Optimal Encoding of Classical Information in a Quantum Medium

Noam Elron and Yonina C. Eldar, *Member, IEEE*

***Abstract*— We investigate optimal encoding and retrieval of digital data, when the storage/communication medium is described by quantum mechanics. We assume an m-ary alphabet with arbitrary prior distribution, and an n-dimensional quantum system. Under these constraints, we seek an encoding-retrieval setup, comprised of code-states and a quantum measurement, which maximizes the probability of correct detection. In our development, we consider two cases. In the first, the measurement is predefined and we seek the optimal code-states. In the second, optimization is performed on both the code-states and the measurement.**

We show that one cannot outperform 'pseudo-classical transmission', in which we transmit n symbols with orthogonal code-states, and discard the remaining symbols. However, such pseudo-classical transmission is not the only optimum. We fully characterize the collection of optimal setups, and briefly discuss the links between our findings and applications such as quantum key distribution and quantum computing. We conclude with a number of results concerning the design under an alternative optimality criterion, the worst-case posterior probability, which serves as a measure of the retrieval reliability.

***Index Terms*— transmitter design, quantum detection, quantum key distribution, semidefinite programming, bilinear matrix inequality.**

I. Introduction

UNDERLYING any scheme for the storage or transmission of information is a physical medium. The encoding and the retrieval of information must therefore involve considerations as to the nature of the medium, with regard to possible corruption of the retrieved data, due to interaction with the environment or to physical limitations of the medium itself. Examples of media and information encoding range from letters printed in ink on paper, through electric charge stored in a capacitor, to photons travelling through an optical fiber. This work is concerned with the encoding of digital information in media, whose physics is described by the laws of quantum mechanics [1].

We concentrate on digital information with a finite alphabet, i.e. the data is one of m possible messages, each one associated with a prior probability p_i. Retrieval of the data is done by performing a measurement, thereby detecting the state of the system.

There are several common criteria for the assessment of information retrieval, which can, for the most part, be divided into two categories. The first is comprised of criteria whose motivation stems from information theory (e.g. mutual information [2]). The second type of criteria aim to measure the reliability of the "per symbol" retrieval, without taking into account any pre- or post-processing (channel coding). The criteria we address in this work are of the second kind.

The laws of quantum mechanics state that the outcome of a measurement, in our case an attempt to retrieve the encoded symbol, is random. Thus, a quantum encoding-retrieval setup is characterized by the transition probabilities

$$\Pr\{i|j\} \triangleq \Pr\{\text{out = symbol } i|\text{in = symbol } j\}.$$

This is reminiscent of the more common classical setups, but whereas the randomness there is induced by noise from the environment, in the quantum case, the randomness is inherent in the system itself.

The state of a quantum system is mathematically represented by a unit trace positive semidefinite operator ρ on an n-dimensional Hilbert space $\mathcal{H}$. Encoding digital information in a quantum system is done by preparing the system in one of m predefined states $\{\rho_i\}_{i=1}^m$, each associated with one of the possible messages. Retrieval is achieved by performing a measurement, and determining in which of these predetermined states the system has been prepared.

However, if a quantum system is in one of several states whose range spaces are not orthogonal, i.e. $\rho_i\rho_j \neq 0$, then no measurement permitted in quantum mechanics can determine without fail which of the states is present; there is a non-zero probability of detection error, i.e. $\Pr\{i|j\} > 0$ for $i \neq j$. The question is then, what valid quantum measurement would yield favorable detection performance.

A popular measure of performance, and the one which is the main interest of this work, is the probability of correct detection

$$P_d = \sum_{i=1}^{m} p_i \Pr\{i|i\}.$$

One of the contributions of this work is a complete characterization of the encoding-retrieval setups which maximize P_d under the constraints imposed by the postulates of quantum mechanics. We also present several new results concerning a different performance measure, the worst-case posterior probability, which is defined in Section VI.

The focus of this paper is the design of a complete digital communications channel (or memory unit), in which the designer can choose both the code-states ρ_i and the detection measurement. We assume that the nature of the data, which is designated by the number of possible symbols m and their prior probabilities p_i, is known. We also assume that the dimension n of the quantum system is given. The dimension of the quantum system determines the ability of the medium

Manuscript received January 66, 2666; revised November 66, 2666.

The authors are with the Technion - Israel Institute of Technology, Haifa 32000, Israel (e-mail: nelron@tx.technion.ac.il; yonina@ee.technion.ac.il).

to store data reliably, much like the signal to noise ratio in classical systems.

Thus, we seek the optimal setup, comprised of code-states and a measurement, that maximize P_d under a constraint of the dimension n of the system. We find the maximum attainable value of P_d for data with an arbitrary prior distribution, and completely characterize the optimal setups, which achieve this value of P_d.

When the number of symbols m is no larger than the dimension of the Hilbert space n, one can simply choose ρ_i as orthogonal pure states and attain perfect detection. When $m > n$ this is no longer possible, and quantum encoding becomes non-trivial.

Motivation for using many symbols in a quantum system of low dimension may stem from benefits, which a protocol provides, for which one is willing to sacrifice the probability of detection or the information rate. For instance, in protocols of quantum key distribution [3] the use of many states enables the detection of eavesdropping on the communication. In Section V-C we elaborate on this point. Another possible scenario is a quantum computation, which has a finite number of possible outputs m, and where for reasons of implementation complexity one cannot create a system large enough (with enough qubits such that $m \leq n$).

The problem of distinguishing among a collection of *specified* quantum states, i.e. when the code-states ρ_i are a given, is regularly referred to as *quantum detection* or *quantum state discrimination*, and has been studied in detail. Necessary and sufficient conditions for an optimal measurement, which maximizes the probability of correct detection P_d, have been derived [4], [5], [6]. Explicit solutions to the problem are known in some particular cases [7], [8], [9], [10], [11], including ensembles obeying a large class of symmetries [12]. The optimal measurement can also be calculated numerically, to within arbitrary accuracy, and in polynomial complexity [6].

Several alternative approaches have also been investigated. These include optimization with regard to other performance criteria, such as mutual information [4] or the worst-case posterior probability [13]. Another approach is *unambiguous detection* [14], [15], [16] in which one allows for an inconclusive result but does not allow for error. More recently, interest has grown in detection in a noisy environment [17], [18], [19], and in situations where the states are only partially known [20] or the prior probabilities not specified [21].

In Section II, the problem is presented in more detail. Then, in Section III, we show that the optimal code-states for a predetermined measurement are states which lie in the eigenspaces of the measurement operators associated with the maximal eigenvalues. This result is of interest both in its own right, and as part of the design of complete optimal encoding-retrieval setups.

Sections IV and V are the heart of this work. In Section IV we show that when encoding digital information in a quantum system of dimension n, the maximum attainable probability of correct detection is achieved by simply discarding $m - n$ of the symbols and using an orthonormal set to encode the remaining n symbols with perfect reconstruction. We dub this method *pseudo-classical* transmission. This is, however, not the only possible encoding-retrieval setup which achieves the maximal value of P_d. In Section V we show that all setups that attain the maximum are composed of pure code-states and of rank-1 measurement operators, and fully characterize the collection of optimal setups. The importance of finding all the optimal setups is discussed in Subsection V-C, where we outline possible use of our results in the analysis of quantum communication and computation protocols.

In Section VI we explore performance in relation to a measure of the reliability of the outcome. We introduce the *worst-case posterior probability*, denoted P_p. Again, when $m > n$ and perfect communication is impossible, the output can never be fully reliable. We provide a simple method for finding an upper bound on P_p for arbitrary states ρ_i and prior probabilities p_i.

Regrettably, for a large family of encoding-retrieval setups, P_p is ill-defined. For this reason, we also define a variation on P_p that we name the *effective* worst-case posterior probability. We investigate how one should choose the code-states, which represent discarded symbols in pseudo-classical transmission, in order to increase the reliability of the output, while still attaining maximal P_d. We develop an upper bound on P_p^{eff}, and present a choice which attains it.

II. Problem Formulation

A. Notation

According to the postulates of quantum mechanics [1], a physical system is mathematically represented by an n-dimensional complex Hilbert space $\mathcal{H}$. The state of the system ρ is represented by a positive semidefinite (PSD) Hermitian operator on $\mathcal{H}$, such that $\mathrm{Tr}(\rho) = 1$. Throughout, we shall use the notation $A \geq 0$ to indicate that an operator A is PSD, and the notation $A \geq B$ to imply that $A - B$ is PSD. If $\mathrm{rank}(\rho) = 1$, then it is known as a pure state.

As is customary in work relating to quantum theory, we shall use Dirac's notation of linear algebra, wherein a vector is denoted by $|u\rangle$, its Hermitian conjugate by $\langle u|$, and inner and outer products are signified by $\langle u|v\rangle$ and $|u\rangle\langle v|$ respectively. We do not assume that $|u\rangle$ is normalized. We denote by $\mathcal{R}(A)$ the range space of a Hermitian operator A, and by $\mathcal{M}(A)$ the eigenspace of its maximal eigenvalue.

