ENTANGLEMENT COST OF GENERALISED MEASUREMENTS

RICHARD JOZSA[1], MASATO KOASHI[1,3], NOAH LINDEN[2], SANDU POPESCU[4],
STUART PRESNELL[1], DAN SHEPHERD[5] and ANDREAS WINTER[1]

[1] *Department of Computer Science, University of Bristol,*
Merchant Venturers Building, Bristol BS8 1UB U.K.
[2] *Department of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW U.K.*
[3] *School of Advanced Studies, The Graduate University for Advanced Studies*
(SOKENDAI), Shonan Village, Hayama, Kanagawa, 240-0193, Japan.
[4] *Department of Physics, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, U.K.*
[5] *CESG, PO Box 144, Cheltenham, Gloucestershire, GL52 5UE U.K.*

Bipartite entanglement is one of the fundamental quantifiable resources of quantum infor-
mation theory. We propose a new application of this resource to the theory of quantum
measurements. According to Naimark's theorem any rank 1 generalised measurement
(POVM) M may be represented as a von Neumann measurement in an extended (tensor
product) space of the system plus ancilla. By considering a suitable average of the entan-
glements of these measurement directions and minimising over all Naimark extensions,
we define a notion of entanglement cost $E_{\min}(M)$ of M.

We give a constructive means of characterising all Naimark extensions of a given
POVM. We identify various classes of POVMs with zero and non-zero cost and explicitly
characterise all POVMs in 2 dimensions having zero cost. We prove a constant upper
bound on the entanglement cost of any POVM in any dimension. Hence the asymptotic
entanglement cost (i.e. the large n limit of the cost of n applications of M, divided by
n) is zero for all POVMs.

The trine measurement is defined by three rank 1 elements, with directions symmet-
rically placed around a great circle on the Bloch sphere. We give an analytic expression
for its entanglement cost. Defining a normalised cost of any d-dimensional POVM by
$E_{\min}(M)/\log_2 d$, we show (using a combination of analytic and numerical techniques)
that the trine measurement is more costly than any other POVM with $d > 2$, or with
$d = 2$ and ancilla dimension 2. This strongly suggests that the trine measurement is the
most costly of all POVMs.

Keywords: Entanglement, Naimark extensions, generalised measurements, trine mea-
surement

Communicated by: I Cirac & C Fuchs

1 Introduction

Bipartite entanglement is one of the fundamental quantifiable resources of quantum informa-
tion theory. In this paper we will propose a new application of this resource to the theory of
quantum measurements, motivated by the following considerations.

Let $\mathcal{M} = \{A_\mu : A_\mu \geq 0, \sum A_\mu = I\}_{\mu=1}^m$ be a positive operator valued measure (POVM) or
"generalised measurement" on the space $\mathcal{H}_d$ of dimension d. We will be interested in POVMs
with rank 1 elements $A_\mu = |k_\mu\rangle\langle k_\mu|$, where the $|k_\mu\rangle$'s are sub-normalised and generally non-

orthogonal vectors. In quantum theory only orthogonal states have the fundamental property of being reliably distinguishable and in a measurement the outcomes correspond to classical information which is similarly reliably distinguishable. This connection between orthogonality and measurement outcomes is embodied in the definition of a (complete) von Neumann measurement, as being based on mutually orthogonal directions. For a (rank 1) generalised measurement the vectors $|k_\mu\rangle$ are not orthogonal yet the outcome information is again classical. In this case the appearance of orthogonality is restored by Naimark's theorem [1, 2] which asserts that any generalised measurement may be viewed as a von Neumann measurement on a suitable enlarged space comprising the target system together with an auxiliary system or ancilla. In the present work we wish to think of von Neumann measurements and their in-built orthogonality as the fundamental substrate of all quantum measurement. The formalism of (non-orthogonal) POVMs is then merely a mathematical artifice that expresses the residual effect on the system when we perform an orthogonal measurement in an extended setting. We will see in section 2 below that the adjoining of an ancilla to represent $\mathcal{M}$ as a von Neumann measurement provides a natural way of associating an entanglement cost to the measurement $\mathcal{M}$. For direct von Neumann measurements on the target system (where no ancilla is required), the entanglement cost will be zero.

A possible motivation or application of such a notion of entanglement cost for a generalised measurement arises in the study of information acquisition from a quantum source. Let $\mathcal{E} = \{|\psi_i\rangle; p_i\}$ be a source of pure quantum signal states $|\psi_i\rangle$ with prior probabilities p_i. Write $\Pi_i = |\psi_i\rangle\langle\psi_i|$. If $\mathcal{M} = \{A_\mu : A_\mu \geq 0, \sum A_\mu = I\}_{\mu=1}^m$ is any POVM on the signal space then its measurement on a signal state results in a joint probability distribution $p(i\&\mu) = p(\mu|i)p(i) = (\operatorname{tr} A_\mu \Pi_i)p_i$. We can then calculate the associated mutual information $I(\mathcal{M} : \mathcal{E})$ which is interpreted as the amount of information that the measurement outcome provides about the identity of the state $|\psi_i\rangle$. For a given source $\mathcal{E}$ we are often interested in determining $\mathcal{M}$ so that $I(\mathcal{M} : \mathcal{E})$ is maximised. It is known [3] that for any source an optimal measurement may be regarded as having all rank 1 elements. Indeed in generality suppose that A_μ has spectral decomposition $A_\mu = \sum \lambda_i |\lambda_i\rangle\langle\lambda_i|$ and in the measurement $\mathcal{M}$ we replace each A_μ by the set of rank 1 operators $\{\lambda_i|\lambda_i\rangle\langle\lambda_i|\}$, giving a refined measurement $\mathcal{M}'$ that is rank 1. Then for any source $\mathcal{E}$ we have $I(\mathcal{M}' : \mathcal{E}) \geq I(\mathcal{M} : \mathcal{E})$. Thus it is natural to restrict attention to only rank 1 POVMs, with elements of the form $A_\mu = |k_\mu\rangle\langle k_\mu|$ and we have

$$p(\mu|i) = |\langle\psi_i|k_\mu\rangle|^2 \qquad p(\mu) = \sum_i p_i |\langle\psi_i|k_\mu\rangle|^2.$$

It is known that some sources have a direct von Neumann measurement as an optimal measurement but there exist other sources for which no such direct von Neumann measurement is optimal. For example the binary source $\mathcal{B} = \{|\psi_1\rangle, |\psi_2\rangle; p_1, p_2\}$ always has an optimal von Neumann measurement [4] whereas the so-called trine source of three symmetric equiprobable states in two dimensions, defined by $\mathcal{T} = \{|\alpha_1\rangle, |\alpha_2\rangle, |\alpha_3\rangle; 1/3, 1/3, 1/3\}$ with

$$|\alpha_j\rangle = \cos\frac{j\pi}{3}|0\rangle + \sin\frac{j\pi}{3}|1\rangle,$$

has the property that no von Neumann measurement is optimal [5]. Let $|\beta_i\rangle$ be three vectors orthogonal to $|\alpha_i\rangle$ respectively and subnormalised to have lengths $\sqrt{2/3}$. Then it may be shown [6] that the POVM $\mathcal{M}_\mathcal{T} = \{|\beta_i\rangle\langle\beta_i|\}_{i=1}^3$ is an optimal measurement for $\mathcal{T}$.

Thus we see that some sources (such as $\mathcal{T}$ having no direct von Neumann measurement as optimal) require an investment of entanglement to reveal the full quota of their available information, while others (such as any binary source) require no such entanglement input. Intuitively (for $\mathcal{T}$) the information needs to be "spread out" in a larger space in order for it to become "fully visible". This framework of ideas leads to many interesting new questions. For example we can ask: which source (of all qubit sources, say) requires the most entanglement investment to see its information content? We will study these issues at a slightly more fundamental level, focussing on the generalised measurements themselves, rather than the sources, using the precise formulations given in the next section.

This paper is organised as follows. In section 2 we develop a precise formulation of our notion of entanglement cost for generalised measurements. Our adopted approach is based on the formalism of Naimark extensions which we briefly review. In section 3 we give a constructive characterisation of all possible Naimark extensions of a given POVM which provides a basis for developing further properties of our notion of entanglement cost in section 4: firstly we identify various classes of POVMs with zero and non-zero cost, and explicitly characterise all POVMs in 2 dimensions having zero cost. Secondly we derive an analytic expression for the entanglement cost of the trine measurement (defined above). Thirdly we derive various upper bounds on the entanglement cost including a constant upper bound (independent of the dimension) for any POVM. Next, in section 5 we consider the question of which POVM has the greatest cost. Introducing the normalised entanglement cost of $\mathcal{M}$ (in d dimensions) as the cost of $\mathcal{M}$ divided by $\log_2 d$ (i.e. the cost per qubit of source space) we show that the trine measurement is more costly than any other POVM with $d \geq 3$ and conjecture that it is the most costly of all POVMs. In section 6 we introduce the notion of the *asymptotic* entanglement cost of a POVM (defined as the large n limit of the cost of n applications of $\mathcal{M}$ divided by n) and show that it is zero for all POVMs. Finally in section 7 we make some further remarks on the significance of our definition of entanglement cost and briefly discuss various other possible approaches.

