Accessible information and optimal strategies for real symmetrical quantum sources

Masahide Sasaki,[1] Stephen M. Barnett,[2] Richard Jozsa,[3] Masao Osaki,[4] and Osamu Hirota[4]
[1]Communications Research Laboratory, Ministry of Posts and Telecommunications, Koganei, Tokyo 184-8795, Japan
[2]Department of Physics and Applied Physics, University of Strathclyde, Glasgow G4 0NG, Scotland
[3]School of Mathematics and Statistics, University of Plymouth, Plymouth, Devon PL4 8AA, England
[4]Research Center for Quantum Communications, Tamagawa University, Tamagawa-gakuen, Machida, Tokyo 194-8610, Japan

(Received 22 December 1998)

We study the problem of optimizing the Shannon mutual information for sources of real quantum states, i.e., sources for which there is a basis in which all of the states have only real components. We consider in detail the sources $\mathcal{E}_M$ of M equiprobable quantum bit (qubit) states lying symmetrically around the great circle of real states on the Bloch sphere and give a variety of explicit optimal strategies. We also consider general real group-covariant sources for which the group acts irreducibly on the subset of all real states and prove the existence of a real group-covariant optimal strategy, extending a theorem of Davies [E. B. Davies, IEEE. Inf. Theory **IT-24**, 596 (1978)]. Finally, we propose an optical scheme to implement our optimal strategies, simple enough to be realized with present technology. [S1050-2947(99)11005-9]

PACS number(s): 03.67.−a, 03.65.Bz, 42.79.Sz, 89.80.+h

I. INTRODUCTION

There are two principal measures of quality in the quantum detection problem for a given finite number of quantum states with fixed prior probabilities. One is the minimization of a specified Bayes cost, and the other is the maximization of the Shannon mutual information [1–3]. The former is useful if one has to reach a decision after performing a single quantum measurement whereas the latter is more relevant for the problem of transmitting as much classical information as possible using the given ensemble of states. In this paper we will consider the problem of maximizing the Shannon mutual information for a certain class of quantum ensembles.

In a general communication setting, let $\{x_i \in X\}$ be input letters and let $\{\xi_i\}$ be their prior probabilities. Let us denote output letters by $\{y_j \in Y\}$. Both the Bayes cost and the Shannon mutual information are defined in terms of the conditional probability $P(j|i)$ of obtaining output y_j provided that the letter sent was x_i. The former is defined as

$$B(X:Y)=\sum_{ij} C_{ij}\xi_i P(j|i) \tag{1}$$

for a Bayes cost matrix $[C_{ij}]$, while the latter is defined as

$$I(X:Y)=\sum_{i,j} P_{XY}(i,j)\log\frac{P_{XY}(i,j)}{P_X(i)P_Y(j)}, \tag{2}$$

where $P_{XY}(i,j)$ $[\equiv \xi_i P(j|i)]$ is the joint probability, and $P_X(i)$ $(\equiv \xi_i)$ and $P_Y(j)$ $[\equiv \Sigma_k \xi_k P(j|k)]$ are the marginal probabilities of the input and output letters x_i and y_j, respectively. (Since all the results in this paper are valid for any logarithm base, we shall specify the base only where necessary.) In classical information theory, the channel matrix $[P(j|i)]$ is given and fixed, characterizing the noise in the channel. In contrast, in a quantum information theoretic context where signal carriers are to be quantum states transmitted without noise, the channel matrix generally becomes a variable. This is because the act of quantum detection itself generally has a probabilistic output so the channel matrix is dependent on the choice of quantum detection strategy. More precisely, the input letters correspond to a set of positive trace class operators of trace one $\{\hat{\rho}_i\}$ on a Hilbert space $\mathcal{H}_s$. A quantum detection strategy is described by a positive operator-valued measure (POVM) on $\mathcal{H}_s$. A POVM is any set $\{\hat{\pi}_j\}$ of Hermitian positive operators forming a resolution of the identity

$$\hat{\pi}_j^\dagger=\hat{\pi}_j, \quad \hat{\pi}_j \geqslant 0 \quad \forall j, \quad \sum_j \hat{\pi}_j=\hat{I}. \tag{3}$$

The detection operator $\hat{\pi}_j$ corresponds to the output letter y_j and the conditional probabilities are given by

$$P(j|i)=\mathrm{Tr}(\hat{\pi}_j\hat{\rho}_i).$$

Thus in the quantum context the optimization of $I(X:Y)$ is carried out with respect to the choice of POVM $\{\hat{\pi}_j\}$ for fixed ensemble $\mathcal{E}=\{\hat{\rho}_i;\xi_i\}$ (i.e., with fixed letter states $\hat{\rho}_i$ and fixed prior probabilities ξ_i). The maximum value of $I(X:Y)$ is called the accessible information of the ensemble $\mathcal{E}$.

The set $\mathcal{P}$ of all POVM's is a convex set and $I(X:Y)$ enjoys the following fundamental property:

(CONV): For a fixed ensemble $\mathcal{E}=\{\hat{\rho}_i;\xi_i\}$, $I(X:Y)$ is a convex function on $\mathcal{P}$.

A proof of (CONV) is given in theorem 2.7.4 of [4]. Let $I(\mathcal{E}:\mathcal{A})$ denote the mutual information obtained from the POVM $\mathcal{A}$ applied to the ensemble $\mathcal{E}$. Then if $\mathcal{A}$ is a convex combination of POVMs $\mathcal{A}_i$,

$$\mathcal{A}=p_1\mathcal{A}_1+\cdots p_n\mathcal{A}_n,$$

it follows from (CONV) that

$$I(\mathcal{E}:\mathcal{A})\leqslant\sum_i p_i I(\mathcal{E}:\mathcal{A}_i)\leqslant\max_i I(\mathcal{E}:\mathcal{A}_i). \tag{4}$$

The Bayes cost $B(X:Y)$ is an affine concave function on the convex set $\mathcal{P}$. Therefore the Bayes cost minimization problem is a kind of linear programming problem and is expected to have a unique solution. A necessary and sufficient condition for specifying the optimum solution is known [1,2]. On the other hand, the Shannon mutual information $I(X:Y)$ is a nonlinear and convex function on $\mathcal{P}$. The maximization of this quantity is a much harder problem and only a necessary condition for the optimum is known [1]. Thus the maximization of $I(X:Y)$ with respect to the detection strategy $\{\hat{\pi}_j\}$ is a basic and open problem in quantum information theory.

In this problem, the number of outputs is not necessarily the same as the number of the inputs. The optimum solution is not necessarily unique either. However, it is known that there must be at least one optimum solution which corresponds to an extreme point of the convex set $\mathcal{P}$. This is due to the convexity of the function $I(X:Y)$. Such an extreme point is a set of rank one elements, which means that each $\hat{\pi}_j$ has the form $\kappa|v\rangle\langle v|$, where $|v\rangle$ is a pure state and $0\leqslant\kappa\leqslant 1$. The number of elements, N, can be bounded by $d\leqslant N\leqslant d^2$, where d is the dimension of the Hilbert space $\mathcal{H}_s$ of which the input state ensemble $\{\hat{\rho}_i\}$ is made [5]. $I(X:Y)$ is also possibly maximized at some interior points of $\mathcal{P}$ as well. In that case the number of outcomes may exceed d^2. Explicit examples of optimal solutions have been given for binary ensembles [6–8] and for the ensemble of four qubit states with tetrahedral symmetry [5]. The latter is a specific example of a general result of Davies [5] characterizing the form of an optimal strategy for any symmetrical ensemble whose symmetry group acts *irreducibly* on the *whole* state space.

In this paper we will study the accessible information and corresponding optimal strategies for an ensemble $\mathcal{E}_M$ of M qubit states with symmetry group Z_M, the group of integers modulo M. Some of our results will also apply to more general ensembles. $\mathcal{E}_M$ may be explicitly described as follows. Let $\{\binom{1}{0},\binom{0}{1}\}$ be the z-spin eigenstates and write $|\psi_0\rangle=\binom{1}{0}$. Let

$$\hat{V}\equiv\exp\left(-i\frac{\pi}{M}\hat{\sigma}_y\right)=\begin{pmatrix}\cos\frac{\pi}{M} & -\sin\frac{\pi}{M}\\ \sin\frac{\pi}{M} & \cos\frac{\pi}{M}\end{pmatrix}. \tag{5}$$

Then $\mathcal{E}_M$ consists of the M states

$$|\psi_k\rangle=\hat{V}^k|\psi_0\rangle=\begin{pmatrix}\cos\frac{k\pi}{M}\\ \sin\frac{k\pi}{M}\end{pmatrix},\quad k=0,\ldots,M-1, \tag{6}$$

taken with equal prior probabilities $\xi_k=1/M$. Note that these states (in the z-spin basis) involve only *real* components. On the Bloch sphere they are equally spaced around a great circle C in the x-z plane consisting of all real states. The antipodal points which have C as an equator are the two $\hat{\sigma}_y$ eigenstates. Thus $\mathcal{E}_M$ is clearly symmetrical with respect to the group Z_M whose generator is represented by $2\pi/M$ rotation about the axis joining the $\hat{\sigma}_y$ eigenstates. At the Hilbert space level the operators $\hat{V}^k$ in Eq. (5) provide a projective unitary representation of Z_M [e.g., $\hat{V}^M=-I$ and cf Eq. (7) later].