B. Encoding Data in Quantum Media

We wish to encode digital information in a quantum medium. The information is represented by an m-ary alphabet, where each symbol has a prior probability p_i. Without loss of generality, we assume that the prior distribution obeys $p_1 \geq p_2 \geq \cdots \geq p_m > 0$. The encoding is achieved by associating with each symbol a predefined quantum state ρ_i, and preparing the system in the appropriate state. We shall refer to the states ρ_i as *code-states*. To a set of code-states $\{\rho_i\}_{i=1}^m$ we refer as an *ensemble*. Whenever an ensemble is arbitrary, we assume that it spans[1] $\mathcal{H}$.

[1] If it does not span $\mathcal{H}$, the problem can always be projected onto the subspace which it spans.

Retrieval of the information is accomplished by using a positive operator valued measurement (POVM), which is a set of m operators $\Pi = \{\Pi_i\}_{i=1}^m$, which satisfy

$$\Pi_i \geq 0, \qquad 1 \leq i \leq m$$
$$\sum_{i=1}^m \Pi_i = I.$$

This is the most general type of measurement allowed by the laws of quantum physics.

The measurement results in one of m possible outcomes, where, given that the state of the system is ρ, the probability of the i-th outcome is

$$\Pr\{i\} = \mathrm{Tr}(\Pi_i \rho).$$

Thus, the probability of correctly detecting the encoded message is

$$P_d = P_d(\Pi_i, \rho_i) = \sum_{i=1}^m p_i \mathrm{Tr}(\Pi_i \rho_i).$$

In this work we use P_d as the main criterion for measuring the quality of an encoding-retrieval setup.

In the next section we find the optimal code-states, in the sense of maximal P_d, for a given measurement. We then characterize, in Sections IV and V, all optimal encoding-retrieval setups, when the design specifications are the nature of the data (the prior probabilities p_i), and the dimension n of the quantum system.

In Section VI we develop several results concerning an alternative measure of performance, the *worst-case posterior probability*. This criterion is an indicator of the reliability of the output, and is defined at the beginning of Section VI.

III. Designing Code-States for an Arbitrary Measurement

In this section we answer the following question. If the detector, i.e. the measurement Π, and the prior probabilities of the data p_i are predetermined, what would be a good choice of code-states ρ_i to encode the data in a quantum medium of dimension n, in terms of P_d? This question is of interest, due to possible implementation restrictions on the detector. As indicated in the introduction, the reverse situation, that of designing a measurement to discriminate among arbitrary states, has been thoroughly studied.

Our result is stated formally in Theorem 1.

Theorem 1: Let $\{p_i\}_{i=1}^m$ be a probability distribution, and let $\{\Pi_i\}_{i=1}^m$ be the measurement operators of a detector. An ensemble of quantum states $\{\rho_i\}_{i=1}^m$ maximizes P_d if and only if

$$\mathcal{R}(\rho_i) \subseteq \mathcal{M}(\Pi_i) \qquad i : \Pi_i \neq 0.$$

For all i such that $\Pi_i = 0$, any choice of ρ_i is optimal. Denoting the maximal eigenvalue of Π_i as $\sigma_{\Pi_i}^{\max}$, the maximal probability of correct detection is given by

$$P_d^{\mathrm{opt}} = \sum_{i=1}^m p_i \sigma_{\Pi_i}^{\max}.$$

Proof: The optimal states $\hat{\rho}_i$ are a solution to

$$\max_{\rho_i} \sum_{i=1}^m p_i \mathrm{Tr}(\Pi_i \rho_i) \quad \text{s.t.} \begin{cases} \rho_i \geq 0, \\ \mathrm{Tr}(\rho_i) = 1. \end{cases} \tag{1}$$

The objective function in (1) is additive in the variables ρ_i, and the constraints on each of the ρ_i are independent. Hence, (1) is separable in i, i.e. the states $\hat{\rho}_i$ are optimal if and only if they are also the solutions to m problems of the form (one for each i)

$$\max_{\rho} \mathrm{Tr}(\Pi \rho) \quad \text{s.t.} \begin{cases} \rho \geq 0, \\ \mathrm{Tr}(\rho) = 1. \end{cases} \tag{2}$$

Any quantum state ρ, such that $\rho \geq 0$ and $\mathrm{Tr}(\rho) = 1$, has an eigendecomposition of the form

$$\rho = \sum_{j=1}^n g_j |u_j\rangle\langle u_j|,$$

where $g_j \geq 0$, $\sum_{j=1}^n g_j = 1$, and $\langle u_j | u_j \rangle = 1$. Since $\Pi \geq 0$, we have that

$$\begin{aligned} \mathrm{Tr}(\Pi\rho) &= \sum_{j=1}^n g_j \langle u_j | \Pi | u_j \rangle \\ &\leq \langle \hat{u} | \Pi | \hat{u} \rangle \sum_{j=1}^n g_j \\ &= \langle \hat{u} | \Pi | \hat{u} \rangle \\ &\leq \sigma_\Pi^{\max}, \end{aligned}$$

where $\langle \hat{u} | \Pi | \hat{u} \rangle = \max_j \langle u_j | \Pi | u_j \rangle$, and $\sigma_\Pi^{\max}$ is the largest eigenvalue of Π. If $\Pi = 0$ then the upper bound is zero and any $\rho \geq 0$ is optimal. When $\Pi \neq 0$, equality is achieved if $\mathrm{Tr}(\Pi\rho) = \sigma_\Pi^{\max}$, i.e. *only* when ρ lies in the eigenspace corresponding to $\sigma_\Pi^{\max}$. ∎

Note that the optimal code-states $\hat{\rho}_i$ are independent of each other and of the prior probabilities p_i. Also note that the optima (the solutions of the problem (1)) form a convex set.

Corollary 1.1: If for all i, $\dim \mathcal{M}(\Pi_i) = 1$, then the ensemble which maximizes P_d is unique.

Proof: When $\dim \mathcal{M}(\Pi_i) = 1$ then ρ_i must be the pure state which spans $\mathcal{M}(\Pi_i)$, and which is unique (due to the requirement of normalization). If this is true for all i, then the entire set of code-states is unique. ∎

In applications, one may have the freedom to choose which symbol will be detected by which of the detection operators. Recalling that we assumed the prior probabilities p_i to be sorted in descending order, maximal P_d can be attained when the detection operators are sorted such that $\sigma_{\Pi_1}^{\max} \geq \sigma_{\Pi_2}^{\max} \geq \cdots \geq \sigma_{\Pi_m}^{\max}$. Doing this, and selecting the optimal code-states as above, would lead to the maximal value of $P_d = \sum_i p_i \sigma_{\Pi_i}^{\max}$.

IV. Optimal Quantum Encoding

We now find the maximal attainable value of P_d when encoding data in a quantum medium. We assume that the nature of the data itself, which is manifested in the prior probabilities p_i, is predetermined, and so is the quantum system itself (i.e. the dimension n). We aim to find an encoding-retrieval setup that maximizes P_d.

Thus, our goal is to find the solutions to

$$\max_{\Pi_i,\rho_i} \sum_{i=1}^{m} p_i \mathrm{Tr}(\Pi_i \rho_i) \tag{3}$$
$$\text{s.t.} \begin{cases} \rho_i \geq 0, & \mathrm{Tr}(\rho_i) = 1, \\ \Pi_i \geq 0, & \sum_{i=1}^{m} \Pi_i = I. \end{cases}$$

This optimization problem is of a class known as *Bilinear Matrix Inequality* (BMI) optimization problems [22]. BMIs are non-convex, and in general, finding a global optimum is an NP-hard problem [23]. Nonetheless, for this particular BMI (3), we are able to formulate a closed form solution, and to completely specify the optimal set.

When the dimension n of the quantum system is equal to the number of possible messages m, then perfect retrieval ($P_d = 1$) is achievable by choosing the code-states ρ_i to be mutually orthogonal pure states, and the measurement such that $\Pi_i = \rho_i$. When $n < m$ this is no longer possible. The most straightforward approach to quantum encoding when $n < m$ is to simply disregard $m - n$ of the messages and aim to perfectly retrieve the remaining n messages. It is clear that the smallest probability of error would occur if the disregarded messages were the ones with smallest prior probabilities. Thus, this approach is embodied in the ensemble-detector setup

$$\Pi_i = \begin{cases} |u_i\rangle\langle u_i| & 1 \leq i \leq n \\ 0 & n < i \leq m \end{cases}$$
$$\rho_i = \begin{cases} |u_i\rangle\langle u_i| & 1 \leq i \leq n \\ \text{Don't care} & n < i \leq m \end{cases} \tag{4}$$

where $\{|u_i\rangle\}_{i=1}^{n}$ is some orthonormal system. When using this setup $P_d = \sum_{i=1}^{n} p_i$.

The distinction between classical and quantum systems is very strongly linked to the fact that non-orthogonality between two quantum states affects the ability to distinguish between them. There is no classical analogue of this property. When the states that a quantum system may be in are mutually orthogonal, it is said to be in "the classical limit". The fact that the setup (4) is comprised only of pure mutually orthogonal states implies that it is classical in nature and that the losses encountered are not due to the fact that the system is governed by quantum mechanics, but to a lossy preprocessing (disregarding some of the messages). In the sequel we refer to (4) as *pseudo-classical* transmission.

It would, at first glance, seem that one may somehow be able to utilize the "quantumness" of the system, i.e. non-orthogonal code-states and measurements, in order to improve on the probability of correct detection P_d. We now formulate and prove a theorem which shows this to be impossible.