2 Formulation of the problem

In our definition of entanglement cost, Naimark's theorem will play a fundamental role so we begin by elaborating its statement here. Let $\mathcal{H}_d$ denote the signal space (of dimension d) and $\mathcal{H}_e$ the ancilla space (of dimension e). Let $|a_0\rangle$ be any fixed chosen state of the ancilla. Let $\mathcal{M} = \{A_\mu\}_{\mu=1}^m$ be any rank 1 POVM and write $p(\mu|\psi) = \langle\psi|A_\mu|\psi\rangle$ for the probability of obtaining outcome μ on the state $|\psi\rangle \in \mathcal{H}_d$. Naimark's theorem states that $\mathcal{M}$ may always be realised as a von Neumann measurement in $\mathcal{H}_d \otimes \mathcal{H}_e$ for e sufficiently large to have $de \geq m$, in the following sense. Let $|\psi, a_0\rangle = |\psi\rangle|a_0\rangle$ be the state obtained from $|\psi\rangle$ by adjoining the ancilla in its standard state. Then there is an orthonormal basis $\{|w_\nu\rangle : \nu = 1, \ldots, ed\}$ in $\mathcal{H}_d \otimes \mathcal{H}_e$ such that for all $|\psi\rangle$:

$$p(\nu|\psi) = |\langle\psi, a_0|w_\nu\rangle|^2 \quad \text{when } \nu = 1, \ldots, m.$$

Any such orthonormal basis (for any choice of $\mathcal{H}_e$) is called a Naimark extension of $\mathcal{M}$. Any generalised measurement admits many inequivalent Naimark extensions and in section 3 below we will see how to classify and construct them all. Since probabilities must sum to 1

it follows that

$$p(\nu|\psi) = |\langle\psi, a_0|w_\nu\rangle|^2 = 0 \quad \text{for } m < \nu \le ed \tag{1}$$

i.e. the POVM is represented in the first m dimensions of the von Neumann measurement $\{|w_\nu\rangle\}$ and then the remaining von Neumann measurement directions must be orthogonal to the subspace $\mathcal{H}_d \otimes |a_0\rangle \subseteq \mathcal{H}_d \otimes \mathcal{H}_e$. These latter measurement directions lie in the span of $\{\mathcal{H}_d \otimes |a_i\rangle : i \ge 1\}$ where $\{|a_0\rangle, |a_1\rangle, \ldots |a_{e-1}\rangle\}$ is an orthonormal extension of $|a_0\rangle$ in $\mathcal{H}_e$. (More generally if the POVM is not rank 1 then Naimark's theorem states that it may similarly be represented as a von Neumann measurement that is incomplete i.e. not rank 1).

The states $|w_\nu\rangle$ are generally entangled across the decomposition $\mathcal{H}_d \otimes \mathcal{H}_e$ and their entanglements will form the basis of our definition of entanglement cost for $\mathcal{M}$. Note that if $\mathcal{M}$ is a von Neumann measurement in $\mathcal{H}_d$ then $m = d$, so we can take $e = 1$ and set $|w_\nu\rangle = |k_\nu\rangle|a_0\rangle$. Thus our entanglement cost will be zero but for more general POVMs we will obtain non-zero values.

In some treatments of Naimark's theorem (such as [1]) the basic signal space $\mathcal{H}_d$ is enlarged not via a tensor product $\mathcal{H}_d \otimes \mathcal{H}_e$ but instead via an embedding into a larger dimensional space $\mathcal{H}_d \hookrightarrow \mathcal{H}_N$ of suitable size to accommodate the number of elements of the POVM. For us it is important that the extension arises from a tensor product and indeed, from a physical point of view the adjoining of an additional ancilla system is the only available physical means of extending a space while retaining the original system intact.

Turning now to the formulation of our problem, we begin with the following ingredients: (i) a source $\mathcal{E} = \{|\psi_i\rangle; p_i\}$ in $\mathcal{H}_d$; (ii) a rank 1 POVM $\mathcal{M} = \{|k_\mu\rangle\langle k_\mu|\}_{\mu=1}^m$ with m outcomes, on $\mathcal{H}_d$; (iii) any Naimark extension $\mathcal{N} = \{|w_\nu\rangle\}_{\nu=1}^{de}$ of $\mathcal{M}$, which is an orthonormal basis of $\mathcal{H}_d \otimes \mathcal{H}_e$. Let E_ν denote the entanglement of $|w_\nu\rangle$ (defined as the von Neumann entropy of the reduced state of $\mathcal{H}_d$ or $\mathcal{H}_e$) and let

$$p(\nu) = \sum_i p_i p(\nu|i) = \sum_i p_i |\langle\psi_i a_0|w_\nu\rangle|^2 = \sum_i p_i |\langle\psi_i|k_\nu\rangle|^2 \tag{2}$$

be the posterior probability of obtaining outcome ν on $\mathcal{E}$ (where in the last expression we take $|k_\nu\rangle = 0$ for $\nu > m$). Then introduce $E = \sum_\nu p(\nu) E_\nu$, the average entanglement of the Naimark extension directions, weighted by their posterior probability of occurrence on $\mathcal{E}$. We define the entanglement cost of $\mathcal{M}$ on $\mathcal{E}$ to be $E_{\min}$, the minimum of E over all choices of Naimark extension of $\mathcal{M}$.

Note that if m (the number of POVM elements) is not an integer multiple of d then any Naimark extension will involve extra directions not associated with any POVM element. However eq. (1) shows that the entanglements of these extra states $|w_\nu\rangle : \nu > m$ do not enter into E. A further desirable feature of our definition is the fact that any von Neumann measurement will have entanglement cost zero (as noted previously).

Our definition of $E_{\min}$ depends on both $\mathcal{M}$ and $\mathcal{E}$. To obtain a characteristic of a generalised measurement alone we note that the POVM condition $\sum |k_\mu\rangle\langle k_\mu| = I$ implies $\sum \langle k_\mu|k_\mu\rangle = d$ so that $\{q_\mu = \langle k_\mu|k_\mu\rangle/d\}_{\mu=1}^m$ is a probability distribution canonically defined by $\mathcal{M}$. From eq. (2) we see that $p(\nu) = \sum_i p_i \langle k_\nu|\psi_i\rangle\langle\psi_i|k_\nu\rangle$ will equal q_ν if $\sum_i p_i|\psi_i\rangle\langle\psi_i| = \frac{1}{d}I$ so we can interpret the canonical distribution q_μ as the posterior probability of obtaining outcome μ on any *maximally mixed* source i.e. any source having $\sum_i p_i|\psi_i\rangle\langle\psi_i| = \frac{1}{d}I$ (e.g. the trine source for $d = 2$).

Using these ingredients we can state our central problem:

Problem: Given a POVM $\mathcal{M} = \{|k_\mu\rangle\langle k_\mu|\}_{\mu=1}^m$ on $\mathcal{H}_d$ let $q_\mu = \frac{1}{d}\langle k_\mu|k_\mu\rangle$. Let $\{|w_\nu\rangle\}$ on $\mathcal{H}_d \otimes \mathcal{H}_e$ be any Naimark extension of $\mathcal{M}$ and let E_ν be the entanglement of $|w_\nu\rangle$. Let $E = \sum_{\mu=1}^m q_\mu E_\mu$. Then we wish to determine $E_{\min}(\mathcal{M})$, the minimum value of E over all choices of Naimark extension of $\mathcal{M}$. $\square$

Note that our definition of entanglement cost (based on the entanglements of the Naimark extension states) is actually a measure of entanglement *production* by the POVM, rather than a measure of "input" entanglement cost, necessary to realise the POVM in some suitable sense. Further discussion of this point will be given in section 7.1 below.

3   Characterising all Naimark extensions of $\mathcal{M}$

In [1] Naimark's theorem is proved by showing how to construct a Naimark extension of any rank 1 POVM. However the extension of the space is achieved by subspace embedding $\mathcal{H}_d \hookrightarrow \mathcal{H}_N$, rather than via a tensor product, adjoining an ancilla. Below we adapt this method to work in the system–plus–ancilla formalism (c.f. also [7]) and show further that it provides a constructive classification of all possible Naimark extensions of the POVM.

Let $\mathcal{M} = \{|k_\mu\rangle\langle k_\mu|\}_{\mu=1}^m$ be any rank 1, m-element POVM in $\mathcal{H}_d$. Let the orthonormal basis $\{|w_\nu\rangle\}_{\nu=1}^{de}$ in $\mathcal{H}_d \otimes \mathcal{H}_e$ be any Naimark extension of it. The action of the POVM on states $|\psi\rangle$ in $\mathcal{H}_d$ corresponds to the Naimark extension acting on $|\psi\rangle|a_0\rangle$ which in turn requires a choice of ancilla state $|a_0\rangle$. When we need explicit mention of the chosen ancilla state we will refer to a "Naimark extension with basic ancilla state $|a_0\rangle$".

We begin by identifying a relationship between the vectors $|k_\mu\rangle$ and $|w_\mu\rangle$. If $\{|a_i\rangle\}_{i=0}^{e-1}$ is any orthonormal basis of $\mathcal{H}_e$ that extends $|a_0\rangle$, we can write

$$|w_\nu\rangle = |v_\nu^0\rangle|a_0\rangle + |v_\nu^1\rangle|a_1\rangle + \ldots + |v_\nu^{(e-1)}\rangle|a_{(e-1)}\rangle$$

for suitable (sub-normalised) coefficient vectors $|v_\nu^i\rangle$ in $\mathcal{H}_d$. The Naimark extension condition is $|\langle\psi, a_0|w_\nu\rangle|^2 = |\langle\psi|k_\nu\rangle|^2$ for $\nu = 1, \ldots, m$ and $|\langle\psi, a_0|w_\nu\rangle|^2 = 0$ for $\nu > m$. Thus we have

$$|\langle\psi|v_\nu^0\rangle| = |\langle\psi|k_\nu\rangle| \quad \text{for all } |\psi\rangle \text{ and } \nu = 1, \ldots, m$$

and by lemma 1 below we get $|v_\nu^0\rangle = e^{i\theta_\nu}|k_\nu\rangle$. These phases do not affect the entanglement that we are studying and without loss of generality we can take $\theta_\nu = 0$ (e.g. they can be absorbed into overall phase choices for the $|w_\nu\rangle$'s). Indeed, the phase of $|k_\nu\rangle$ is not even defined by the POVM.