This symmetry group does not act irreducibly on the whole state space. Indeed the $\hat{\sigma}_y$ eigenstates are left invariant by the group action. (Irreducibility on the whole state space requires that the only invariant point is the maximally mixed state $\frac{1}{2}\hat{I}$.) Hence we cannot apply Davies' theorem [5] to provide an optimal strategy for $\mathcal{E}_M$. Nevertheless we will prove that the conclusion of Davies' theorem remains true in this case, i.e., that there exists a pure state $|a_0\rangle$ such that the Z_M-symmetric POVM

$$\mathcal{A}_M=\left\{\frac{2}{M}|a_k\rangle\langle a_k|:k=0,\ldots,M-1\right\},$$

$$\text{where}\quad |a_k\rangle=\hat{V}^k|a_0\rangle$$

is an optimal strategy for $\mathcal{E}_M$. Furthermore, we will show that $|a_0\rangle$ may be taken to be the state orthogonal to $|\psi_0\rangle$.

The case $M=3$ is of particular interest. It is the so-called trine ensemble which has been much studied [9–11]. Holevo in 1973 [9] showed that no von Neumann measurement in $\mathcal{H}_2$ can be an optimal strategy, demonstrating the necessity of considering more general POVMs in quantum detection theory. Since that time it has been conjectured that the strategy $\mathcal{A}_3$ above is optimal for the trine source. Our results resolve this conjecture affirmatively.

The strategy $\mathcal{A}_M$ has M elements. However, as noted above, for ensembles in $d=2$ dimensions there is always an optimal strategy with at most $d^2=4$ elements (which does not increase with M). We will show that the ensembles $\mathcal{E}_M$ always have an optimal strategy with at most three elements and explicit strategies of this form will be described for all M. If M is even, then $\mathcal{E}_M$ consists of $M/2$ pairs of orthogonal states. Let $\{|\xi\rangle,|\eta\rangle\}$ be any one of these pairs. We will show that the two-element POVM $\{|\xi\rangle\langle\xi|,|\eta\rangle\langle\eta|\}$ (a regular von Neumann measurement) is always an optimal strategy when M is even. We will also describe further optimal K-element POVMs where K lies between 3 and M.

II. GROUP-THEORETIC APPROACH

We begin by setting up a group-theoretic formalism for symmetric ensembles, leading to a main result (theorem 1) which applies to symmetric ensembles of real states in any dimension $d\geqslant 2$. An essential requirement in many of our results will be that various states and unitary operators be *real*. The requirement that a state or operator be real has, of course, no intrinsic physical meaning. When we speak of real states and real operators we will always mean simply that *there exists a basis of the Hilbert space relative to which all the required objects simultaneously have real components or real matrix elements*.

A projective unitary representation of a group G is an assignment of a unitary operation $\hat{U}_g$ to each member of G satisfying

$$\hat{U}_{g_1}\hat{U}_{g_2}=e^{i\phi(g_1,g_2)}\hat{U}_{g_1g_2}, \tag{7}$$

where the phases $\phi(g_1,g_2)$ may be chosen arbitrarily. A finite ensemble $\mathcal{E}$ of equiprobable (generally mixed) states is said to be symmetric with respect to the group G, or G-covariant, if the following condition is satisfied: there is a projective unitary representation $\{\hat{U}_g\}$ of G such that for all g, $\hat{U}_g\hat{\rho}\hat{U}_g^\dagger$ is in $\mathcal{E}$ whenever $\hat{\rho}$ is in $\mathcal{E}$. We write

$$g\hat{\rho}=\hat{U}_g\hat{\rho}\hat{U}_g^\dagger \tag{8}$$

for the action of g on the state $\hat{\rho}$. The phases $\phi(g_1,g_2)$ do not appear in Eq. (8) and $g_1(g_2(\hat{\rho}))=(g_1g_2)(\hat{\rho})$. Note that, in contrast to Davies [5], we do not require that G parametrises $\mathcal{E}$, i.e., G need not act transitively on the set of states of $\mathcal{E}$. For example, $\mathcal{E}_M$ is Z_M-covariant and the action is transitive, but $\mathcal{E}_{2N}$ is also Z_2- and Z_N-covariant via nontransitive actions.

A G-covariant POVM $\mathcal{A}$ (for the projective unitary representation $\{\hat{U}_g\}$) is a POVM such that $\hat{U}_g\hat{A}\hat{U}_g^\dagger$ is in $\mathcal{A}$ whenever $\hat{A}$ is in $\mathcal{A}$. We write

$$g\hat{A}\equiv\hat{U}_g\hat{A}\hat{U}_g^\dagger \tag{9}$$

for the action of g on a POVM element $\hat{A}$. From Eqs. (8) and (9) we see that $\mathrm{Tr}(\hat{A}\hat{\rho})=\mathrm{Tr}(g\hat{A}\cdot g\hat{\rho})$, i.e., the probability of outcome $\hat{A}$ on state $\hat{\rho}$ is G-invariant. Hence

$$\mathrm{Tr}(g\hat{A}\cdot\hat{\rho})=\mathrm{Tr}(\hat{A}\cdot g^{-1}\hat{\rho}), \tag{10}$$

so that the set of probabilities of the G-shifted outputs $g\hat{A}$ on a fixed input $\hat{\rho}$ is obtained as a permutation of the set of probabilities of the unshifted output $\hat{A}$ acting on suitably shifted inputs.

Let $\mathcal{E}$ be a G-covariant ensemble with projective unitary representation $\{\hat{U}_g\}$. We aim to find conditions on $\{\hat{U}_g\}$ which will guarantee the existence of a G-covariant POVM $\mathcal{A}=\{\hat{A}_g : g\in G\}$ with elements parametrized by G, and having group action $g\hat{A}_h=\hat{A}_{gh}$. Thus if e is the identity of G we have

$$\hat{A}_g=\hat{U}_g\hat{A}_e\hat{U}_g^\dagger, \tag{11}$$

and we require

$$\hat{M}\equiv\sum_{g\in G}\hat{A}_g=\hat{I}. \tag{12}$$

(Later we will take the elements of $\mathcal{A}$ to be rank 1 and consider the question of when $\mathcal{A}$ is an optimal strategy for $\mathcal{E}$.) From Eq. (11) we see that $\hat{M}$ commutes with all the $\hat{U}_g$'s:

$$\hat{U}_g\hat{M}=\hat{M}\hat{U}_g. \tag{13}$$

Thus if the set $\{\hat{U}_g\}$ acts irreducibly on the state space (i.e., there is no proper invariant subspace), Schur's lemma will guarantee that Eq. (12) holds. This fact is used by Davies [5] to characterize an optimal strategy for any G-covariant ensemble whose symmetry group acts irreducibly on the whole state space. However this condition of full irreducibility on the whole state space is not necessary for Eq. (12) to hold. We will use the following more general form of Schur's lemma.

Lemma 1. Let $\{\hat{M}_g\}$ be any set of nonsingular d by d matrices over some field F which acts irreducibly on the vector space $V=F^d$ (i.e., there is no proper subspace mapped to itself by all the $\hat{M}_g$'s). Suppose that $\hat{K}$ is any matrix that commutes with all the $\hat{M}_g$'s:

$$\hat{K}\hat{M}_g=\hat{M}_g\hat{K}. \tag{14}$$

Then (a) either $\hat{K}=0$ or $\hat{K}$ is nonsingular; (b) if $\hat{K}$ has a nonzero eigenvalue λ in F, then $\hat{K}=\lambda\hat{I}$.

Proof. (a) Let $\hat{K}(V)$ denote the image of V under the map $\hat{K}$ and similarly for $\hat{M}_g(V)$. Since $\hat{M}_g$ is nonsingular we have $\hat{M}_g(V)=V$. By Eq. (14) we have $\hat{M}_g\hat{K}(V)=\hat{K}\hat{M}_g(V)=\hat{K}(V)$, i.e., $\hat{K}(V)$ is an invariant subspace. Hence either $\hat{K}(V)=0$ (in which case $\hat{K}=0$) or else $\hat{K}(V)=V$ (in which case $\hat{K}$ is nonsingular). (b) If $\hat{K}$ has eigenvalue λ in F, then $\hat{B}=\hat{K}-\lambda\hat{I}$ is singular. Also $\hat{B}\hat{M}_g=\hat{M}_g\hat{B}$ for all g. Hence by (a), $\hat{B}$ must be zero, i.e., $\hat{K}=\lambda\hat{I}$. ■

We will apply this lemma with $F=\mathbb{R}$ to obtain useful results about G-covariant ensembles of *real* states whose group G acts irreducibly only on the restricted set $\mathbb{R}^d$ of real states (but not necessarily irreducibly on the full state space). This is the case for our ensembles $\mathcal{E}_M$. Let $|G|$ denote the size of G and let $d=\mathrm{Tr}\,\hat{I}$ be the dimension of the Hilbert space.

Lemma 2. Suppose that $\{\hat{U}_g\}$ is a projective unitary representation of G such that $\hat{U}_g$ are all *real* matrices and $\{\hat{U}_g\}$ acts irreducibly on $\mathbb{R}^d$. Let $|v\rangle\in\mathbb{R}^d$ be any real state. Write

$$\hat{A}_g=\frac{d}{|G|}\hat{U}_g|v\rangle\langle v|\hat{U}_g^\dagger.$$

Then $\{\hat{A}_g : g\in G\}$ is a G-covariant POVM, i.e., $\Sigma_{g\in G}\hat{A}_g=\hat{I}$.

