Fast rate estimation of a unitary operation in SU(d)

Jonas Kahn
Université Paris-Sud 11, Département de Mathématiques, Bâtiment 425, 91405 Orsay Cedex, France
(Received 13 March 2006; revised manuscript received 2 October 2006; published 23 February 2007)

We give an explicit procedure based on entangled input states for estimating a SU(d) operation U with rate of convergence $1/N^2$ when sending N particles through the device. We prove that this rate is optimal. We also evaluate the constant C such that the asymptotic risk is C/N^2. However, other strategies might yield a better constant C.

DOI: 10.1103/PhysRevA.75.022326 PACS number(s): 03.67.-a

I. INTRODUCTION

The question that we are investigating in this paper is: "What is the best way of estimating a unitary operation U?"

By "unitary operation," we mean a device (or a *channel*) that sends a density operator ρ_0 on $\mathbb{C}^d$ to another density operator $\rho = U\rho_0 U^*$, where $U \in \mathrm{SU}(d)$, a special unitary matrix.

We immediately stress that the solution to this estimation problem can be divided into two parts: what is the input state, and which measurement (POVM, see Def. 1) should we apply on the output state? Indeed, in order to estimate the channel U, we have to let it act on a state (the input state). And once we have the output state, the problem consists in discriminating states in the family of possible output states.

This estimation of unitary operation has been extensively studied over the last few years.

The first invitation was Ref. [1], featuring numerous special cases. In most of those, the unitary U is known to belong to some subset of SU(2).

Then Ref. [2] provided the form of an optimal state to be sent in with non-specified coefficients depending on the cost function [we give the formula of this state in Eq. (2.2)]. In that paper the authors consider the situation where the unitary operation is performed independently on N systems. That study applied to any SU(d), and any covariant loss function, in particular fidelity, in a Bayesian framework. The proposed input state uses an ancilla, that is, an auxiliary system that is not sent through the unitary channel with Hilbert space $(\mathbb{C}^d)^{\otimes N}$. The state is prepared as a superposition of maximally entangled states, one for each irreducible representation of SU(d) appearing in $(\mathbb{C}^d)^{\otimes n}$. We emphasize that the state is an entangled state of $(\mathbb{C}^d)^{\otimes N} \otimes (\mathbb{C}^d)^{\otimes N}$: we do not send N copies of an entangled state through the device, but all the N systems that are sent through the channel together with the N particles of the ancilla are part of the same entangled state, yielding the most general possible strategy. There was no evaluation of the rate of convergence, though.

Subsequent works mainly focused on SU(2), as the case is simpler and yields many applications, e.g., transmission of reference frames in quantum communication. Indeed, the latter is equivalent to the estimation of a SU(2) operation. The first strategy to be proved to converge (in fidelity) at $1/N^2$ rate was not covariant [3]. It made no use of an ancilla. Later, the same rate was achieved for a covariant measurement with an ancilla [4] through a judicious choice of the coefficients left free in the state proposed in Ref. [2]. The optimal constant (π^2/N^2 for the fidelity) was also computed. It was almost simultaneously noticed [5,6] that asymptotically the ancilla is unnecessary. Indeed, what we need is entangling different copies of the same irreducible representation. Now each irreducible representation appears with multiplicity in $(\mathbb{C}^d)^{\otimes N}$, most of them with higher multiplicity than dimension, which is the condition we need. This method was dubbed "self-entanglement." The advantage is that we need to prepare half the number of particles, as we do not need an ancilla. In all these articles, the Bayesian paradigm with uniform prior was used. The same $1/N^2$ rate was shown to hold true in a minimax sense, in pointwise estimation [7]. We stress the importance of this $1/N^2$ rate, proving how useful entanglement can be. Indeed, in classical data analysis, we cannot expect a better rate than $1/N$. Similarly, the $1/N$ bound holds for any strategy where the N particles we send through the device are not entangled "among themselves" (that is, even if there is an ancilla for each of these N particles).

Another popular theme has been the determination of the phase ϕ for unitaries of the form $U_\phi = e^{i\phi H}$. This very special case already has many applications, especially in interferometry or measurement of small forces, as featured in the review article [8] and references therein. A common feature of the most efficient techniques is the need for entangled states of many particles, and much experimental work has aimed at generating such states. These methods essentially involve either manipulation of photons obtained through parametric down-conversion (for example, Ref. [9]), ions in ion traps (for example, Ref. [10]) or atoms in cavity QED (for example, Ref. [11]).

In recent years, there has been renewed interest in the SU(d) case. Notably, Ref. [12] takes off from Ref. [2], allowing for more general symmetries and making explicit for natural cost functions both the free coefficients—as the coordinates of the eigenvector of a matrix—and the POVM (see theorem 1 below). With a completely different strategy, aiming rather at pointwise estimation (and, therefore, minimax theorems), an input state for $U^{\otimes n}$ was found [13,14] such that the quantum Fisher information matrix is scaling like $1/N^2$, yielding hopes of getting as fast an estimator for SU(d). No associated measurement was found in that paper.

Given the state of the art, a natural question is whether we can obtain, as for SU(2), this dramatic increase in performance when using entanglement for general SU(d). That is,

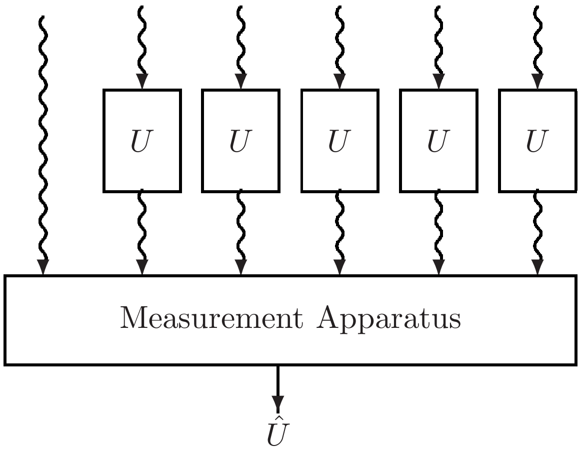

FIG. 1. Most general estimation scheme of U when n copies are available at the same time, and using entanglement.

do we have an estimation procedure whose rate is $1/N^2$, instead of $1/N$? Neither Ref. [12], where the asymptotics are not studied for SU(d), nor Ref. [13], where no measurement is given, answer this question.