Lemma 1 *If* $|\langle a|x\rangle| = |\langle b|x\rangle|$ *for all* $|x\rangle$ *then* $|a\rangle = |b\rangle e^{i\theta}$.

Proof Putting $x = a$ and b gives $|\langle a|b\rangle| = \langle a|a\rangle = \langle b|b\rangle$ and then the projection of $|a\rangle$ perpendicular to $|b\rangle$ is zero. $\square$

For $\nu > m$ we have $\langle\psi, a_0|w_\nu\rangle = 0$ so $|v_\nu^0\rangle = 0$ for $\nu > m$. Hence we have proved:

Proposition 1 *If* $\{|w_\nu\rangle\}_{\nu=1}^{de}$ *is any Naimark extension of* $\{|k_\mu\rangle\langle k_\mu|\}_{\mu=1}^m$ *with basic ancilla state* $|a_0\rangle$ *then the* $|w_\nu\rangle$ *have the form*

$$
\begin{aligned}
|w_\nu\rangle &= |k_\nu\rangle|a_0\rangle + |v_\nu^1\rangle|a_1\rangle + \ldots + |v_\nu^{(e-1)}\rangle|a_{(e-1)}\rangle \quad \text{for } \nu = 1, \ldots, m \\
|w_\nu\rangle &= |v_\nu^1\rangle|a_1\rangle + \ldots + |v_\nu^{(e-1)}\rangle|a_{(e-1)}\rangle \quad \text{for } \nu = m+1, \ldots, de
\end{aligned}
$$

i.e. $|v_\nu^0\rangle = |k_\nu\rangle$ *for* $\nu = 1, \ldots, m$ *and* $|v_\nu^0\rangle = 0$ *for* $\nu > m$.

We also have a converse result:

Proposition 2 *Let $\{|w_\nu\rangle\}_{\nu=1}^{de}$ be any orthonormal basis of $\mathcal{H}_d \otimes \mathcal{H}_e$. Using the basis $\{|a_i\rangle\}_{i=0}^{e-1}$ write*

$$|w_\nu\rangle = \sum_{i=0}^{e-1} |v_\nu^i\rangle |a_i\rangle.$$

Then $\{A_\nu = |v_\nu^0\rangle\langle v_\nu^0|\}_{\nu=1}^{de}$ is a POVM on $\mathcal{H}_d$. Similarly for each i, $\{A_\nu^{(i)} = |v_\nu^i\rangle\langle v_\nu^i|\}_{\nu=1}^{de}$ is a POVM on $\mathcal{H}_d$.

Proof For any $|\xi\rangle$ in $\mathcal{H}_d$ we have $|\langle \xi, a_0 | w_\nu\rangle|^2 = |\langle \xi|v_\nu^0\rangle|^2$ so

$$\sum_\nu \langle\xi|v_\nu^0\rangle\langle v_\nu^0|\xi\rangle = 1.$$

Writing $P = \sum_\nu |v_\nu^0\rangle\langle v_\nu^0|$ we thus have $\langle\xi|P|\xi\rangle = 1$ and P is also Hermitian. Taking $|\xi\rangle$ as the eigenstates of P we see that all eigenvalues must be 1 so $P = I$ and $\{|v_\nu^0\rangle\langle v_\nu^0|\}_{\nu=1}^d$ is a POVM. Similarly by first applying a unitary swap operation on $\mathcal{H}_e$ that interchanges the roles of $|a_0\rangle$ and $|a_i\rangle$ we get the corresponding result for each i. $\square$

Returning to our question of constructing all possible Naimark extensions of $\mathcal{M} = \{|k_\mu\rangle\langle k_\mu|\}_{\mu=1}^m$, proposition 1 shows that we need to characterise all possible sets $|v_\nu^i\rangle$ for $i \geq 1$ which make the $|w_\nu\rangle$'s orthonormal. This is achieved as follows. We will work in terms of components with respect to a chosen basis $\{|e_j\rangle\}$ of $\mathcal{H}_d$ and $\{|a_i\rangle\}$ of $\mathcal{H}_e$. Let M be the $m \times d$ matrix whose m rows are the components of the $|k_\mu\rangle$'s for $\mu = 1, \ldots, m$.

Proposition 3 *The POVM condition $\sum_\mu A_\mu = I$ is equivalent to the following condition: the d columns of M are an orthonormal set of vectors in $\mathbb{C}^m$.*

Proof We have $|k_\mu\rangle = \sum_j M_{\mu j}|e_j\rangle$ so the POVM condition $\sum_\mu |k_\mu\rangle\langle k_\mu| = I$ is

$$\sum_\mu M_{\mu j} M_{\mu k}^* = \delta_{jk}$$

i.e. for each j and k the jth and k^{th} columns of M are orthonormal. $\square$

We will write any vector $|\Phi\rangle \in \mathcal{H}_d \otimes \mathcal{H}_e$ as a row vector of components lexicographically ordered in the product basis $|e_j\rangle|a_k\rangle$. Thus we have de components and if

$$|\Phi\rangle = \sum_{i=0}^{e-1} |v^i\rangle |a_i\rangle$$

then the first d numbers in the row vector are the components of $|v^0\rangle$, the next d are the components of $|v^1\rangle$ and so on; the last d are the components of $|v^{(e-1)}\rangle$.

Above we introduced the $m \times d$ matrix M (having the $|k_\mu\rangle$'s as rows). Let Z be the $(de - m) \times d$ matrix with all zero entries. Then by proposition 3 the columns of

$$P = \left[\begin{array}{c} M \\ Z \end{array} \right]$$

are d orthonormal vectors in $\mathbb{C}^{de}$. We can extend these d columns to a full orthonormal basis of de vectors in $\mathbb{C}^{de}$ by adjoining further columns to get a $de \times de$ matrix

$$L = \left[\begin{array}{cc} M & A \\ Z & B \end{array} \right].$$

Here A has size $m \times (de - d)$ and B has size $(de - m) \times (de - d)$. The freedom in the choice of this extension is the set of all maximal orthonormal families in $\mathbb{C}^{de}$ that are orthogonal to the first d columns of L. This set is isomorphic to the set of all orthonormal bases of $\mathbb{C}^{de-d}$.

Proposition 4 *For any square matrix R, R has orthonormal columns iff R has orthonormal rows.*

Proof This is a standard fact from linear algebra. $\square$

Now L has orthonormal columns so by proposition 4 L has orthonormal rows too. (Indeed L is a unitary matrix.) Hence the rows of A and B read out in segments of length d give the vectors $|v_\nu^i\rangle$, $i \geq 1$, of a Naimark extension of $\mathcal{M}$ (as in propositions 1 and 2). Conversely we see that every Naimark extension arises in this way: taking any Naimark extension as in proposition 1 we write out the $de \times de$ matrix L' with the components of the $|w_\nu\rangle$'s as rows. Then the rows of L' are orthonormal so by proposition 4 the columns are orthonormal too. Also by proposition 1 the entries of L' in the positions of Z are all zero.

Hence we have a prescription for generating all possible Naimark extensions of the POVM $\{|k_\mu\rangle\langle k_\mu|\}_{\mu=1}^m$ in dimension d with an ancilla of dimension e:

- Write down the matrix M whose rows are the components of the vectors $|k_\mu\rangle$.

- Extend the columns of M by zeroes to get a total length of de.

- Extend these d orthonormal columns to a full orthonormal basis of $\mathbb{C}^{de}$ (by adjoining further columns) giving a square matrix L.

- Look at the rows of L to read off the Naimark extension states.

3.1 Bounding the ancilla size

Although unboundedly large ancillas (i.e. e values) can be used in the above prescription, they do not help in reducing the entanglement cost of $\mathcal{M}$. It suffices to consider only $e \leq md$ as shown next.

Theorem 1 *For the m element POVM $\mathcal{M} = \{|k_\mu\rangle\langle k_\mu|\}_{\mu=1}^m$ let $\{|w_\nu\rangle\}_{\nu=1}^{dn}$ be any Naimark extension with ancilla dimension n and entanglement cost E. Then there exists a Naimark extension of $\mathcal{M}$ with ancilla dimension $e \leq md$ which also has entanglement cost E.*

Proof Each $|w_\nu\rangle$ is an entangled state in $\mathcal{H}_d \otimes \mathcal{H}_n$ and hence is supported on a subspace of dimension at most d in the ancilla space. So in the ancilla space at most md dimensions are utilised by the m states $|w_\nu\rangle$ for $\nu = 1, \ldots, m$. Hence by a local unitary transformation on the ancilla (which does not affect the entanglements), we can rotate these dimensions into the first md positions. Hence an ancilla of dimension md suffices to allow a Naimark extension with the same entanglement cost. $\square$

4 Some properties of the entanglement cost of $\mathcal{M}$

4.1 Measurements with zero and nonzero entanglement cost

In this subsection we characterise a class of POVMs for any dimension d, with zero entanglement cost and also a class which has non-zero cost. For $d = 2$ we give a full characterisation of zero cost POVMs.