Proof. Let $\hat{M}=\Sigma_{g\in G}\hat{A}_g$. Then $\hat{M}$ is a real matrix and $\hat{M}\hat{U}_g=\hat{U}_g\hat{M}$ for all $g\in G$. Also $\hat{M}$ is a Hermitian positive matrix (being a sum of projectors with positive coefficients) so it has a real positive eigenvalue $\lambda>0$. By the previous lemma, $\hat{M}=\lambda\hat{I}$. Since $\mathrm{Tr}\,\hat{A}_g=d/|G|$ for all g, we get $\mathrm{Tr}\,\hat{M}=d=\mathrm{Tr}\,\hat{I}$ so $\lambda=1$. ■

Theorem 1. Let $\mathcal{E}$ be any ensemble of equiprobable real states in dimension d. Suppose that $\mathcal{E}$ is G-covariant with respect to a projective unitary representation $\{\hat{U}_g\}$ of real matrices which acts irreducibly on $\mathbb{R}^d$. Then there exists a real pure state $|v\rangle$ such that the G-covariant POVM $\mathcal{D}=\{\hat{D}_g : g\in G\}$ defined by

$$\hat{D}_g=\frac{d}{|G|}\hat{U}_g|v\rangle\langle v|\hat{U}_g^\dagger$$

is an optimal strategy for $\mathcal{E}$.

Proof. We will work in the basis with respect to which the states of $\mathcal{E}$ and the matrices $\hat{U}_g$ have real entries. Let $\mathcal{A}$

$=\{\hat{A}_1, \ldots ,\hat{A}_n\}$ be *any* optimal POVM for $\mathcal{E}$. We will transmogrify $\mathcal{A}$ into the required form while preserving optimality. First strip off all imaginary parts of the entries of the matrices $\hat{A}_k$. Let $\tilde{A}_k = \mathrm{Re}(\hat{A}_k)$ and $\tilde{\mathcal{A}} = \{\tilde{A}_1, \ldots ,\tilde{A}_n\}$. Then $\tilde{\mathcal{A}}$ is again a POVM and has real symmetric matrices as elements. [To see that $\tilde{A}_k$ is a positive matrix note that A_k positive implies that the complex conjugate $\hat{A}_k^*$ is positive so $\tilde{A}_k = \frac{1}{2}(\hat{A}_k + \hat{A}_k^*)$ must be positive. Also $\Sigma\hat{A}_k = \hat{I}$ and $\hat{I}$ is real so $\Sigma\tilde{A}_k = I$ too.] Next note that $\mathrm{Tr}\,\hat{A}_k\hat{\rho} = \mathrm{Tr}\,\tilde{A}_k\hat{\rho}$ for any real state $\hat{\rho}$ [since $\mathrm{Im}(\hat{A}_k)$ is antisymmetric] so $\tilde{\mathcal{A}}$ remains an optimal strategy.

In general $\tilde{\mathcal{A}}$ will not have rank 1 elements even if $\mathcal{A}$ had rank 1 elements. Thus decompose each $\tilde{A}_k$ into its rank 1 eigenprojectors (multiplied by the corresponding eigenvalues) which are necessarily real as the eigenvalues/vectors of any real symmetric matrix are real. Then form the larger POVM $\mathcal{B} = \{\hat{B}_1, \ldots ,\hat{B}_m\}$ comprising all the scaled rank 1 eigenprojectors above. Such a refinement of a POVM can never decrease the mutual information, so $\mathcal{B}$ with real rank 1 elements is still optimal.

Now look at

$$\hat{C}_{kg} = \frac{1}{|G|} g\hat{B}_k \quad \text{for} \quad g \in G \quad \text{and} \quad k = 1, \ldots ,m. \tag{15}$$

Note that $\Sigma_{kg}\hat{C}_{kg} = \hat{I}$ since $\Sigma\hat{B}_k = \hat{I}$ and $g\hat{I} = \hat{I}$ for all g. Let $\mathcal{C} = \{\hat{C}_{kg}\}$ be the corresponding POVM with $|G|m$ elements. Thus $\mathcal{C}$ is G-covariant but the action of G is not transitive. We finally aim to cut down $\mathcal{C}$ to a smaller optimal G-covariant POVM with elements labeled by G.

Let $I(\mathcal{E}:\mathcal{A})$ denote the mutual information obtained from any POVM $\mathcal{A}$ applied to any ensemble $\mathcal{E}$. First we show that $I(\mathcal{E}:\mathcal{C}) = I(\mathcal{E}:\mathcal{B})$ so that $\mathcal{C}$ remains optimal. Let us label the inputs by $i \in \mathcal{I}$ and denote conditional probabilities for $\mathcal{C}$ by $P_{\mathcal{C}}(kg|i)$. Denote the conditional probabilities for $\mathcal{B}$ by $P_{\mathcal{B}}(k|i)$ and let ξ be the constant prior input probability. Then

$$P_{\mathcal{C}}(kg|i) = \mathrm{Tr}\,\hat{C}_{kg}\hat{\rho}_i = \frac{1}{|G|}\mathrm{Tr}\,g\hat{B}_k \cdot \hat{\rho}_i .$$

According to Eq. (10), for each fixed g and k the resulting set of probabilities labeled by $i \in \mathcal{I}$ will be just a *permutation* of the set $P_{\mathcal{B}}(k|i)$, rescaled by $1/|G|$. Thus the joint probabilities

$$P_{\mathcal{E}\mathcal{C}}(i,kg) = \xi P_{\mathcal{C}}(kg|i) = \frac{1}{|G|}\xi P_{\mathcal{B}}(k|i) = \frac{1}{|G|}P_{\mathcal{E}\mathcal{B}}(i,k)$$

and the marginal probabilities for $\mathcal{C}$ and $\mathcal{B}$,

$$P_{\mathcal{C}}(kg) = \sum_l \xi P_{\mathcal{C}}(kg|l) = \frac{1}{|G|}\sum_l \xi P_{\mathcal{B}}(k|l) = \frac{1}{|G|}P_{\mathcal{B}}(k),$$

will be independent of g, and also

$$\sum_i P_{\mathcal{E}\mathcal{C}}(i,kg)\log\frac{P_{\mathcal{E}\mathcal{C}}(i,kg)}{\xi P_{\mathcal{C}}(kg)} = \frac{1}{|G|}\sum_i P_{\mathcal{E}\mathcal{B}}(i,k)\log\frac{P_{\mathcal{E}\mathcal{B}}(i,k)}{\xi P_{\mathcal{B}}(k)}$$

will be independent of g. The mutual information $I(\mathcal{E}:\mathcal{C})$ and $I(\mathcal{E}:\mathcal{B})$ is given by [cf. Eq. (2)]

$$I(\mathcal{E}:\mathcal{C}) = \sum_{i,kg} P_{\mathcal{E}\mathcal{C}}(i,kg)\log\frac{P_{\mathcal{E}\mathcal{C}}(i,kg)}{\xi P_{\mathcal{C}}(kg)},$$

$$I(\mathcal{E}:\mathcal{B}) = \sum_{i,k} P_{\mathcal{E}\mathcal{B}}(i,k)\log\frac{P_{\mathcal{E}\mathcal{B}}(i,k)}{\xi P_{\mathcal{B}}(k)}.$$

On substituting the above G-invariant expressions into $I(\mathcal{E}:\mathcal{C})$ we readily get $I(\mathcal{E}:\mathcal{C}) = I(\mathcal{E}:\mathcal{B})$. (Our argument is actually an explicit example of the claim in lemma 5 of [5].) Hence $\mathcal{C}$ remains optimal.

Finally note that for each i, $\hat{B}_i/(\mathrm{Tr}\,\hat{B}_i)$ is a real pure state so, by lemma 2,

$$\mathcal{D}_i = \left\{ \frac{d}{|G|}\frac{g\hat{B}_i}{\mathrm{Tr}\,\hat{B}_i} : g \in G \right\}$$

is a POVM for each i. Now $(\mathrm{Tr}\,\hat{B}_i/d)\mathcal{D}_i = \{(1/|G|)g\hat{B}_i : g \in G\}$, so $\mathcal{C}$ is a convex combination

$$\mathcal{C} = \sum_{i=1}^{m}\frac{\mathrm{Tr}\,\hat{B}_i}{d}\mathcal{D}_i .$$

Hence, by Eq. (4),

$$I(\mathcal{E}:\mathcal{C}) \leq \max_i I(\mathcal{E}:\mathcal{D}_i).$$

Since $\mathcal{C}$ was optimal it follows that at least one of the $\mathcal{D}_i$'s is optimal. This gives an optimal G-covariant POVM with real rank 1 elements, parametrized by G, completing the proof. ■