In this article, we first prove that we cannot expect a better rate than $1/N^2$. This kind of bound based on the laws of quantum physics, without any *a priori* on the experimental device, is traditionally called the Heisenberg limit of the problem. Then we choose a completely explicit input state of the form (2.2) (as in Ref. [2]), by specifying the coefficients. By using the associated POVM, the estimator of a unitary quantum operation $U \in \mathrm{SU}(d)$ converges at rate $1/N^2$. The constant is not optimal, but is briefly studied at the end of the paper. We obtain these results with fidelity as a cost function, both in a Bayesian setting, with a uniform prior, and in a minimax setting. Notice that we shall not need an ancilla.

The next section consists in formulating the problem and restating theorem 2 of Ref. [12] within our framework. Section III then shows that it is impossible to converge at a rate faster than $O(N^{-2})$. In Sec. IV, we write a general formula for the risk of a strategy as described in theorem 1, and in Sec. V we specify our estimators by choosing our coefficients in (4). We then prove that the risk of this estimator is $O(N^{-2})$. The last section (Sec. VI) consists in finding the precise asymptotic speed of our procedure, that is the constant C in CN^{-2}. We finish by stating in theorem 2 the results of the paper.

II. DESCRIPTION OF THE PROBLEM

We are given an unknown unitary operation $U \in \mathrm{SU}(d)$ and must estimate it "as precisely as possible." We are allowed to let it act on N particles, so that we are discriminating between the possible $U^{\otimes N}$. We shall work both with pointwise estimation (as preferred by mathematicians) and with a Bayes uniform prior (a favorite of physicists).

Any estimation procedure can be described as follows (see Fig. 1): the unitary channel $U^{\otimes N}$ acts as

$$U^{\otimes N} \otimes 1 : (\mathbb{C}^d)^{\otimes N} \otimes \mathcal{K} \to (\mathbb{C}^d)^{\otimes N} \otimes \mathcal{K},$$

on the space of the N systems together with a possible ancilla. The input state $\rho_n \in M((\mathbb{C}^d)^{\otimes n} \otimes \mathcal{K}_n)$ is mapped into an output state on which we perform a measurement M whose result is the estimator $\hat{U} \in \mathrm{SU}(d)$.

Recall that a measurement is mathematically defined by a positive operator-valued measure:

Definition 1: A positive operator-valued measure (POVM for short) on a Hilbert space $\mathcal{H}$ with values in a probability space $(\mathcal{X}, \mathcal{A})$ is a collection of operators $M(A)$ on $\mathcal{H}$ for $A \in \mathcal{A}$ such that:

$M(A)$ is a nonnegative for any A,

$M(\mathcal{X}) = \mathbf{1}_{\mathcal{H}}$,

for any $(A_i)_{i\in\mathcal{N}}$ with A_i two by two disjoint, $M(\cup A_i) = \Sigma M(A_i)$.

In order to evaluate the quality of an estimator $\hat{U}$, we fix a cost function $\Delta(U,V)$. The global pointwise risk of the estimator is

$$R_P(\hat{U}) = \sup_{U\in\mathrm{SU}(d)} \mathbb{E}_U[\Delta(U,\hat{U})].$$

The probability distribution of $\hat{U}$ depends on U, and we take expectation with respect to this probability distribution.

On the other hand, the Bayes risk with uniform prior is

$$R_B(\hat{U}) = \int_{\mathrm{SU}(d)} \mathbb{E}_U[\Delta(U,\hat{U})]\mathrm{d}\mu(U),$$

where μ is the Haar measure on SU(d).

As cost function, we choose the fidelity F (or rather $1-F$), which for an element of SU(d) is defined as

$$\Delta(U,\hat{U}) = 1 - \frac{|\mathrm{Tr}(U^{-1}\hat{U})|^2}{d^2} = 1 - \frac{|\chi_\square(U^{-1}\hat{U})|^2}{d^2},$$

where $\chi_\square$ is the character of the defining representation of SU(d), whose Young tableau consists in only one box. In other words, $\chi_\square(U) = \mathrm{Tr}(U)$.

Before really addressing the problem, we make a few remarks on why this choice of distance is suitable for mathematical analysis.

First, this cost function is covariant, i.e., $\Delta(U,\hat{U}) = \Delta(1_{\mathbb{C}^d}, U^{-1}\hat{U})$.

Second, a useful feature within the Bayesian framework is that Δ is of the form (2.1), as required in theorem 1. Indeed we can rewrite $\Delta(U,\hat{U})$ as $1-\chi_\square(U^{-1}\hat{U})\chi^*_\square(U^{-1}\hat{U})/d^2$. Now the conjugate of a character is the character of the adjoint representation, the product of two characters is again the character of a possibly reducible representation π. This character is equal to the sum of the characters of the irreducible representations appearing in the Clebsch-Gordan development of π, in which all coefficients are non-negative. Therefore, $\Delta = 1 - (\Sigma_{\vec{\lambda}} a_{\vec{\lambda}} \chi^*_{\vec{\lambda}})$ where $a_{\vec{\lambda}} \geqslant 0$ and $\vec{\lambda}$ runs over all irreducible representations of SU(d). That is the condition (2.1) that we shall need for applying Theorem 1, given at the end of the section.

On the other hand, the theory of pointwise estimation deals usually with the variance of the estimated parameters when we use a smooth parameterization of SU(d). As we want to use the quantum Cramér-Rao bound (3.4), we need

Δ to be quadratic in the parameters to the first order, and positive lower bounded for $\hat{U}$ outside a neighborhood of U. As Δ is covariant, it is sufficient to check this with $U=\mathbf{1}_{\mathbb{C}^d}$. Now an example of a smooth parameterization in a neighborhood of the identity is $U(\theta)=\exp(\Sigma_\alpha \theta_\alpha T_\alpha)$ where $\theta \in \mathbb{R}^{d^2-1}$ and the T_α are generators of the Lie algebra, so that $\mathrm{Tr}(T_\alpha)=0$. Now $\mathrm{Tr}[\exp(\Sigma_\alpha \theta_\alpha T_\alpha)]=d+\Sigma_\alpha \theta_\alpha \mathrm{Tr}(T_\alpha)+O(\|\theta\|^2)$, so that the trace minus d and, consequently, Δ, is quadratic in θ to the first order. We shall write $\vec{\lambda}_1 \otimes \vec{\lambda}_2$ for the tensor representation of two irreducible representations $\vec{\lambda}_1$ and $\vec{\lambda}_2$.

