Amount of information obtained by a quantum measurement

S. Massar[1] and S. Popescu[2,3]
[1]*Service de Physique Théorique, Université Libre de Bruxelles, Case Postale 225, Boulevard du Triomphe, B1050 Bruxelles, Belgium*
[2]*H. H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TW, United Kingdom*
[3]*BRIMS, Hewlett-Packard Laboratories, Stoke Gifford, Bristol BS12 SQZ, United Kingdom*
(Received 21 July 1999; revised manuscript received 23 December 1999; published 12 May 2000)

In this paper we address the problem of how much can be learned about an unknown quantum state by a measurement. To this end we consider optimal measurements for the state estimation problem, that is measurements that maximize the expectation of a fidelity function. We then enlarge the class of optimal measurements to measurements that act collectively on blocks of input states, and in addition we only require that the fidelity of the measurement be arbitrarily close to the optimal fidelity. We then consider the Shannon information of the outputs of optimal measurements, which is the amount of data produced by the measurements. We show that in the enlarged class of optimal measurements described above one can always construct an optimal measurement so that the Shannon information of its outputs equals the von Neumann entropy of the unknown states. Since this result is valid for all choices of fidelity functions and all distributions of input states, it provides a model independent answer to the question of how much can be learned about a quantum state by a measurement. Namely, this result shows that a measurement can extract at most one meaningful bit from every qubit carried by the unknown state.

PACS number(s): 03.67.−a, 03.65.Bz, 03.67.Hk

I. INTRODUCTION

Quantum mechanics has at its core a fundamental statistical aspect. Suppose you are given a single quantum particle in a state $|\Psi\rangle$ unknown to you. There is no way to find what $|\Psi\rangle$ is—to find it out you need an infinite ensemble of quantum particles, all prepared in the same state. Indeed, the different properties that characterize the state are, in general, complementary to one another; measuring one disturbs the rest. Only if an infinite ensemble is given can one find out the state. But infinite ensembles do not exist in practice. Given a finite ensemble of identically prepared particles, how well can one estimate the state? The problem is a fundamental one for understanding the very basis of quantum mechanics. It has been investigated by many authors, see for instance [1,2], and it constitutes probably the oldest problem in what is at present called ''quantum information.'' Here we approach this problem from a different point of view which, we think, leads to a deeper understanding.

What is the optimal way to estimate the quantum state given a finite ensemble? As such the question is not well posed. Indeed, since we cannot completely determine the state, i.e., completely determine all its properties, we must decide which particular property we want to determine. For an ensemble of spins, for example, estimating as well as possible the mean value of the z spin component is, obviously, a different question than estimating as well as possible the mean value of the x spin component.

But things are in fact even more complicated. The apparent benign words ''as well as possible'' in the previous paragraph are not well defined. Indeed, ''as well as possible'' actually means ''as well as possible given a specific measure of what 'well' means.'' Obviously, one can imagine many different measures. For example, suppose that a source emits states $|\psi_i\rangle$ with probability p_i. The problem is to design a measurement at the end of which we must guess which state was emitted. Let the guess be $|\phi_g\rangle$, and let the measure of success (fidelity) be

$$f(i,g)=|\langle\phi_g|\psi_i\rangle|^2, \tag{1}$$

i.e., the absolute value square of the scalar product in between the true state $|\psi_i\rangle$ and the guess $|\phi_g\rangle$. The goal is to optimize the measurement such that it yields the highest average fidelity

$$F=\sum_{i,j} p_i f(i,g(j))p(j|i), \tag{2}$$

where $p(j|i)$ is the probability for the measurement to have outcome j if the state is $|\psi_i\rangle$ and $g(j)$ is the guessed value if the outcome of the measurement is j. On the other hand, one can imagine another fidelity function, such as

$$f'(i,g)=|\langle\phi_g|\psi_i\rangle|^4. \tag{3}$$

Or one could try to optimize the mutual information

$$I=-\sum_i p_i \log_2 p_i+\sum_j p_j \sum_i p(i|j)\log_2 p(i|j) \tag{4}$$

or any other measure.

The important point to notice about the above different problems is that the different fidelities (2)–(4) not only define different scales according to which we measure the degree of success in estimating the state, but also, implicitly, define which property of the state we are actually estimating. If all the different fidelities were to lead to the same optimal measurements, we could say that we learn the same property about the state but just expressed in a different way. However, the different fidelities will in general lead to different optimal measurements which means that in each case we learn a different property about the system.

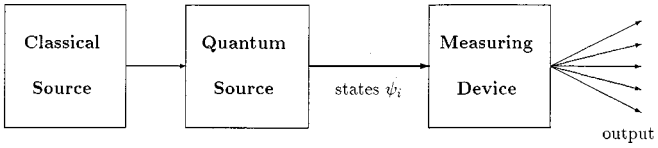

FIG. 1. Chain of events leading to a quantum state estimation problem. The classical source specifies which state should be sent. The quantum source then emits the corresponding state. Finally the measuring device tries to identify the emitted state.

To summarize, in general each particular estimation problem is completely different from the other; they measure different properties and their degree of success is measured on different scales, with the scales also defining implicitly what exactly is the property we estimate.

That one can learn different properties is a fact of life inherent to quantum mechanics. But there is no reason not to use the same scale to gauge how successful we have been in learning the property we decided to measure. The aim of this paper is to propose such a universal scale, and in the process to introduce a unique approach to quantum state estimation.

II. MAIN IDEA

The central point of our approach starts from a simple but fundamental question: what do we actually learn from a measurement on a state? Let us illustrate this question by an example. We shall contrast two situations. Consider a source which emits spin-1/2 particles. In the first case the particles are polarized with equal probability along either the $+z$ ($|\uparrow_z\rangle$) or $-z$ ($|\downarrow_z\rangle$) directions. In the second case the states are polarized along random directions uniformly distributed on the sphere. Suppose we want to identify the states as well as possible according to the fidelity equation (2). In the first case it is obvious that a measurement along σ_z perfectly identifies the state, hence the fidelity is $F=1$. In the second case, it has been shown [3] that the measurement along σ_z is also optimal. But in this case the states cannot be identified perfectly, and the fidelity is only $F=2/3$. On the other hand, what actually happened is that in both cases before we perform the measurement we know that the outcomes of the measurement are either $+1$ or -1, and the *a priori* probabilities of the two outcomes are equal. When we perform the measurement this uncertainty is resolved. Hence in both cases the measurement yields 1 bit of information. Our main idea is to interpret this quantity as the information we extract from the state.