III. OPTIMAL STRATEGIES FOR $\mathcal{E}_M$

We now return to the Z_M-covariant ensemble $\mathcal{E}_M$ in two dimensions, comprising the states

$$|\psi_k\rangle = \begin{pmatrix} \cos\dfrac{k\pi}{M} \\ \sin\dfrac{k\pi}{M} \end{pmatrix}, \quad k = 0, \ldots ,M-1,$$

with equal prior probabilities $1/M$. According to theorem 1, there must exist an optimal Z_M-covariant POVM $\mathcal{A} = \{\hat{A}_0, \ldots ,\hat{A}_{M-1}\}$ with M real rank 1 elements. The elements will have the form $\hat{A}_j = |a_j\rangle\langle a_j|$ with

$$|a_j\rangle = \hat{V}^j|a_0\rangle = \sqrt{\frac{2}{M}}\begin{pmatrix} \cos\left(\theta + \dfrac{j\pi}{M}\right) \\ \sin\left(\theta + \dfrac{j\pi}{M}\right) \end{pmatrix}, \quad j = 0, \ldots ,M-1, \tag{16}$$

and $\hat{V}$ is given in Eq. (5). The conditional probabilities $p(j|k) = |\langle a_j|\psi_k\rangle|^2$ may be readily computed and after some rearrangement we obtain the mutual information $I(\theta)$ explicitly as

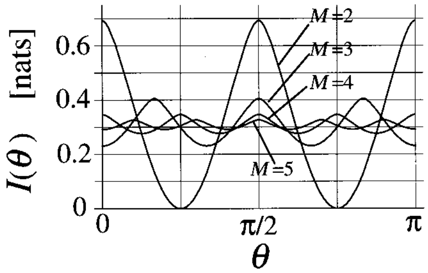

FIG. 1. Shannon mutual information $I(\theta)$ in nats versus the optimization parameter θ for $M=2$, 3, 4, and 5.

$$I(\theta)=\frac{1}{M}\sum_{k=0}^{M-1}\left[1+\cos\left(2\theta-\frac{2k\pi}{M}\right)\right]\times\ln\left[1+\cos\left(2\theta-\frac{2k\pi}{M}\right)\right]. \tag{17}$$

In this section, the base of the logarithm is taken as e. [For this base the numerical value of Eq. (17) is the amount of information in nats (natural units) rather than bits (binary digits).] From the symmetry, $I(\theta)$ is a periodic function with period π/M. Figure 1 shows numerical plots of $I(\theta)$ for $M=2$, 3, 4, and 5 and illustrates the following basic property.

Lemma 3. For each M, $I(\theta)$ has a global maximum at $\theta=\pi/2$.

The proof of this lemma is given in the Appendix.

Hence in general an optimal strategy for $\mathcal{E}_M$ consists of choosing a real rank 1 POVM with elements $\hat{A}_k$ lying in directions orthogonal to the input states $|\psi_k\rangle$. This POVM will be denoted by $\mathcal{A}_M$. The output $\hat{A}_k$ signifies with certainty that the input was not $|\psi_k\rangle$ but leaves a residual uncertainty in the remaining signal states.

For a given ensemble $\mathcal{E}$ the optimal strategy is not unique and in practice it may be of interest to find optimal POVMs with the minimum number of elements. The G-covariant optimal POVM above has M elements and we note here some ways of reducing this number using the group theoretic approach. In the next section, with different methods, we will show that three elements always suffice for any real qubit source, and develop corresponding strategies for the $\mathcal{E}_M$'s.

Lemma 4. Suppose that $k\neq 1$ divides M exactly. Then there is a Z_k-covariant optimal POVM for $\mathcal{E}_M$ with k real rank 1 elements.

Proof. Since k divides M, Z_M has a subgroup isomorphic to Z_k and so $\mathcal{E}_M$ is Z_k-covariant. Since $k\neq 1$, the action of Z_k contains a nontrivial rotation so it acts irreducibly on $\mathbb{R}^2$. Thus theorem 1 immediately gives the required result. ■

Remark. Lemma 4 may also be obtained by a convexity argument as follows. We will illustrate the idea with the specific example of $M=15$ and $k=3$. The general case is a straightforward generalization. $Z_{15}=\{0,1,\ldots,14\}$ has the subgroup $\{0,5,10\}$ isomorphic to Z_3. Let $\mathcal{A}_{15}=\{\hat{A}_0,\hat{A}_1,\ldots,\hat{A}_{14}\}$ be the optimal strategy given by theorem 1 and lemma 3, with the direction of $\hat{A}_k$ being orthogonal to the kth state of $\mathcal{E}_{15}$. According to lemma 2, the three directions 0,5,10 corresponding to the subgroup may be used to define a POVM. We just need to rescale $\hat{A}_0,\hat{A}_5$, and $\hat{A}_{10}$ so that they add up to $\hat{I}$. The scaling factor is $M/k=5$. Thus $\mathcal{B}_0=\{5\hat{A}_0,5\hat{A}_5,5\hat{A}_{10}\}$ is a POVM. Now $\hat{I}$ is always G invariant so we can apply the group elements $l=1,\ 2,\ 3$, and 4 of Z_{15} to $\mathcal{B}_0$ to obtain POVMs,

$$\mathcal{B}_l\equiv l\mathcal{B}_0=\{5\hat{A}_l,5\hat{A}_{l+5},5\hat{A}_{l+10}\}\quad\text{for}\quad l=0,1,2,3,4.$$

Note that the $\mathcal{B}_l$'s have elements parametrized by the *cosets* of Z_3 in Z_{15}. Also by symmetry of the construction, $I(\mathcal{E}_{15}:\mathcal{B}_l)$ is independent of l. Furthermore $\mathcal{A}_{15}$ is a uniform convex combination of the $\mathcal{B}_l$'s,

$$\mathcal{A}_{15}=\sum_{l=0}^{4}\frac{1}{5}\mathcal{B}_l,$$

so by Eq. (4),

$$I(\mathcal{E}_{15}:\mathcal{A}_{15})\leq\max_{l} I(\mathcal{E}_{15}:\mathcal{B}_l).$$

Since $\mathcal{A}_{15}$ was optimal we see that $\mathcal{B}_l$ is optimal for each l. This gives the result of lemma 4 and also identifies the directions of the k element POVM as being any chosen symmetrical set of k directions orthogonal to corresponding states of $\mathcal{E}_M$. ■

An immediate special case is the following.

Corollary. If M is even, then $\mathcal{E}_M$ is made up of $M/2$ pairs of orthogonal states. The von Neumann measurement defined by any one of these orthogonal pairs is an optimal strategy for $\mathcal{E}_M$. ■

Thus if M is composite we can significantly reduce the number of elements in our optimal strategy, but if M is prime then this number remains large. In the next section we give a different approach to reducing the number of elements, showing that just three elements always suffice for any ensemble of real qubit states.

IV. OPTIMAL POVMs WITH THREE ELEMENTS

Davies [5] has shown that any ensemble in d dimensions has an optimal strategy with N elements where $d\leq N\leq d^2$. This is directly based on (CONV), that is, $I(X:Y)$ is a convex function on the convex set $\mathcal{P}$ of all POVMs. Because of this, $I(X:Y)$ will always take its maximum value at an extreme point of the convex set $\mathcal{P}$ (and also possibly at some interior points as well). Each extreme point of $\mathcal{P}$ consists of N rank 1 elements bounded by $d\leq N\leq d^2$. If we restrict attention to only *real* ensembles, then this upper bound on N can be improved as follows [12].

Lemma 5. Let $\mathcal{E}$ be any ensemble of real states in d dimensions. Then the Shannon mutual information can be maximized by a POVM with N elements where $d\leq N\leq d(d+1)/2$.

Proof. The proof proceeds along the same lines as the original one in Ref. [5] with a slight replacement. For any POVM $\{\hat{\pi}_j\}$ write $\hat{\pi}_j=\mu_j d\overline{\pi}_j$, where Tr $\overline{\pi}_j=1$, so

$$\sum_j \mu_j\overline{\pi}_j=\hat{I}/d,\quad \sum_j \mu_j=1. \tag{18}$$

Let $\mathcal{X}$ be the (compact convex) set of all positive Hermitian operators with trace 1 (such as the $\overline{\pi}_j$'s). Since $I(\mathcal{E}:\mathcal{A})$ is a convex function on the set $\mathcal{P}$ of all POVMs, its maximum is attained at an extreme point of $\mathcal{P}$. The essential point of the original proof in Ref. [5] is that every extreme point of $\mathcal{P}$ has $D+1$ rank 1 elements where D is the real dimension of $\mathcal{X}$. In the case of general ensembles $D=d^2-1$. In our case of real ensembles the members of $\mathcal{X}$ and $\mathcal{P}$ can be restricted to real matrices so $\mathcal{X}$ comprises real symmetric trace 1 matrices and $D=d(d+1)/2-1$. Hence the extreme points of $\mathcal{P}$ have $\leqslant d(d+1)/2$ elements. ■

Thus for the real ensembles $\mathcal{E}_M$ with $d=2$, POVMs with three real elements suffice to provide an optimal strategy. To describe such a POVM, we first introduce the three real (unnormalized) vectors

$$|\omega_0\rangle = c\begin{pmatrix}1\\0\end{pmatrix}, \tag{19a}$$

$$|\omega_1\rangle = a\begin{pmatrix}\cos\varphi_a\\ \sin\varphi_a\end{pmatrix}, \tag{19b}$$

$$|\omega_2\rangle = b\begin{pmatrix}\cos\varphi_b\\ \sin\varphi_b\end{pmatrix}, \tag{19c}$$

where the first vector lies along the first basis direction and the remaining two are in a general position. Imposing the condition $\Sigma_j|\omega_j\rangle\langle\omega_j|=\hat{I}$, we get

$$c=\sqrt{2-a^2-b^2}, \tag{20a}$$

$$a^2=\frac{\cos\varphi_b}{\sin\varphi_a\sin(\varphi_a-\varphi_b)}, \tag{20b}$$

$$b^2=\frac{\cos\varphi_a}{\sin\varphi_b\sin(\varphi_b-\varphi_a)}, \tag{20c}$$

and

$$0\leqslant a^2+b^2\leqslant 2. \tag{20d}$$

Once the angles φ_a and φ_b have been chosen, a, b, and c are fixed. Finally we rotate these vectors around the y axis through an angle θ to make the general POVM with three real rank 1 elements:

$$\hat{\omega}_j(\theta)\equiv|\omega_j(\theta)\rangle\langle\omega_j(\theta)|, \tag{21a}$$

$$|\omega_j(\theta)\rangle=\hat{V}(\theta)|\omega_j\rangle, \quad \hat{V}(\theta)\equiv\exp(-i\theta\hat{\sigma}_y). \tag{21b}$$

This gives the most general POVM $\{\hat{\omega}_0(\theta),\hat{\omega}_1(\theta),\hat{\omega}_2(\theta)\}$ in terms of three independent parameters φ_a, φ_b, and θ.