As stated at the beginning of this section, we are working with $U^{\otimes N}$. The Clebsch-Gordan decomposition of the nth tensor product representation is

$$U^{\otimes N} = \bigoplus_{\vec{\lambda}:|\vec{\lambda}|=N} U^{\vec{\lambda}} \otimes 1_{\mathbb{C}^{\mathcal{M}(\vec{\lambda})}}$$

acting on $\oplus_{\vec{\lambda}:|\vec{\lambda}|=N}\mathcal{H}^{\vec{\lambda}} \otimes \mathbb{C}^{\mathcal{M}(\vec{\lambda})}$, where $\mathcal{H}^{\vec{\lambda}}=\mathbb{C}^{\mathcal{D}(\vec{\lambda})}$ is the representation space of $\vec{\lambda}$, $\mathcal{M}(\vec{\lambda})$ is the multiplicity of $\vec{\lambda}$ in the nth tensor product representation, and $\mathcal{D}(\vec{\lambda})$ the dimension of $\vec{\lambda}$. We refer to $\mathbb{C}^{\mathcal{M}(\vec{\lambda})}$ as the multiplicity space of $\vec{\lambda}$. We have indexed the irreducible representations of $\mathrm{SU}(d)$ by $\vec{\lambda}=(\lambda_1,\ldots,\lambda_d)$, and written $|\vec{\lambda}|=\Sigma_{i=1}^d \lambda_i$. Notice that this labelling of irreducible representations is redundant, but that if $|\vec{\lambda}^1|=|\vec{\lambda}^2|$, then $\vec{\lambda}^1$ and $\vec{\lambda}^2$ are equivalent (denoted $\vec{\lambda}^1 \equiv \vec{\lambda}^2$) if and only if $\vec{\lambda}^1=\vec{\lambda}^2$.

The starting point of our argument will be the following reformulation of the results of Ref. [12], with less generality and without the formula for the risk whose form is not adapted to our subsequent analysis:

Theorem 1. (Ref. [12]) Let $U \in \mathrm{SU}(d)$ be a unitary operation to be estimated, through its action on N particles. We may use entanglement and/or an ancilla.

Then, for a uniform prior and any cost function of the form

$$c(U,\hat{U}) = a_0 - \sum_{\vec{\lambda}} a_{\vec{\lambda}} \chi^*_{\vec{\lambda}}(U^{-1}\hat{U}), \tag{2.1}$$

we can find as optimal input state a pure state of the form

$$|\Psi\rangle = \bigoplus_{\vec{\lambda}:|\vec{\lambda}|=N} \frac{c(\vec{\lambda})}{\sqrt{\mathcal{D}(\vec{\lambda})}} \sum_{i=1}^{\mathcal{D}(\vec{\lambda})} |\psi_i^{\vec{\lambda}}\rangle \otimes |\phi_i^{\vec{\lambda}}\rangle \tag{2.2}$$

with $c(\vec{\lambda}) \geq 0$, and the normalization condition,

$$\sum_{\vec{\lambda}} c(\vec{\lambda})^2 = 1. \tag{2.3}$$

Moreover, $|\psi_i^{\vec{\lambda}}\rangle$ is an orthonormal basis of $\mathcal{H}^{\lambda}$ and $|\phi_i^{\vec{\lambda}}\rangle$ are orthonormal vectors of the multiplicity space, which may be augmented by an ancilla if necessary (see remark below on the dimensions).

The corresponding measurement is the covariant POVM with seed $\Xi=|\eta\rangle\langle\eta|$ given by:

$$|\eta\rangle = \bigoplus_{\vec{\lambda}|c(\vec{\lambda})\neq 0} \sqrt{\mathcal{D}(\vec{\lambda})} \sum_{i=1}^{\mathcal{D}(\vec{\lambda})} |\psi_i^{\vec{\lambda}}\rangle \otimes |\phi_i^{\vec{\lambda}}\rangle, \tag{2.4}$$

that is a POVM whose density with respect to the Haar measure is given by $m(U)=U|\eta\rangle\langle\eta|U^*$ with

$$U|\eta\rangle = \bigoplus_{\vec{\lambda}|c(\vec{\lambda})\neq 0} \sqrt{\mathcal{D}(\vec{\lambda})} \sum_{i=1}^{\mathcal{D}(\vec{\lambda})} U^{\vec{\lambda}}|\psi_i^{\vec{\lambda}}\rangle \otimes |\phi_i^{\vec{\lambda}}\rangle.$$

Remark. We use $\mathcal{D}(\vec{\lambda})$ orthonormal vectors in the multiplicity space of $\vec{\lambda}$. This requires $\mathcal{M}(\vec{\lambda}) \geq \mathcal{D}(\vec{\lambda})$. If this is not the case, we must increase the dimension of the multiplicity space by using an ancilla in $\mathbb{C}^{\delta}$. Then the action of U is $U^{\otimes N} \otimes \mathbf{1}_{\mathbb{C}^\delta}$ whose Clebsch-Gordan decomposition is $\oplus_{\vec{\lambda}||\vec{\lambda}|=N} U^{\vec{\lambda}} \otimes \mathbf{1}_{\mathbb{C}^{\delta\mathcal{M}(\vec{\lambda})}}$. With big enough δ, we have $\delta\mathcal{M}(\vec{\lambda}) \geq \mathcal{D}(\vec{\lambda})$. Notice that an ancilla is not necessary if $c(\vec{\lambda})=0$ for all $\vec{\lambda}$ such that $\mathcal{D}(\vec{\lambda}) > \mathcal{M}(\vec{\lambda})$.

Another remark is that, as defined, our POVM is not properly normalized: $M(\mathrm{SU}(d)) \neq \mathbf{1}$, but is equal to the projection on the space spanned by the $U|\Psi\rangle$. As this is the only subspace of importance, we can complete the POVM (through the seed, for example) *ad libitum*.

Our estimator $\hat{U}$ is the result of the measurement with POVM defined by (2.4) and input state of the form (2.2), with specific $c(\vec{\lambda})$. Such an estimator is covariant, that is $p_U(\hat{U})=p_{1_{\mathbb{C}^d}}(U^{-1}\hat{U})$, where p_U is the probability distribution of $\hat{U}$ when we are estimating U. The cost function is also covariant, so that $\mathbb{E}_U[\Delta(U,\hat{U})]$ does not depend on U. This implies that the Bayesian risk and the pointwise risk coincide. With the second equality true for all $U \in \mathrm{SU}(d)$, we have

$$R_B(\hat{U}) = R_P(\hat{U}) = \mathbb{E}_U[\Delta(U,\hat{U})]. \tag{2.5}$$

Theorem 1 states that there exists an optimal (Bayes uniform) estimator $\hat{U}_o$ of this form [corresponding to the optimal choice of $c(\vec{\lambda})$], so that it obeys (2.5). From this we first prove that no estimator whatsoever can have a better rate than $1/N^2$.