Theorem 2: Let $\{p_i\}_{i=1}^{m}$ be a probability distribution with $p_1 \geq p_2 \geq \cdots \geq p_m > 0$. Denoting by $\hat{P}_d$ the maximal probability of correct detection for a quantum system of dimension $n \leq m$, we have that

$$\hat{P}_d = \sum_{i=1}^{n} p_i.$$

Proof: Let $\tilde{P}_d = \sum_{i=1}^{n} p_i$. Since the pseudo-classical setup (4) achieves $P_d(\Pi_i, \rho_i) = \tilde{P}_d$, we have that $\hat{P}_d \geq \tilde{P}_d$. We prove the theorem by showing that $\hat{P}_d \leq \tilde{P}_d$.

The maximal value of P_d is the solution of (3). From Theorem 1, after maximizing with respect to ρ_i, (3) reduces to

$$\max_{\Pi_i} \sum_{i=1}^{m} p_i \sigma_{\Pi_i}^{\max} \tag{5}$$
$$\text{s.t.} \begin{cases} \Pi_i \geq 0, & \text{(a)} \\ \sum_{i=1}^{m} \Pi_i = I. & \text{(b)} \end{cases}$$

The constraint (5a) implies that

$$\sigma_{\Pi_i}^{\max} \geq 0, \qquad 1 \leq i \leq m, \tag{6}$$

and from (5b)

$$\sigma_{\Pi_i}^{\max} \leq 1, \qquad 1 \leq i \leq m,$$
$$\sum_{i=1}^{m} \sigma_{\Pi_i}^{\max} \leq n. \tag{7}$$

(The bottom expression in (7) is obtained by taking the trace of (5b)). We now replace (5) by a scalar program,

$$\max_{\sigma_i} \sum_{i=1}^{m} p_i \sigma_i \tag{8}$$
$$\text{s.t.} \begin{cases} 0 \leq \sigma_i \leq 1, \\ \sum_{i=1}^{m} \sigma_i \leq n. \end{cases}$$

Problem (8) was created by relaxing the constraints of problem (5) - we keep only the constraints on the eigenvalues and disregard the original matrix-inequality constraints. Therefore, the solution of (8) is always larger or equal to the solution of (5), and thus, serves as an upper bound.

The optimization problem (8) is a linear programme. Its *Lagrange dual problem* [24] is given by

$$\min_{\eta_i,\nu_i,\mu} g(\eta_i, \mu) \tag{9}$$
$$\text{s.t.} \begin{cases} \eta_i, \nu_i, \mu \geq 0, & \text{(a)} \\ p_i - \eta_i + \nu_i - \mu = 0, & \text{(b)} \end{cases}$$

where $1 \leq i \leq m$ and

$$g(\eta_i, \mu) = \sum_{i=1}^{m} \eta_i + n\mu$$

Using the constraint (9b), the variables ν_i can be eliminated, yielding

$$\min_{\eta_i,\mu} g(\eta_i,\mu) \tag{10}$$
$$\text{s.t.} \begin{cases} \eta_i, \mu \geq 0, \\ \eta_i + \mu \geq p_i. \end{cases}$$

From Lagrange duality theory, for any point in the feasibility set of (10), the objective $g(\eta_i,\mu)$ is greater or equal to the solution of the primal problem (8). In other words, for any dual feasible point (η,μ), $g(\eta_i,\mu)$ is an upper bound on the solution of (5). Consider

$$\hat{\eta}_i = \begin{cases} p_i - p_{n+1} & 1 \leq i \leq n \\ 0 & n < i \leq m \end{cases}$$
$$\hat{\mu} = p_{n+1}. \tag{11}$$

Because $p_1 \geq \cdots \geq p_m$ it is dual feasible. For this choice

$$g(\hat{\eta}_i,\hat{\mu}) = \sum_{i=1}^{m} \hat{\eta}_i + n\hat{\mu} = \sum_{i=1}^{n} (\hat{\eta}_i + \hat{\mu}) = \sum_{i=1}^{n} p_i.$$

In conclusion, we have shown that

$$\max(3) = \max(5) \leq \max(8) \leq \min(10) \leq \sum_{i=1}^{n} p_i,$$

which implies that for any valid ensemble and detector $P_d = \sum_{i=1}^m p_i \mathrm{Tr}(\Pi_i\rho_i) \leq \tilde{P}_d$. ■

The implication of Theorem 2 is that one can achieve the optimal probability of correct detection by using orthogonal pure states and von Neumann measurements, which are easy to implement. Nevertheless, there may be setups $\{\rho_i, \Pi_i\}_{i=1}^m$ other then (4) which attain $P_d(\Pi_i,\rho_i) = \hat{P}_d$. In the next section we identify all the ensemble-detector setups which achieve maximum probability of correct detection. The importance of characterizing the set of optima is that we may be able to select an optimum that has preferable performance with regard to other quality of service measures. Also, there may be communication protocols which require using a "non-classical" ensemble. These aspects are discussed in greater detail in Section V-C.

V. Characterization of Optimal Setups

In this section we introduce the notion of *tight frame encoding setups*, and show that all optima are of this form (Theorem 3). We then fully characterize the set of optima for a given prior probability distribution (Theorem 4 and corollaries).

A. Tight Frame Encoding Setups

A *tight frame* [25] is a set of m vectors $\{|u_i\rangle\}_{i=1}^m$ which satisfy

$$\sum_{i=1}^{m} |u_i\rangle\langle u_i| = I. \tag{12}$$

We define a "Tight Frame Encoding Setup" (TFES) to be an ensemble-detector setup of the form

$$\Pi_i = |u_i\rangle\langle u_i|,$$
$$\rho_i = \begin{cases} \frac{1}{\langle u_i|u_i\rangle} |u_i\rangle\langle u_i| & \langle u_i|u_i\rangle > 0 \\ \text{Don't care,} & \langle u_i|u_i\rangle = 0 \end{cases}$$

where the vectors $|u_i\rangle$ obey (12). The pseudo-classical setup (4) is an example of a TFES. The probability of correct detection when using a TFES is $P_d = \sum_{i=1}^m p_i \langle u_i|u_i\rangle$.

The constraint (12) on the vectors ensures that Π is a valid POVM. It also implies several properties of the vectors $|u_i\rangle$, which are summarized in the following lemma.

Lemma 1: Let $\{|u_i\rangle\}_{i=1}^m$ be a set of vectors which satisfy (12). Then,

$$\langle u_i|u_i\rangle \leq 1, \tag{13}$$

$$\text{if } \langle u_i|u_i\rangle = 1 \text{ then } \langle u_i|u_j\rangle = \delta_{i,j}, \tag{14}$$

$$\sum_{i=1}^{m} \langle u_i|u_i\rangle = n. \tag{15}$$

Proof: See Appendix A. ■

Tight frames are of interest in many fields and applications where one seeks a set of vectors whose mutual "interference" is minimal. Specifically, in classical communication, they play an important role in Syncronous CDMA systems [26], [27]. Also, the simplex constellation, which is known to be optimal under certain energy constraints [28], [29], is a tight frame.

The significance of TFESs to quantum encoding is established by the following result:

Theorem 3: All ensemble-detector setups $\{\hat{\rho}_i, \hat{\Pi}_i\}_{i=1}^m$ which achieve $P_d(\hat{\Pi}_i,\hat{\rho}_i) = \hat{P}_d$ are TFESs.

The proof of Theorem 3, relies on the following lemma, whose proof is given in Appendix B.

Lemma 2: For any ensemble-detector setup $(\hat{\rho}_i, \hat{\Pi}_i)$ which achieves $P_d(\hat{\Pi}_i,\hat{\rho}_i) = \hat{P}_d$, the largest eigenvalues of the detection operators satisfy

$$\sum_{i=1}^{m} \sigma_{\hat{\Pi}_i}^{\max} = n.$$

Proof: (of Thm. 3) For a given ensemble $\tilde{\rho} = \{\tilde{\rho}_i\}$, denote the set of all detectors $\Pi = \{\Pi_i\}$ which are optimal for its detection

$$S_{\tilde{\rho}} \triangleq \left\{ \Pi \,\middle|\, P_d(\Pi,\tilde{\rho}) = \max_{\Psi \text{ is POVM}} P_d(\Psi,\tilde{\rho}) \right\}.$$

In [6] it is shown that if $\Pi \in S_{\tilde{\rho}}$ then

$$\mathrm{rank}(\Pi_i) \leq \mathrm{rank}(\tilde{\rho}_i).$$

For a given detector $\tilde{\Pi} = \{\tilde{\Pi}_i\}$, denote the set of all ensembles $\rho = \{\rho_i\}$ for which $\tilde{\Pi}$ is the optimal detector by

$$D_{\tilde{\Pi}} \triangleq \left\{ \rho \,\middle|\, P_d(\tilde{\Pi},\rho) = \max_{\varrho \text{ are QS}} P_d(\tilde{\Pi},\varrho) \right\}.$$

By the expression "ϱ are QS" we imply that for each i, ϱ_i is a valid quantum state. From Theorem 1, any $\rho \in D_{\tilde{\Pi}}$ must satisfy[2]

$$\mathcal{R}(\rho_i) \subseteq \mathcal{M}(\tilde{\Pi}_i) \subseteq \mathcal{R}(\tilde{\Pi}_i), \qquad i : \tilde{\Pi}_i \neq 0 \tag{16}$$

implying that

$$\text{rank}(\rho_i) \leq \text{rank}(\tilde{\Pi}_i). \qquad i : \tilde{\Pi}_i \neq 0$$