Proposition 5 *Let $\mathcal{M}_0 = \{|k_\mu\rangle\langle k_\mu|\}_{\mu=1}^m$ and $\mathcal{M}_1 = \{|l_\nu\rangle\langle l_\nu|\}_{\nu=1}^n$ be two POVMs. If $\{p_0, p_1\}$ is a probability distribution define the convex combination POVM (with $m + n$ elements) $\mathcal{M} = \bigcup_i p_i \mathcal{M}_i = \{p_0|k_\mu\rangle\langle k_\mu|, p_1|l_\nu\rangle\langle l_\nu|\}_{\mu=1,\nu=1}^{m,n}$. Then*

$$E_{\min}(\mathcal{M}) \le \sum_i p_i E_{\min}(\mathcal{M}_i).$$

Proof Let $B_0 = \{|w_\mu^0\rangle\}$ and $B_1 = \{|w_\nu^1\rangle\}$ be Naimark extensions of $\mathcal{M}_0$ and $\mathcal{M}_1$ respectively, that realise the minimal entanglement costs. We can enlarge the smaller ancilla space so without loss of generality we may assume that the ancilla spaces have the same dimension e and that the fundamental state $|a_0\rangle$ (c.f. proposition 1) coincides for B_0 and B_1. Now consider the set $B = \{|w_\mu^0\rangle|0\rangle, |w_\nu^1\rangle|1\rangle\}_{\mu=1,\nu=1}^{me,ne}$ where $|0\rangle$ and $|1\rangle$ are orthonormal states of a new system which we view as an extension of the ancilla. This is clearly an orthonormal set and if we take the fundamental ancilla state $|a_0'\rangle = |a_0\rangle(\sqrt{p_0}|0\rangle + \sqrt{p_1}|1\rangle)$ then B is a Naimark extension of $\mathcal{M}$. Since the ancilla in each state of B has been extended as a product state, the entanglements are unchanged so the entanglement cost of $\mathcal{M}$ is at most $\sum_i p_i E_{\min}(\mathcal{M}_i)$. □

Note that in the above construction, if $\mathcal{M}_0$ and $\mathcal{M}_1$ have elements which are proportional, they remain as separate elements in the convex combination POVM. Since von Neumann measurements have zero cost we immediately get:

Corollary 1 *Let $\mathcal{V}_1, \ldots, \mathcal{V}_K$ be von Neumann measurements and $\mathcal{P} = \{p_1, \ldots, p_K\}$ a probability distribution. Then the POVM $\mathcal{M} = \bigcup_i p_i \mathcal{V}_i$ has zero entanglement cost.*

$\mathcal{M}$ in the corollary has the physical interpretation of a probabilistic mixture of von Neumann measurements i.e. we first take a classical sample from $\mathcal{P}$ and then perform the corresponding von Neumann measurement. If $\mathcal{M} = \{|k_\mu\rangle\langle k_\mu|\}_{\mu=1}^m$ is any POVM on qubits (i.e. $d = 2$) then we may represent the elements $|k_\mu\rangle\langle k_\mu|$ as subnormalised Bloch vectors v_μ of length $r_\mu = \sqrt{\langle k_\mu|k_\mu\rangle}$ in the Bloch sphere. Then $\mathcal{M}$ is a mixture of von Neumann measurements iff the set of Bloch vectors is invariant under inversion through the centre of the sphere.

Note that we do not generally have equality in proposition 5. For example if $\mathcal{M} = \{|k_\mu\rangle\langle k_\mu|\}$ is any POVM in $d = 2$ define the POVM $\overline{\mathcal{M}}$ to be the measurement based on subnormalised states $|n_\mu\rangle$ where $|k_\mu\rangle$ and $|n_\mu\rangle$ are orthogonal and have equal lengths. Then generally $E_{\min}(\mathcal{M}) = E_{\min}(\overline{\mathcal{M}}) \ne 0$ yet $\frac{1}{2}\mathcal{M} \cup \frac{1}{2}\overline{\mathcal{M}}$ has zero cost, being a mixture of pairs of orthogonal directions i.e. a mixture of von Neumann measurements in $d = 2$.

The construction $\mathcal{M} = \bigcup_i p_i \mathcal{M}_i$ in proposition 5 corresponds to first making a classical probabilistic choice i and then applying the POVM corresponding to the seen choice. We now consider reversing the order of these operations. Let $\mathcal{M} = \{|k_\mu\rangle\langle k_\mu|\}_{\mu=1}^m$ be any POVM and let $\mathcal{P}^\mu = \{p_1^\mu, \ldots, p_{K(\mu)}^\mu\}$ be a classical probability distribution with $K(\mu)$ outcomes for each $\mu = 1, \ldots, m$. Consider the POVM $\mathcal{M}(\mathcal{P}^\mu)$ defined as follows: first apply $\mathcal{M}$. If the outcome μ is seen then sample $\mathcal{P}^\mu$ to get a value l. Output the pair (μ, l). In mathematical terms we take each element $|k_\mu\rangle\langle k_\mu|$ of $\mathcal{M}$ and replace it by $K(\mu)$ elements $p_1^\mu|k_\mu\rangle\langle k_\mu|, \ldots, p_{K(\mu)}^\mu|k_\mu\rangle\langle k_\mu|$ which are identical except for their normalisations. Geometrically for POVMs on qubits, we replace each Bloch vector v_μ by $K(\mu)$ vectors parallel to v_μ whose squared lengths sum to the squared length of v_μ.

Proposition 6 *For any POVM $\mathcal{M}$ and collection $\mathcal{P}^\mu$ of probability distributions as above we have*

$$E_{\min}(\mathcal{M}(\mathcal{P}^\mu)) \le E_{\min}(\mathcal{M}).$$

Thus if $\mathcal{V} = \bigcup_i q_i \mathcal{V}_i$ is any mixture of von Neumann measurements then $E_{\min}(\mathcal{V}(\mathcal{P}^\mu)) = 0$.

Proof Let $B = \{|w_\mu\rangle\}$ (with basic ancilla state $|a_0\rangle$) be any Naimark extension of $\mathcal{M}$ realising its minimal cost. For each μ let $|\alpha_1\rangle, \ldots, |\alpha_{K(\mu)}\rangle$ be an orthonormal set of states in a further auxiliary space with $\langle 0|\alpha_i\rangle = \sqrt{p_i^\mu}$ (where $|0\rangle$ is any choice of fixed state in the auxiliary space). To construct such a set of states we write the $\sqrt{p_i^\mu}$'s as the first column of a $K(\mu) \times K(\mu)$ matrix and fill out the rest of the matrix with orthonormal columns. Then the $|\alpha_i\rangle$'s are given (via components in a basis whose first member is $|0\rangle$) by the rows of the matrix. With this construction, $B' = \{|w_\mu\rangle|\alpha_i^\mu\rangle\}_{i,\mu}$ is clearly an orthonormal set and it provides a Naimark extension of $\mathcal{M}(\mathcal{P}^\mu)$ if we use basic ancilla state $|a_0\rangle|0\rangle$. Since the ancilla has been extended by product states the entanglement of $|w_\mu\rangle|\alpha_i^\mu\rangle$ is the same as that of $|w_\mu\rangle$ so the entanglement cost of $\mathcal{M}$ with B equals the entanglement cost of $\mathcal{M}(\mathcal{P}^\mu)$ with B' giving the result. $\square$

Hence if a POVM $\mathcal{M}$ can be viewed as a mixture of von Neumann measurements followed by a classical probabilistic choice (whose distribution may depend on the chosen von Neumann measurement's outcome) then $\mathcal{M}$ has zero cost.

Example The above characterisation does not generally exhaust all POVMs with zero cost. Consider the 5 outcome POVM $\mathcal{N}$ in $d = 3$ defined by

$$|k_1\rangle = |0\rangle \quad |k_2\rangle = \tfrac{1}{\sqrt{2}}|1\rangle \quad |k_3\rangle = \tfrac{1}{\sqrt{2}}|2\rangle$$
$$|k_4\rangle = \tfrac{1}{2}(|1\rangle + |2\rangle) \quad |k_5\rangle = \tfrac{1}{2}(|1\rangle - |2\rangle).$$

One may write down a Naimark extension of product states showing that $E_{\min}(\mathcal{N}) = 0$. However $\mathcal{N}$ is not a mixture of Von Neumann measurements (as the number of outcomes is not a multiple of d) and it is not of the form $\mathcal{M}(\mathcal{P}^\mu)$ for any $\mathcal{M}$ or $\mathcal{P}^\mu$ (as it has no parallel outcomes $|k_\mu\rangle$). But if we replace the element $|k_1\rangle$ by two elements $|k_{1,1}\rangle = |k_{1,2}\rangle = \frac{1}{\sqrt{2}}|0\rangle$ then the resulting 6 element POVM $\mathcal{N}'$ is a mixture of von Neumann measurements defined by $\{|k_{1,1}\rangle, |k_2\rangle, |k_3\rangle\}$ and $\{|k_{1,2}\rangle, |k_4\rangle, |k_5\rangle\}$. Thus $E_{\min}(\mathcal{N}') = 0$ and $\mathcal{N}$ arises by an identification of parallel elements. Later (after proposition 8) we will see that in the special case of $d = 2$, all POVMs having zero cost are of the form $\mathcal{V}(\mathcal{P}^\mu)$ of proposition 6 (i.e. where $\mathcal{V}$ is a mixture of von Neumann measurements).

