\rho^{\otimes N} P_Y)$ over $Y/N \in \Sigma$. The explicit computation uses the Weyl character formula ([18] Chap. 9, Sec. 9.1), which we do not need elsewhere in the paper.

Clearly, the distribution is peaked at the true spectrum and our claim is that this will become exact in the limit $N \rightarrow \infty$. To prove convergence we will use large deviation methods which give us not only the convergence just stated but an *exponential error estimate* of the form

$$K_N(\Delta) \approx \exp(-N \inf_{s \in \Delta} I(s)), \tag{11}$$

where I denotes a positive function on Σ, called the *rate function*, which vanishes only for $s = r$.

For the statement of the main theorem we say that a measurable subset $\Delta \subset \Sigma$ has "small boundary" if its interior is

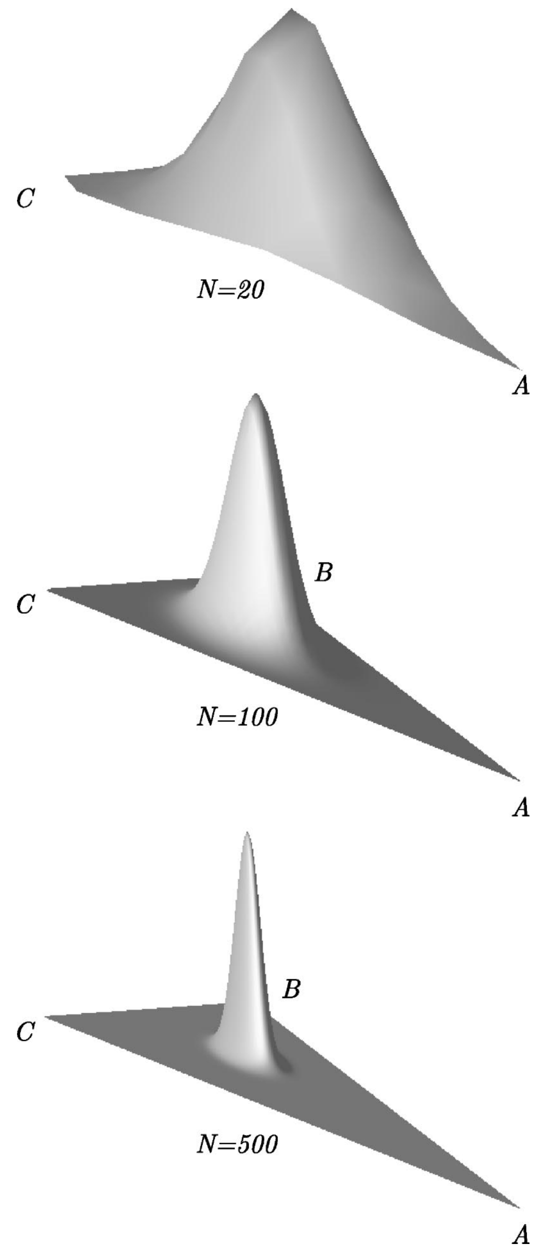

FIG. 1. Probability distribution $\mathrm{Tr}(\rho^{\otimes N}P_Y)$ for $d=3$, $N=20,100,500$, and $r=(0.6,0.3,0.1)$. The set Σ is the triangle with corners $A=(1,0,0)$, $B=(1/2,1/2,0)$, $C=(1/3,1/3,1/3)$.

dense in its closure. A typical choice for Δ is the complement of a ball around the true spectrum.

Theorem. The estimator defined in Eqs. (8) and (9) is asymptotically exact. Moreover, we have the error estimate

$$\lim_{N\to\infty}\frac{1}{N}\ln K_N(\Delta)=\inf_{s\in\Delta}I(s) \tag{12}$$

for any set $\Delta\subset\Sigma$ with small boundary, where the rate function $I:\Sigma\to[0,\infty]$ is

$$I(s)=\sum_j s_j(\ln s_j-\ln r_j). \tag{13}$$

The expression for I is the relative entropy [19] of the probability vectors s and r. Relative entropies occur also as the rate functions in large deviation properties of independent identically distributed (classical [20] or quantum [21]) random variables, although there seems to be no direct way to reduce the above theorem to these standard setups.