We are now in a position to maximize the Shannon mutual information of $\mathcal{E}_M$ with (at most) three-element POVMs. We first give a useful preliminary lemma.

Lemma 6. Let $\mathcal{A}=\{\lambda_a^2|a\rangle\langle a|\}$ be any POVM with rank 1 elements labeled by a, where $0<\lambda_a\leqslant 1$ is real and

$$|a\rangle=\begin{pmatrix}\cos\theta_a\\ \sin\theta_a\end{pmatrix}$$

in the z-spin basis. Then the mutual information for $\mathcal{E}_M$ is given by

$$I(\mathcal{E}_M:\mathcal{A})=\sum_a\frac{\lambda_a^2}{2}I(\theta_a), \tag{22}$$

where $I(\theta)$ is the function given in Eq. (17).

Proof. The states $|\psi_k\rangle$ of $\mathcal{E}_M$ given in Eq. (6) lead to the conditional probabilities

$$P(a|k)=\lambda_a^2|\langle\psi_k|a\rangle|^2=\frac{\lambda_a^2}{2}\left[1+\cos\left(2\theta_a-\frac{2k\pi}{M}\right)\right].$$

Substituting these into Eq. (2) readily yields the formula Eq. (22) after a little algebra. ■

Theorem 2. The Shannon mutual information of $\mathcal{E}_M$ (for $M>2$) is maximized by the POVM $\mathcal{W}=\{\hat{\omega}_j^*=|\omega_j^*\rangle\langle\omega_j^*|:j=0,1,2\}$, where

$$|\omega_0^*\rangle=\begin{pmatrix}0\\ \sqrt{2-a^2-b^2}\end{pmatrix}, \tag{23a}$$

$$|\omega_1^*\rangle=a\begin{pmatrix}-\sin\left(\dfrac{m\pi}{M}\right)\\ \cos\left(\dfrac{m\pi}{M}\right)\end{pmatrix}, \tag{23b}$$

$$|\omega_2^*\rangle=b\begin{pmatrix}\sin\left(\dfrac{n\pi}{M}\right)\\ \cos\left(\dfrac{n\pi}{M}\right)\end{pmatrix}, \tag{23c}$$

and

$$a^2=\frac{\cos\left(\dfrac{n\pi}{M}\right)}{\sin\left(\dfrac{m\pi}{M}\right)\sin\left(\dfrac{(m+n)\pi}{M}\right)}\geqslant 0, \tag{24a}$$

$$b^2=\frac{\cos\left(\dfrac{m\pi}{M}\right)}{\sin\left(\dfrac{n\pi}{M}\right)\sin\left(\dfrac{(m+n)\pi}{M}\right)}\geqslant 0. \tag{24b}$$

Here m and n are any positive integers satisfying

$$0\leqslant a^2+b^2\leqslant 2. \tag{24c}$$

In some cases one of a, b, and $\sqrt{2-a^2-b^2}$ is zero and the POVM has only two elements.

Proof. For the three-element POVM $\mathcal{W}(\theta,\varphi_a,\varphi_b)=\{\hat{\omega}_0(\theta),\hat{\omega}_1(\theta),\hat{\omega}_2(\theta)\}$ with rank 1 elements, lemma 6 immediately gives

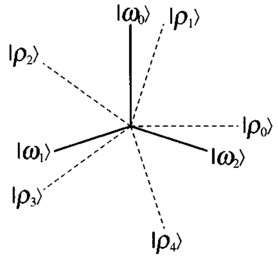

FIG. 2. Optimal POVM directions (thick solid lines) given by theorem 2 in the case of $M=5$. The input states are represented as $(-1)^k|\psi_k\rangle$ by the dashed lines whose lengths correspond to a unit state vector. The lengths of the thick solid lines are scaled according to the normalization factors of the corresponding POVM elements.

$$I(\mathcal{E}_M:\mathcal{W})=\left(1-\frac{a^2}{2}-\frac{b^2}{2}\right)I(\theta)+\frac{a^2}{2}I(\theta+\varphi_a)$$

$$+\frac{b^2}{2}I(\theta+\varphi_b).$$

Hence $I(\mathcal{E}_M:\mathcal{W})\le\max_\theta I(\theta)$. By lemma 3 this maximum is $I(\pi/2)$, the accessible information of $\mathcal{E}_M$. Furthermore, $I(\theta)$ is periodic in θ with period π/M. Hence we can achieve $I(\mathcal{E}_M:\mathcal{W})=I(\pi/2)$ by setting $\theta=\pi/2$ and choosing ϕ_a and ϕ_b to be any integer multiples of π/M. This gives Eqs. (23). Equations (24) are just the condition for $\{\hat{\omega}_j^*\}$ to be a POVM. ■

From this theorem we can develop various kinds of optimal strategies. We noted previously in corollary 1 that if M is even, then there exists an optimal strategy based on a pair of orthogonal directions. This also follows from theorem 2: if $M=4L-2$ with $L=1,2,\ldots$, then we may take $n=2L-1$ giving $a=0$ and a two-element POVM based on the directions $\binom{0}{1}$ and $\binom{1}{0}$. If $M=4L$ with $L=1,2,\ldots$, we may take $m=n=L$ giving $\sqrt{2-a^2-b^2}=0$ and an optimal POVM based on the directions $\binom{-1}{1}$ and $\binom{1}{1}$. In both cases the pair of directions coincides with an orthogonal pair of states of $\mathcal{E}_M$.

If M is odd, at least three outputs are required. In the case of $M=3$, we get an optimum strategy with three elements of equal norm. This coincides with our previous result $\mathcal{A}_3$ of theorem 1 and lemma 3. The cases of $M=5$ and $M=7$ are more interesting. In both cases, the optimum strategies consist of the three elements with the two different norms (in contrast to the Z_M-covariant strategies of theorem 1). A solution for $M=5$ is shown in Fig. 2. The POVM elements are represented by the thick solid lines and the dashed lines represent the input states. [Note that, for ease of presentation, these dashed lines representing the states of $\mathcal{E}_M$—symmetrically distributed around a whole circle—correspond to the vectors $(-1)^k|\psi_k\rangle$ rather than the original vectors in Eq. (6)]. According to choices of parameters (m,n) in theorem 2, there can be several configurations of the POVM directions. But by the symmetry of $\mathcal{E}_5$ they all lie in the same position relative to the ensemble as a whole, characterized by $a^2=b^2=1/[2\sin^2(2\pi/5)]$ as shown in Fig. 2.

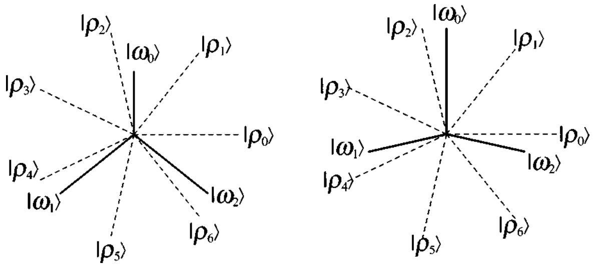

FIG. 3. Two inequivalent optimal POVMs in the case of $M=7$. The POVM directions and input states are represented by thick and dashed lines, respectively, according to the conventions of Fig. 2.

Figure 3 shows the case of $M=7$. There are now two inequivalent classes of POVM element directions. One corresponds to $a^2=b^2=1/[2\sin^2(2\pi/7)]$ where the angle between the two measurement vectors directed downward is $2\pi/7$ (the left figure), and the other corresponds to $a^2=b^2=1/[2\sin^2(3\pi/7)]$ where the angle between the two measurement vectors directed downward is $3\pi/7$ (the right figure).

Lemma 6 and theorem 2 may be used to provide a further variety of optimal K-element POVMs for $\mathcal{E}_M$, where K is between 3 and M.

Lemma 7. Let $\mathcal{A}$ be any POVM as described in lemma 6 for which all angles θ_a have the form

$$\theta_a=\frac{\pi}{2}+k_a\frac{\pi}{M},\quad \text{where } k_a \text{ is an integer.} \tag{25}$$

Then $\mathcal{A}$ is an optimal strategy for $\mathcal{E}_M$.