Proposition 7 *Suppose that a POVM $\mathcal{M}' = \{|k_\mu\rangle\langle k_\mu|\}_{\mu=1}^m$ has zero entanglement cost and $|k_1\rangle$ and $|k_2\rangle$ are parallel. Let $|k_{12}\rangle\langle k_{12}| = |k_1\rangle\langle k_1| + |k_2\rangle\langle k_2|$. Then the rank 1 POVM $\mathcal{M} = \{|k_{12}\rangle\langle k_{12}|, |k_3\rangle\langle k_3|, \ldots, |k_m\rangle\langle k_m|\}$ (with $m - 1$ elements) has zero cost too.*

Proof Let $\{|w_\mu\rangle\}$ with basic ancilla state $|a_0\rangle$ be a Naimark extension of $\mathcal{M}'$ with zero cost. Thus each $|w_\mu\rangle$ for $\mu = 1, \ldots, K$ is a product state of the form

$$|w_\mu\rangle = |\hat{k}_\mu\rangle|\alpha_\mu\rangle$$

where $|\hat{k}_\mu\rangle$ is $|k_\mu\rangle$ normalised and $|\langle a_0|\alpha_\mu\rangle|^2 = \langle k_\mu|k_\mu\rangle$. Since $|\hat{k}_1\rangle = |\hat{k}_2\rangle$ we deduce that $|\alpha_1\rangle$ and $|\alpha_2\rangle$ are orthonormal. Let $|\alpha_{12}\rangle$ be the normalised projection of $|a_0\rangle$ into the plane of $|\alpha_1\rangle$ and $|\alpha_2\rangle$. Then $|\langle a_0|\alpha_{12}\rangle|^2 = |\langle a_0|\alpha_1\rangle|^2 + |\langle a_0|\alpha_2\rangle|^2$. Write $|w_{12}\rangle = |\hat{k}_1\rangle|\alpha_{12}\rangle$. Then the orthonormal set $\{|w_{12}\rangle, |w_3\rangle, \ldots\}$ is a Naimark extension for $\mathcal{M}$ with zero cost. $\square$

Above we have characterised a class of POVMs in any dimension d with zero cost. Next we characterise a class that is guaranteed to have non-zero cost and fully characterise zero cost POVMs in $d = 2$.

Proposition 8 *Let $\mathcal{M} = \{|k_\mu\rangle\langle k_\mu|\}_{\mu=1}^m$ be any POVM on $\mathcal{H}_d$ having zero entanglement cost. Then, for any μ we have the operator inequality*

$$|k_\mu\rangle\langle k_\mu| + \sum_{\nu:\langle k_\mu|k_\nu\rangle=0} |k_\nu\rangle\langle k_\nu| \geq \langle k_\mu|k_\mu\rangle\mathbf{1}. \tag{3}$$

Proof We may assume $\mu = 1$ without loss of generality. Let A be the set of integers ν satisfying $2 \leq \nu \leq m$ and $\langle k_1|k_\nu\rangle \neq 0$, and let $\bar{A} \equiv \{1,\ldots,m\} - A$. Note that the LHS of Eq. (3) is equal to $\sum_{\nu\in\bar{A}} |k_\nu\rangle\langle k_\nu|$. Since $\mathcal{M}$ has zero entanglement cost, it has a Naimark extension $\{|w_\nu\rangle\}_{\nu=1}^m$ in $\mathcal{H}_d \otimes \mathcal{H}_e$ of the form $|w_\nu\rangle = |\hat{k}_\nu\rangle|\alpha_\nu\rangle$, where $|\langle\alpha_\nu|a_0\rangle|^2 = \langle k_\nu|k_\nu\rangle$. For $\nu \in A$, $\langle w_1|w_\nu\rangle = 0$ implies $\langle\alpha_1|\alpha_\nu\rangle = 0$. The operator $P \equiv \sum_{\nu\in A} |w_\nu\rangle\langle w_\nu|$ is thus a projector onto a subspace of $\mathcal{H}_d \otimes (\mathcal{H}_e - |\alpha_1\rangle)$, and $P \leq \mathbf{1} \otimes (\mathbf{1} - |\alpha_1\rangle\langle\alpha_1|)$. Then we have $\sum_{\nu\in A} |k_\nu\rangle\langle k_\nu| = \langle a_0|P|a_0\rangle \leq \mathbf{1}(1 - |\langle\alpha_1|a_0\rangle|^2) = \mathbf{1}(1 - \langle k_1|k_1\rangle)$. Since $\sum_{\nu\in\bar{A}} |k_\nu\rangle\langle k_\nu| = \mathbf{1} - \sum_{\nu\in A} |k_\nu\rangle\langle k_\nu|$, we obtain Eq. (3). $\square$

This proposition states that for each element $|k_\mu\rangle$ in a zero-cost POVM, we can find elements orthogonal to $|k_\mu\rangle$ of a sufficient number to satisfy Eq. (3). Hence we get the following corollaries.

Corollary 2 *If a POVM $\mathcal{M}$ has an element $|k_\mu\rangle$ such that no other element $|k_\nu\rangle$ is orthogonal to $|k_\mu\rangle$ then $\mathcal{M}$ has non-zero entanglement cost.*

Corollary 3 *Let $\mathcal{M}$ be POVM in dimension $d = 2$. Then $\mathcal{M}$ has zero cost iff $\mathcal{M}$ is a mixture of von Neumann measurements followed by classical probabilistic choices i.e. $\mathcal{M}$ has the form $\mathcal{V}(\mathcal{P}^\mu)$ of proposition 6.*

Proof Suppose that $\mathcal{M}$ with $d = 2$ has zero cost. By identifying all parallel sets of elements of $\mathcal{M}$ we can construct a POVM $\mathcal{M}'$ satisfying $|\langle\hat{k}_\mu|\hat{k}_{\mu'}\rangle| \neq 1$ for $\mu \neq \mu'$. By proposition 7 $\mathcal{M}'$ has zero cost too. Since $d = 2$, Proposition 8 requires that any element $|k_\mu\rangle$ in $\mathcal{M}'$ should be paired to another element $|k_{\mu'}\rangle$ satisfying $\langle k_\mu|k_{\mu'}\rangle = 0$ and $\langle k_\mu|k_\mu\rangle = \langle k_{\mu'}|k_{\mu'}\rangle$. Hence $\mathcal{M}'$ is a mixture of von Neumann measurements $\mathcal{V}$ and $\mathcal{M}$ has the form $\mathcal{V}(\mathcal{P}^\mu)$. The converse implication is given by proposition 6. $\square$

4.2   The entanglement cost of the trine measurement

The minimisation needed to determine $E_{\min}(\mathcal{M})$ cannot generally be performed analytically except for some special POVMs. Amongst the latter is the trine measurement (defined below), which is perhaps the simplest generalised measurement that is not a von Neumann measurement.

The trine measurement is defined as a POVM $\{|\beta_\mu\rangle\langle\beta_\mu|\}_{\mu=1,2,3}$ on $\mathcal{H}_2$, where

$$\begin{aligned} \langle\beta_\mu|\beta_\mu\rangle &= 2/3 \\ |\langle\beta_\mu|\beta_{\mu'}\rangle| &= 1/3 \text{ for } \mu \neq \mu' \end{aligned} \tag{4}$$

The three directions $|\beta_\mu\rangle$ are equally spaced around a great circle on the Bloch sphere. Recall that this measurement is optimal for the trine source whose states are respectively orthogonal to the $|\beta_\mu\rangle$'s.

Theorem 2 *The entanglement cost of the trine measurement is $\frac{2}{3}H\left(\frac{1}{2}(1 - \frac{1}{\sqrt{3}})\right) = 0.496...$ ebits (where $H(x) = -x\log_2 x - (1-x)\log_2(1-x)$ is the binary entropy function).*

The proof of this theorem is given in the appendix.

4.3   Upper bounds on the cost of any measurement $\mathcal{M}$

While determining the minimum entanglement cost $E_{\min}(\mathcal{M})$ for a general POVM $\mathcal{M}$ is difficult, we can construct a Naimark extension having a fairly small entanglement cost as follows.

Proposition 9 *Let $\mathcal{M} = \{|k_\mu\rangle\langle k_\mu|\}_{\mu=1}^m$ be any POVM on $\mathcal{H}_d$. Then, for any normalized state $|\zeta\rangle$ in $\mathcal{H}_d$, there exists a Naimark extension $\{|w_\nu\rangle\}_{\nu=1}^{de}$ of $\mathcal{M}$ having the following form, with e at most $m+1$:*

$$|w_\nu\rangle = |k_\nu\rangle|a_0\rangle + |\zeta\rangle|\xi_\nu\rangle \ \text{ for } \nu = 1, \ldots, m \tag{5}$$

where $\{|\xi_\nu\rangle\}_{\nu=1}^m$ lie in the span of $\{|a_i\rangle : i \geq 1\}$ (denoted $\mathcal{H}_e - |a_0\rangle$) and $\{|w_\nu\rangle\}_{\nu=m+1}^{de}$ lie in $\mathcal{H}_d \otimes (\mathcal{H}_e - |a_0\rangle)$. This extension has entanglement cost

$$E = \sum_{\mu=1}^m \frac{r_\mu^2}{d} H\left(\frac{1}{2} - \frac{1}{2}\sqrt{\alpha_\mu^2 + (1 - \alpha_\mu^2)|\langle\zeta|\hat{k}_\mu\rangle|^2}\right), \tag{6}$$

where $r_\mu^2 = \langle k_\mu|k_\mu\rangle$, $|\hat{k}_\mu\rangle = r_\mu^{-1}|k_\mu\rangle$, and $\alpha_\mu = 1 - 2r_\mu^2$.