III. WHY WE CANNOT EXPECT BETTER RATE THAN $1/N^2$

For proving this result, we need the Bayesian risk for prior π other than the uniform prior:

$$R_\pi(\hat{U}) = \mathbb{E}_\pi[\mathbb{E}_U[\Delta(U,\hat{U})]].$$

As $\hat{U}_o$ is Bayesian optimal for the uniform prior, we only have to prove that $R_B(\hat{U}_o)=O(N^{-2})$. This is also sufficient for pointwise risk as, for any estimator $\hat{U}$, we have $R_B(\hat{U}) \leq R_P(\hat{U})$. Moreover, as $\mathbb{E}_U[\Delta(U,\hat{U}_o)]$ does not depend on U, $R_\pi(\hat{U}_o)=R_B(\hat{U}_o)$. It is then sufficient to prove, for a π of our choice, that

$$R_\pi(\hat{U}_o) = O(N^{-2}). \tag{3.1}$$

The idea is to find a Cramér-Rao bound that we can apply to some π. We shall combine the Braunstein and Caves information inequality (3.3) and the Van Trees inequality (3.2) to obtain the desired quantum Cramér-Rao bound, much in the spirit of Ref. [15]. This bound will yield an explicit rate through a result of [13].

Van Trees' inequality states that given a classical statistical model smoothly parameterized by $\theta \in \Theta \subset \mathbb{R}^p$, and a smooth prior with compact support $\Theta_0 \subset \Theta$, then for any estimator $\hat{\theta}$, we have

$$\mathbb{E}_\pi[\mathrm{Tr}\{V_\theta(\hat{\theta})\}] \geq \frac{p^2}{\mathbb{E}_\pi[\mathrm{Tr}\{I(\theta)\}] - \mathcal{I}_\pi}, \tag{3.2}$$

where $I(\theta)$ is the Fisher information matrix of the model at point θ, $\mathcal{I}_\pi$ is a finite (for reasonable π) constant depending on π (quantifying in some way the prior information), and $V_\theta(\hat{\theta}) \in M_p(\mathbb{R})$ is the mean square error (MSE) of the estimator $\hat{\theta}$ at point θ given by

$$V_\theta(\hat{\theta})_{\alpha,\beta} = \mathbb{E}[(\theta_\alpha - \hat{\theta}_\alpha)(\theta_\beta - \hat{\theta}_\beta)].$$

This form of Van Trees inequality is obtained by setting $N=1$, $G=C=Id$, and $\psi=\theta$ in Eq. (12) of Ref. [15].

Now the Braunstein and Caves information inequality [16] yields an upper bound on the information matrix $I_M(\theta)$ of any classical statistical model obtained by applying the measurement M to a quantum statistical model. For any family of quantum states parameterized by a p-dimensional parameter $\theta \in \Theta \in \mathbb{R}^p$, for any measurement M on these states, the following holds:

$$I_M(\theta) \leq H(\theta), \tag{3.3}$$

where $H(\theta)$ is the quantum Fisher information matrix at point θ.

Now it was proved in Ref. [13] that for a smooth parameterization of an open set of SU(d), and for any input state, the quantum Fisher information of the output states fulfills

$$H(\theta) = O(N^2).$$

Inserting in (3.2) together with (3.3), we get as quantum Cramér-Rao bound

$$\mathbb{E}_\pi[\mathrm{Tr}(V_\theta(\hat{\theta}))] = O\left(\frac{1}{N^2}\right). \tag{3.4}$$

We now want to apply this bound to obtain Eq. (3.1). There are a few small technical difficulties. First of all, we cannot use the uniform prior for π as SU(d) is not homeomorphic to an open set of $\mathbb{R}^p$. We then have to define two neighborhoods of the identity $\Theta_0 \subset \Theta$, allowing to use of the Van Trees inequality. Now our estimator $\hat{U}_o$ need not be in Θ, so that we shall in fact apply Van Trees inequality to a modified estimator $\tilde{U}$. Finally, this bound is on the variance, and we must relate it to Δ.

Our first task consists in restricting our attention to a neighborhood Θ of $\mathbf{1}_{\mathbb{C}^d}$. It corresponds to a neighborhood Θ (we use the same notation) of $0 \in \mathbb{R}^p$ through $U = \exp(\Sigma_\alpha \theta_\alpha T_\alpha)$. This holds if the neighborhood is small enough, so we define it by $U \in \Theta$ if and only if $\Delta(1_{\mathbb{C}^d}, U) < \epsilon$ for a fixed small enough ϵ. We define Θ_0 through $U \in \Theta_0$ for $\Delta(1_{\mathbb{C}^d}, U) \leq \epsilon/3$, and take a smooth fixed prior π with support in Θ_0, such that $\mathcal{I}_\pi < \infty$.

Now we modify our estimator $\hat{U}_o$ into an estimator $\tilde{U}$ given by $\tilde{U} = \hat{U}_o$ for $\hat{U}_o \in \Theta$ and $\tilde{U} = \mathbf{1}_{\mathbb{C}^d}$ for $\hat{U}_o \in \Theta$. Then, by the triangle inequality, for any $U \in \Theta_0$, we have $\Delta(U, \hat{U}_o) \geq \Delta(U, \tilde{U})$.

The fundamental point of the reasoning [used at Eq. (3.5)] is that, as Δ is quadratic at the first-order, there is a positive constant c such that, for any $U^1, U^2 \in \Theta$, corresponding to θ^1, θ^2, we have $\Delta(U_1, U_2) \geq c \Sigma_\alpha (\theta^1_\alpha - \theta^2_\alpha)^2$.

Finally, we get

$$R_\pi(\hat{U}_o) = \mathbb{E}_\pi[\mathbb{E}_U[\Delta(U, \hat{U}_o)]] \geq \mathbb{E}_\pi[\mathbb{E}_U[\Delta(U, \tilde{U})]] \geq c \mathbb{E}_\pi[V_{\tilde{\theta}}] = O(N^{-2}). \tag{3.5}$$

We have thus proved Eq. (3.1), and hence our bound on the efficiency of any estimator.