This idea might seem paradoxical at first sight because in one case we completely recognize the state whereas in the other case we recognize it badly. To understand let us introduce a classical source that decides which quantum state is emitted from the quantum source (see Fig. 1). In the first case the classical source must only specify one bit (either $+z$ or $-z$) to determine which state is emitted. In the second case it must provide a direction n_{in} (i.e., an infinite number of bits) in order to specify the state $|\uparrow_{n_{in}}\rangle$. In both cases one extracts one bit of information. In the first case this means that the classical information supplied by the source is completely recovered. In the second case information is lost. However, it is now clear that the loss does not occur during the measurement, but during the first step, where classical information is converted into quantum information.

To summarize, the quantum state estimation problem as presented in Fig. 1 consists of a chain of events that starts with a classical source which tells the quantum source what state to emit, and ends with the measurement. The fidelity measures the overall performance of the chain since it is proportional to the scalar product $n_{in}\cdot n_{guess}$. On the other hand, the number of bits in the output characterizes how much information is extracted by the measurement. Therefore in this paper we shall focus on the latter quantity.

III. QUALITATIVE STATEMENT OF THE MAIN RESULT

The preceding discussion suggests that the Shannon information of the outcomes

$$I^S_{output}=-\sum_j p_j \log_2 p_j , \tag{5}$$

where $p_j=\Sigma_i p(j|i)$ is the probability of outcome j, measures how much information is extracted from the state. This idea, however, has to be refined.

The main problem is that there may be redundancies in the outputs of the measurement. As a trivial example, a measurement could be accompanied by the flip of a coin, and the outcomes of the measurement would consist of both the outcomes of the measurement proper and the outcomes of the coin flip. This adds one bit to the entropy of the outputs without telling anything about the system. In less trivial examples involving positive operator valued measures (POVMs) and ancillas, redundancies can arise in a less obvious way, and it is not immediate how they can be identified and eliminated.

Our main result is that no matter what property of the system one wants to measure, when the redundancy is eliminated, the remaining Shannon information of the outputs has a universal upper bound which is the von Neumann entropy of the quantum source,

$$I^S_{output}(\text{no redundancy})\leqslant I^{VN}_{input}, \tag{6}$$

where $I^{VN}_{input}=-\operatorname{Tr}\rho\log_2\rho$ is the Shannon information of the quantum source and ρ is the density matrix of the quantum source $\rho=\Sigma_i p_i|\psi_i\rangle\langle\psi_i|$. A more precise formulation of Eq. (6) will be given in Sec. V.

One does not always attain equality in Eq. (6). Indeed some questions are more informative about the system than others. Less informative questions can be answered by measurements whose output entropy is smaller. More informative questions require measurements with more entropy. But the most detailed questions can always be answered in I^{VN}_{input} bits.

IV. STRATEGY

The main problem we face in deriving Eq. (6) is to eliminate the redundancy. In order to do this we shall proceed in several steps.

(1) The first step is to decide which property we are interested in. We may fix the property directly (for instance decide to measure the average of σ_z) or implicitly by choosing a fidelity. In the rest of this paper we shall adopt the second approach.

(2) We then look at optimal measurements, that is measurements that maximize the fidelity. In general there is an entire class of such measurements.

(3) We perform a second optimization. Namely among the optimal measurements we look for the measurements which minimize I^S_{output}.

This double optimization strategy has already been considered for some particular cases in [4,5].

One expects that this strategy yields measurements that have no spurious redundancy. However, as we will show below through some examples, redundancies cannot be completely eliminated by the above procedure and we will have to further modify it.

These further modifications are motivated by the classical and quantum theory of information [6,7] which suggest the idea of performing measurements on blocks of quantum states, rather than on individual particles. Thus we shall allow the measuring device to accumulate a large number L of input states before making a collective measurement on the L states simultaneously. It is in the context of these collective measurements that we make the two optimizations [points (2) and (3) above] and thereby eliminate the spurious redundancies.

We want to emphasize that this procedure cannot increase the fidelity since the subsequent particles are completely uncorrelated. However, by considering measurements on large blocks we can hope to reduce the redundancy of the measurement, i.e., the entropy of the outcomes, by making ''better use'' of each outcome.

Two technicalities have to be taken into account. First of all we must take care not to modify the definition of fidelity. That is, the fidelity must still be the fidelity of each state individually, rather than the fidelity for the whole block. Second we should not require the measurement to absolutely maximize the fidelity, since then using block measurements does not help to reduce the entropy (this follows once more from the fact that the subsequent states are completely uncorrelated). However, following the ideas of information theory, we shall only require that the measurement has a fidelity approaching arbitrarily closely the optimum. In this framework we shall prove Eq. (6).

In Sec. V we show how to formulate mathematically the above strategy, and in particular the two technical points just mentioned.

V. PRECISE STATEMENT OF THE MAIN RESULT

Consider the general setup of Fig. 1. The states emitted by the quantum source $|\psi_i\rangle$ belong to a Hilbert space of dimension d. They occur with probability p_i. Their density matrix is $\rho=\Sigma_i p_i|\psi_i\rangle\langle\psi_i|$ with $\mathrm{Tr}\,\rho=1$. The most general measurement on the input states is a POVM with M element: $a_j \geqslant 0$, $\Sigma_{j=1}^M a_j = I_d$.

We define a general fidelity as follows. To each outcome j of the measurement we associate a guessed value g through a function $g(j)$. g may be a quantum state $|\phi_g\rangle$ as in Eq. (1), but it can also be something completely different. In addition we are given a fidelity $f(i,g)$ which measures how good g is for the input i. The fidelity f could be the function defined in Eqs. (1) or (3), but it could also be something completely different. The fidelity F of the measurement $\{a_j\}$ is the average value of $f(i,g)$:

$$F(\{a_j\})=\sum_i p_i \sum_j p(j|i) f(i,g(j)), \tag{7}$$

where $p(j|i)$ is the probability that the outcome of the measurement is j if the input state is $|\psi_i\rangle$

$$p(j|i)=\langle\psi_i|a_j|\psi_i\rangle. \tag{8}$$

We will be particularly interested in the optimal measurements for which F is maximal. Denote F_{max} the maximum of F,

$$F_{max}=\max_{\{a_j\}} F(\{a_j\}). \tag{9}$$

Suppose now that the source emits large blocks of L input states $|\psi_{i_1}\cdots\psi_{i_L}\rangle$. The successive states of the block are distributed independently with the same distribution p_i. Consider a measurement A_j acting on the whole block of L input states. To each outcome j we can associate L values of g, one for each input state, through L functions $g_1(j),\ldots,g_L(j)$. The fidelity for outcome j, averaged over the L input states, is $(1/L)\Sigma_{k=1}^L f(i_k,g_k(j))$. We therefore define the fidelity F_L of a measurement $\{A_j\}$ acting on large blocks as the average over the outcomes of the measurement and over the L input states of $f(i,g)$,

$$\begin{aligned} F_L(\{A_j\}) &= \sum_{i_1,\ldots,i_L} p_{i_1}\cdots p_{i_L} \sum_j \langle\psi_{i_1}\cdots\psi_{i_L}|A_j|\psi_{i_1}\cdots\psi_{i_L}\rangle \\ &\quad\times\frac{1}{L}\sum_{k=1}^{L} f(i_k,g_k(j)). \end{aligned} \tag{10}$$