Proof Existence of a Naimark extension immediately assures us that there exists a set of m vectors $\{|\tilde{\xi}_\nu\rangle\}_{\nu=1}^m$ satisfying $\langle k_{\nu'}|k_\nu\rangle + \langle\tilde{\xi}_{\nu'}|\tilde{\xi}_\nu\rangle = \delta_{\nu\nu'}$. (Choose $|\tilde{\xi}_\nu\rangle \equiv |v_\nu^1\rangle|a_1\rangle + \ldots + |v_\nu^{(e-1)}\rangle|a_{(e-1)}\rangle$ in Proposition 1, for example.) If we take $e = m+1$, we can then find m vectors $\{|\xi_\nu\rangle\}_{\nu=1}^m$ lying in $\mathcal{H}_e - |a_0\rangle$, satisfying $\langle k_{\nu'}|k_\nu\rangle + \langle\xi_{\nu'}|\xi_\nu\rangle = \delta_{\nu\nu'}$. Using these vectors, define $\{|w_\nu\rangle\}_{\nu=1}^m$ through Eq. (5). The set $\{|w_\nu\rangle\}_{\nu=1}^m$ is then orthonormal, and can hence be extended to an orthonormal basis $\{|w_\nu\rangle\}_{\nu=1}^{de}$ of $\mathcal{H}_d \otimes \mathcal{H}_e$ with $\{|w_\nu\rangle\}_{\nu=m+1}^{de}$ being orthogonal to the subspace $\mathcal{H}_d \otimes |a_0\rangle$. Proposition 1 then guarantees that this basis is a Naimark extension of $\mathcal{M}$.

The state $|w_\nu\rangle$ ($\nu \leq m$) can be written in terms of normalized vectors as

$$r_\mu|\hat{k}_\mu\rangle|0\rangle + \sqrt{1 - r_\mu^2}|\zeta\rangle|1_\mu\rangle$$

where $\langle 0|1_\mu\rangle = 0$. Then the derivation of Eq. (6) is straightforward[4]. □

This construction shows that for any POVM on any size of Hilbert space, the minimum entanglement cost is less than 1 ebit (as $H(x) \leq 1$ and $\sum_\mu r_\mu^2/d = 1$) but we can derive more detailed upper bounds for $E_{\min}(\mathcal{M})$. For any POVM, we can assume $\langle k_1|k_1\rangle \geq \langle k_\mu|k_\mu\rangle (\mu = 2, \ldots, m)$ without loss of generality. Then, choose $|\zeta\rangle = |\hat{k}_1\rangle$ in the above construction. Since $|w_1\rangle$ has no entanglement and the other $\{|w_\nu\rangle\}$ have less than one ebit, $E \leq \sum_{\mu=2}^m r_\mu^2/d = 1 - r_1^2/d$ for this construction. Noting that $r_1^2 m \geq \sum_\mu r_\mu^2 = d$, we obtain an upper bound as a function of the number of elements m:

Corollary 4 *Let $\mathcal{M} = \{|k_\mu\rangle\langle k_\mu|\}_{\mu=1}^m$ be any POVM on $\mathcal{H}_d$. Then,*

$$E_{\min}(\mathcal{M}) \leq 1 - \frac{1}{m}.$$

We also have an upper bound as a function of d, by considering the expectation value $\overline{E}$ of E when $|\zeta\rangle$ is randomly chosen among all pure states in H_d. The expectation value of $|\langle \zeta | \hat{k}_\mu \rangle|^2$ is given by $\overline{|\langle \zeta | \hat{k}_\mu \rangle|^2} = 1/d$. We will use the following lemma.

Lemma 2 $H[(1 - \sqrt{x})/2]$ *is a concave function of x for $0 \le x \le 1$.*

Proof For $0 \le x < 1$, the first derivative of this function can be expanded as

$$-(\ln 4)^{-1} \sum_{m=0}^{\infty} x^m/(2m+1)$$

which is obviously monotone decreasing. □

Using this lemma and Eq. (6), we have

$$\overline{E} \le \sum_{\mu=1}^{m} \frac{r_\mu^2}{d} H\left(\frac{1}{2} - \frac{1}{2}\sqrt{\alpha_\mu^2 + (1 - \alpha_\mu^2)\overline{|\langle \zeta | \hat{k}_\mu \rangle|^2}}\right) \le \sum_{\mu=1}^{m} \frac{r_\mu^2}{d} H\left[\frac{1}{2}\left(1 - \frac{1}{\sqrt{d}}\right)\right].$$

Since $E_{\min}(\mathcal{M})$ is not larger than this averaged value, we obtain the following upper bound.

Theorem 3 *Let $\mathcal{M} = \{|k_\mu\rangle\langle k_\mu|\}_{\mu=1}^{m}$ be any POVM on $\mathcal{H}_d$. Then,*

$$E_{\min}(\mathcal{M}) \le H\left[\frac{1}{2}\left(1 - \frac{1}{\sqrt{d}}\right)\right].$$

Some values of this upper bound are given in the following table.

d	$H\left[\frac{1}{2}\left(1 - \frac{1}{\sqrt{d}}\right)\right]$
2	0.600876
3	$0.744008 = 0.469417 \log_2 3$
4	$0.811278 = 0.405639 \log_2 4$

5　The most costly generalised measurement

To any (discrete) POVM $\mathcal{M}$ we have assigned its minimal entanglement cost $E_{\min}(\mathcal{M})$ and it is interesting to ask: which POVM $\mathcal{M}$ (for any d) has the *largest* such cost? Furthermore we may prefer to scale the cost by $\log_2 d$, the number of qubits in the source space and ask for $\mathcal{M}$ that maximises the "normalised entanglement cost" $F(\mathcal{M}) = E_{\min}(\mathcal{M})/\log_2 d$ i.e. the POVM which requires the maximal entanglement investment per qubit of source support.

According to theorem 3, for source dimension d we have

$$F(\mathcal{M}) \le \frac{1}{\log_2 d} H\left[\frac{1}{2}(1 - \frac{1}{\sqrt{d}})\right].$$

This upper bound is a decreasing function of d with value 0.46... for $d = 3$. Recalling that the trine measurement has minimum cost 0.49... in $d = 2$ we see that $F(\mathcal{M})$ is maximal for $d = 2$ and $\mathcal{M}$ must have entanglement cost at least as large as the trine measurement. Although we have been unable to exhaustively search the POVM space of $d = 2$ (either numerically or analytically), we do have a numerical scheme for evaluating the entanglement cost of any POVM in $d = 2$ under the restriction that the ancilla has dimension 2. Using this scheme we have found that no other POVM with $d = 2$ and $e = 2$ has a cost as large as the trine cost.

Details of the numerical scheme and further numerical studies will be given in a forthcoming publication[9].

Hence we have shown:

Theorem 4 *The trine measurement is the unique POVM with maximal normalised entanglement cost* $E_{\min}(\mathcal{M})/\log_2 d$ *amongst all POVMs with*
(a) $d \geq 3$, *or*
(b) $d = 2$ *and ancilla size* $e = 2$.
This provides strong evidence for

Conjecture The trine measurement is the unique POVM with maximal normalised entanglement cost, amongst all (discrete) POVMs.

To decide the conjecture it remains only to treat the case of $d = 2$ with $e \geq 3$ (including POVMs with an unboundedly large number of outcomes).

6 Asymptotic entanglement cost

So far we have been considering the entanglement cost of a single instance of a POVM $\mathcal{M}$. In quantum information theory it is often advantageous and interesting to consider an asymptotic scenario in which information is collected from more and more instances of the target system and we assess the minimal cost per system that results from allowing global operations on all instances jointly. In this spirit, for a POVM $\mathcal{M}$ we may consider the entanglement cost of $n\mathcal{M}$, the POVM defined by applying $\mathcal{M}$ n times. We introduce the *asymptotic entanglement cost of* $\mathcal{M}$ as being $\lim_{n\to\infty} E_{\min}(n\mathcal{M})/n$. If we implement $n\mathcal{M}$ simply as n independent instances of $\mathcal{M}$ then the resulting entanglement cost per measurement will be $nE(\mathcal{M})/n = E(\mathcal{M})$. However if we consider global Naimark extensions of the single POVM defined by $n\mathcal{M}$ on $\mathcal{H}_d^{\otimes n}$ then theorem 3 immediately gives:

Theorem 5 *For any POVM the asymptotic entanglement cost is zero.*

Further discussion of the significance of this result is given below in section 7.2.

7 Concluding remarks

7.1 Other formulations?

Our notion of entanglement cost – based on the entanglements of the Naimark extension states – is "retrospective" in the sense that it quantifies the entanglement *produced* by the measurement. In contrast it would be interesting to develop a quantitative notion of "active" cost i.e. the entanglement *input* required to *perform* the POVM, in some suitable sense. A similar situation occurs in the question of quantifying the amount of entanglement in a bipartite mixed state. We have E_D, the entanglement of distillation i.e. the amount of entanglement that can be produced, analogous to our notion of POVM cost, and E_F, the entanglement of formation i.e. the amount of entanglement that needs to be invested to prepare the state. Just as $E_D \leq E_F$ for all states, we would expect that our $E_{\min}(\mathcal{M})$ would be a lower bound on any reasonable notion of input entanglement cost required to implement the POVM.