Assume that an ensemble-detector setup $\{\hat{\rho}_i, \hat{\Pi}_i\}$ achieves $\hat{P}_d$. Then

$$P_d(\hat{\Pi}, \hat{\rho}) = \max_{\Psi \text{ is POVM}} P_d(\Psi, \hat{\rho}),$$

implying that $\hat{\Pi} \in S_{\hat{\rho}}$ and that therefore

$$\text{rank}(\hat{\Pi}_i) \leq \text{rank}(\hat{\rho}_i). \tag{17}$$

In addition

$$P_d(\hat{\Pi}, \hat{\rho}) = \max_{\varrho \text{ are QS}} P_d(\hat{\Pi}, \varrho),$$

indicating, in turn, that $\hat{\rho} \in D_{\hat{\Pi}}$ and that

$$\text{rank}(\hat{\rho}_i) \leq \text{rank}(\hat{\Pi}_i), \qquad \hat{\Pi}_i \neq 0. \tag{18}$$

From (17) and (18), we get

$$\text{rank}(\hat{\rho}_i) = \text{rank}(\hat{\Pi}_i), \qquad \hat{\Pi}_i \neq 0.$$

Together with (16), this indicates that

$$\mathcal{R}(\hat{\rho}_i) = \mathcal{R}(\hat{\Pi}_i), \qquad \hat{\Pi}_i \neq 0. \tag{19}$$

Thus, for each i, all the non-zero eigenvalues of $\hat{\Pi}_i$ are equal.

Denoting $r_i = \text{rank}(\hat{\Pi}_i)$, we now have that for all i

$$\text{Tr}(\Pi_i) = r_i \sigma^{\max}_{\hat{\Pi}_i}.$$

Taking the trace of the requirement $\sum_{i=1}^m \hat{\Pi}_i = I$, one gets

$$\sum_{i=1}^m r_i \sigma^{\max}_{\hat{\Pi}_i} = \sum_{i=1}^m \sigma^{\max}_{\hat{\Pi}_i} + \sum_{i=1}^m (r_i - 1)\sigma^{\max}_{\hat{\Pi}_i} = n.$$

If there is some i for which both $r_i > 1$ and $\sigma^{\max}_{\hat{\Pi}_i} > 0$, then

$$\sum_{i=1}^m \sigma^{\max}_{\hat{\Pi}_i} < n.$$

Since we have assumed that $\hat{\Pi}$ is part of an optimal setup, this is a contradiction with Lemma 2. Thus, for all i such that $\hat{\Pi}_i \neq 0$, $\text{rank}(\hat{\Pi}_i) = 1$. In conjunction with (19), this proves the theorem. ■

An interesting aspect of the above result is that $\hat{P}_d$ can only be attained by setups in which the detected code-states (those for which the corresponding measurement operator is not zero) are pure states. This is hardly surprising, since obviously, for mixed states the chances of "interference" between code-states is greater.

[2] If A is PSD and $A \neq 0$ then $\mathcal{M}(A) \subseteq \mathcal{R}(A)$.

B. Choice of TFES

From Theorem 3 we know that all optima are TFESs. Not all choices of TFES are, however, necessarily optimal. We now show that the set of optimal TFESs is dependant on the prior probabilities $\{p_i\}_{i=1}^m$, and on the dimension of the quantum medium n, and characterize this dependance. The following results (Theorem 4 and corollaries) fully characterize all optimal solutions for a given prior distribution and dimension n.

In order to formulate our results we introduce a classification of the symbols into three distinct subsets, according to the prior probability distribution and the dimension n. Recalling that we assume $p_1 \geq p_2 \geq \cdots \geq p_m$, we define

- $\mathcal{I}_1 = \{i \,|\, p_i > p_n\}$,
- $\mathcal{I}_2 = \{i \,|\, p_i = p_n\}$,
- $\mathcal{I}_3 = \{i \,|\, p_i < p_n\}$.

Note that $\mathcal{I}_1$ and $\mathcal{I}_3$ may be empty.

Theorem 4: Let $\{p_i\}_{i=1}^m$ be a non-increasing distribution of probabilities, and let $\{|u_i\rangle\}_{i=1}^m$ be the vectors of a TFES in a quantum system of dimension n. This TFES is optimal in the sense of probability of correct detection if and only if (i) for all $i \in \mathcal{I}_1$, $\langle u_i|u_i\rangle = 1$, and (ii) for all $i \in \mathcal{I}_3$, $\langle u_i|u_i\rangle = 0$.

Before proving Theorem 4, we point out the following important corollaries:

Corollary 4.1: Let $\{p_i\}_{i=1}^m$ be a non-increasing distribution of probabilities, and let $\{\hat{\rho}_i, \hat{\Pi}_i\}_{i=1}^m$ be an *optimal* encoding setup in a quantum system of dimension n. Then

1) $\Pr\{j|i\} = \delta_{i,j} \qquad i \in \mathcal{I}_1$,
2) $\Pr\{\text{det } i\} = 0 \qquad i \in \mathcal{I}_3$.

Proof: From Theorem 3, we know that the ensemble-detector setup is a TFES. From Theorem 4, for all $i \in \mathcal{I}_1$, $\hat{\Pi}_i = |u_i\rangle\langle u_i|$, such that $\langle u_i|u_i\rangle = 1$. Together with (14), it is easy to see that

$$\Pr\{j|i\} = \text{Tr}(\hat{\Pi}_j \hat{\rho}_i) = \frac{1}{\langle u_i|u_i\rangle} |\langle u_i|u_j\rangle|^2 = \delta_{i,j}.$$

Also from Theorem 4, for all $i \in \mathcal{I}_3$, $\hat{\Pi}_i = 0$, indicating that the probability of detecting the i-th message is $\Pr\{\text{det } i\} = \sum_j p_j \text{Tr}(\hat{\Pi}_i \hat{\rho}_j) = 0$. ■

Corollary 4.2: Let $\{p_i\}_{i=1}^m$ be a non-increasing distribution of probabilities. If $p_n > p_{n+1}$ then any optimal setup $\{\hat{\rho}_i, \hat{\Pi}_i\}_{i=1}^m$ must be of the form (4) (pseudo-classical).

Proof: From Theorem 3, the optimal setup must be a TFES. When $p_n > p_{n+1}$ we have $\mathcal{I}_3 = \{n+1, \ldots, m\}$, which, using Theorem 4, indicates that

$$\langle u_i|u_i\rangle = 0 \qquad n+1 \leq i \leq m.$$

Together with (12) this implies

$$\sum_{i=1}^n |u_i\rangle\langle u_i| = I. \tag{20}$$

A set of n vectors in n-dimensional space can satisfy (20) if and only if they form an orthonormal set. Thus, the only optimal setup when $p_n > p_{n+1}$ is (4). ■

Corollary 4.3: If $p_i = \frac{1}{m}$ for all i, then all TFESs achieve $\hat{P}_d$.

Proof: The Corollary follows directly from Theorem 4, for $\mathcal{I}_1 = \mathcal{I}_2 = \emptyset$. ■

We now prove Theorem 4.

Proof: (of Thm. 4) Assume that $\{|u_i\rangle\}_{i=1}^m$ are the vectors of a TFES which is optimal in the sense of P_d.

Assume that $\mathcal{I}_1 \neq \emptyset$ and denote by k the largest index in $\mathcal{I}_1$. (i.e. $\mathcal{I}_1 = \{1,\ldots,k\}$). This means that $p_k > p_{k+1} = p_{k+2} = \cdots = p_n$ (from the definition of $\mathcal{I}_1$, we have that $k < n$). For any TFES we can write

$$\begin{aligned} P_d &= \sum_{i=1}^m p_i\langle u_i|u_i\rangle \\ &= \sum_{i=1}^k p_i\langle u_i|u_i\rangle + \sum_{i=k+1}^m p_i\langle u_i|u_i\rangle \\ &\leq \sum_{i=1}^k p_i\langle u_i|u_i\rangle + p_{k+1}\sum_{i=k+1}^m \langle u_i|u_i\rangle && (21) \\ &= \sum_{i=1}^k p_i\langle u_i|u_i\rangle + p_{k+1}\left(n - \sum_{i=1}^k\langle u_i|u_i\rangle\right) && (22) \\ &= \sum_{i=1}^k \left[p_i\langle u_i|u_i\rangle + p_{k+1}(1-\langle u_i|u_i\rangle)\right] + (n-k)p_{k+1}, && (23) \end{aligned}$$

where the transition from (21) to (22) relies on (15).

Recall that for all $i \in \mathcal{I}_1$ we have $p_i > p_{k+1}$. If for some $1 \leq i \leq k$, $\langle u_i|u_i\rangle < 1$, then from (23)

$$\begin{aligned} P_d &< \sum_{i=1}^k \left[p_i\langle u_i|u_i\rangle + p_i(1-\langle u_i|u_i\rangle)\right] + (n-k)p_{k+1} \\ &= \sum_{i=1}^k p_i + (n-k)p_{k+1} = \sum_{i=1}^n p_i = \hat{P}_d. \end{aligned}$$

Here we have relied on the fact that $p_{k+1} = p_{k+2} = \cdots = p_n$. Therefore, in order to achieve $\hat{P}_d$, the vectors $|u_i\rangle$ must satisfy

$$\langle u_i|u_i\rangle = 1, \qquad i \in \mathcal{I}_1$$

This concludes the proof of the first statement of the 'only if' direction.