The key property of F_L which justifies this definition is obtained by a rewriting of Eq. (10). To this end we define the operators $A_j^{(k)}$ as the operators A_j restricted to the space of particle k,

$$A_j^{(k)}=\mathrm{Tr}_{k'\neq k}\left(\prod_{l'\neq k}\rho_{l'}\right)A_j. \tag{11}$$

The operators $A_j^{(k)}$ are positive and sum to the identity on the space of particle k, hence they constitute a POVM acting on the space of particle k. We can then rewrite F_L as

$$F_L(\{A_k\}) = \frac{1}{L}\sum_{l=1}^{L}\sum_{k}\sum_{i_l} p_{i_l}\langle\psi_{i_l}|A_j^{(k)}|\psi_{i_l}\rangle f(i_l, j_l(k))$$

$$= \frac{1}{L}\sum_{l=1}^{L} F(\{A_j^{(k)}\}). \tag{12}$$

Thus F_L is just the average of the fidelities for the restricted measurements $A_j^{(k)}$. From this it follows that F_L is less or equal to F_{max}:

$$F_L \leq F_{max}. \tag{13}$$

Equality can be attained by a measurement which is the product of optimal measurements on each input state,

$$A_j = a_{j_1} \otimes \cdots \otimes a_{j_L}, \tag{14}$$

where $\{a_j\}$ is a measurement with M outcomes which maximize Eq. (7). Note that this optimal measurement $\{A_j\}$ has M^L outcomes.

It is these properties that justify our definition of F_L. Indeed they guarantee that one can compare in a meaningful way the fidelity for measurements on individual input states and measurements on large blocks.

We are now in a position to formulate our main result with precision. We state it as a theorem.

Theorem. Consider a state estimation problem in which the unknown state $|\psi_i\rangle$ have density matrix $\rho = \Sigma_i p_i |\psi_i\rangle\langle\psi_i|$ and von Neumann entropy $I_{input}^{VN} = -\mathrm{Tr}\,\rho \log_2 \rho$. Consider a bounded fidelity function $f(i,g)$ and the corresponding optimal fidelity F_{max} given by Eq. (9). Given any $\epsilon > 0$ and $\eta > 0$, then there exists L_0 such that for any $L \geq L_0$, and any N larger than $2^{L(I_{input}^{VN}+\eta)}$, there exists a measurement on sequences of L input states which has N outcomes and attains a fidelity $F_L \geq F_{max} - \epsilon$ where F_L is defined in Eq. (10).

To summarize, there is no best way of estimating an unknown quantum state. Different measurements will learn about different properties of the state, and it is up to us to choose which property we want to learn about. However, once we fix the property we want to learn about, one cannot learn more than $I_{input}^{VN} = -\mathrm{Tr}\,\rho \log_2 \rho$ bits about this property. That is, a measurement can extract at most one meaningful bit from each qubit coming from the source.

Before presenting the proof of this result, we discuss its generalization to other fidelities, its connection with other aspects of quantum information theory, and we present two examples that show the nontrivial character of the result.

VI. OTHER FIDELITIES

Our main result, as stated with precision at the end of the preceding section, applies to fidelities in the form of Eq. (7). This is very general since the fidelity function $f(i,g)$ is arbitrary. Nevertheless one can consider other more general fidelities.

As a first generalization, we consider fidelities in the form Eq. (1), but for which both the POVM elements and the guessed states are undetermined and both must be varied to find the optimum estimation strategy. That is, whereas in Sec. V the specification of an estimation strategy consisted only of the POVM elements $\{a_j\}$, it now consists of the set $\{a_j, \phi_g(j)\}$ which comprises both the POVM elements and the guessed states. An example of such a fidelity was considered in [3]. The unknown states $|\psi_i\rangle$ were taken to be n spin-1/2 particles all polarized along the same direction Ω and the fidelity was taken to be the scalar product of one spin polarized along Ω with one spin polarized along the guessed direction $f = |\langle\uparrow_\Omega|\uparrow_{\Omega_g}\rangle|^2$.

It is easy to show that our main result Eq. (6) also applies to such more general fidelities for which both the POVM elements and the guessing strategy can be optimized. Indeed allowing the guess to vary just corresponds to considering a family of fidelities. Since our result holds for each fidelity separately, it applies to the optimal one of the family.

One can, however, construct even more general fidelities (for instance, by taking the fidelity to be non linear in the POVM elements). For such more general fidelities it is an open question whether our claim also applies. One example of such more general fidelities is the mutual information equation (4).

VII. RELATION TO THE CLASSICAL CAPACITY OF A QUANTUM CHANNEL

The result presented here finds its origin in a reflexion on the recent developments of quantum information theory, in particular on the classical capacity of a quantum channel. The two problems are related, as the following discussion shows.

Consider the result [9] that the Holevo bound [10] for the classical capacity of a pure state quantum channel can be attained. An important ingredient in this proof was to derive certain properties of random sequences of states $|\psi_i\rangle$, each chosen with probability p_i. Namely it was shown that if the number N of sequences of length L is less than $N < L I_{input}^{VN} = -L\,\mathrm{Tr}\,\rho \log_2 \rho$, with $\rho = \Sigma_i p_i |\psi_i\rangle\langle\psi_i|$, then each sequence is with high probability almost orthogonal to the other sequences. It is this property that ensures that with high probability such sequences constitute good code words that can be reliably distinguished by the receiver.

In the present case we consider random sequences of POVM elements a_j. We show that when the number N of such sequences of length L is greater than $N > L I_{input}^{VN}$ (where I_{input}^{VN} is the von Neumann entropy of the input states on which the measurement should act), then the sequences (projected onto the probable subspace of the input states) almost sum to the identity operator. Thus with minor tinkering one can almost always transform this random sequence of POVM elements into a genuine POVM acting on sequences of L input states.

In both cases the properties of the sequences change dramatically when the number of sequences becomes equal to the von Neumann entropy of the states $N = L I_{input}^{VN}$. In the first case the code words cease to be mutually orthogonal. In the second case the sequences of POVM elements cease to

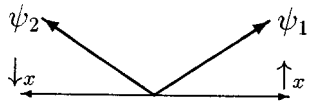

FIG. 2. The two input states $|\psi_1\rangle, |\psi_2\rangle = \alpha|\uparrow\rangle \pm \beta|\downarrow\rangle$. The optimal measurement is a measurement of the spin in the x direction.

sum to the identity. But in one case the limit is approached from below and in the other case from above. Thus the present analysis and that of [9] explore complementary properties of random sequences of operators.