III. SKETCH OF PROOF

Rather than giving a proof of every detail, our aim here is to explain why the scaled Young frames Y/N appear in the estimation problem. The crucial observation is that the Young frame $(Y_1,\dots,Y_d)$ is the *highest weight* of the representation π_Y in the ordering $\rhd$ and this ordering is directly related to picking out the fastest growing exponential in certain integrals of the measures K_N.

The integrals we need to study are the Laplace transforms of the measures K_N. We introduce the "scaled cumulant generating function"

$$c(\eta)=\lim_{N\to\infty}\frac{1}{N}\ln\int_\Sigma K_N(ds)e^{N\eta\cdot s}, \tag{14}$$

where $\eta\in\mathbf{R}^d$, and $\eta\cdot s$ is the scalar product. If the measures K_N behave like Eq. (11) the integrand near s behaves like $\exp N[\eta\cdot s-I(s)]$, and the largest contribution comes from the fastest growing exponential:

$$c(\eta)=\sup_s[\eta\cdot s-I(s)]. \tag{15}$$

This is an instance of Varadhan's theorem [22], which has a converse, the Gärtner-Ellis theorem ([16], Theorem II.6.1): if the limit (14) exists and is differentiable then the estimate in the theorem holds, with the rate function determined from Eq. (15) by inverse Legendre transformation. We will follow Duffield [17] by computing the limit (14) from group theoretical data.

Consider the "maximally Abelian subgroup" $C\subset\mathrm{GL}(d,\mathbf{C})$ of diagonal matrices

$$\rho_h=\mathrm{diag}[\exp(h_1),\dots,\exp(h_d)] \tag{16}$$

for $h\in\mathbf{C}^d$. Since these commute, all the operators $\pi_Y(\rho_h)$ commute in every representation π_Y, and can hence be simultaneously diagonalized. The vectors $\mu=(\mu_1,\dots,\mu_d)$ such that $\pi_Y(\rho_h)\psi=\exp(\mu\cdot h)\psi$ for some nonzero vector ψ are called *weights* of the representation π_Y. The dimension $m(\mu)$ of the corresponding eigenspace is called the *multiplicity* of μ. One particular weight (with multiplicity 1) is the Young frame Y itself (interpreted as an element of $\mathbf{R}^d$) and it turns out that Y is the maximum (the "heighest weight") among all weights of π_Y, in the $\rhd$ ordering from Eq. (6). Representation theory of semisimple Lie algebras [18] shows that each irreducible, analytic representation of $\mathrm{GL}(d,\mathbf{C})$ is uniquely characterized (up to unitary equivalence) by its highest weight Y.

In order to estimate the integral (14), we need the quantities $\mathrm{tr}(\rho^{\otimes N}P_Y)$. For simplicity we assume that ρ is nonsingular, i.e., an element of $\mathrm{GL}(d,\mathbf{C})$. By Eq. (3) we have

$$\mathrm{Tr}(\rho^{\otimes N}P_Y)=\mathrm{Tr}[\pi_Y(\rho)\otimes\mathbf{1}]=\chi_Y(\rho)\dim(\mathcal{S}_Y), \tag{17}$$

where

$$\chi_Y(\rho):=\mathrm{tr}[\pi_Y(\rho)] \tag{18}$$

is the *character* of the representation π_Y. Since χ_Y is unitarily invariant $[\chi_Y(U\rho U^*)=\chi_Y(\rho)]$ we may assume without loss of generality that ρ is diagonal and its matrix elements are arranged in descending order. Using the notation from Eq. (16) this assumption reads

$$\rho=\rho_h\in C \quad \text{with} \quad h\triangleright 0 \quad \text{and} \quad \sum_j \exp(h_j)=1. \tag{19}$$