Proof. Since $I(\theta)$ is periodic with period π/M, we have $I(\theta_a)=I(\pi/2)$ for all a. Also $\Sigma\lambda_a^2=2$ so that Eq. (22) immediately gives $I(\mathcal{E}_M:\mathcal{A})=I(\pi/2)$ i.e., $\mathcal{A}$ is optimal. ■

Now note the following facts.

(a) All POVMs in theorem 2 satisfy Eq. (25).

(b) If $\mathcal{A}=\{\hat{A}_i\}$ is any POVM satisfying Eq. (25), then any Z_M-shifted version $\mathcal{A}_l$ of $\mathcal{A}$, defined for each $l\in Z_M$ by

$$\mathcal{A}_l=\{\hat{V}^l\hat{A}_i\hat{V}^{\dagger l}\}$$

is a POVM also satisfying Eq. (25). (The angles θ_a are just shifted by $l\pi/M$).

(c) If $\mathcal{A}_1,\ldots,\mathcal{A}_N$ is any list of POVMs satisfying Eq. (25), then any convex combination of the $\mathcal{A}_i$'s will satisfy Eq. (25). (In forming convex combinations we naturally amalgamate POVM elements from different $\mathcal{A}_i$'s that lie in the same direction.)

Hence any convex combination of any Z_M-shifted versions of the POVMs in theorem 2 will be an optimal strategy. For example, let us consider a convex combination between two POVM's in the case of $M=5$. The following $\{\hat{\omega}_j\}$ is one of the optimum detection strategies from theorem 2:

$$\hat{\omega}_0=(1-a^2)(\hat{I}-\hat{\sigma}_z), \tag{26a}$$

$$\hat{\omega}_1=\frac{a^2}{2}\left[\hat{I}-\sin\left(\frac{4\pi}{5}\right)\hat{\sigma}_x-\cos\left(\frac{4\pi}{5}\right)\hat{\sigma}_z\right], \tag{26b}$$

$$\hat{\omega}_2 = \frac{a^2}{2}\left[\hat{I} - \sin\left(\frac{6\pi}{5}\right)\hat{\sigma}_x - \cos\left(\frac{6\pi}{5}\right)\hat{\sigma}_z\right], \tag{26c}$$

where $a^2 = 1/[2\sin^2(2\pi/5)]$. The convex combination between $\{\hat{\omega}_j\}$ and $\{\hat{V}^2\hat{\omega}_j\hat{V}^{\dagger 2}\}$ forms the resolution of the identity

$$(1-\lambda)\sum_{j=0}^{2}\hat{\omega}_j + \lambda\sum_{k=0}^{2}\hat{V}^2\hat{\omega}_k\hat{V}^{\dagger 2} = \hat{I} \quad (\lambda \geqslant 0), \tag{27}$$

and we define

$$\hat{\mu}_0 = (1-\lambda)\hat{\omega}_0 + \lambda\hat{V}^2\hat{\omega}_2\hat{V}^{\dagger 2}, \tag{28a}$$

$$\hat{\mu}_1 = (1-\lambda)\hat{\omega}_1 + \lambda\hat{V}^2\hat{\omega}_0\hat{V}^{\dagger 2}, \tag{28b}$$

$$\hat{\mu}_2 = (1-\lambda)\hat{\omega}_2, \tag{28c}$$

$$\hat{\mu}_3 = \lambda\hat{V}^2\hat{\omega}_1\hat{V}^{\dagger 2}. \tag{28d}$$

(Note that $\hat{\omega}_0 \propto \hat{V}^2\hat{\omega}_2\hat{V}^{\dagger 2}$ and $\hat{\omega}_1 \propto \hat{V}^2\hat{\omega}_0\hat{V}^{\dagger 2}$.) This gives a four-element POVM $\{\hat{\mu}_j\}$ which maximizes the Shannon mutual information for $\mathcal{E}_5$.

The strategies in theorem 2 are not generally Z_M-covariant but they correspond to extreme points of $\mathcal{P}$. On the other hand, the Z_M-covariant strategy of theorem 1 is generally not an extreme point of $\mathcal{P}$. The Z_M-covariant POVM of theorem 1 can be related to the asymmetrical three-element POVM of theorem 2 as follows. Note first that if $\mathcal{W} = \{\hat{\omega}_j\}$ is any optimal POVM, then so is $m\mathcal{W} = \{\hat{V}^m\hat{\omega}_j\hat{V}^{\dagger m}\}$ for any $m \in Z_M$. Indeed,

$$I(\mathcal{E}_M : \mathcal{W}) = I(\mathcal{E}_M : m\mathcal{W}) \tag{29}$$

since the set of states of $\mathcal{E}_M$ is invariant under the action of Z_M. Given any one of the $N(=2,3)$-element POVMs $\{\hat{\omega}_j^*\}$ defined in theorem 2, one can consider the resolution of the identity

$$\frac{1}{M}\sum_{m=0}^{M-1}\sum_{j=0}^{N-1}\hat{V}^m\hat{\omega}_j^*\hat{V}^{\dagger m} = \hat{I}. \tag{30}$$

But the MN elements $\{\hat{V}^m\hat{\omega}_j^*\hat{V}^{\dagger m}\}$ are proportional to each other in groups of N and these groups may each naturally be summed and assigned a single element. This leads to the covariant M-element POVM which is just $\mathcal{A}_M$ of theorem 1 and lemma 3. In this sense $\mathcal{A}_M$ may be thought of as a convex combination

$$\mathcal{A}_M = \sum_{k\in Z_M}\frac{1}{M}k\mathcal{W},$$

where $\mathcal{W}$ is any one of the POVMs in theorem 2. If we know that $\mathcal{A}_M$ is optimal, then Eqs. (29) and (4) will imply that $\mathcal{W}$ is optimal too. This provides an alternative proof of theorem 2 if we already know theorem 1 and lemma 3. On the other hand, if conversely we are given the result of theorem 2 (which uses lemma 3), then the accessible information of $\mathcal{E}_M$ must be $I(\pi/2)$ so $\mathcal{A}_M$ must be optimal [since $I(\mathcal{E}_M : \mathcal{A}_M) = I(\pi/2)$ by definition of $I(\theta)$ and $\mathcal{A}_M$].

V. IMPLEMENTATION

The optimal POVMs $\mathcal{A}_M$ and $\mathcal{W}$ given in theorems 1 and 2 may be of interest from the viewpoint of putting quantum detection theory to the test. None of the POVMs for attaining maximum mutual information have been demonstrated by experiment yet. So far, only two kinds of optimal quantum detection scenarios have been confirmed experimentally. One is the Helstrom bound as the minimum *average* error probability [2], and the other is the Ivanovic-Dieks-Peres bound which gives the maximum probability for error-free detection, sometimes referred to as the unambiguous measurement [13–16]. (A concise review of both criteria can be found in Ref. [17].) The former scenario was first demonstrated experimentally by Barnett and Riis [18]. The latter has been demonstrated in the laboratory by Huttner *et al.* [19]. Both of these are concerned with discrimination between binary nonorthogonal states, in which case the optimal detection strategy consists of a projection onto the orthogonal pair of measurement vectors, that is, von Neumann measurement. In our case of $\mathcal{A}_M$ and $\mathcal{W}$ for $\mathcal{E}_M$ with M odd, we are dealing with essentially *nonorthogonal* measurement vectors in $\mathcal{H}_2$, which is called a *generalized* measurement. No von Neumann measurement can be an optimal strategy for $\mathcal{E}_M$ with M odd. This case is of particular interest here. It is already well known that this kind of generalized measurement can be converted into a standard von Neumann measurement in a larger Hilbert space by introducing an ancillary system. This so-called Naimark extension ensures that any POVM can be physically implemented in principle [2,3].

In this section we propose an optical scheme to demonstrate the optimal POVMs specified by $\mathcal{W}$ for $\mathcal{E}_M$ made of single-mode photon polarization states. As seen in the preceding section, $\mathcal{W}$ has three outcomes at most and suffices to provide an optimal strategy for all $\mathcal{E}_M$'s. For M odd, it is always possible to find the optimal strategy with $m=n$, that is, $a^2 = b^2 = 1/[2\sin^2(m\pi/M)]$ in theorem 2 if m is taken as $M/4 < m < M/2$. We consider the implementation of this particular detection strategy. The measurement vectors can be represented by

$$|\omega_0^*\rangle = -\sin\frac{\gamma}{2}|\downarrow\rangle, \tag{31a}$$

$$|\omega_1^*\rangle = -\frac{1}{\sqrt{2}}|\uparrow\rangle + \frac{1}{\sqrt{2}}\cos\frac{\gamma}{2}|\downarrow\rangle, \tag{31b}$$

$$|\omega_2^*\rangle = \frac{1}{\sqrt{2}}|\uparrow\rangle + \frac{1}{\sqrt{2}}\cos\frac{\gamma}{2}|\downarrow\rangle, \tag{31c}$$

where

$$\cos\frac{\gamma}{2} \equiv \cot\frac{m\pi}{M}, \quad \sin\frac{\gamma}{2} \equiv -\sqrt{1-\cot^2\frac{m\pi}{M}}, \tag{32}$$