Nevertheless, even thought the limiting case $N = LI^{VN}_{input}$ is approached from different directions in [9] and in the present paper, the mathematical techniques that we use in this paper are related to the techniques used in [9]. These techniques are the subject of Sec. XI.

VIII. EXAMPLES

Before embarking on a proof of our result, we give two examples which illustrate the main points that must be taken into account in the proof.

In the first example there are two possible input states $|\psi_1\rangle = \alpha|\uparrow\rangle + \beta|\downarrow\rangle$ and $|\psi_2\rangle = \alpha|\uparrow\rangle - \beta|\downarrow\rangle$ which occur with equal probability. The density matrix of the source is $\rho = \alpha^2|\uparrow\rangle\langle\uparrow| + \beta^2|\downarrow\rangle\langle\downarrow|$ which is different from the identity for $\alpha \neq \beta$. Therefore the von Neumann entropy of the input states $I^{VN}_{input} < 1$ qubit.

The fidelity is defined as follows: after each measurement one must guess whether the state is $|\psi_1\rangle$ or $|\psi_2\rangle$. In case of a correct guess one receives a score of $+1$, and for an incorrect guess one receives a score of -1. The aim is to maximize the average score. The techniques of Sec. X can be used to show that the optimal measurement is a von Neumann measurement of σ_x, see Fig. 2. The two outcomes of this measurement occur with equal probability, and hence $I^S_{output} = 1 > I^{VN}_{input}$.

In this example, a natural first step in eliminating the redundancy is to project blocks of input states onto their probable subspace [7,8]. This projection succeeds with arbitrarily high probability, and affects the input states arbitrarily little. But it reduces the dimensionality of the Hilbert space of the input states from 2^N to $2^{NI^{VN}_{input}}$. Hence if we can prove that there is a von Neumann measurement restricted to the probable subspace that is optimal, we will have proved our claim. However, the construction of such a von Neumann measurement is nontrivial, as is illustrated in the next example.

In our second example there is no ''most probable'' subspace because the density matrix of the inputs is completely random. In this example there are three input states $|\psi_1\rangle = |\uparrow\rangle$, $|\psi_2\rangle = \frac{1}{2}|\uparrow\rangle + \sqrt{3}/2|\downarrow\rangle$, $|\psi_3\rangle = \frac{1}{2}|\uparrow\rangle - \sqrt{3}/2|\downarrow\rangle$, each occurring with equal probability $p_i = 1/3$. The density matrix of these states is $\rho = I/2$ and their entropy is $I^{VN}_{input} = 1$ qubit. The fidelity is defined as above: after the measurement one must guess which was the input state. If the guess is correct one scores $+1$ point, if the guess is incorrect, one scores -1 points. The aim is to maximize the average score (fidelity).

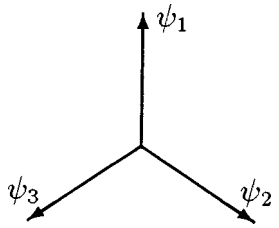

FIG. 3. The three input states ψ_1, ψ_2, ψ_3 in the second example. The optimal measurement is a POVM whose elements are projectors onto the three states ψ_1, ψ_2, ψ_3.

Using the techniques of Sec. X, one can show that the elements of an optimal POVM are necessarily proportional to the three projectors $|\psi_1\rangle\langle\psi_1|, |\psi_2\rangle\langle\psi_2|, |\psi_3\rangle\langle\psi_3|$, see Fig. 3. Therefore the optimal POVM whose output entropy is minimum is $\{\frac{2}{3}|\psi_1\rangle\langle\psi_1|, \frac{2}{3}|\psi_2\rangle\langle\psi_2|, \frac{2}{3}|\psi_3\rangle\langle\psi_3|\}$. In this case $I^S_{output} = \log_2 3 > 1$ bits. Any other optimal measurements has larger output entropy $I^S_{output} \geqslant \log_2 3$ bits. One can also show that there is no measurement on blocks of L input states whose fidelity is strictly equal to the optimum and whose output entropy is less then $L \log_2 3$ bits. However, if one only requires that the fidelity is arbitrarily close to the maximum, then in the asymptotic limit ($L \to \infty$) the output entropy can be made arbitrarily close to L bits, thereby attaining the bound Eq. (6). The main difficulty of the proof will be to construct such a measurement on large blocks whose output entropy is equal to L bits and whose fidelity is arbitrarily close to the optimal fidelity.

IX. PLAN OF THE PROOF

The next sections are devoted to proving the bound Eq. (6). In Sec. X we derive some properties of the optimal measurements. In Sec. XI we show how to construct a measurement on large blocks which has little redundancy. In Sec. XII we derive an intermediate result concerning the fidelity of the measurement constructed in Sec. XI. If the states are uniformly distributed in Hilbert space (i.e., the density matrix is proportional to the identity, $\rho = I/d$), then this intermediate result already proves our main claim Eq. (6). When the states are not uniformly distributed in Hilbert space, we must first project blocks of states onto the probable subspace before using the intermediate result of Sec. XII. This is done in Sec. XIII and completes the proof of Eq. (6).

X. PROPERTIES OF OPTIMAL MEASUREMENTS

Consider a fidelity as defined in Sec. V. We summarize here the main properties of optimal measurements, i.e., those that maximize the fidelity, see also [2,3].

First of all note that we can always take the optimal POVM to consist of one-dimensional projectors $b_j = |b_j\rangle\langle b_j|$ (the b_j are not normalized). Indeed, refining a POVM can only increase the fidelity. This can be seen formally in the following way: suppose the a_j are an optimal POVM, but not necessarily made out of one-dimensional projectors. Then each a_j can always be decomposed as $a_j = \Sigma_k |b_{jk}\rangle\langle b_{jk}|$ since it is a positive operator. Inserting this

into the expression for F one sees that the b_{jk} [to which we associate the same value of g: $g(jk)=g(j)$] are also optimal.