There are various possible approaches to formulating a notion of an input entanglement cost. For example, we could demand that the only allowed measurement is a standard product von Neumann measurement of the form $\{|e_j\rangle|l\rangle\}$ on $\mathcal{H}_d \otimes \mathcal{H}_e$ (for any desired ancilla space with orthonormal basis $\{|l\rangle\}$). Then to implement a POVM $\mathcal{M}$ we first apply a unitary

transformation U on $\mathcal{H}_d \otimes \mathcal{H}_e$ so that the measurement of $\mathcal{M}$ on $|\psi_i\rangle|a_0\rangle$ is equivalent to the product measurement on $|\tilde{\psi}_i\rangle = U|\psi_i\rangle|a_0\rangle$. Finally we compute the entanglement E_i of $|\tilde{\psi}_i\rangle$ and ask for the minimal average $E = \sum p_i E_i$, minimised over all choices of Naimark extension for $\mathcal{M}$. The average here is constructed using the prior probabilities of the signal states. Although this approach is appealing – in that we explicitly consider the amount of entanglement E_i that needs to be "put into" the source states – it is unfortunately trivial as the answer is always $E_i^{(\min)} = 0$! To see this, note that we can always begin the action of U on $|\psi_i\rangle|a_0\rangle$ by swapping the signal state out entirely into the ancilla and leaving the signal space in a standard state $|s_0\rangle$ (independent of i). Then any generalised measurement on the source can be simulated by subsequent actions entirely in $\mathcal{H}_e$ with the system plus ancilla always being in a product state, having $|s_0\rangle$ in the signal space i.e. $E_i = 0$. (This kind of trivial entanglement–avoiding construction was also noticed in [8]).

To avoid this problem of the minimal E being identically zero, we could instead ask about the "amount of coupling" needed to implement the unitary operation U on $\mathcal{H}_d \otimes \mathcal{H}_e$. There are a number of possible meanings to "amount of coupling", and this is a subject of considerable current interest [10]. For example one can talk about the maximum amount of entanglement that can be produced by U or the amount of entanglement needed to implement U. Or one can consider how long it takes to implement U using a fixed Hamiltonian. Interestingly, in the context of the above discussion, the swap operation on $\mathcal{H}_d \otimes \mathcal{H}_d$ typically has a *high* cost for these quantities. Thus the method in the previous paragraph for having $E_i = 0$ (making essential use of a swap operation) would become expensive (and would presumably no longer be a minimal strategy). However progress in understanding non-local aspects of bi-partite interactions has not yet reached the stage where the possible measures of interaction are well understood, or indeed straightforward to calculate for typical operations. Therefore further development of this approach must await the introduction of a (suitably tractable) definition of interaction cost for bipartite transformations.

Although we do not have a general notion of entanglement cost of a bipartite unitary operation, it is interesting to note that a definition of entanglement cost for a POVM, based on such a notion, would be expected to be *inequivalent* to the definition adopted in this paper. To see this consider the product basis of two qubits given by:

$$|\psi_0\rangle = |0\rangle|0\rangle \quad |\psi_1\rangle = |1\rangle|0\rangle \quad |\psi_2\rangle = |+\rangle|1\rangle \quad |\psi_3\rangle = |-\rangle|1\rangle$$

where $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$. If we take the initial ancilla state to be $|a_0\rangle = |+\rangle$ then this product basis is actually a Naimark extension of the 4 element POVM $\{|k_\mu\rangle\langle k_\mu|\}$ with

$$|k_1\rangle = \frac{1}{\sqrt{2}}|0\rangle \quad |k_2\rangle = \frac{1}{\sqrt{2}}|1\rangle \quad |k_3\rangle = \frac{1}{\sqrt{2}}|+\rangle \quad |k_4\rangle = \frac{1}{\sqrt{2}}|-\rangle. \tag{7}$$

(This POVM may be interpreted as the 50/50 probabilistic mixture of two von Neumann measurements $\{|0\rangle, |1\rangle\}$ and $\{|+\rangle, |-\rangle\}$: we first measure the ancilla $|a_0\rangle$ in the basis $\{|0\rangle, |1\rangle\}$ (cf the second qubit of the $|\psi_i\rangle$ states) and then we choose the von Neumann measurement for the first qubit according to whether the outcome is 0 or 1.) Since we have a Naimark extension of product states this POVM has zero entanglement cost by our adopted definition. However it can be shown that no product unitary transformation on two qubits can transform the bi-orthogonal product basis $\{|j\rangle|l\rangle\}_{j,l=0}^1$ to the basis $\{|\psi_i\rangle\}_{i=0}^3$. i.e. the POVM would be expected

to require interaction between system and ancilla and thus have non-zero entanglement cost in the new sense. It is interesting to note that if, as a form of interaction, we allow free classical communication between system and ancilla then the POVM in eq. (7) again has zero cost (using the protocol described after eq. (7), with the ancilla measurement result being communicated to the system).

7.2 Entanglement or space?

Although we have introduced a notion of entanglement cost for POVMs, theorem 5 indicates that in an asymptotic sense, entanglement between system and ancilla is not important since the (suitably normalised) average entanglement of the measurement outcomes is zero. This leads to the question of what role the ancilla actually plays i.e. we wish to characterise a resource that is being consumed. If the number M of POVM elements exceeds the dimension d of the source space, then in the asymptotic scenario of n applications of $\mathcal{M}$, the Naimark extension requires M^n orthogonal directions which is exponentially larger than the source dimension d^n. Here the ancilla provides the extra space. Hence we are lead to the intriguing suggestion that "unentangled space" itself might be a quantifiable and useful resource in quantum information processing. Further exploration of this idea will be given in a future paper.

One may also consider a similar asymptotic formulation in the context of extracting information from sources. Let $\mathcal{E}$ be a source in dimension d and let $\mathcal{E}^{\otimes n}$ be the source (in dimension d^n) whose signal states are blocks of n independent emissions from $\mathcal{E}$ (with prior probabilities for $\mathcal{E}^{\otimes n}$ being products of those from $\mathcal{E}$). The accessible information $\mathrm{Acc}(\mathcal{E}^{\otimes n})$ of $\mathcal{E}^{\otimes n}$ is determined by optimising the mutual information over all POVMs in dimension d^n and it is known that $\mathrm{Acc}(\mathcal{E}^{\otimes n}) = n\mathrm{Acc}(\mathcal{E})$. But to focus on the role of the ancilla we introduce firstly, $\mathrm{Acc}^{\perp}(\mathcal{E}^{\otimes n})$, the maximal mutual information obtainable by applying only von Neumann measurements in dimension d^n (i.e. no ancilla needed) and secondly, $\mathrm{Acc}^0(\mathcal{E}^{\otimes n})$, the maximal mutual information obtainable by applying only POVMs in dimension d^n that have zero entanglement cost (i.e. an ancilla is needed but it remains unentangled with the source). Then we can ask: in the limit of large n, does either $\mathrm{Acc}^{\perp}(\mathcal{E}^{\otimes n})/n$ or $\mathrm{Acc}^0(\mathcal{E}^{\otimes n})/n$ approach $\mathrm{Acc}(\mathcal{E})$? In [11] a related question has been considered, where the object function is not the mutual information but a measure of success probability of identifying some desired property of the signal states. In this context it was shown that we may choose optimal POVMs that resemble von Neumann measurements more and more, in the asymptotic limit.

Acknowledgements

References

1. Peres, A. (1993) *Quantum theory: concepts and methods* (especially §9.6) Kluwer academic publishers.
2. Naimark, M. A. (1940) *Izv. Akad. Nauk SSSR, Ser. Mat.* **4**, p277-318.

3. Davies, E. B. (1978) *IEEE Trans. Inform. Theory* **IT24**, p596-599.
4. Levitin, L. (1995) *Proc. Quantum communication and measurement*, ed. V. Belavkin, O. Hirota and R. Hudson, p439-448, Plenum Press, New York.
5. Holevo, A. (1973) *Prob. Peredachi Inform.* **9** no. 2, p31-42.
6. Sasaki, M., Barnett, S., Jozsa, R., Osaki, M. and Hirota, O. (1999) *Phys. Rev.* **A59**, p3325-3335.
7. Preskill, J. (1998) *Lecture Notes for Physics 229: Quantum Information and Computation*, California Institute of Technology.
8. Busch, P. (2002) *Note on non-entangling measurements*, preprint available at quant-ph/0209090.
9. Presnell, S. (2003) Ph.D. thesis, in preparation.
10. Cirac, J.I., Dür, W., Kraus, B. and Lewenstein, M.(2001) *Phys Rev Lett* **86**, 544; Collins, D., Linden, N. and Popescu, S. (2001) *Phys Rev* **A 64**, 032302; Eisert, J., Jacobs, K., Papadopoulos, P. and Plenio, M. (2001) *Phys Rev* **A 62**, 052317; Leifer, M., Henderson, L. and Linden, N. (2003) *Phys Rev* **A67**, 012306; Bennett, C., Harrow, A., Leung, D. and Smolin, J. (2002) quant-ph/0205057; Khaneja, N., Brockett, R. and Glaser, S. (2001) *Phys Rev* **A63**, 032308; Bennett, C., Cirac, J. I., Leifer, M., Leung, D., Linden, N., Popescu, S. and Vidal, G. (2002) *Phys Rev* **A66**, 012305.
11. Massar, S. and Popescu, S. (2000) *Phys. Rev.* **A61**, p062303; Winter, A. and Massar, S. (2001) *Phys. Rev.* **A64**, p012311.