We go on to prove the second statement. Assume that $k' = \max \mathcal{I}_2 < m$ (i.e. $\mathcal{I}_3 = \{k'+1,\ldots,m\} \neq \emptyset$). By definition $p_{k'} > p_{k'+1}$. We again have

$$P_d = \sum_{i=1}^m p_i\langle u_i|u_i\rangle = \sum_{i=1}^{k'} p_i\langle u_i|u_i\rangle + \sum_{i=k'+1}^m p_i\langle u_i|u_i\rangle.$$

If for some $i \in \mathcal{I}_3$, $\langle u_i|u_i\rangle > 0$, then

$$\begin{aligned} P_d &< \sum_{i=1}^{k'} p_i\langle u_i|u_i\rangle + p_{k'}\sum_{i=k'+1}^m \langle u_i|u_i\rangle && (24) \\ &= \sum_{i=1}^n p_i\langle u_i|u_i\rangle + p_n\sum_{i=n+1}^m \langle u_i|u_i\rangle && (25) \\ &= \sum_{i=1}^n p_i\langle u_i|u_i\rangle + p_n\left(n - \sum_{i=1}^n\langle u_i|u_i\rangle\right) && (26) \\ &= \sum_{i=1}^n p_i\langle u_i|u_i\rangle + p_n\sum_{i=1}^n\big(1-\langle u_i|u_i\rangle\big) \\ &\leq \sum_{i=1}^n p_i\langle u_i|u_i\rangle + \sum_{i=1}^n p_i\big(1-\langle u_i|u_i\rangle\big) \\ &= \sum_{i=1}^n p_i = \hat{P}_d, \end{aligned}$$

where the transitions from (24) to (26) rely on the fact that $k' \in \mathcal{I}_2$ and on (15). Thus, for any TFES which achieves maximal P_d

$$\langle u_i|u_i\rangle = 0, \qquad i \in \mathcal{I}_3.$$

We continue by proving the 'if' direction. Assume that for all $i \in \mathcal{I}_1$, $\langle u_i|u_i\rangle = 1$, and that for all $i \in \mathcal{I}_3$, $\langle u_i|u_i\rangle = 0$. We must first note that under these conditions, using (15) yields

$$\sum_{i=1}^{k'}\langle u_i|u_i\rangle = n. \qquad (27)$$

If $\mathcal{I}_1 = \emptyset$, then $\mathcal{I}_2 = \{1,\ldots,k'\}$. We can then write

$$\begin{aligned} P_d &= \sum_{i=1}^m p_i\langle u_i|u_i\rangle && (28) \\ &= p_1\sum_{i=1}^{k'}\langle u_i|u_i\rangle && (29) \\ &= np_1 = \sum_{i=1}^n p_i = \hat{P}_d, && (30) \end{aligned}$$

where the transition from (28) to (29) relies on the facts that for all $i > k'$, $\langle u_i|u_i\rangle = 0$, and $p_1 = p_2 = \cdots = p_n$. The transition from (29) to (30) is based on (27).

If $\mathcal{I}_1 = \{1,\ldots,k\}$ and $\mathcal{I}_2 = \{k+1,\ldots,k'\}$ then, similarly,

$$\begin{aligned} P_d &= \sum_{i=1}^m p_i\langle u_i|u_i\rangle \\ &= \sum_{i=1}^k p_i\langle u_i|u_i\rangle + p_{k+1}\sum_{i=k+1}^{k'}\langle u_i|u_i\rangle \\ &= \sum_{i=1}^k p_i + (n-k)p_{k+1} = \sum_{i=1}^n p_i = \hat{P}_d, \end{aligned}$$

thereby completing the proof. ■

Theorem 5 below summarizes the assertions of Theorems 2, 3 and 4, in concise form, and completely characterizes all optimal transmitter-receiver setups.

Theorem 5: Let $\{p_i\}_{i=1}^m$ be a probability distribution with $p_1 \geq p_2 \geq \cdots \geq p_m > 0$. For a given number $n \leq m$, define the index sets

- $\mathcal{I}_1 = \{i \,|\, p_i > p_n\}$,
- $\mathcal{I}_2 = \{i \,|\, p_i = p_n\}$,
- $\mathcal{I}_3 = \{i \,|\, p_i < p_n\}$.

The maximal probability of correct detection for a quantum system of dimension $n \leq m$ is

$$\hat{P}_d = \sum_{i=1}^{n} p_i.$$

The optimum is achieved if and only if the ensemble-detector setup is of the form

$$\Pi_i = |u_i\rangle\langle u_i|,$$
$$\rho_i = \begin{cases} \frac{1}{\langle u_i|u_i\rangle}|u_i\rangle\langle u_i| & \langle u_i|u_i\rangle > 0 \\ \text{Don't care}, & \langle u_i|u_i\rangle = 0 \end{cases}$$

where the vectors $\{|u_i\rangle\}_{i=1}^m$ obey

$$\sum_{i=1}^{m} |u_i\rangle\langle u_i| = I,$$
$$\langle u_i|u_i\rangle = 1, \qquad i \in \mathcal{I}_1$$
$$\langle u_i|u_i\rangle = 0. \qquad i \in \mathcal{I}_3$$

C. Application to the Analysis of Communication Protocols

In many applications, additional constraints, other then the ones imposed by the physics, are placed on the encoding-retrieval setup. In quantum key distribution [3], for example, constraints arise due to the need for security against eavesdropping. Further constraints may occur due to technical (implementation) issues. The work at hand can then serve for two purposes. The first is to quantify the degradation in P_d due to the need to meet the extra design constraints. This can be done by simply comparing the performance of the constrained system to the theoretical upper bound $\hat{P}_d$. The second possible use of this work, in this context, is to search within the set of optimal TFESs for a setup, which is close to meeting the demands posed by the application. When taking the latter approach we are assured optimal performance with regard to P_d.

Consider the BB84 protocol [30]. In this QKD protocol, Alice wishes to send Bob secure binary information. In order to counter possible eavesdropping, she sends one of $m = 4$ messages with $p_i = \frac{1}{4}$ over a 2-dimensional quantum channel. The code-states used are denoted $|u_{ij}\rangle$, where $i, j = 0, 1$, and they obey the relations

$$|\langle u_{ij}|u_{i'j'}\rangle|^2 = \begin{cases} \delta_{j,j'} & i = i' \\ 1/2 & i \neq i' \end{cases}$$
$$\frac{1}{2}\sum_{i,j=0}^{1} |u_{ij}\rangle\langle u_{ij}| = I \tag{31}$$

Note that (31) indicates that this collection of vectors is a tight frame.

Bob utilizes the POVM (of order 4) $\Pi_{ij} = \frac{1}{2}|u_{ij}\rangle\langle u_{ij}|$, in order to retrieve Alice's message. They then exchange knowledge on which "pair of states" was received (by, for example, comparing the i index). If both the sent and the detected symbols originate from the same pair, then the transferred bit of information is taken as the member of the pair that was detected (the j index). If the symbols originate from different pairs, the received symbol is discarded. In order to promote security, Alice and Bob use $m > n$, at a cost of reduced data rate. The security of this protocol has been extensively studied.

The probability of correct detection achieved by Bob prior to the exchange of the i index is $P_d = 1/2$. This is equal to the upper bound $\hat{P}_d$ for this case, meaning that under the requirement of countering eavesdropping, Bob achieves the maximal possible performance. The fact that the upper bound is reached would hardly surprise most readers, in the context of a protocol as simple as BB84. It does, however, serve to illustrate the possible use of the unconstrained upper bound $\hat{P}_d$ in quantifying the efficacy of more complex communication protocols.

VI. Optimal Worst-Case Posterior Probability

An alternative quality of service measure for systems of digital communication/storage is the *worst-case posterior probability* [31], [13]. The posterior probability, defined as

$$P_p(i) \triangleq \frac{\Pr\{\text{message } i \text{ detected correctly}\}}{\Pr\{\text{message } i \text{ detected}\}} = \frac{p_i \text{Tr}(\Pi_i \rho_i)}{\sum_j p_j \text{Tr}(\Pi_i \rho_j)}, \tag{32}$$

is the answer to the question: "Given that the detected message is i, what is the probability that it is the right answer?". The worst-case posterior probability is then

$$P_p \triangleq \min_{i=1,\ldots,m} P_p(i).$$

The higher the value of P_p, the more reliable the output of the measurement. Theorem 6 below provides a simple method for finding an upper bound on P_p for a given set of code-states ρ_i and prior probabilities p_i.

Note that for the pseudo-classical TFES (4), and in fact for any setup in which one of the POVM elements is zero, the posterior probability is ill-defined, since the denominator in (32) is zero. We therefore introduce a surrogate measure of the reliability of the outcome, designed to replace P_p in this case.

Since we seek a measure of reliability of the output, there is no point in taking into account outputs which never occur. Hence we choose to measure the most unreliable outcome, of the set of *possible* outcomes. The *effective worst-case posterior probability* is defined as

$$P_p^{\text{eff}} = \min_{i|\ \Pr\{\det\ i\}>0} P_p(i),$$

where $\Pr\{\det\ i\} = \sum_j p_j \text{Tr}(\Pi_i \rho_j)$ is the probability of detecting the i-th outcome. Whenever P_p is well defined, then $P_p^{\text{eff}} = P_p$.