Thus we can optimize F in the class of POVM's whose elements all have rank one $|b_j\rangle\langle b_j|$. These elements are subject to the unitarity condition $\Sigma_j|b_j\rangle\langle b_j|=I_d$. This can be implemented by introducing d^2 Lagrange multipliers $\lambda_{\mu\nu}$ which we group into one operator $\hat{\lambda}$:

$$F=\sum_i p_i\sum_j \langle\psi_i|b_j\rangle\langle b_j|\psi_i\rangle f(i,g(j)) - \mathrm{Tr}\left[\hat{\lambda}\left(\sum_j |b_j\rangle\langle b_j|-I_d\right)\right] = \sum_j \mathrm{Tr}[(\hat{F}_j-\hat{\lambda})|b_j\rangle\langle b_j|]+\mathrm{Tr}\,\hat{\lambda}, \tag{15}$$

where $\hat{F}_j=\Sigma_i p_i|\psi_i\rangle\langle\psi_i|f(i,g(j))$. If we vary this with respect to $\langle b_j|$, we obtain the equations

$$(\hat{F}_j-\hat{\lambda})|b_j\rangle=0. \tag{16}$$

Inserting this into Eq. (15) shows that $F_{max}=\mathrm{Tr}\,\hat{\lambda}$.

Equation (16) is the essential equation to find optimal measurements explicitly. For instance, consider the first example of Sec. VIII. There are two input states ψ_1 and ψ_2 and two guessed states $\phi_1^{guess}=|\psi_1\rangle$ and $\phi_2^{guess}=|\psi_2\rangle$. If the input state is $|\psi_1\rangle$ and one guesses ϕ_1^{guess}, then $f=+1$, whereas if the input state is $|\psi_2\rangle$ and one guesses ϕ_1^{guess}, then $f=-1$, hence $\hat{F}_1=\frac{1}{2}(|\psi_1\rangle\langle\psi_1|-|\psi_2\rangle\langle\psi_2|)=+\alpha\beta\sigma_x$. Similarly $\hat{F}_2=\frac{1}{2}(|\psi_2\rangle\langle\psi_2|-|\psi_1\rangle\langle\psi_1|)=-\alpha\beta\sigma_x$. The task is then to find an operator $\hat{\lambda}$ such that null eigenvectors of $\hat{F}_{1,2}-\hat{\lambda}$ can satisfy the completeness relation. The only possibility is $\hat{\lambda}=\alpha\beta I$. Therefore the optimal measurement is along the x axis, and $F_{max}=2\alpha\beta$. The second example of Sec. VIII can be treated along similar lines.

An important consequence of Eq. (16) is an explicit expression for the value of F if the measurement is not optimal. Consider a measurement a_j' which is not optimal, but each positive operator a_j' is ''close'' to the corresponding operator b_j of the optimal measurement. We then decompose the operator a_j' in terms of its components along $|b_j\rangle$: $a_j'=X_j|b_j\rangle\langle b_j|+Y_j|b_j\rangle\langle b_j^\perp|+Y_j^*|b_j^\perp\rangle\langle b_j|+z_j$ where the state $|b_j^\perp\rangle$ is orthogonal to $|b_j\rangle$ and the operator z_j obeys $z_j|b_j\rangle=0$, $\langle b_j|z_j=0$. Inserting this decomposition into the expression for F, we obtain

$$F(a')=\mathrm{Tr}\,\hat{\lambda}+\sum_j \mathrm{Tr}[(\hat{F}_j-\hat{\lambda})a_j'] = F_{max}+\sum_j \mathrm{Tr}[(\hat{F}_j-\hat{\lambda})z_j] \geqslant F_{max}-C\,\mathrm{Tr}\,z_j, \tag{17}$$

where we have used Eq. (16) and C is some positive constant independent of j. This expresses in a simple way how much the fidelity differs from its maximal value in terms of how much the measurement differs from the optimal measurement.

XI. ELIMINATING REDUNDANCY

Our aim in this section is to construct a measurement with less outcomes than the optimal measurement Eq. (14). The next two sections will be devoted to proving that this measurement does not diminish the fidelity. This measurement is very similar to the measurement used in [9] to decode a classical message sent through a quantum communication channel.

We start from the optimal POVM acting on one input state and decomposed into one-dimensional projectors $b_i=|b_i\rangle\langle b_i|$. We express it in terms of the normalized operator $\tilde{b}_i=|\tilde{b}_i\rangle\langle\tilde{b}_i|=b_i/\mathrm{Tr}(b_i)$ as $b_i=\beta_i\tilde{b}_i$. (Throughout the text we shall denote normalized operators by $\tilde{}$). The β_i sum to $\Sigma_i\beta_i=d$ obtained by taking the trace of the completeness relation.

We now construct N operators acting on the space of L input states,

$$\tilde{B}_j=|\tilde{B}_j\rangle\langle\tilde{B}_j|=\tilde{b}_{j_1}\otimes\cdots\otimes\tilde{b}_{j_L}, \tag{18}$$

where each $\tilde{b}_{j_k}$ is chosen randomly and independently from the set $\tilde{b}_1,\ldots,\tilde{b}_M$ with probabilities $p_1=\beta_1/d,\ldots,p_M=\beta_M/d$.

The $|\tilde{B}_j\rangle$ span a subspace H_B of the Hilbert space of the L input states. In this subspace the operator $B=\Sigma_j\tilde{B}_j$ is strictly positive, hence we can construct the operators

$$C_j=|C_j\rangle\langle C_j|=B^{-1/2}\tilde{B}_jB^{-1/2}. \tag{19}$$

The C_j are positive operators, which sum up the identity in H_B: $\Sigma_{j=1}^N C_j=\Pi_B$ where Π_B is the projector onto H_B. The POVM we shall use consists of the C_j and the projector onto the complementary subspace $C_0=I_{d^L}-\Pi_B$ (I_{d^L} is the identity on the Hilbert space of the L input states).

Our strategy in the next sections will be to compute the average fidelity $\overline{F_L}$, where the average is taken over possible choices of B_j in Eq. (18). We shall show that the average of F_L satisfies our main result stated at the end of Sec. V. Therefore there necessarily are some choices of B_j that also satisfy our main result. But first we derive some important properties of the C_j. We shall obtain mean properties, where the mean is the average over choices of B_j in Eq. (18).

(i) The mean of $\tilde{B}_j$ is $\overline{\tilde{B}_j}=I_{d^L}/d^L$.

(ii) The mean of B is

$$\overline{B}=\sum_{j=1}^N \overline{\tilde{B}_j}=\frac{N}{d^L}I_{d^L}. \tag{20}$$

This motivates our writing

$$B=\frac{N}{d^L}(I_{d^L}+\Delta) \tag{21}$$

and subsequently making expansions in Δ.