and $|\uparrow\rangle$ and $|\downarrow\rangle$ are orthonormal bases of polarization. The first step is to make orthogonal measurement vectors by embedding $\{|\omega_0^*\rangle, |\omega_1^*\rangle, |\omega_2^*\rangle\}$ into a three- or higher-dimensional Hilbert space. One possible physical prescription is to make an optical circuit with two input ports, say, ''a'' and ''b.'' The signal state is guided into the port ''a,'' while the port ''b'' is initialized as the vacuum state. We can then consider the four-dimensional Hilbert space spanned by the orthonormal basis $\{|E_j\rangle\}$,

$$|E_0\rangle \equiv |\uparrow\rangle_a|0\rangle_b, \tag{33a}$$

$$|E_1\rangle \equiv |\downarrow\rangle_a|0\rangle_b, \tag{33b}$$

$$|E_2\rangle \equiv |0\rangle_a|\uparrow\rangle_b, \tag{33c}$$

$$|E_3\rangle \equiv |0\rangle_a|\downarrow\rangle_b, \tag{33d}$$

where $|0\rangle$ is the vacuum state and the subscripts a and b indicate the port ''a'' and ''b,'' respectively. A natural orthogonalization is

$$|\Omega_0\rangle \equiv |\omega_0^*\rangle_a|0\rangle_b + \cos\frac{\gamma}{2}|0\rangle_a|\uparrow\rangle_b, \tag{34a}$$

$$|\Omega_1\rangle \equiv |\omega_1^*\rangle_a|0\rangle_b + \frac{1}{\sqrt{2}}\sin\frac{\gamma}{2}|0\rangle_a|\uparrow\rangle_b, \tag{34b}$$

$$|\Omega_2\rangle \equiv |\omega_2^*\rangle_a|0\rangle_b + \frac{1}{\sqrt{2}}\sin\frac{\gamma}{2}|0\rangle_a|\uparrow\rangle_b, \tag{34c}$$

$$|\Omega_3\rangle \equiv |0\rangle_a|\downarrow\rangle_b, \tag{34d}$$

or equivalently,

$$|\Omega_0\rangle \equiv -\sin\frac{\gamma}{2}|\downarrow\rangle_a|0\rangle_b + \cos\frac{\gamma}{2}|0\rangle_a|\uparrow\rangle_b, \tag{35a}$$

$$|\Omega_1\rangle \equiv \frac{1}{\sqrt{2}}\left(-|\uparrow\rangle_a|0\rangle_b + \cos\frac{\gamma}{2}|\downarrow\rangle_a|0\rangle_b + \sin\frac{\gamma}{2}|0\rangle_a|\uparrow\rangle_b\right), \tag{35b}$$

$$|\Omega_2\rangle \equiv \frac{1}{\sqrt{2}}\left(|\uparrow\rangle_a|0\rangle_b + \cos\frac{\gamma}{2}|\downarrow\rangle_a|0\rangle_b + \sin\frac{\gamma}{2}|0\rangle_a|\uparrow\rangle_b\right), \tag{35c}$$

$$|\Omega_3\rangle \equiv |0\rangle_a|\downarrow\rangle_b. \tag{35d}$$

It is easy to check that $\{|\Omega_0\rangle, |\Omega_1\rangle, |\Omega_2\rangle\}$ give the same channel matrix as $\{|\omega_0^*\rangle, |\omega_1^*\rangle, |\omega_2^*\rangle\}$, that is, $\langle\omega_j|\psi_i\rangle = \langle\Omega_j|(|\psi_i\rangle_a|0\rangle_b)$ $(j=0,1,2)$. The second step is to decompose the von Neumann measurement $\{|\Omega_j\rangle\}$ into a unitary transformation followed by a measurement in the basis $\{|E_j\rangle\}$ in order to find a practical detector structure. We may write

$$\langle\Omega_0| \equiv \langle E_2|\hat{U}_2\hat{U}_1, \tag{36a}$$

$$\langle\Omega_1| \equiv \langle E_1|\hat{U}_2\hat{U}_1, \tag{36b}$$

$$\langle\Omega_2| \equiv \langle E_0|\hat{U}_2\hat{U}_1, \tag{36c}$$

$$\langle\Omega_3| \equiv \langle E_3|\hat{U}_2\hat{U}_1, \tag{36d}$$

where $\hat{U}_1$ and $\hat{U}_2$ are given by the matrices

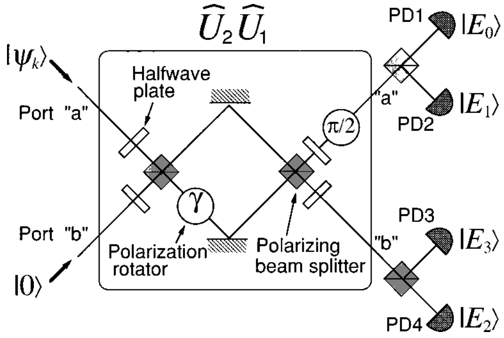

FIG. 4. Optical circuit implementing $\mathcal{W}=\{\hat{\omega}_0^*, \hat{\omega}_1^*, \hat{\omega}_2^*\}$. It consists of the unitary transformation $\hat{U}_2\hat{U}_1$ followed by the measurement $\{|E_j\rangle\}$. $\hat{U}_2\hat{U}_1$ is effected by four half-wave plates, two polarizing beam splitters, and two polarization rotators. The measurement $\{|E_j\rangle\}$ is made by photon counting at the four output ports.

$$\hat{U}_1 \equiv \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\frac{\gamma}{2} & \sin\frac{\gamma}{2} & 0 \\ 0 & -\sin\frac{\gamma}{2} & \cos\frac{\gamma}{2} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{37}$$

$$\hat{U}_2 \equiv \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{38}$$

in the $\{|E_0\rangle, |E_1\rangle, |E_2\rangle, |E_3\rangle\}$ -basis representation. Equations (36) mean that in the detector, the signal state $|\psi_i\rangle_a|0\rangle_b$ is first transformed by $\hat{U}_2\hat{U}_1$, and is then measured in the basis $\{|E_j\rangle\}$ which corresponds to the simultaneous measurement with respect to *which path* and *which polarization.* The final step is to translate $\hat{U}_2\hat{U}_1$ into a practical circuit. In fact, this unitary transformation can be effected by the simple circuit consisting of passive linear optical devices such as polarizing beam splitters, polarization rotators, and half-wave plates [20]. The circuit is shown in Fig. 4. (The bases $|\uparrow\rangle$ and $|\downarrow\rangle$ are assumed to be linearly polarized states.) The $\hat{U}_2\hat{U}_1$ part consists of four half-wave plates, two polarizing beam splitters, and two polarization rotators. Each half-wave plate is oriented at 45° to the horizontal so that each component of polarization is rotated to 90°. The polarization rotator represented by the circle with the rotation angle γ performs

$$\hat{R}_y(\gamma) = \begin{pmatrix} \cos\frac{\gamma}{2} & \sin\frac{\gamma}{2} \\ -\sin\frac{\gamma}{2} & \cos\frac{\gamma}{2} \end{pmatrix}. \tag{39}$$

The polarizing beam splitter represented by the square functions as a perfect mirror only for ↓ polarization (fast axis polarization). Light polarized along ↑ polarization (slow axis polarization) passes straight through it perfectly. The measurement $\{|E_j\rangle\}$ is made by photon counting at the four output ports. Note that only a single photon count at one of the three ports is expected and the outcome $|E_3\rangle$ is never expected. This structure is valid for any M (the number of the signals) if one tunes the rotation angle γ in $\hat{R}_y(\gamma)$ according to the value of M [see Eq. (32)]. The circuit is simple enough to be implemented with present technology.

VI. CONCLUDING REMARKS

We have considered optimal strategies for symmetrical sources of real quantum states, treating in detail the sources $\mathcal{E}_M$ of M real qubit states placed symmetrically in the x-z plane around the Bloch sphere. Davies [5] has provided a general theorem characterizing an optimal strategy for any G-covariant source whose group acts irreducibly on the whole state space. The symmetry group Z_M of $\mathcal{E}_M$ does not act irreducibly on that state space so Davies's theorem cannot be directly applied. However, we proved an extension of this theorem which applies to G-covariant sources of real states for which the group acts irreducibly on the subset of real states (as is the case for $\mathcal{E}_M$). This led to a Z_M-covariant optimal strategy $\mathcal{A}_M$ for $\mathcal{E}_M$.

We also derived alternative optimal strategies $\mathcal{W}$ which contain at most three real POVM elements. In deriving this strategy $\mathcal{W}$ we exploited the convexity of $I(X:Y)$ on the convex set $\mathcal{P}$ of all POVMs. These strategies are not G-covariant in general but correspond to extreme points of $\mathcal{P}$. The small number of elements can be advantageous for practical implementation of the detection strategies as seen in the preceding section. The G-covariant strategy is not generally an extreme point of $\mathcal{P}$ but for higher dimensions it would seem easier to derive explicit G-covariant solutions rather than extreme point solutions.