We now write formulas for the risk of any estimator of the form given in Theorem 1.

IV. FORMULAS FOR THE RISK

By (2.5), our risk $R_P(\hat{U})$ is equal to the pointwise risk at $\mathbf{1}_{\mathbb{C}^d}$, with which we shall work

$$\int_{\mathrm{SU}(d)} p_{1_{\mathbb{C}^d}}(\hat{U}) \left\{ 1 - \frac{|\chi_\square(\hat{U})|^2}{d^2} \right\} d\mu(\hat{U}). \tag{4.1}$$

Now we compute the probability distribution of $\hat{U}$ for a given $|\Psi\rangle$ of the form (2.2), that is

$$p_{1_{\mathbb{C}^d}}(\hat{U}) = \langle \Psi | \hat{U} \Xi \hat{U}^* | \Psi \rangle = \left| \sum_{\vec{\lambda}: |\vec{\lambda}| = N} \frac{c(\vec{\lambda})}{\mathcal{D}(\vec{\lambda})} \mathcal{D}(\vec{\lambda}) \sum_{i=1}^{\mathcal{D}(\vec{\lambda})} \langle \psi_i^{\vec{\lambda}} | U | \psi_i^{\vec{\lambda}} \rangle \right|^2 = \left| \sum_{\vec{\lambda}: |\vec{\lambda}| = N} c(\vec{\lambda}) \chi_{\vec{\lambda}}(\hat{U}) \right|^2,$$

where we have used the character $\chi_{\vec{\lambda}}$ of $\vec{\lambda}$ as the trace of U in the representation.

Then, using (4.1), recalling that $p_{1_{\mathbb{C}^d}}$ is a probability density with respect to the Haar measure μ on SU(d), and that $\chi_{\vec{\lambda}^1} \chi_{\vec{\lambda}^2} = \chi_{\vec{\lambda}^1 \otimes \vec{\lambda}^2}$, we get

$$R_P(\hat{U}) = 1 - \frac{1}{d^2} \int_{\mathrm{SU}(d)} \left| \sum_{\vec{\lambda}: |\vec{\lambda}| = N} c(\vec{\lambda}) \chi_{\vec{\lambda} \otimes \square}(\hat{U}) \right|^2 d\mu(\hat{U}). \tag{4.2}$$

In order to evaluate the second term, we use the following orthogonality relations for characters:

$$\int_{\mathrm{SU}(d)} d\mu(U) \chi_{\vec{\lambda}^1}(U) \chi_{\vec{\lambda}^2}(U)^* = \delta_{\vec{\lambda}^1 \equiv \vec{\lambda}^2}. \tag{4.3}$$

To do so we need the Clebsch-Gordan series of $\vec{\lambda} \otimes \square$,

$$\vec{\lambda} \otimes \square = \oplus_{\{1 \leq i \leq d | \lambda_i > \lambda_{i+1}\}} \vec{\lambda} + e_i, \tag{4.4}$$

where conventionally $\lambda_{d+1}=0$. Here we see $\vec{\lambda}$ as a d-dimensional vector and e_i as the ith basis vector.

We then reorganize the sum of characters as

$$\sum_{\vec{\lambda}:|\vec{\lambda}|=N} c(\vec{\lambda}) \chi_{\vec{\lambda}\otimes\square}(\hat{U}) = \sum_{\vec{\lambda}':|\vec{\lambda}'|=N+1} \sum_{i \in \mathcal{S}(\vec{\lambda}')} c(\vec{\lambda}'-e_i) \chi_{\vec{\lambda}'}(\hat{U}),$$

where $\mathcal{S}(\vec{\lambda}')$ is the set of i between 1 and d such that $\vec{\lambda}-e_i$ is still a representation, that is $\lambda'_i > \lambda'_{i+1}$. We shall write $\#\mathcal{S}(\vec{\lambda}')$ for its cardinality.

Inserting in (4.2) and remembering (4.3), we are left with

$$R_P(\hat{U}) = 1 - \frac{\sum_{\vec{\lambda}':|\vec{\lambda}'|=N+1} \left| \sum_{i \in \mathcal{S}(\vec{\lambda}')} c(\vec{\lambda}'-e_i) \right|^2}{d^2}. \tag{4.5}$$

To go any further, we must work with specific $c(\vec{\lambda})$.

V. CHOICE OF THE COEFFICIENTS $c(\vec{\lambda})$ AND PROOF OF THEIR EFFICIENCY

We now have to choose the coefficients $c(\vec{\lambda})$ so that the right-hand side of (4.5) is small.

It appears useful to introduce subsets of the set of all irreducible representations. Let $\mathcal{P}_N=\{\vec{\lambda} \| \vec{\lambda}|=N; \lambda_1 > \cdots > \lambda_d > 0\}$. Obviously, if $\vec{\lambda} \in \mathcal{P}_{N+1}$, then $\#\mathcal{S}(\vec{\lambda})=d$, and the converse is true. We can see them intuitively as points on $a(d-1)$-dimensional surface and, with this picture in mind, we shall speak of the border of $\mathcal{P}_N$ (when $\lambda_i=\lambda_{i+1}+1$ for some i), or of being far from the border (without precise mathematical meaning).

We are ready to give heuristic arguments on how good coefficients should behave.

We must try to get the fraction in (4.5) close to one. Now

$$\begin{aligned}
&\frac{\sum_{\vec{\lambda}':|\vec{\lambda}'|=N+1} \left| \sum_{i \in \mathcal{S}(\vec{\lambda}')} c(\vec{\lambda}'-e_i) \right|^2}{d^2} \\
&\leq \sum_{\vec{\lambda}':|\vec{\lambda}'|=N+1} \frac{\#\mathcal{S}(\vec{\lambda}')}{d} \frac{\sum_{i \in \mathcal{S}(\vec{\lambda}')} |c(\vec{\lambda}'-e_i)|^2}{d} \\
&\leq \sum_{\vec{\lambda}':|\vec{\lambda}'|=N+1} \frac{\sum_{i \in \mathcal{S}(\vec{\lambda}')} |c(\vec{\lambda}'-e_i)|^2}{d} \\
&\leq \sum_{\vec{\lambda}:|\vec{\lambda}|=N} |c(\vec{\lambda})|^2 = 1.
\end{aligned}$$