In the second part of this section we consider ensemble-detector setups which attain optimal P_d. According to Theorem 5, in many instances, these setups are a TFES, whose detector has zero elements. For all i, such that $\Pi_i = 0$, we can choose the code-states freely, without degrading the performance in P_d. This raises the question, how should one choose the 'don't care' states, so that the output of the system would be reliable? We prove an upper bound on P_p^{eff} over the set of optimal TFESs (Theorem 7), and show that for the pseudo-classical TFES (4), there is a choice of 'don't care' states which attains this bound (Corollary 7.1).

A. An Upper Bound on P_p for a Given Ensemble

Theorem 6: Let $\{\rho_i\}_{i=1}^m$ be m arbitrary quantum states of dimension n, with prior probabilities p_i. Define the operators

$$A_i(\delta) = (1-\delta)\sum_{k=1}^{m} p_k\rho_k - p_i\rho_i,$$

where $\delta \in \mathcal{R}$. If for some $1 \le i \le m$, $A_i(\delta) \ge 0$, then $P_p \le 1-\delta$.

Proof: Showing that $P_p \le 1-\delta$, is equivalent to showing that $||e_{\text{post}}||_{\text{wc}} \ge \delta$, where

$$||e_{\text{post}}||_{\text{wc}} \triangleq \max_i \left(1 - \frac{p_i \text{Tr}(\Pi_i\rho_i)}{\sum_i p_i \text{Tr}(\Pi_i\rho_i)}\right).$$

We introduce $||e_{\text{post}}||_{\text{wc}}$, known as the *worst-case posterior error*, in order to adhere with the notation in reference [13].

Assume that $\{\rho_i\}_{i=1}^m$ is an arbitrary ensemble of quantum states with prior probabilities p_i. In [13] it was shown that if the value of

$$\min_{\Pi_i, s} s \tag{33}$$

$$\text{s.t.} \begin{cases} \Pi_i \ge 0, \\ \sum_{i=1}^{m} \Pi_i = I_n \\ \text{Tr}[\Pi_i A_i(\delta)] \le s, \end{cases}$$

is non-negative (i.e. $s^{\text{opt}} \ge 0$) for a specific choice of δ, then $||e_{\text{post}}||_{\text{wc}} \ge \delta$.

The dual program of (33) is

$$\max_{Y, \lambda_i} \text{Tr}(Y)$$

$$\text{s.t.} \begin{cases} \lambda_i \ge 0, \\ \sum_{i=1}^{m} \lambda_i = 1, \\ \lambda_i A_i(\delta) - Y \ge 0. \end{cases}$$

This means that if, for a specific value of δ, one can find real scalars λ_i and an operator Y, such that

$$\lambda_i \ge 0,$$
$$\sum_{i=1}^{m} \lambda_i = 1,$$
$$\lambda_i A_i(\delta) - Y \ge 0, \tag{34}$$

then $\text{Tr}(Y) \le s^{\text{opt}}$. Therefore, if in addition to the requirements (34), Y also satisfies $\text{Tr}(Y) \ge 0$, then we are assured that $s^{\text{opt}} \ge 0$, and that δ is a lower bound on the *optimal* worst-case posterior error $||e_{\text{post}}||_{\text{wc}}^{\text{opt}}$.

Define the index subset

$$Q(\delta) = \{i | A_i(\delta) \ge 0\},$$

and denote its cardinality by $|Q(\delta)|$. If $Q(\delta)$ is non-empty, then we can choose

$$Y = 0, \qquad \lambda_i = \begin{cases} 0, & i \notin Q(\delta) \\ \frac{1}{|Q(\delta)|}, & i \in Q(\delta) \end{cases}$$

which satisfy all the above requirements (34). Thus, whenever $Q(\delta)$ is non-empty, δ is a lower bound on $||e_{\text{post}}||_{\text{wc}}^{\text{opt}}$, and $P_p \le 1-\delta$. ∎

As an example of the application of Theorem 6, consider a set of pure states $\{|\alpha_i\rangle\}_{i=1}^m$ in an n-dimensional space, such that

$$\sum_{i=1}^{m} |\alpha_i\rangle\langle\alpha_i| = \frac{m}{n} I$$

is satisfied. This condition implies that $|\alpha_i\rangle$ form a tight frame whose vectors all have equal length[3]. When using this ensemble for encoding equiprobable data $p_i = \frac{1}{m}$, the operators $A_i(\delta)$, defined in Theorem 6, are

$$A_i(\delta) = \frac{1-\delta}{n} I - \frac{1}{m}|\alpha_i\rangle\langle\alpha_i|.$$

For $\delta = 1 - \frac{n}{m}$, we get that $A_i(\delta) \ge 0$ for all i. In conjunction with Theorem 6, this implies that when using this ensemble, $P_p \le \frac{n}{m}$. For this particular ensemble, the bound is tight, and is attained by the detector $\Pi_i = \frac{n}{m}|\alpha_i\rangle\langle\alpha_i|$.

B. Choosing the 'Don't Care' States of Optimal TFESs

The ensemble $\{|\alpha_i\rangle\}_{i=1}^m$ described above, and the detector which attains the bound on P_p form a TFES. From Corollary 4.3, we know that this setup also attains maximal P_d. In many situations, though, setups which attain maximum P_d, have $\Pi_i = 0$ for some i. When this is the case, there are undecided degrees of freedom to the TFES - the 'don't care' states. We would like to be able to choose these states so that the outcome of the measurement is reliable. We measure the reliability using P_p^{eff} (since P_p is ill-defined). We now introduce some notation, and formulate an upper bound on P_p^{eff}, over the set of optimal TFESs. We then show that this bound can be attained when using the pseudo-classical TFES (4).

Assume that for some arbitrary prior probability distribution $\{p_i\}_{i=1}^m$, the TFES described by $\{|u_i\rangle\}_{i=1}^m$ is optimal in the sense of P_d. We define the new index sets

$$\tilde{\mathcal{I}}_2 = \{i \,|\, p_i = p_n,\ |u_i\rangle \ne 0\} \subseteq \mathcal{I}_2,$$
$$\tilde{\mathcal{I}}_3 = \{i \,|\, |u_i\rangle = 0\} \supseteq \mathcal{I}_3$$

Using this notation, $\{\rho_i\}_{i\in\tilde{\mathcal{I}}_3}$ are the don't care states. In addition, note that $\mathcal{I}_1 \cup \tilde{\mathcal{I}}_2 \cup \tilde{\mathcal{I}}_3 = \{1, \ldots, m\}$.

[3]The existence of such tight frames for any finite n and m, is addressed in [32], [33] and references therein.

Theorem 7: Let $\{p_i\}_{i=1}^m$ be a non-increasing probability distribution, and let $\{|u_i\rangle\}_{i=1}^m$ be the vectors describing a TFES, which is optimal in the sense of P_d. Denote

$$c = \Pr\{\text{'don't care'}\} = \sum_{i\in\tilde{\mathcal{I}}_3} p_i.$$

Then

$$P_p^{\text{eff}} \le \frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n p_i + c}.$$

Proof: We rely on the following lemma, whose proof is given in Appendix C.

Lemma 3: Let $\{p_i\}_{i=1}^m$ be a probability distribution, and let $\{|u_i\rangle\}_{i=1}^m$ be the vectors describing a TFES, which is optimal in the sense of P_d. For any $i \in \mathcal{I}_1 \cup \tilde{\mathcal{I}}_2$

$$P_p(i) \le \frac{p_i\langle u_i|u_i\rangle}{p_i\langle u_i|u_i\rangle + \sum_{j\in\tilde{\mathcal{I}}_3} p_j\langle u_i|\rho_j|u_i\rangle}.$$

Equality for all $i \in \mathcal{I}_1 \cup \tilde{\mathcal{I}}_2$ is attained only for the pseudo-classical TFES (4).

By definition, and using Lemma 3,

$$P_p^{\text{eff}} \le \min_{i\in\mathcal{I}_1\cup\tilde{\mathcal{I}}_2} \frac{p_i\langle u_i|u_i\rangle}{p_i\langle u_i|u_i\rangle + \sum_{j\in\tilde{\mathcal{I}}_3} p_j\langle u_i|\rho_j|u_i\rangle}.$$

This implies that the value of P_p^{eff} is no greater than the value of

$$\max_{\rho_j, t} \quad t \tag{35}$$
$$\text{s.t.} \begin{cases} \rho_j \ge 0, \\ \operatorname{Tr}(\rho_j) = 1, \\ t \le \dfrac{p_i\langle u_i|u_i\rangle}{p_i\langle u_i|u_i\rangle + \sum_{j\in\tilde{\mathcal{I}}_3} p_j\langle u_i|\rho_j|u_i\rangle} \quad i \in \mathcal{I}_1 \cup \tilde{\mathcal{I}}_2 \end{cases}$$

Define

$$\lambda_i \triangleq \sum_{j\in\tilde{\mathcal{I}}_3} p_j\langle u_i|\rho_j|u_i\rangle \qquad i \in \mathcal{I}_1 \cup \tilde{\mathcal{I}}_2$$

From the restrictions posed in (35) on ρ_j, and using (12) together with the definition of $\tilde{\mathcal{I}}_3$, we know that λ_i must satisfy

$$\lambda_i \ge 0, \qquad i \in \mathcal{I}_1 \cup \tilde{\mathcal{I}}_2 \tag{36}$$
$$\sum_{i\in\mathcal{I}_1\cup\tilde{\mathcal{I}}_2} \lambda_i = c. \tag{37}$$

We now relax the constraints of the optimization problem (35), replacing the design variables ρ_j by the variables λ_i and retaining the constraints (36) and (37). To obtain the desired bound we must solve

$$\max_{\lambda_i, t} \quad t \tag{38}$$
$$\text{s.t.} \begin{cases} \lambda_i \ge 0, \\ \sum_{i\in\mathcal{I}_1\cup\tilde{\mathcal{I}}_2} \lambda_i = c, \\ t \le \dfrac{p_i\langle u_i|u_i\rangle}{p_i\langle u_i|u_i\rangle + \lambda_i} \quad i \in \mathcal{I}_1 \cup \tilde{\mathcal{I}}_2 \end{cases}$$

Since (38) is a relaxation of (35) its solution is an upper bound on P_p^{eff}.