(iii) The dimension of H_B (denoted $\dim_{H_B}$) is

$$\dim_{H_B}=\sum_j \mathrm{Tr}\, C_j=\sum_j \mathrm{Tr}\, B^{-1}\tilde{B}_j=\frac{d^L}{N}\sum_j \mathrm{Tr}\frac{1}{I_{d^L}+\Delta}\tilde{B}_j$$
$$\geqslant\frac{d^L}{N}\sum_j \mathrm{Tr}(I_{d^L}-\Delta)\tilde{B}_j. \tag{22}$$

Furthermore,

$$\mathrm{Tr}\,\Delta\tilde{B}_j=\mathrm{Tr}\left(\frac{d^L}{N}B-I_{d^L}\right)\tilde{B}_j=\mathrm{Tr}\left[\frac{d^L}{N}\left(\tilde{B}_j+\sum_{k\neq j}\tilde{B}_k\tilde{B}_j\right)-\tilde{B}_j\right], \tag{23}$$

where we have used the fact that $\tilde{B}_j^2=\tilde{B}_j$. We now take the average of this expression. Using the fact that for $k\neq j$, $\tilde{B}_k$, and $\tilde{B}_j$ are independent, the average of $\tilde{B}_k\tilde{B}_j(k\neq j)$ is the product of the averages $\overline{\tilde{B}_k\tilde{B}_j}=\overline{\tilde{B}_j}\;\overline{\tilde{B}_k}=I_{d^L}/d^{2L}$. And hence $\overline{\sum_{k\neq j}\tilde{B}_k\tilde{B}_j}=(N-1)I_{d^L}/d^{2L}$. Putting all together, we find $\overline{\mathrm{Tr}\,\Delta\tilde{B}_j}=(d^L-1/N)$ and

$$d^L\geqslant\overline{\dim H_B}\geqslant d^L\left(1-\frac{d^L-1}{N}\right). \tag{24}$$

This shows that if N is slightly larger than the dimension of the Hilbert space d^L, then the $C_j(j\neq 0)$ fill the Hilbert space.

(iv) Finally we need to know how much the C_j differ from the $\tilde{B}_j$. We write $|C_j\rangle=\alpha_j|\tilde{B}_j\rangle+|B_j^\perp\rangle$ and compute α_j^2,

$$\alpha_j^2=\mathrm{Tr}\,C_j\tilde{B}_j=\mathrm{Tr}\,\tilde{B}_jB^{-1/2}\tilde{B}_jB^{-1/2}=(\mathrm{Tr}\,\tilde{B}_jB^{-1/2})^2$$
$$\geqslant\frac{d^L}{N}\left(1-\frac{1}{2}\mathrm{Tr}\,\tilde{B}_j\Delta\right)^2. \tag{25}$$

Hence

$$\overline{\alpha_j^2}\geqslant\frac{d^L}{N}(1-\overline{\mathrm{Tr}\,\tilde{B}_j\Delta})=\frac{d^L}{N}\left(1-\frac{d^L-1}{N}\right). \tag{26}$$

This is then used to compute the average of $\langle B_j^\perp|B_j^\perp\rangle$:

$$\overline{\langle B_j^\perp|B_j^\perp\rangle}=\overline{\mathrm{Tr}\,C_j}-\overline{\mathrm{Tr}\,C_jB_j}\leqslant\frac{d^L}{N}\,\frac{d^L-1}{N}, \tag{27}$$

which shows that the C_j are arbitrarily close to the $\tilde{B}_j$ when $N>d^L$.

XII. AN INTERMEDIATE RESULT

In this section we shall prove the following intermediate result. Suppose that the input states $|\psi_i\rangle$ belong to a Hilbert space of dimension d and have a density matrix $\rho=\Sigma_i p_i|\psi_i\rangle\langle\psi_i|$. Denote by ρ_{max} the largest eigenvalue of ρ. Consider measurements on blocks of L input states. Give yourself any positive number $\eta>0$. Let N be any integer larger than $2^{L(2\log_2 d+\log_2\rho_{max}+\eta)}$. Then there exist measurements with N outcomes with a fidelity $F_L\geqslant F_{max}-R2^{-L\eta}$ where R is a positive constant.

In the next section we shall combine this intermediate result with the concept of probable subspace of a long sequence of states to prove our claim in full generality.

To prove this intermediate result, we proceed as follows.

Let $\{b_j=|bj\rangle\langle b_j|\}$ be a POVM that maximizes the fidelity F Eq. (7). Using the algorithm of Eqs. (18) to (19) we construct a measurement C_j, $j=0,\ldots,N$, acting on the space of L copies of the input states.

Let us consider the fidelity for the measurement C_j:

$$F_L=\sum_{j=0}^{N}\frac{1}{L}\sum_{k=1}^{L}\sum_{i_k}p_{i_k}\langle\psi_{i_k}|C_j^{(k)}|\psi_{i_k}\rangle f(i_k,g_k(j)), \tag{28}$$

where the $C_j^{(k)}=\mathrm{Tr}_{l\neq k}(\Pi_{l'\neq k}\rho_{l'})C_j$ are defined as in Eq. (11).

We can decompose $C_j^{(k)}$ (for $j\neq 0$) according to its components along $|\tilde{b}_{jk}\rangle$: $C_j^{(k)}=X_{jk}|\tilde{b}_{jk}\rangle\langle\tilde{b}_{jk}|+Y_{jk}|\tilde{b}_{jk}\rangle\langle b_{jk}^\perp|+Y_{jk}^*|b_{jk}^\perp\rangle\langle\tilde{b}_{jk}|+z_{jk}$ where $z_{jk}|\tilde{b}_{jk}\rangle=0$, $\langle\tilde{b}_{jk}|z_{jk}=0$. Inserting this expression in Eq. (28), and using Eq. (17), yields

$$F_L\geqslant\frac{1}{L}\sum_{k=1}^{L}\left(F_{max}-C\sum_{j=1}^{N}\mathrm{Tr}\,z_{jk}-C\,\mathrm{Tr}\,C_0^{(k)}\right), \tag{29}$$

where the last term comes from the $C_0=I_{d^L}-\Pi_B$ outcome.

It remains to calculate $\mathrm{Tr}\,C_0^{(k)}$ and $\mathrm{Tr}\,z_{jk}$. We start with the former

$$C_0^{(k)}=\mathrm{Tr}_{l\neq k}\left(\prod_{l'\neq k}\rho_{l'}\right)(I_{d^L}-\Pi_B)$$
$$\leqslant(\rho_{max})^{L-1}\mathrm{Tr}(I_{d^L}-\Pi_B)=(\rho_{max})^{L-1}(d^L-\dim H_B), \tag{30}$$

where ρ_{max} is the largest eigenvalue of ρ.