Our results have added to the relatively small number of quantum sources for which optimal strategies are explicitly known. They may be extended in various straightforward ways (which we have omitted for clarity of presentation). For example, the optimal strategies $\mathcal{A}_M$ and $\mathcal{W}$ for $\mathcal{E}_M$ remains optimal for the M-state source

$$\left\{(1-\epsilon)|\psi_k\rangle\langle\psi_k|+\epsilon\frac{1}{2}\hat{I}_2 : k\in Z_M\,;\frac{1}{M}\right\},$$

where each pure signal has been corrupted by noise given by the maximally mixed state $\frac{1}{2}\hat{I}_2$. This mixed state ensemble is clearly also G-covariant and the process of deriving the optimal strategy for this ensemble is quite the same as in the pure state case ($\epsilon=0$) but just multiplying the cosine terms in Eq. (17) by $(1-\epsilon)$. Then the same strategy remains optimal for the G-covariant mixed state ensemble although the accessible information decreases with ϵ as expected.

It is perhaps worth briefly contrasting our results of maximizing the mutual information with the problem of minimizing the average error probability. The latter is defined for $\mathcal{E}_M$ and any M-element POVM by

$$P_{\rm e}=1-\frac{1}{M}\sum_{k=0}^{M-1}P(k|k). \tag{40}$$

The $P_{\rm e}$-optimal strategy is $\{\hat{\pi}_k\}=\{2/M|\psi_k\rangle\langle\psi_k| : k\in Z_M\}$, that is, the POVM based on the state directions themselves. This is true also for the above mixed state ensemble. (The necessary and sufficient conditions for P_e optimality, as given in [1,2], are easily verified for $\{\hat{\pi}_k\}$.) Generally P_e minimization is an essentially different type of optimization problem from $I(X:Y)$ maximization.

Within the confines of our formalism, various interesting issues remain unresolved. For example we would like to know an optimal strategy for the real Z_M-covariant source ''double-$\mathcal{E}_M$'' in four dimensions comprising the 2-qubit signal states $\{|\psi_k\rangle|\psi_k\rangle : k\in Z_M\,;1/M\}$. In this case the symmetry group Z_M does not act irreducibly even on the subset of all real 2-qubit states. Interesting properties of double $\mathcal{E}_3$ have been considered in [11] from the viewpoint of coding gain of transmittable information.

It is also a remaining difficult problem to optimize a quantum channel over both the *a priori* probability distribution of signals *and* the detection strategy for a fixed set of quantum states. The solution is known only for the binary pure state channel.

ACKNOWLEDGMENTS

The authors would like to thank A. S. Holevo and T. S. Usuda for giving crucial comments on this work. They would also like to thank C. A. Fuchs, C. H. Bennett, and A. Chefles for helpful discussions. R.J. is grateful to W. K. Wootters who drew his attention to the reality of the source $\mathcal{E}_3$ as an important property, which ultimately led to theorem 1. M.S. and S.M.B. thank the Great Britain Sasakawa Foundation and the British Council for financial support. S.M.B. and R.J. thank the U.K. Engineering and Physical Science Research Council for financial support.

APPENDIX: PROOF OF THE LEMMA 3

Since $|\cos(2\theta-2k\pi/M)|<1$, $I(\theta)$ can be expanded by using the formula

$$(1+x)\ln(1+x)=x+\sum_{n=2}^{\infty}\frac{(-1)^n}{n(n-1)}x^n,\quad |x|<1. \tag{A1}$$

We get

$$\begin{aligned}I(\theta)=&\frac{1}{M}\sum_{k=0}^{M-1}\left[\cos\left(2\theta-\frac{2k\pi}{M}\right)\right.\\&\left.+\sum_{n=2}^{\infty}\frac{(-1)^n}{n(n-1)}\cos^n\left(2\theta-\frac{2k\pi}{M}\right)\right]\\=&\frac{1}{M}\sum_{n=2}^{\infty}\frac{(-1)^n}{n(n-1)}\sum_{k=0}^{M-1}\cos^n\left(2\theta-\frac{2k\pi}{M}\right),\end{aligned}$$

since $\Sigma_{k=0}^{M-1}\cos(2\theta-2k\pi/M)=0$. Next we separate out the even and odd parts of the series and replace powers of cosines by multiple angle cosines to get

$$
\begin{aligned}
I(\theta)=&\frac{1}{M}\sum_{n=1}^{\infty}\frac{(-1)^{2n}}{2n(2n-1)}\sum_{k=0}^{M-1}\cos^{2n}\left(2\theta-\frac{2k\pi}{M}\right)\\
&+\frac{1}{M}\sum_{n=1}^{\infty}\frac{(-1)^{2n+1}}{(2n+1)2n}\sum_{k=0}^{M-1}\cos^{2n+1}\left(2\theta-\frac{2k\pi}{M}\right)\\
=&\frac{1}{M}\sum_{n=1}^{\infty}\frac{(-1)^{2n}}{2n(2n-1)}\sum_{k=0}^{M-1}\frac{1}{2^{2n-1}}\left\{\frac{1}{2}\binom{2n}{n}\right.\\
&\left.+\sum_{l=0}^{n-1}\binom{2n}{l}\cos\left[(2n-2l)\left(2\theta-\frac{2k\pi}{M}\right)\right]\right\}\\
&+\frac{1}{M}\sum_{n=1}^{\infty}\frac{(-1)^{2n+1}}{(2n+1)2n}\sum_{k=0}^{M-1}\frac{1}{2^{2n}}\\
&\times\left\{\sum_{l=0}^{n}\binom{2n+1}{l}\cos\left[(2n+1-2l)\left(2\theta-\frac{2k\pi}{M}\right)\right]\right\}.
\end{aligned}
\tag{A2}
$$

Then recall that

$$
\sum_{k=0}^{M-1}\cos\left(L\frac{2k\pi}{M}\right)=\begin{cases}M & \text{for } L/M=q \quad \text{(integer)}\\ 0 & \text{for } L/M\neq\text{integer}.\end{cases}
\tag{A3}
$$

Applying this to Eq. (A2) with $L=2n-2l$ and $L=2n+1-2l$ in the even and odd series, we get

$$
\begin{aligned}
I(\theta)=&\sum_{n=1}^{\infty}\frac{(-1)^{2n}}{2n(2n-1)2^{2n-1}}\left[\frac{1}{2}\binom{2n}{n}\right.\\
&\left.+\sum_{l=0}^{n-1}\sum_{q=0}^{\infty}\binom{2n}{l}\cos(2\theta qM)\,\delta_{2n-2l,qM}\right]\\
&+\sum_{n=1}^{\infty}\frac{(-1)^{2n+1}}{(2n+1)2n2^{2n}}\\
&\times\left[\sum_{l=0}^{n}\sum_{q=0}^{\infty}\binom{2n+1}{l}\cos(2\theta qM)\,\delta_{2n+1-2l,qM}\right]\\
=&\sum_{n=1}^{\infty}\frac{1}{2n(2n-1)2^{2n}}\binom{2n}{n}\\
&+\sum_{q=0}^{\infty}f(qM)(-1)^{qM}\cos(2\theta qM),
\end{aligned}
\tag{A4}
$$

where

$$
\begin{aligned}
f(qM)=&\sum_{n=1}^{\infty}\sum_{l=0}^{n-1}\frac{\binom{2l+qM}{l}}{(2l+qM)(2l+qM-1)2^{2l+qM-1}}\\
&\times(\delta_{2n-2l,qM}+\delta_{2n+1-2l,qM}).
\end{aligned}
\tag{A5}
$$

Since $f(qM)>0$, $I(\theta)$ is maximized when $(-1)^{qM}\cos(2\theta qM)=1$, that is, $\theta=\pi/2$ for all M. ■

[1] A. S. Holevo, J. Multivariate Anal. **3**, 337 (1973).

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

[3] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer Academic Publishers, Dordrecht, 1993), pp. 279–289.

[4] T. Cover and J. Thomas, *Elements of Information Theory* (John Wiley and Sons, New York, 1991).

[5] E. B. Davies, IEEE Trans. Inf. Theory **IT-24**, 596 (1978).

[6] C. A. Fuchs and A. Peres, Phys. Rev. A **53**, 2038 (1996).

[7] M. Ban, K. Yamazaki, and O. Hirota, Phys. Rev. A **55**, 22 (1997).

[8] M. Osaki, M. Ban, and O. Hirota, J. Mod. Opt. **45**, 269 (1998).

[9] A. S. Holevo, Probl. Peredachi Inf. **9**, 31 (1973).

[10] P. Hausladen and W. K. Wootters, J. Mod. Opt. **41**, 2385 (1994).

[11] A. Peres and W. Wootters, Phys. Rev. Lett. **66**, 1119 (1992).

[12] The authors are indebted to A. S. Holevo for a private communication of the proof of lemma 5.

[13] I. D. Ivanovic, Phys. Lett. A **123**, 257 (1987).

[14] D. Dieks, Phys. Lett. A **126**, 303 (1988).

[15] A. Peres, Phys. Lett. A **128**, 19 (1988).

[16] A. Chefles and S. M. Barnett, Phys. Lett. A **250**, 223 (1998).

[17] S. M. Barnett, Philos. Trans. R. Soc. London, Ser. A **255**, 2279 (1997).

[18] S. M. Barnett and E. Riis, J. Mod. Opt. **44**, 1061 (1997).

[19] B. Huttner, A. Muller, J. D. Gautier, H. Zbinden, and N. Gisin, Phys. Rev. A **54**, 3783 (1996).

[20] N. J. Cerf, C. Adami, and P. G. Kwiat, Phys. Rev. A **57**, R1477 (1998).