Consider the choice

$$\hat{\lambda}_i = \frac{cp_i\langle u_i|u_i\rangle}{\sum_{i=1}^n p_i}$$
$$\hat{t} = \frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n p_i + c}$$

Recalling the fact that for any TFES which is optimal in the sense of P_d

$$\sum_{i\in\mathcal{I}_1\cup\tilde{\mathcal{I}}_2} p_i\langle u_i|u_i\rangle = \sum_{i=1}^n p_i,$$

one can easily verify that it is a feasible point. In addition, we have that for all $i \in \mathcal{I}_1 \cup \tilde{\mathcal{I}}_2$

$$\frac{p_i\langle u_i|u_i\rangle}{p_i\langle u_i|u_i\rangle + \hat{\lambda}_i} = \hat{t}.$$

This implies that for this choice of the variables λ_i, $\hat{t}$ is the largest value that the objective t can take. Due to the equality constraint, for any other choice of λ_i, there must be some $i \in \mathcal{I}_1 \cup \tilde{\mathcal{I}}_2$ such that $\lambda_i > \hat{\lambda}_i$. For this particular i we would have

$$t \le \frac{p_i\langle u_i|u_i\rangle}{p_i\langle u_i|u_i\rangle + \lambda_i} < \frac{p_i\langle u_i|u_i\rangle}{p_i\langle u_i|u_i\rangle + \hat{\lambda}_i} = \hat{t}$$

and thus, for any other choice of λ_i, all feasible t are strictly smaller than $\hat{t}$. In conclusion, $\hat{t}$ is the optimum value of problem (38), and therefore bounds P_p^{eff} from above. ■

The bound presented by Theorem 7 is not always tight, since it depends on the tightness of the bound given by Lemma 3. However, for the case of the pseudo-classical TFES, there is a choice of the 'don't care' states which attains the upper bound, and is therefore optimal in the sense of P_p^{eff}.

Corollary 7.1: When using the pseudo-classical TFES (4), with the choice

$$\rho_j = \frac{1}{\sum_{i=1}^n p_i} \sum_{i=1}^n p_i|u_i\rangle\langle u_i|, \qquad j = n+1, \dots, m$$

for the 'don't care' states, the value of P_p^{eff} attains the upper bound given by Theorem 7.

Proof: For the case of the pseudo-classical TFES, we have that

$$c = \sum_{i=n+1}^m p_i$$

and from Theorem 7

$$P_p^{\text{eff}} \le \frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n p_i + \sum_{i=n+1}^m p_i} = \sum_{i=1}^n p_i.$$

For all $i \leq n$ we get

$$\begin{aligned} P_p(i) &= \frac{p_i}{p_i + \sum_{j=n+1}^{m} \frac{p_j}{\sum_{i=1}^{n} p_i} \sum_{k=1}^{n} p_k |\langle u_i | u_k \rangle|^2} \\ &= \frac{p_i}{p_i + \sum_{j=n+1}^{m} \frac{p_i p_j}{\sum_{i=1}^{n} p_i}} \\ &= \frac{\sum_{i=1}^{n} p_i}{\sum_{i=1}^{n} p_i + \sum_{j=n+1}^{m} p_j} \\ &= \sum_{i=1}^{n} p_i. \end{aligned}$$

Thus $P_p^{\text{eff}} = \sum_{i=1}^{n} p_i$, which is equal to the upper bound. ■

VII. Conclusion

We have addressed the question of retrieval of digital data encoded in a quantum medium, using as our main performance criterion the probability of correct detection. We have found the optimal code-states for an arbitrary detector, and the optimal encoding-retrieval setups for an arbitrary prior distribution.

In terms of P_d one cannot do better then pseudo-classical transmission (orthonormal code-states and measurement operators). We have also shown that of all the setups which attain maximal P_d, the pseudo-classical TFES can be made to have optimal effective worst-case posterior probability. We have, however, indicated that under certain circumstances, there are benefits for using fully quantum setups (non-orthogonal code-states).

The natural extension of this work is the design of optimal setups with added constraints. Such constraints may arise due to requirements other than reliable communication, such as the need for security discussed above. Constraints may also stem from implementation issues which are typical to specific quantum systems that regularly serve for transmission and storage of information.

Acknowledgements

We would like to thank Moshe Nazarathy, who helped spark our interest in the subjects of this paper. We also benefited from discussions with Tal Mor.

Appendix A
Proof of Lemma 1

For all $1 \leq i \leq m$ we have that

$$|\langle u_i | u_i \rangle|^2 \leq \sum_{k=1}^{m} |\langle u_i | u_k \rangle|^2 = \langle u_i | \sum_{k=1}^{m} |u_k\rangle\langle u_k | u_i \rangle.$$

Using (12), this implies that

$$|\langle u_i | u_i \rangle|^2 \leq \langle u_i | I | u_i \rangle = \langle u_i | u_i \rangle.$$

Thereupon $\langle u_i | u_i \rangle \leq 1$, proving the property (13).

If $\langle u_i | u_i \rangle = 1$ then

$$\langle u_i | u_i \rangle = \langle u_i | \sum_{k=1}^{m} |u_k\rangle\langle u_k | u_i \rangle = \sum_{k=1}^{m} |\langle u_i | u_k \rangle|^2 = 1$$

making

$$\sum_{k \neq i} |\langle u_i | u_k \rangle|^2 = 0$$

Since this is a sum of nonnegative numbers, then for all $k \neq i$ we have

$$|\langle u_i | u_k \rangle|^2 = 0 \qquad \Rightarrow \qquad \langle u_i | u_k \rangle = 0$$

proving (14). Property (15) follows from taking the trace of (12).

Appendix B
Proof of Lemma 2

Assume that $(\bar{\eta}_i, \bar{\mu})$ is a feasible point of the programme (10), such that $\bar{\mu} = 0$. From the constraint (10b), $\bar{\eta}_i$ must satisfy $\bar{\eta}_i \geq p_i$ and then

$$g(\bar{\eta}_i, \bar{\mu}) \geq \sum_{i=1}^{m} p_i > \sum_{i=1}^{n} p_i = g(\hat{\eta}_i, \hat{\mu}),$$

where $(\hat{\eta}_i, \hat{\mu})$ are defined in (11). Thus $(\bar{\eta}_i, \bar{\mu})$ cannot be a dual optimal point. All dual optimal points must satisfy $\mu \neq 0$.

One of the KKT conditions for the solution to problem (8) is

$$\mu \left(n - \sum_{i=1}^{m} \sigma_i \right) = 0.$$

Since the dual optimal $\mu \neq 0$, then any optimal values of σ_i must satisfy.

$$\sum_{i=1}^{m} \sigma_i = n. \tag{B.1}$$

Let $\{\Pi_i\}$ be a POVM, which is part of an optimal ensemble-detector setup, i.e. $\sum_i p_i \sigma_{\Pi_i}^{\max} = \hat{P}_d$. By choosing

$$\hat{\sigma}_i = \sigma_{\Pi_i}^{\max} \tag{B.2}$$

we get $\sum_i p_i \hat{\sigma}_i = \hat{P}_d$, ensuring that $\hat{\sigma}_i$ are an optimum of (8), and thus satisfy (B.1). In conjunction with (B.2), this proves the Lemma.

