Converses from non-signalling codes and their relationship to converses from hypothesis testing

Author(s): Matthews, William

Publication Date: 2016

Permanent Link: https://doi.org/10.3929/ethz-a-010645629

Rights / License: In Copyright - Non-Commercial Use Permitted
Converses from non-signalling codes and their relationship to converses from hypothesis testing

William Matthews
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA.
Email: will@northala.net

Abstract—Finite blocklength converses for classical and quantum channel coding can be obtained by relaxing the optimization over independent encoding and decoding procedures to procedures which are merely “non-signalling”. This approach, inspired by quantum information theory, results in converses which are closely related to the hypothesis testing-based converse of Polyaniski-Poor-Verdu. Indeed, in the classical case they are equivalent. I will give an overview of the non-signalling codes method and describe its relationship to the hypothesis testing approach.

I. LINEAR TRANSFORMATIONS OF CONDITIONAL DISTRIBUTIONS

Consider the following situation. Given a symbol $M$, Alice applies some, possibly randomised, process to produce symbols $X$ and $F$. $F$ is sent to Bob over a noiseless channel. Alice uses $X$ as the input to some discrete channel, from which Bob receives output $Y$. Bob applies some process to $F$ and $Y$ to obtain a symbol $W$. We assume that $M=(F,X)\rightarrow (F,Y)\rightarrow W$ is a Markov chain. Letting $N(y|x)=P_{Y|X}(y|x)$, $E(f,x|m)=P_{F|X|M}(f|x,m)$, and $D(w|f,y)=P_{W|FY}(w|f,y)$, we have

$$P_{WYM}(w,y,x,m)=Z(x,w|m,y)N(y|x)P_{M}(m)$$

where

$$Z(x,w|m,y):=\sum_{f}D(w|f,y)E(f,x|m).$$

The conditional distribution (2) is what one would have if $N(y|x)=Q_{Y}(y)$ for some distribution $Q$. It is non-signalling from Bob to Alice, which means

$$\forall x,m,y,y':\sum_{w}Z(x,w|m,y)\sum_{y'}Z(x,w|m,y')=Z(x|m).$$

in particular

$$\forall x,m,y:\sum_{w}Z(x,w|m,y)=\sum_{f}E(f,x|m).$$

Conversely, any bipartite conditional distribution which is non-signalling from Bob to Alice has a (non-unique) decomposition of the form (2) (see [9]). Operationally, this means that it can be implemented by local operations and one-way communication from Alice to Bob. Note that $P_{WYM}$ depends on $E$ and $D$ only through the distribution $Z$.

The distribution of $W$ given $M$ in the present scenario is

$$P_{W|M}(w|m)=\sum_{x,y}Z(x,w|m,y)N(y|x).$$

Clearly, any linear transformation which takes conditional distributions for $Y$ given $X$ to conditional distributions for $W$ given $M$ can be written in the form (5) if we allow $Z(x,w|m,y)$ to be arbitrary numbers. In fact, the map will have the property that it transforms every conditional distribution to a conditional distribution if and only if $Z$ is a conditional distribution which is non-signalling from Bob to Alice (see [9]).

Naturally, we can write

$$P_{WYM}(w,y,x,m)=\hat{Z}(m,w|x,y)P_{XY}(x,y),$$

where, for $x$ such that $P_{X}(x)>0$ we define

$$\hat{Z}(m,w|x,y):=P_{M|XY}(m,w|x,y)$$

$$=P_{WYM}(w,y,x,m)P_{YX}(x)P_{W}(w|x)$$

$$=Z(x,w|m,y)P_{M}(m)P_{X}(x).$$

The final equality follows from (1). Note that $P_{X}(x)=\sum_{m,n}Z_{n,m}(x|m)P_{M}(m)$, so $\hat{Z}$ depends only on $Z$ and $P_{M}$ (not on $N_{Y|X}$). For $x$ such that $P_{X}(x)=0$ we let

$$\hat{Z}(m,w|x,y):=P_{M}(m)P_{W}(w).$$

It follows that

$$P_{M}(w)=\sum_{m,w}Z(m,w|x,y)\sum_{x}N_{Y|X}(y|x)P_{X}(x).$$

We will make use of this expression in the next section, and give a quantum generalisation of it in Section V. Note that $\hat{Z}(m,w|x,y)$ is non-signalling from Bob to Alice.