To estimate $\mathrm{Tr}\,z_{jk}$ we recall the decomposition of $|C_j\rangle=\alpha_j|\tilde{B}_j\rangle+|B_j^\perp\rangle$. We can further decompose $|B_j^\perp\rangle$ according to whether when restricted to the space of the kth particle, it is equal to $|b_{j_k}\rangle$ or not: $|B_j^\perp\rangle=|\tilde{b}_{j_k}\rangle|\phi\rangle+|\tilde{b}_{j_k}^\perp\rangle|\chi\rangle$. Inserting this into the trace which yields $C_j^{(k)}$, we obtain

$$C_j^{(k)}=\mathrm{Tr}_{l\neq k}\left(\prod_{l'\neq k}\rho_{l'}\right)(\alpha_j|\tilde{B}_j\rangle+|\tilde{b}_{j_k}\rangle|\phi\rangle$$
$$+|\tilde{b}_{j_k}^\perp\rangle|\chi\rangle)(\alpha_j^*\langle\tilde{B}_j|+\cdots)$$
$$=|\tilde{b}_{j_k}\rangle\langle\tilde{b}_{j_k}|X_{jk}+|\tilde{b}_{j_k}\rangle\langle\tilde{b}_{j_k}^\perp|Y_{jk}+|\tilde{b}_{j_k}^\perp\rangle\langle\tilde{b}_{j_k}|Y_{jk}^*+|\tilde{b}_{j_k}^\perp\rangle$$
$$\times\langle\tilde{b}_{j_k}^\perp|Z_{jk}. \tag{31}$$

The coefficients X_{jk}, Y_{jk}, Z_{jk} are easily calculated. The one of interest is $Z_{jk}=\mathrm{Tr}\,z_{jk}$,

$$Z_{jk}=\mathrm{Tr}\prod_{l'\neq k}\rho_{l'}|\chi\rangle\langle\chi|\leqslant(\rho_{max})^{(L-1)}\langle\chi|\chi\rangle$$
$$\leqslant(\rho_{max})^{(L-1)}\langle B_j^{\perp}|B_j^{\perp}\rangle. \quad (32)$$

Inserting these bounds into the expression for F_L we obtain

$$F_L\geqslant\frac{1}{L}\sum_{k=1}^{L}[F_{max}-C(\rho_{max})^{L-1}\langle B_j^{\perp}|B_j^{\perp}\rangle$$
$$-C(\rho_{max})^{L-1}(d^L-\dim H_B)]. \quad (33)$$

We now take the average of this expression over all possible choices of b_{jk} operators in Eq. (18). Inserting Eqs. (24) and (27) yields

$$\overline{F_L}\geqslant F_{max}-2C(\rho_{max})^{L-1}d^L\frac{d^L-1}{N}. \quad (34)$$

Therefore if $N\geqslant 2^{L(2\log_2 d+\log_2\rho_{max}+\eta)}$, then $\overline{F_L}\geqslant F_{max}-R2^{-L\eta}$ where $R=2C/\rho_{max}$. This proves the intermediate result.

Note that if the input states are uniformly distributed in Hilbert space, i.e., $\rho=I/d$, then this intermediate result directly implies our main claim. Indeed when $\rho=I/d$, $\rho_{max}=1/d$, then $\overline{F_L}\geqslant F_{max}-R2^{-L\eta}$ if $N\geqslant 2^{L(\log_2 d+\eta)}=2^{L(I_{input}^{VN}+\eta)}$. When the input states are not uniformly distributed in Hilbert space, we must use the notion of probable Hilbert space of a long sequence to prove our main result. This is done in the next section.

XIII. MEASUREMENTS ON PROBABLE SUBSPACES

We now combine the result of the previous section with the notion of probable subspace of large blocks of states. We first recall the properties of the probable subspace [7,8]. Consider a long sequence of L' input states $|\psi_{i_1}\ldots\psi_{i_{L'}}\rangle$. The density matrix of these states is $\rho=\Pi_{k=1}^{L'}\rho_k$. The projector Π onto the probable subspace has the properties that given $\epsilon'>0$, $\eta'>0$, and for L' sufficiently large, (1) $\mathrm{Tr}\,\Pi\rho\geqslant 1-\epsilon'$, i.e., the probability to be in the probable subspace is arbitrarily close to 1.

(2) Π and ρ commute, i.e., the eigenvectors of ρ are either eigenvectors of Π or of $1-\Pi$. And, furthermore, the eigenvectors which are common to Π and ρ have eigenvalues comprised between $2^{L'(-H-\eta')}\leqslant(\rho_{L'})_i\leqslant 2^{L'(-H+\eta')}$.

(3) From these two properties it follows that the dimension of the probable Hilbert space is bounded by $(1-\epsilon')2^{L'(H-\eta')}\leqslant\mathrm{Tr}\,\Pi\leqslant 2^{L'(H+\eta')}$.

Let us now show that measurements restricted to the probable subspace are arbitrarily close to optimal. Suppose that A_j is a measurement that optimizes the state determination problem Eq. (10) for sequences of L' input states [for instance, the measurement equation (14)]. Consider the POVM consisting of the operators $A_j'=\Pi A_j\Pi$ [to which we associate the unmodified guesses $g(j)$] and the operator $I-\Pi$ (to which we associate the minimal value of the fidelity f_{min}). The fidelity for this measurement is

$$F_{L'}=\sum_{i_1\cdots i_{L'}}p_{i_1}\cdots p_{i_{L'}}\sum_{j=1}^{N}\langle\psi_{i_1}\cdots\psi_{i_{L'}}|\Pi A_j\Pi|\psi_{i_1}\cdots\psi_{i_{L'}}\rangle\frac{1}{L'}\sum_{k=1}^{L'}f(i_k,g_k(j))$$
$$+\sum_{i_1\cdots i_{L'}}p_{i_1}\cdots p_{i_{L'}}\langle\psi_{i_1}\cdots\psi_{i_{L'}}|1-\Pi|\psi_{i_1}\cdots\psi_{i_{L'}}\rangle f_{min}$$
$$\geqslant F_{max}-\sum_{i_1\cdots i_{L'}}p_{i_1}\cdots p_{i_{L'}}\sum_{j=1}^{N}\langle\psi_{i_1}\cdots\psi_{i_{L'}}|A_j-\Pi A_j\Pi|\psi_{i_1}\cdots\psi_{i_{L'}}\rangle\frac{1}{L'}\sum_{k=1}^{L'}f(\psi_{i_k},\phi_{jk})-|f_{min}|\mathrm{Tr}\,\rho(1-\Pi). \quad (35)$$