The first inequality was obtained using Cauchy-Schwarz inequality for each inner sum. There is equality if $c(\vec{\lambda}-e_i)$ does not depend on i. From this we deduce that for most $\vec{\lambda}$, the $c(\vec{\lambda}-e_i)$ must be approximately equal, especially if they are large. The second inequality follows from $\#\mathcal{S}(\vec{\lambda}') \leq d$. From this we deduce that for $\vec{\lambda} \in \mathcal{P}_{N+1}$, the coefficients $c(\vec{\lambda}-e_i)$ must be small. Remark that about $1/N$ of the $\vec{\lambda}'$ such that $|\vec{\lambda}'|=N+1$ are not in $\mathcal{P}_{N+1}$, so that if all $c(\vec{\lambda})$ were equal, these border terms would cause our rate to be $1/N$. The key of the third inequality is to notice that each $c(\vec{\lambda})$ is appearing in the sum once for each term in its Clebsch-Gordan series (4.4), and that there are at most d terms. Please note that there are d terms if $\vec{\lambda} \in \mathcal{P}_N$, and if $\vec{\lambda}$ is in $\mathcal{P}_{N+1}$, far from the border, then $\vec{\lambda}'-e_i$ is in $\mathcal{P}_N$, far from the border.

The conclusion of these heuristics is that we must choose coefficients "locally" approximately equal (at most $1/N$ variation in ratio), and that the coefficients must go to 0 when we are approaching the border of $\mathcal{P}_N$.

One weight satisfying these heuristics is the following:

$$c(\vec{\lambda}) = \mathcal{N} \prod_{i=1}^{d} p_i, \tag{5.1}$$

where $\mathcal{N}$ is a normalization constant to ensure that (2.3) is satisfied and $p_i=\lambda_i-\lambda_{i+1}$. We shall use it below, and prove that it delivers the $1/N^2$ rate.

A first remark about these weights is that $c(\vec{\lambda})=0$ if $\vec{\lambda} \in \mathcal{P}_N$. Now, for any $\vec{\lambda} \in \mathcal{P}_N$, we have $\mathcal{D}(\vec{\lambda}) \geq \mathcal{M}(\vec{\lambda})$, so that we do not need an ancilla.

Indeed, using hook formulas (see p. 131 and p. 215 of Ref. [17]), we get $\mathcal{M}(\vec{\lambda})/\mathcal{D}(\vec{\lambda})=N!\prod_{i=1}^{d}\frac{(\lambda_i+d-i)!}{(d-i)!}$. Now for $\vec{\lambda} \in \mathcal{P}_N$, we know that $\lambda_i \neq 0$. Under this constraint and $\Sigma\lambda_i = N$, the maximum is attained by $\lambda_1=N-d+1$ and $\lambda_i=1$ for $i\neq 1$. We end up with exactly 1.

We shall now use Eq. (5.1) and express the numerator of (4.5) with our choice of p_i. Notice first that if p_j characterize $\vec{\lambda}'$, then those which characterize $\vec{\lambda}'-e_i$ are given by $p_j^{(i)} = p_j+\delta_{j,i-1}-\delta_{j,i}$. So,

$$\mathcal{N}^{-1} c(\vec{\lambda}'-e_i) = \prod_{j=1}^{d} p_j + r_{\vec{\lambda}'}(i),$$

with

$$r_{\vec{\lambda}'}(i) = -\prod_{j \neq i} p_j + \delta_{j>1}\left(\prod_{j \neq i-1} p_j - \prod_{j \neq i, i-1} p_j \right).$$

Introducing another notation will make this slightly more compact. For a vector $\vec{x}$ with d components and $\mathcal{E}$ a subset of $\{1,\ldots,d\}$, define

$$x_{\mathcal{E}} = \prod_{j \neq \mathcal{E}} x_j. \tag{5.2}$$

Then,

$$r_{\vec{\lambda}'}(i) = -p_{\{i\}} + \delta_{j>1}(p_{\{i-1\}} - p_{\{i,i-1\}}).$$

Notice now that for $\vec{\lambda} \in \mathcal{P}_N$, there are exactly d irreducible representations appearing in the Clebsch-Gordan decomposition of $\vec{\lambda} \otimes \square$ (4.4). So that $c(\vec{\lambda})^2$ appears exactly d times in $\Sigma_{\vec{\lambda}':|\vec{\lambda}|=N+1} \Sigma_{i \in \mathcal{S}(\vec{\lambda}')} c(\vec{\lambda}'-e_i)^2$. We may then rewrite the renormalization constant $\mathcal{N}$ as

$$d^{-1}\sum_{\vec{\lambda}':|\vec{\lambda}'|=N+1}\sum_{i\in\mathcal{S}(\vec{\lambda}')}\prod_{j=1}^{d}p_j^{(i)2}.$$

Therefore, rewriting the second term in (4.5) with our values of $c(\vec{\lambda})$, we aim at proving

$$\frac{\sum_{\vec{\lambda}':|\vec{\lambda}'|=N+1}\left(\sum_{i\in\mathcal{S}(\vec{\lambda}')}\prod_{j=1}^{d}p_j+r_{\vec{\lambda}'}(i)\right)^2}{d\sum_{\vec{\lambda}':|\vec{\lambda}'|=N+1}\sum_{i\in\mathcal{S}(\vec{\lambda}')}\left(\prod_{j=1}^{d}p_j+r_{\vec{\lambda}'}(i)\right)^2}=1+O(N^{-2}). \tag{5.3}$$

Let us expand the numerator,

$$\sum_{\vec{\lambda}':|\vec{\lambda}'|=N+1}\left(\sum_{i\in\mathcal{S}(\vec{\lambda}')}\prod_{j=1}^{d}p_j+r_{\vec{\lambda}'}(i)\right)^2=C_t(1+t_1+t_2),$$

with

$$C_t=\sum_{\vec{\lambda}'}(\#\mathcal{S}(\vec{\lambda}'))^2\prod_{j=1}^{d}p_j^2,$$

$$t_1=\frac{2\sum_{\vec{\lambda}'}\sum_{i\in\mathcal{S}(\vec{\lambda}')}\#\mathcal{S}(\vec{\lambda}')r_{\vec{\lambda}'}(i)\prod_{j=1}^{d}p_j}{C_t},$$

$$t_2=\frac{\sum_{\vec{\lambda}'}\left(\sum_{i\in\mathcal{S}(\vec{\lambda}')}r_{\vec{\lambda}'}(i)\right)^2}{C_t}.$$