II. NON-SIGNALLING CODES

We can regard channel coding as a special case of the scenario described in the previous section. Let $M$ and $W$ take values in the same set of size $k$. We can interpret $M$ as the message and $W$ as the estimate of that message made by the decoder. Let $M$ be uniformly distributed. The average probability of error is $Pr(M\neq W)$. With the arbitrary noiseless communication from Alice to Bob allowed in the
previous section one can obviously find zero-error codes of arbitrary size for any channel \( N_{Y|X} \). A conventional code corresponds to the situation where

\[
Z(x, w|m, y) = E(x|m)D(w|y).
\]

(12)

For these types of code, \( Z \) is non-signalling not only from Bob to Alice but also from Alice to Bob, that is

\[
\forall w, y, m, m' : \sum_x Z(x, w|m, y) = \sum_x Z(x, w|m', y) =: Z_{W|Y}(w|y).
\]

(13)

We call any code with this property a non-signalling code [7].

The condition (13) implies that \( \hat{Z} \) satisfies

\[
\sum_x \hat{Z}(m, w|x, y)P_X(x) = P_M(m)Z_{W|Y}(w|y)
\]

(14)

and, if \( \hat{Z} \) satisfies this condition then the corresponding \( Z \) is non-signalling from Alice to Bob. The success probability of the code for channel \( N_{Y|X} \) is

\[
Pr(M = W) = \sum_{m,x,y} Z(m, m|x, y)N_{Y|X}(y|x)P_M(m)
\]

(15)

\[
= \sum_{m,x,y} \hat{Z}(m, m|x, y)N_{Y|X}(y|x)P_X(x).
\]

(16)

Remark 1. Fixing \( N_{Y|X} \), the success probability (15) is a linear functional of \( Z \) and, since the constraints which make \( Z \) non-signalling are linear, maximising the success probability over all non-signalling codes is a linear program.

Using symmetry, this can be simplified to one whose size is independent of \( k \) [9].

If we use a non-signalling code and take a channel \( R_{Y|X} \) where \( Y \) and \( X \) are independent, i.e. \( R_{Y|X}(y|x) = Q_Y(y) \) then, using (14), the distribution of \((M, W)\) is

\[
Q_{MW}(m, w) = \sum_{x,y} \hat{Z}(m, w|x, y)P_X(x)Q_Y(y)
\]

(17)

\[
= P_M(m) \sum_y Z_{W|Y}(w|y)Q_Y(y).
\]

(18)

that is \( W \) and \( M \) are independent. In this situation, for any choice of \( Q_Y \), \( Pr(M = W) = 1/k \), that is

\[
\forall Q_Y : \sum_{m,x,y} \hat{Z}(m, m|x, y)P_X(x)Q_Y(y) = 1/k.
\]

(19)

III. HYPOTHESIS TESTING CONVERSE

Consider the following hypothesis testing problem. The null hypothesis is that \( X \) and \( Y \) are distributed according to \( P_{XY} \). The alternative hypothesis is a composite hypothesis, which states that \( X \) and \( Y \) are distributed according to \( P_{X|Q_Y} \) for some arbitrary \( Q_Y \). A hypothesis test is specified by

\[
T[x, y] := Pr(\text{Accept null}|X = x, Y = y).
\]

(20)

The minimum type-II error which can be attained by a test with type-I error no more than \( \epsilon \) is

\[
\beta^*_\epsilon(P_{XY}) := \min_{Q_Y} \max_{\epsilon > 0} \sum_{x,y} T[x, y] P_X(x)Q_Y(y)
\]

subject to

\[
\sum_{y \neq y} T[x, y] P_X(x) \geq 1 - \epsilon.
\]

(22)

Let us define for distributions \( p \) and \( q \),

\[
\beta_\epsilon(p, q) := \min_{T} \left\{ \sum_x T[z]q(z) : \sum_x T[z]p(z) \geq 1 - \epsilon \right\}.
\]

The set of distributions for \( Y \) and the set of tests are both compact, convex sets and the objective function on the RHS of (21) is a bilinear function of the distribution and test. Therefore, by von Neumann’s minimax theorem

\[
\beta^*_\epsilon(P_{XY}) = \max_{Q_Y} \beta_\epsilon(P_{XY}, \|P_XQ_Y\|).
\]

(24)

Proposition 2. There is a non-signalling code of size \( k \), input distribution \( P_X \), and error probability \( \epsilon \) for channel \( N_{Y|X} \) if and only if there is a test \( T \) with

\[
\sum_{x,y} T[x, y] N_{Y|X}(y|x)P_X(x) = 1 - \epsilon, \quad \text{and} \quad \forall Q_Y : \sum_{x,y} T[x, y]Q_Y(y)P_X(x) = 1/k.
\]

(25)

(26)

Proof. Suppose that we have a non-signalling code of size \( k \) which attains error probability \( \epsilon \) for channel \( N_{Y|X} \). The distribution of \( X \) is fixed by \( Z \) and the fact that \( M \) is uniformly distributed. For the direct part, let \( Z \) be the bipartite conditional distribution for a non-signalling code satisfying the stated properties. If we let

\[
T[x, y] = \sum_{m=1}^k \hat{Z}(m, m|x, y),
\]

(27)

then using (16) we obtain (25) and, using (19) in addition, we obtain (26).