Hence we can express $\chi_Y(\rho)$ in terms of the weights of π_Y:

$$\chi_Y(\rho)=\sum_\mu m(\mu)\exp(\mu\cdot h), \tag{20}$$

where the sum is taken over all weights μ of π_Y. Since $h\triangleright 0$ and $Y\triangleright\mu$ for all μ we see that $\exp(Y\cdot h)$ is the largest exponential. We therefore estimate

$$\exp(Y\cdot h)\leqslant\chi_Y(\rho)\leqslant\dim(\mathcal{R}_Y)\exp(Y\cdot h). \tag{21}$$

Hence, if we introduce for any $h,\eta\in\mathbf{R}^d, h,\eta\triangleright 0$ the two expressions

$$J(h,\eta)=\int_\Sigma K_N(ds)e^{N\eta\cdot s} \tag{22}$$

$$=\sum_Y \mathrm{Tr}(\rho_h^{\otimes N}P_Y)e^{N\eta\cdot Y/N} \tag{23}$$

$$=\sum_Y \chi_Y(\rho_h)e^{\eta\cdot Y}\dim(\mathcal{S}_Y) \tag{24}$$

and

$$J'(h,\eta)=\sum_Y e^{(h+\eta)\cdot Y}\dim(\mathcal{S}_Y) \tag{25}$$

we get

$$J'(h,\eta)\leqslant J(h,\eta)\leqslant\dim(\mathcal{R}_Y)J'(h,\eta). \tag{26}$$

If we combine this with the consequence of Weyl's dimension formula that $\dim(\mathcal{R}_Y)$ is bounded above by a polynomial $p(N)$ in N, uniformly in Y ([17], Lemma 2.2), and take logarithms we get

$$\ln J'(h,\eta)\leqslant\ln J(h,\eta)\leqslant \text{const} \times\ln N+\ln J'(h,\eta). \tag{27}$$

Since $J'(h,\eta)$ grows exponentially in N its logarithm is linear in N and we see that $J(h,\eta)$ and $J'(h,\eta)$ are asymptotically equivalent in the sense that

$$(1/N)[\ln J(h,\eta)-\ln J'(h,\eta)]\rightarrow 0. \tag{28}$$

In the same sense we can continue the chain of equivalences

$$J(h,\eta)\approx J'(h,\eta)=J'(h+\eta,0)\approx J(h+\eta,0) \tag{29}$$

$$=\int_\Sigma K_N(ds). \tag{30}$$

Here we have used Eq. (22) for $J(h+\eta,0)$, and the $h+\eta$ dependence is contained in $K_N(ds)$ via $\rho_{h+\eta}$. Together with the definition of K_N in Eq. (7) this implies

$$J(h,\eta)\approx\int_\Sigma K_N(ds)=K_N(\Sigma)=\mathrm{Tr}(E_N(\Sigma)\rho_{h+\eta}^{\otimes N}) \tag{31}$$

$$=\mathrm{Tr}(\rho_{h+\eta}^{\otimes N})=(\mathrm{tr}\rho_{h+\eta})^N. \tag{32}$$

Hence, if $r_\alpha=\exp(h_\alpha)$ are the eigenvalues of a nonsingular density operator, we get for Eq. (14) the expression

$$c(\eta)=\ln\sum_\alpha r_\alpha\exp(\eta_\alpha). \tag{33}$$

It is then a simple calculus exercise to verify the above rate function as the Legendre transform $I(s)=\sup_\eta[\eta\cdot s-c(\eta)]$.