Similarly the denominator can be read as

$$d\sum_{\vec{\lambda}':|\vec{\lambda}'|=N+1}\sum_{i\in\mathcal{S}(\vec{\lambda}')}\left(\prod_{j=1}^{d}p_j+r_{\vec{\lambda}'}(i)\right)^2=C_u(1+u_1+u_2),$$

with

$$C_u=\sum_{\vec{\lambda}'}d\,\#\mathcal{S}(\vec{\lambda}')\prod_{j=1}^{d}p_j^2,$$

$$u_1=\frac{2d\sum_{\vec{\lambda}'}\sum_{i\in\mathcal{S}(\vec{\lambda}')}r_{\vec{\lambda}'}(i)\prod_{j=1}^{d}p_j}{C_u},$$

$$u_2=\frac{\sum_{\vec{\lambda}'}d\sum_{i\in\mathcal{S}(\vec{\lambda}')}r_{\vec{\lambda}'}(i)^2}{C_u}.$$

With these notations, we aim at proving the set of estimates given in lemma 1. Indeed they imply

$$\frac{\sum_{\vec{\lambda}':|\vec{\lambda}'|=N+1}\left(\sum_{i\in\mathcal{S}(\vec{\lambda}')}\prod_{j=1}^{d}p_j+r_{\vec{\lambda}'}(i)\right)^2}{d\sum_{\vec{\lambda}':|\vec{\lambda}'|=N+1}\sum_{i\in\mathcal{S}(\vec{\lambda}')}\left(\prod_{j=1}^{d}p_j+r_{\vec{\lambda}'}(i)\right)^2}=1+t_2-u_2+O(N^{-3}), \tag{5.4}$$

with (t_2-u_2) of order N^{-2}. By (5.3), the risk of the estimator is then $u_2-t_2+O(N^{-3})$. Thus proving lemma 1 amounts to proving $1/N^2$ rate.

We shall make use of the notation $\Theta(f)$, meaning that there are universal positive constants m and M such that

$$mf\leqslant\Theta(f)\leqslant Mf.$$

Lemma 1. With the above notations,

$$C_u=C_t=d^2\sum_{\vec{\lambda}':|\vec{\lambda}'|=N+1}\left(\prod_{j=1}^{d}p_j\right)^2=\Theta(N^{3d-1}),$$

$$t_1=u_1=O(N^{-1}),$$

$$t_2=O(N^{-2}),$$

$$u_2=O(N^{-2}).$$

Proof. We first prove the first line.

Indeed for $\vec{\lambda}'\in\mathcal{P}_{N+1}$, all i are in $\mathcal{S}(\vec{\lambda}')$, and $(\Sigma_{i\in\mathcal{S}(\vec{\lambda}')}\Pi_{j=1}^{d}p_j)^2=d\Sigma_{i\in\mathcal{S}(\vec{\lambda}')}\Pi_{j=1}^{d}p_j^2=d^2\Pi_{j=1}^{d}p_j^2$. But if $\vec{\lambda}'\in\mathcal{P}_{N+1}$, there is at least one p_j equal to zero, so they do not contribute to the sum. So that $C_u=C_t=d^2\Sigma_{\vec{\lambda}':|\vec{\lambda}'|=N+1}(\Pi_{j=1}^{d}p_j)^2$.

We have then equality of the denominators of t_1 and u_1. The same argument gives equality of the numerators. On $\mathcal{P}_{N+1}$, $\#\mathcal{S}(\vec{\lambda})=d$ so that

$$\sum_{i\in\mathcal{S}(\vec{\lambda}')}\#\mathcal{S}(\vec{\lambda}')r_{\vec{\lambda}'}(i)\prod_{j=1}^{d}p_j=d\sum_{i\in\mathcal{S}(\vec{\lambda}')}r_{\vec{\lambda}'}(i)\prod_{j=1}^{d}p_j,$$

and outside $\mathcal{P}_{N+1}$, $\Pi_{j=1}^{d}p_j=0$ so that the equality still holds. Therefore, $t_1=u_1$.

Now $p_j\leqslant N+1$ so that $\Pi_{j=1}^{d}p_j\leqslant(N+1)^d$ and $|r_{\vec{\lambda}'}(i)|\leqslant 2(N+1)^{d-1}$. Moreover, as $1\leqslant\lambda_i\leqslant N+1$ and λ_d is known if the other λ_i are known, the number of elements $\vec{\lambda}$ in $\mathcal{P}_{N+1}$ satisfies $\#\mathcal{P}_{N+1}\leqslant(N+1)^{d-1}$. Thus the numerator of t_1 and u_1 is $O(N^{3d-2})$ and that of t_2 and u_2 is $O(N^{3d-3})$. To end the proof of the lemma, it is then sufficient to show that $C_u=\Theta(N^{3d-1})$.

Let us write $N+1=a(1+d(d+1))/2+b$ with a and b natural integers and $b<(1+d(d+1))$. We then select h_i for $i=1$ to d such that $\Sigma h_i=a/2$. The number of ways of partitioning $a/2$ in d parts is $\binom{a/2+d-1}{d-1}$, and this is $\Theta(a^{d-1})=\Theta(N^{d-1})$. To each of these partitions, we associate a different $\vec{\lambda}'$ in $\mathcal{P}_{N+1}$ through $\lambda_i=(d-i+1)a+\delta_{i=1}b+h_i$. For each of these $\vec{\lambda}$, we have $p_j=\lambda_j-\lambda_{j+1}\geqslant a/2$, so that $\Pi_{j=1}^{d}p_j^2=\Theta(N^{2d})$. We may lower bound C_u by the sum over these $\vec{\lambda}'$ of $\Pi_{j=1}^{d} p_j^2$, so that we have proved $C_u = \Theta(N^{3d-1})$. ■

VI. EVALUATION OF THE CONSTANT IN THE SPEED OF CONVERGENCE AND FINAL RESULT

The strategy we study is asymptotically optimal up to a constant, but a better constant can probably be obtained. Anything like $c(\vec{\lambda}) = (\Pi p_j)^{\alpha}$ with $\alpha \geq 1/2$ should yield the same rate, though it would be more cumbersome to prove. Polynomials in the p_j could also bring some improvement. All the same we give in this section a quick evaluation of the constant, that may serve as a benchmark for more precise strategies.