The main difficulty is to bound the second term whose absolute value we denote by T_2. We proceed as follows:

$$T_2=\Bigg|\sum_{i_1\cdots i_{L'}}p_{i_1}\cdots p_{i_{L'}}\sum_{j=1}^{N}\langle\psi_{i_1}\cdots\psi_{i_{L'}}|A_j$$
$$-\Pi A_j\Pi|\psi_{i_1}\cdots\psi_{i_{L'}}\rangle\frac{1}{L'}\sum_{k=1}^{L'}f(\psi_{i_k},\phi_{jk})\Bigg|$$
$$\leqslant|f_{max}|\sum_{i_1\cdots i_{L'}}p_{i_1}\cdots p_{i_{L'}}$$
$$\times\sum_{j=1}^{N}|\langle\psi_{i_1}\cdots\psi_{i_{L'}}|A_j-\Pi A_j\Pi|\psi_{i_1}\cdots\psi_{i_{L'}}\rangle|, \quad (36)$$

where f_{max} is the maximum value of the fidelity. Denote by S the operator

$$S=|\psi_{i_1}\cdots\psi_{i_{L'}}\rangle\langle\psi_{i_1}\cdots\psi_{i_{L'}}|-\Pi|\psi_{i_1}\cdots\psi_{i_{L'}}\rangle$$
$$\times\langle\psi_{i_1}\cdots\psi_{i_{L'}}|\Pi. \quad (37)$$

We decompose the Hermitian operator S into its eigenstates $S=\Sigma_i s_i|s_i\rangle\langle s_i|$. We then have

$$\sum_j |\mathrm{Tr}\, SA_j| = \sum_j \left| \sum_i s_i \langle s_i|A_j|s_i\rangle \right| \leqslant \sum_j \sum_i |s_i|\langle s_i|A_j|s_i\rangle = \sum_i |s_i| = \mathrm{Tr}|S|. \tag{38}$$

We now use lemma I.4 of [11] which, applied to our case, states that

$$\mathrm{Tr}|S| \leqslant \sqrt{8(1-\langle \psi_{i_1}\cdots\psi_{i_{L'}}|\Pi|\psi_{i_1}\cdots\psi_{i_{L'}}\rangle)}. \tag{39}$$

Going back to Eq. (36), we have

$$T_2 \leqslant |f_{max}| \sum_{i_1\cdots i_{L'}} p_{i_1}\cdots p_{i_{L'}} \sqrt{8(1-\langle \psi_{i_1}\cdots\psi_{i_{L'}}|\Pi|\psi_{i_1}\cdots\psi_{i_{L'}}\rangle)}. \tag{40}$$

Using the concavity of $\sqrt{1-x}$, we have

$$T_2 \leqslant |f_{max}| \sqrt{8 \sum_{i_1\cdots i_{L'}} p_{i_1}\cdots p_{i_{L'}}(1-\langle \psi_{i_1}\cdots\psi_{i_{L'}}|\Pi|\psi_{i_1}\cdots\psi_{i_{L'}}\rangle)} = |f_{max}| \sqrt{8[1-\mathrm{Tr}(\rho\Pi)]}. \tag{41}$$

Putting everything together we have

$$F_{L'} \geqslant F_{max} - |f_{min}|\epsilon' - |f_{max}|\sqrt{8\epsilon'}. \tag{42}$$

This shows that the restriction of the measurement to the probable Hilbert space diminishes the fidelity by an arbitrarily small amount of order $O(\sqrt{\epsilon'})$.

We can now build a measurement which satisfies our main result as stated at the end of Sec. V. We decompose the input states into blocks of L' states. On each of these blocks we first carry out a partial measurement Π and $I-\Pi$ to know whether it is in the probable subspace or not. If the result is $I-\Pi$ the sequence is discarded. The sequences which pass the test are kept.

We now take the sequences that have passed the test as the input states in the intermediate result. These sequences belong to a Hilbert space of dimension $\dim H_{probable} \leqslant 2^{L'(I^{VN}_{input}+\eta')}$ and the largest eigenvalue of their density matrix is $\rho_{max} \leqslant 2^{L'(-I^{VN}_{input}+\eta')}$. To apply the intermediate result, we take an integer L and an $\eta > 0$. Then there exists a measurement on blocks of L sequences which has a number of possible outcomes equal to any integer N larger than $2^{L[L'(I^{VN}_{input}+3\eta')+\eta]} = 2^{LL'(I^{VN}_{input}+3\eta'+\eta/L')}$ and which has a fidelity larger than $F_{LL'} \geqslant F_{max} - |f_{min}|\epsilon' - |f_{max}|\sqrt{8\epsilon'} - R2^{-L\eta}$, where R is a positive constant.

Let us calculate the entropy $I^S_{outputs}$ of the outputs of this measurement. We need less than $I_{\epsilon'} = -\epsilon' \log_2 \epsilon' - (1-\epsilon')\log_2(1-\epsilon')$ bits to describe whether or not the input state passes the first test of belonging to the probable Hilbert space or not. If it does then we need less than $\log_2 N$ bits to encode the output of the measurement on the L blocks of probable sequences. Therefore the total number of bits we need to describe the outcome of this measurement on LL' elementary input states is $I^S_{output} \leqslant \log_2 N + LI_{\epsilon'}$. Replacing N by its lower bound, we have $I^S_{output} \leqslant LL'(I^{VN}_{input} + (3\eta' + \eta/L' + I_{\epsilon'}/L')$. Since ϵ', η', and η can be chosen arbitrarily small, and L' arbitrarily large, our claim is proven.

XIV. CONCLUSION

In this paper we have obtained a quantitative estimate of how much information can be obtained by a quantum measurement. We considered optimal measurements, that is measurements which maximize a fidelity function. We then enlarged the set of optimal measurements in two ways. First, we considered optimal measurements that act collectively on large blocks of input states rather than measurements restricted to act on each state separately. Second, we did not require the fidelity of the measurements to be exactly equal to the optimal fidelity, but only that it be arbitrarily close to the optimal fidelity. In this context we showed that whatever property of a quantum system one wants to learn about, one can learn at most one bit of information about every qubit of quantum information carried by the unknown quantum system. That is, the Shannon entropy of the outcomes of optimal measurements can always be made equal or less than the von Neumann entropy of the unknown quantum states.

ACKNOWLEDGMENTS

We would like to thank Andreas Winter for helpful discussions. S.M. would like to thank Utrecht University where part of this work was carried out. S.M. is a research associate of the Belgian National Research Fund (FNRS).

[1] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic, New York, 1976).

[2] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).

[3] S. Massar and S. Popescu, Phys. Rev. Lett. **74**, 1259 (1995).

[4] R. Derka, V. Buzek, and A. K. Ekert, Phys. Rev. Lett. **80**, 1571 (1998).

[5] J. I. Latorre, P. Pascual, and R. Tarrach, Phys. Rev. Lett. **81**,

1351 (1998).
[6] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).
[7] B. Schumacher, Phys. Rev. A **51**, 2738 (1995).
[8] R. Jozsa and B. Schumacher, J. Mod. Opt. **41**, 2343 (1994).
[9] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wooters, Phys. Rev. A **54**, 1869 (1996).
[10] A. S. Holevo, Probl. Peredachi Inf. **9**, 3 (1973) [Probl. Inf. Transm. **9**, 177 (1973)].
[11] A. Winter, e-print quant-ph/9907077.