This concludes our sketch of proof. In order to expand it into a full proof, one needs to extend the computation of $c(\eta)$ to $\eta\not\triangleright 0$, and prove that this extension has the required regularity properties for the application of the converse of Varadhan's theorem cited above. This has been carried out by Duffield [17] in a context that is, on the one hand wider, because it includes tensor powers of much more general representations of semisimple Lie groups, but on the other hand narrower, because it contains only the case $\rho=d^{-1}\mathbf{1}$ of our theorem. However, one can extend Duffield's result by multiplying his measures K_N by the factor $\chi_Y(\rho)/\chi_Y(\mathbf{1})$ and using for this factor the estimate (21).

IV. DISCUSSION

Although the estimate we discuss is asymptotically exact, it is not at all clear whether and in what sense it might be *optimal*, even for finite N. We have experimented with various figures of merit for estimation and found different "optimal" estimators for low N, rarely coinciding with the E_N determined by Eq. (9). It is also not at all clear how much could be gained by optimization here.

An interesting extension will also be the construction of estimators for the full density operator. It is very suggestive to compose this out of the above estimator for the spectrum, and to use for each Young frame a covariant observable to estimate the eigenbasis of ρ. The density of the covariant observable might be based on the highest weight vector of π_Y.

ACKNOWLEDGMENTS

Funding by the European Union project EQUIP (Contract No. IST-1999-11053) and financial support from the DFG (Bonn) are gratefully acknowledged.

[1] W.K. Wootters and W.H. Zurek, Nature (London) **299**, 802 (1982).
[2] R. Werner, Phys. Rev. A **58**, 1827 (1998).
[3] M. Keyl and R.F. Werner, J. Math. Phys. **40**, 3283 (1999).
[4] D. Bruß and C. Macchiavello, Phys. Lett. A **253**, 249 (1999).
[5] B. Efron and R.J. Tibshirani, *An Introduction to the Bootstrap* (Chapman and Hall, New York, 1993).
[6] S. Weigert, in *Trends in Quantum Mechanics*, edited by H.D. Doebner, S.T. Ali, M. Keyl, and R.F. Werner (World Scientific, Singapore, 2000), pp. 146–156.
[7] U. Leonhardt, *Measuring the Quantum State of Light* (Cambridge University Press, Cambridge, 1997).
[8] D.G. Fischer and M. Freyberger, Phys. Lett. A **273**, 293 (2000).
[9] G. Vidal, J.I. Latorre, P. Pascual, and R. Tarrach, Phys. Rev. A **60**, 126 (1999).
[10] R.D. Gill and S. Massar, Phys. Rev. A **61**, 042312 (2000).
[11] H. Weyl, *The Classical Groups* (Princeton University, Princeton, NJ, 1946).
[12] J.I. Cirac, A.K. Ekert, and C. Macchiavello, Phys. Rev. Lett. **82**, 4344 (1999).
[13] A. Acín, R. Tarrach, and G. Vidal, Phys. Rev. A **61**, 062307 (2000).
[14] N. Gisin, e-print quant-ph/0010036, 2001.
[15] M. Keyl and R.F. Werner, Ann. Inst. Henri Poincaré **2**, 1 (2001).
[16] R.S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics* (Springer, Berlin, 1985).
[17] N.G. Duffield, Proc. Am. Math. Soc. **109**, 503 (1990).
[18] B. Simon, *Representations of Finite and Compact Groups* (American Mathematical Society, Providence, RI, 1996).
[19] M. Ohya and D. Petz, *Quantum Entropy and Its Use* (Springer, Berlin, 1993).
[20] H. Cramér, *Colloque Consacré à la Théorie des Probabilités*, Vol. 736 of *Actualités Scientifique et Industrielles* (Hermann, Paris, 1938), pp. 5–23.
[21] R. Werner, in *Quantum Probability and Related Topics*, vol. VII, edited by L. Accardi (World Scientific, Singapore, 1992), pp. 349–381.
[22] S.R.S. Varadhan, Commun. Pure Appl. Math. **19**, 261 (1966).