Appendix C
Proof of Lemma 3

From Theorem 5, for any TFES which is optimal in the sense of P_d,

$$\begin{aligned} \Pr\{\text{det } i\} &= \sum_{j=1}^{m} p_j \text{Tr}(\Pi_i \rho_j) \\ &= \sum_{j \in \mathcal{I}_1 \cup \tilde{\mathcal{I}}_2} \frac{p_j}{\langle u_j | u_j \rangle} |\langle u_i | u_j \rangle|^2 + \sum_{j \in \tilde{\mathcal{I}}_3} p_j \langle u_i | \rho_j | u_i \rangle. \end{aligned} \tag{C.1}$$

By definition, for such a TFES,

$$P_p(i) = \frac{p_i \langle u_i | u_i \rangle}{\Pr\{\text{det } i\}}.$$

In order to prove the lemma, we must show that for any such TFES, and for every $i \in \mathcal{I}_1 \cup \tilde{\mathcal{I}}_2$

$$\Pr\{\text{det } i\} \geq p_i \langle u_i | u_i \rangle + \sum_{j \in \tilde{\mathcal{I}}_3} p_j \langle u_i | \rho_j | u_i \rangle,$$

From (C.1), this is equivalent to showing that for $i \in \mathcal{I}_1 \cup \tilde{\mathcal{I}}_2$

$$\sum_{j \in \mathcal{I}_1 \cup \tilde{\mathcal{I}}_2} \frac{p_j}{\langle u_j | u_j \rangle} |\langle u_i | u_j \rangle|^2 \geq p_i \langle u_i | u_i \rangle \tag{C.2}$$

Theorem 5 states that for all $i \in \mathcal{I}_1$, $\langle u_i | u_i \rangle = 1$. Recalling (14), this means that for every $i \in \mathcal{I}_1$, and $j \in \mathcal{I}_1 \cup \tilde{\mathcal{I}}_2$, $\langle u_i | u_j \rangle = \delta_{i,j}$. Thus, for every $i \in \mathcal{I}_1$ (C.2) is met with equality, and for every $i \in \tilde{\mathcal{I}}_2$,

$$\begin{aligned}\sum_{j \in \mathcal{I}_1 \cup \tilde{\mathcal{I}}_2} \frac{p_j}{\langle u_j | u_j \rangle} |\langle u_i | u_j \rangle|^2 &= p_n \sum_{j \in \tilde{\mathcal{I}}_2} \frac{1}{\langle u_j | u_j \rangle} |\langle u_i | u_j \rangle|^2 \\ &\geq p_n \sum_{j \in \tilde{\mathcal{I}}_2} |\langle u_i | u_j \rangle|^2. \end{aligned} \tag{C.3}$$

The last inequality stems from the fact that for all j, $\frac{1}{\langle u_j | u_j \rangle} \geq 1$ (using (13)).

Consider the subspace $\mathcal{G} = \text{span}\{|u_j\rangle\}_{j \in \tilde{\mathcal{I}}_2}$, and denote by $P_{\mathcal{G}}$ the orthogonal projection onto $\mathcal{G}$. From the orthogonality between $\{|u_j\rangle\}_{j \in \mathcal{I}_1}$ and $\mathcal{G}$, and from (12), it is clear to see that $\{|u_j\rangle\}_{j \in \tilde{\mathcal{I}}_2}$ is a tight frame on the subspace $\mathcal{G}$, i.e.

$$\sum_{j \in \tilde{\mathcal{I}}_2} |u_j\rangle \langle u_j| = P_{\mathcal{G}}.$$

Thus, for all $i \in \tilde{\mathcal{I}}_2$,

$$\sum_{j \in \tilde{\mathcal{I}}_2} |\langle u_i | u_j \rangle|^2 = \langle u_i | u_i \rangle. \tag{C.4}$$

Incorporating this with (C.3), we find that (C.2) is met for all $i \in \mathcal{I}_1 \cup \tilde{\mathcal{I}}_2$, thus proving the upper bound. This bound is tight for all $i \in \mathcal{I}_1 \cup \tilde{\mathcal{I}}_2$ only when there is equality in (C.3). This occurs when $\langle u_j | u_j \rangle = 1$ for all $j \in \tilde{\mathcal{I}}_2$. The only TFES for which this is true is the pseudo-classical setup (4).

REFERENCES

[1] A. Peres, *Quantum Theory: Concepts and Methods*. Waterloo, Canada: Kluwer Academic Publishers, 1993.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & sons, 1991.

[3] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, "Quantum cryptography," *Rev. Mod. Phys.*, vol. 74, pp. 145–195, Jan 2002.

[4] A. S. Holevo, "Statistical decision theory for quantum systems," *J. Multivar. Anal.*, vol. 3, pp. 337–394, Dec 1973.

[5] H. P. Yuen, R. S. Kennedy, and M. Lax, "Optimum testing of multiple hypotheses in quantum detection theory," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 125–134, 1975.

[6] Y. C. Eldar, A. Megretski, and G. C. Verghese, "Designing optimal quantum detectors via semidefinite programming," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1012–1017, 2003.

[7] C. W. Helstrom, *Quantum Detection and Estimation Theory*. New York: Academic Press, 1976.

[8] M. Charbit, C. Bendjaballah, and C. W. Helstrom, "Cutoff rate for the m-ary psk modulation channel with optimal quantum detection," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1131–1133, Sep 1989.

[9] M. Osaki, M. Ban, and O. Hirota, "Derivation and physical interpretation of the optimum detection operators for coherent-state signals," *Phys. Rev. A*, vol. 54, pp. 1691–1701, Aug 1996.

[10] M. Ban, K. Kurokawa, R. Momose, and O. Hirota, "Optimum measurements for discrimination among symmetric quantum states and parameter estimation," *Int. J. Theor. Phys.*, vol. 36, pp. 1269–1288, 1997.

[11] Y. C. Eldar and G. D. Forney, Jr., "On quantum detection and the square-root measurement," *IEEE Trans. Inform. Theory*, vol. 47, pp. 858–872, 2001.

[12] Y. C. Eldar, A. Megretski, and G. C. Verghese, "Optimal detection of symmetric mixed quantum states," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1198–1207, 2004.

[13] R. L. Kosut, I. Walmsley, Y. C. Eldar, and H. Rabitz, "Quantum state detector design: Optimal worst-case a posteriori performance," 2004, submitted for publication. [Online]. Available: http://arXiv.org/abs/quant-ph/0403150

[14] I. D. Ivanovic, "How to differentiate between non-orthogonal states," *Phys. Lett. A*, vol. 123, pp. 257–259, Aug 1987.

[15] Y. C. Eldar, "A semidefinite programming approach to optimal unambiguous discrimination of quantum states," *IEEE Trans. Inform. Theory*, vol. 49, pp. 446–456, Feb 2003.

[16] Y. C. Eldar, M. Stojnic, and B. Hassibi, "Optimal quantum detectors for unambiguous detection of mixed states," *Phys. Rev. A*, vol. 69, no. 6, p. 062318, 2004.

[17] M. Sasaki, R. Momose, and O. Hirota, "Quantum detection for on-off keyed mixed-state signals with a small amount of thermal noise," *Phys. Rev. A.*, vol. 55, no. 4, pp. 3222–3225, Apr 1997.

[18] V. Vilnrotter and C. W. Lau, "Quantum detection of binary and ternary signals in the presence of thermal noise fields," *The InterPlanetary Network Progress Report 42-152, October-Decmber 2002*, Feb 2003. [Online]. Available: http://ipnpr.jpl.nasa.gov/tmo/progress_report/42-152/152B.pdf

[19] J. I. Concha and H. V. Poor, "Multiaccess quantum channels," *IEEE Trans. Inform. Theory*, vol. 50, pp. 725–747, May 2004.

[20] N. Elron and Y. C. Eldar, "Quantum detection with uncertain states," *Phys. Rev. A.*, vol. 72, p. 032338, 2005. [Online]. Available: http://arxiv.org/abs/quant-ph/0501084

[21] G. M. D'Ariano, M. F. Sacchi, and J. Kahn, "Minimax quantum state discrimination," 2005, submitted to Phys. Rev. A. [Online]. Available: http://arXiv.org/abs/quant-ph/0504048

[22] Y. Xiao, F. Crusca, and E. K.-w. Chu, "Bilinear matrix inequalities in robust control: Phase I - problem formulation," Monash University, Victoria, Australia, Tech. Rep. MECSE-3-1996, Apr 1996. [Online]. Available: http://www.ds.eng.monash.edu.au/techrep/reports/

[23] O. Toker and H. Özbay, "On the $\mathcal{NP}$-hardness of solving bilinear matrix inequalities and simultaneous stabilization with static output feedback," in *Proc. of the 1995 American Control Conference*, Seattle, Washington, 1995, pp. 2525–2526.

[24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, Mar 2004.

[25] Y. C. Eldar and G. D. Forney, Jr., "Optimal tight frames and quantum measurement," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 599–610, Mar 2002.

[26] S. Verdú, *Multiuser Detection*. Cambridge University Press, 1998.

[27] P. Viswanath, V. Anantharam, and D. N. C. Tse, "Optimal sequences, power control, and user capacity of syncronous CDMA systems with linear MMSE multiuser receivers," *IEEE Trans. Inform. Theory*, vol. 45, no. 6, pp. 1968–1983, Sep 1999.

[28] A. V. Balakrishnan, "A contribution to the sphere-packing problem of communication systems," *J. Math. Anal. Appl.*, vol. 3, pp. 485–506, Dec 1961.

[29] M. Steiner, "The strong simplex conjecture is false," *IEEE Trans. Inform. Theory*, vol. 40, no. 3, pp. 721–731, May 1994.

[30] C. H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," in *Proc. of IEEE International Conference on Computers, Systems and Signal Processing*, Bangalore, India, Dec 1984, pp. 175–179.

[31] D. R. Cox and D. V. Hinkley, *Theoretical Statistics*. London, UK: Chapman and Hall, 1974.

[32] R. Reams and S. Waldron, "Isometric tight frames," *Electr. J. Lin. Algeb.*, vol. 9, pp. 122–128, Jul 2002.

[33] J. A. Tropp, I. S. Dhillon, R. W. Heath, Jr., and M. A. Sustik, "Designing structured tight frames via an alternating projection method," *IEEE Trans. Inform. Theory*, vol. 51, no. 1, pp. 188–209, 2005.