Appendix: proof of theorem 2

Any Naimark extension $\{w_\nu\}_{\nu=1}^{de}$ of the trine POVM $\{|\beta_\mu\rangle\langle\beta_\mu|\}_{\mu=1,2,3}$ in $\mathcal{H}_2 \otimes \mathcal{H}_e$ has the following form

$$|w_\mu\rangle = |\beta_\mu\rangle|a_0\rangle + \sqrt{\frac{1}{3}}|\chi_\mu\rangle \ \text{ for } \ \mu = 1,2,3,$$

where $\{|\chi_\mu\rangle\}_{\mu=1,2,3}$ and $\{|w_\nu\rangle\}_{\nu=4}^{de}$ lie in $\mathcal{H}_2 \otimes (\mathcal{H}_e - |a_0\rangle)$. The orthonormality of $\{|w_\nu\rangle\}_{\nu=1}^3$ leads to $|\langle\chi_\mu|\chi_{\mu'}\rangle| = 1$ for any $1 \le \mu, \mu' \le 3$, implying that $\{|\chi_\mu\rangle\}_{\mu=1,2,3}$ are the same except for their phase factors, and the reduced state of $|\chi_\mu\rangle\langle\chi_\mu|$ on $\mathcal{H}_2$ is independent of μ. Let us take an orthonormal basis $\{|0\rangle, |1\rangle\}$ of $\mathcal{H}_2$ such that this common density operator $\mathrm{tr}_e(|\chi_\mu\rangle\langle\chi_\mu|)$, having rank 2, is given by $p|0\rangle\langle0| + [(1-p)/2]\mathbf{1}$.

The average entanglement E for this extension is

$$E = \sum_\mu \frac{1}{3}S(\rho_\mu),$$

where

$$\rho_\mu \equiv \frac{2}{3}|\hat{\beta}_\mu\rangle\langle\hat{\beta}_\mu| + \frac{p}{3}|0\rangle\langle0| + \frac{1-p}{6}\mathbf{1}$$

and $|\hat{\beta}_\mu\rangle$ is $|\beta_\mu\rangle$ normalized. Let us first consider the dependence of E on p, with $\{|\beta_\mu\rangle\}$ fixed. If we define density operators $\tilde{\rho}_\mu^{(p)}$ by

$$\tilde{\rho}_\mu^{(p)} \equiv \frac{2}{3}|0\rangle\langle0| + \frac{p}{3}|\hat{\beta}_\mu\rangle\langle\hat{\beta}_\mu| + \frac{1-p}{6}\mathbf{1},$$

it is obvious that $S(\rho_\mu) = S(\tilde{\rho}_\mu^{(p)})$. Noting that

$$\rho_1^{(p)} = \frac{1+2p}{3}\rho_1^{(1)} + \frac{1-p}{3}\rho_2^{(1)} + \frac{1-p}{3}\rho_3^{(1)}$$

and similar expressions for $\rho_2^{(p)}$ and $\rho_3^{(p)}$, we obtain $\sum_\mu S(\tilde{\rho}_\mu^{(p)}) \geq \sum_\mu S(\tilde{\rho}_\mu^{(1)})$ from the concavity of S. Hence, in finding the minimum of E, we only need to consider the case $p = 1$. In the following we assume that $p = 1$.

Let us introduce Pauli operators on $\mathcal{H}_2$ defined by $\sigma_x \equiv |1\rangle\langle 0| + |0\rangle\langle 1|$, $\sigma_y \equiv i|1\rangle\langle 0| - i|0\rangle\langle 1|$, and $\sigma_z \equiv |0\rangle\langle 0| - |1\rangle\langle 1|$. Define Bloch vectors for $|\hat{\beta}_\mu\rangle$ by $\boldsymbol{p}^{(\mu)} \equiv \langle\hat{\beta}_\mu|\boldsymbol{\sigma}|\hat{\beta}_\mu\rangle$, where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$. The eigenvalues of ρ_μ are then given by

$$\frac{1}{2} \pm \frac{\sqrt{4p_z^{(\mu)} + 5}}{6}$$

where

$$p_z^{(\mu)} = 2|\langle 0|\hat{\beta}_\mu\rangle|^2 - 1.$$

We can write E as a function of $(p_z^{(1)}, p_z^{(2)}, p_z^{(3)})$ as

$$E(p_z^{(1)}, p_z^{(2)}, p_z^{(3)}) = \sum_\mu \frac{1}{3} f(p_z^{(\mu)}),$$

where

$$f(z) \equiv H\left(\frac{1}{2} + \frac{\sqrt{4z + 5}}{6}\right).$$

As can be seen from Lemma 2 (in section 4.3), $f(z)$ is concave and hence $E(p_z^{(1)}, p_z^{(2)}, p_z^{(3)})$ is concave as a function of the triplet $(p_z^{(1)}, p_z^{(2)}, p_z^{(3)})$.

Next, let us consider the possible range of the triplet $(p_z^{(1)}, p_z^{(2)}, p_z^{(3)})$. From Eq. (4), the Bloch vectors $\boldsymbol{p}^{(i)}$ have unit length and they are equally spaced around a great circle on the Bloch sphere. Hence we can derive the following:

$$\boldsymbol{p}^{(1)} + \boldsymbol{p}^{(2)} + \boldsymbol{p}^{(3)} = \boldsymbol{0} \tag{8}$$

$$|\boldsymbol{p}^{(\mu)}|^2 = 1 \tag{9}$$

$$|\boldsymbol{p}^{(1)} - \boldsymbol{p}^{(2)}|^2 = 2 - 2\cos(2\pi/3) = 3 \tag{10}$$

$$(\boldsymbol{p}^{(1)} - \boldsymbol{p}^{(2)}) \cdot \boldsymbol{p}^{(3)} = 0 \tag{11}$$

We will use the following Lemma.

Lemma 3 *If* $\boldsymbol{p} \cdot \boldsymbol{q} = 0$,

$$\frac{p_z^2}{|\boldsymbol{p}|^2} + \frac{q_z^2}{|\boldsymbol{q}|^2} \leq 1.$$

Proof Let us decompose each of the two vectors into the z component and the xy-plane component *viz.* $\boldsymbol{p} = p_z\boldsymbol{z} + \boldsymbol{p}_\perp$ and $\boldsymbol{q} = q_z\boldsymbol{z} + \boldsymbol{q}_\perp$. Then, $\boldsymbol{p} \cdot \boldsymbol{q} = p_zq_z + \boldsymbol{p}_\perp \cdot \boldsymbol{q}_\perp = 0$, implying that $(p_zq_z)^2 \leq |\boldsymbol{p}_\perp|^2|\boldsymbol{q}_\perp|^2 = (|\boldsymbol{p}|^2 - p_z^2)(|\boldsymbol{q}|^2 - q_z^2)$. $\square$

Applying this Lemma to $\boldsymbol{p}^{(1)} - \boldsymbol{p}^{(2)}$ and $\boldsymbol{p}^{(3)}$, we have

$$(p_z^{(1)} - p_z^{(2)})^2/3 + (p_z^{(3)})^2 \leq 1. \tag{12}$$

From Eq. (8), we also have

$$p_z^{(1)} + p_z^{(2)} + p_z^{(3)} = 0. \tag{13}$$

If we change the variables by

$$
\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} p_z^{(1)} \\ p_z^{(2)} \\ p_z^{(3)} \end{pmatrix},
$$

then the conditions (13) and (12) become

$$
Z = 0 \tag{14}
$$
$$
X^2 + Y^2 \le 3/2. \tag{15}
$$

The region specified above is a convex set, with the extreme points forming a circle. Note that relabeling of $(1,2,3)$ in Eq. (12) merely reproduces the same conditions. Let us parameterize the extreme points by an angle θ, such that $X = \sqrt{3/2}\cos(\theta + \pi/6)$, $Y = \sqrt{3/2}\sin(\theta + \pi/6)$, and $Z = 0$. Then E is written as

$$
E(\theta) = \frac{1}{3} \sum_{k=-1,0,1} f\left(\cos(\theta + \frac{2k\pi}{3}) \right)
$$

Since $E(\theta + 2\pi/3) = E(\theta)$ and $E(-\theta) = E(\theta)$,

$$
\min_\theta E(\theta) = \min_{0 \le \theta \le \pi/3} E(\theta).
$$

Now we can numerically check that $E'(\theta) \le 0$ for $0 \le \theta \le \pi/3$. This implies

$$
\min_{0 \le \theta \le \pi/3} E(\theta) = E(0) = \frac{2}{3} H\left(\frac{1}{2}(1 - \frac{1}{\sqrt{3}}) \right) = 0.496..
$$

Since $E(p_z^{(1)}, p_z^{(2)}, p_z^{(3)})$ is concave,

$$
\min E(p_z^{(1)}, p_z^{(2)}, p_z^{(3)}) \ge \min_\theta E(\theta)
$$

Since there exists a valid point $(p_z^{(1)}, p_z^{(2)}, p_z^{(3)})$ corresponding to $\theta = 0$ (e.g., take $|\psi\rangle = |\hat{\beta}_1\rangle$), we conclude that

$$
\min E = \frac{2}{3} H\left(\frac{1}{2}(1 - \frac{1}{\sqrt{3}}) \right) = 0.496.. \qquad \square
$$