Write $p_j = (N+1)x_j$. Then, recalling our notation [Eq. (5.2)],

$$\prod_{j=1}^{d} p_j^2 = (N+1)^{2d} \prod_{j=1}^{d} x_j^2,$$

$$r_{\vec{\lambda}'}(i) = (N+1)^{d-1}(-x_{\{i\}} + \delta_{i>1} x_{\{i-1\}} + O(N^{-1})).$$

Similarly, the set of allowed $\vec{x} = (x_1, \ldots, x_n)$ may be described as

$$\mathcal{S}_{N+1} = \left\{ \vec{x} | x_j(N+1) \in \mathbb{N}; \sum_{j=1}^{d} (d-j+1)x_j = 1 \right\}.$$

We may then rewrite

$$u_2 = \frac{\sum_{\vec{x} \in \mathcal{S}_{N+1}} d \sum_{i=1}^{d} (x_{\{i\}} - \delta_{i>1} x_{\{i-1\}})^2}{d^2 (N+1)^2 \sum_{\vec{x} \in \mathcal{S}_{N+1}} \prod_{j=1}^{d} x_j^2} + O(N^{-3}),$$

$$t_2 = \frac{\sum_{\vec{x} \in \mathcal{S}_{N+1}} (x_{\{i\}} - \delta_{i>1} x_{\{i-1\}})^2}{d^2 (N+1)^2 \sum_{\vec{x} \in \mathcal{S}_{N+1}} \prod_{j=1}^{d} x_j^2} + O(N^{-3}).$$

Subtracting, we obtain (the first sums being on $\mathcal{S}_{N+1}$)

$$u_2 - t_2 + O(N^{-3}) \tag{6.1}$$

$$= \frac{\sum_{\vec{x}} 2d \left(\sum_{i=1}^{d} (x_{\{i\}})^2 - \sum_{i=2}^{d} x_{\{i\}} x_{\{i-1\}} \right) - (d+1)(x_{\{d\}})^2}{n^2 d^2 \sum_{\vec{x}} \prod_{j=1}^{d} x_j^2}. \tag{6.2}$$

Now $\mathcal{S}_{N+1}$ is the intersection $\mathcal{S}$ of the lattice in $[0,1]^d$ with mesh size $1/(N+1)$ with the hyperplane given by the equation $\Sigma(d-j+1)x_j = 1$. Therefore, the points of $\mathcal{S}_{N+1}$ are a regular paving of a flat $(d-1)$-dimensional volume, with more and more points [we know that $\#\mathcal{S}_{N+1} = O(N^{d-1})$]. Therefore, both denominator and numerator of Eq. (6.1) are Riemannian sums with respect to the Lebesgue measure, with a multiplicative constant that is the same for both. Therefore we have proved

Theorem 2. The estimator $\hat{U}$ corresponding to (5.1) has the following risk:

$$R_B(\hat{U}) = R_P(\hat{U}) = \mathbb{E}_{1_{\mathbb{C}^d}}[\Delta(1_{\mathbb{C}^d}, \hat{U})] = CN^{-2} + O(N^{-3}),$$

where C is the fraction

$$\frac{\int_{\mathcal{S}} 2d \left(\sum_{i=1}^{d} (x_{\{i\}})^2 - \sum_{i=2}^{d} x_{\{i\}} x_{\{i-1\}} \right) - (d+1)(x_{\{d\}})^2 \mathrm{d}\vec{x}}{d^2 \int_{\mathcal{S}} \prod_{j=1}^{d} x_j^2 \mathrm{d}\vec{x}}.$$

Up to a multiplicative constant, this risk is asymptotically optimal, both for a Bayes uniform prior and for global pointwise estimation.

Numerical estimation, up to two digits, for the low dimensions yields:

$$10 \quad \text{for } d=2$$

$$75 \quad \text{for } d=3$$

$$2.7 \times 10^2 \quad \text{for } d=4.$$

VII. CONCLUSION

We have given a strategy for estimating an unknown unitary channel $U \in \mathrm{SU}(d)$, and proved that the convergence rate of this strategy is $1/N^2$. We have further proved that this rate is optimal, even if the constant may be improved.

The interest of this result lies in that such rates are much faster than the $1/N$ achieved in classical estimation and, though they had already been obtained for SU(2), they were never before shown to hold for general SU(d).

ACKNOWLEDGMENTS

We are indebted to Manuel Ballester for an introduction to this question and for kindly providing the figure, and to Mădălin Guţă for numerous suggestions and extensive rereading.

[1] M. Childs, J. Preskill, and J. Renes, J. Mod. Opt. **47**, 155 (2000).

[2] A. Acin, E. Jane, and G. Vidal, Phys. Rev. A **64**, 050302(R) (2001).

[3] A. Peres and P. F. Scudo, Phys. Rev. Lett. **87**, 167901 (2004).

[4] E. Bagan, M. Baig, and R. Munoz-Tapia, Phys. Rev. A **69**, 050303(R) (2004).

[5] E. Bagan, M. Baig, and R. Munoz-Tapia, Phys. Rev. A **70**, 030301(R) (2004).

[6] G. Chiribella, G. M. D'Ariano, P. Perinotti, and M. F. Sacchi, Phys. Rev. Lett. **93**, 180503 (2004).

[7] M. Hayashi, Phys. Lett. A **354**, 183 (2006).

[8] V. Giovannetti, S. Lloyd, and L. Maccone, Science **306**, 1330 (2004).

[9] H. S. Eisenberg, J. F. Hodelin, G. Khoury, and D. Bouwmeester, Phys. Rev. Lett. **94**, 090502 (2005).

[10] D. A. R. Dalvit, R. L. de Matos Filho, and F. Toscano, New J. Phys. **8**, 276 (2006).

[11] D. Vitali, S. Kuhr, M. Brune, and J. M. Raimond, e-print quant-ph/0602006.

[12] G. Chiribella, G. M. D'Ariano, and M. F. Sacchi, Phys. Rev. A **72**, 042338 (2005).

[13] M. A. Ballester, e-print quant-ph/0507073.

[14] M. A. Ballester, Ph.D. thesis, available at http://homepages.cwi.nl/balleste/phdthesis.html

[15] R. Gill, e-print math.ST/0512443, under revision for *Annals of Statistics* (2005).

[16] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994).

[17] I. V. Schensted, *A Course on the Application of Group Theory to Quantum Mechanics* (Neo Press, Peaks Island, 1976).