For the converse, let \( T \) be a test satisfying (25) and (26), and let

\[
\hat{Z}(m, m|x, y) = \frac{1}{k} \delta_{mw} T[x, y]
\]

(28)

\[
\frac{1}{k(k-1)} \left( 1 - \delta_{mw} \right) \left( 1 - T[x, y] \right).
\]

This clearly satisfies (3). Using (26) we have

\[
\sum_x \hat{Z}(m, m|x, y)P_X(x) = \frac{1}{k} \delta_{mw} \sum_x T[x, y]P_X(x)
\]

(29)

\[
+ \frac{1}{k(k-1)} \left( 1 - \sum_x T[x, y]P_X(x) \right) = 1/k^2
\]

(30)

so \( \hat{Z} \) also satisfies (14). It follows that \( Z \) satisfies (3) and (13), so it is a non-signalling code. Furthermore, by (25),

\[
Pr(M = W) = \sum_{m,x,y} \hat{Z}(m, m|x, y)N_{Y|X}(y|x)P_X(x)
\]

(31)

\[
= 1 - \epsilon.
\]
A constraint on tests of the form (26) is a rather unusual
in the context of hypothesis testing. In [10], tests with this
property (or more generally, property (35)) are called “$P_X$
balanced”), and as noted there, we may relax this condition
without changing the minimax type-II error probability: Suppose we have a test $T'$ which satisfies
\[ \sum_{x,y} T'(x,y)N_{Y|X}(y|x)P_X(x) \geq 1 - \epsilon, \] and
\[ \forall y : \sum_{x,y} T'(x,y)Q_Y(y|x)P_X(x) \leq \beta. \] (32) (33)
The later condition is equivalent to
\[ \forall y : \sum_{x} T'(x,y)P_X(x) =: c_y \leq \beta. \] (34)
If we let
\[ T(x,y) = (1 - \lambda_y)T'(x,y) + \lambda_y, \]
where $\lambda_y = \frac{\beta - c_y}{1 - \epsilon}$, then
\[ \forall y : \sum_{x} T(x,y)P_X(x) = \beta, \] and since $T'(x,y) \leq T(x,y) \leq 1$ for all $x,y$
\[ \sum_{x,y} T(x,y)N_{Y|X}(y|x)P_X(x) \geq 1 - \epsilon. \] (35) (36)
It follows that there is a non-signalling code of size $k$ and input distribution $P_X$ with error probability $\epsilon$ for $N_{Y|X}$ if and only if
\[ 1/k \geq \beta^{+}(N_{Y|X}(y|x)P_X(x)). \] (37)

**Theorem 3.** There is a non-signalling code of size $k$ and input distribution $P_X$ and error probability $\epsilon$ for $N_{Y|X}$ if and only if
\[ k \leq \min_{Q_Y} \beta^{+}(N_{Y|X}P_X)||Q_Y P_X||^{-1}. \] (38)
There is a non-signalling code of size $k$ and error probability $\epsilon$ for $N_{Y|X}$ if and only if
\[ k \leq \max_{P_Y} \min_{Q_Y} \beta^{+}(N_{Y|X}P_X)||Q_Y P_X||^{-1}. \] (39)
As an upper-bound this is exactly the “minimax” converse given (for conventional codes) in [6] and further studied in [10].

**IV. A LITTLE BACKGROUND**

For any two systems $Q$ and $Q$ of equal dimension $d$ we define $|\Phi^+\rangle_{QQ} := \sum_{j<k} \langle j|_{Q} \otimes |j|_{Q}$ and $\Phi^+_{QQ} = |\Phi^+\rangle_{QQ}$. The vector $|\Phi^+\rangle_{QQ}$ has the property that for any operator $L_{\bar{Q}}$
\[ L_{\bar{Q}}|\Phi^+\rangle_{QQ} = L_{\bar{Q}}^{T}|\Phi^+\rangle_{QQ}. \] (40)
where $L_{\bar{Q}}^{T}$ is the transpose of $L_{\bar{Q}} := \text{id}^{Q_{-\bar{Q}}} L_{\bar{Q}}$ in the computational basis ($\text{id}^{Q_{-\bar{Q}}}$ is the linear map which takes the computational basis for operators on $\bar{Q}$ to that for $Q$, i.e. $\text{id}^{Q_{-\bar{Q}}} : |i\rangle_{\bar{Q}} \mapsto |i\rangle_{\bar{Q}}$). This fact is sometimes referred

\[ \square \]

do to as the ‘transpose trick’. We also note that $\text{Tr}_{Q} \Phi^+_{QQ} = \mathbb{I}_{Q}$
and $\text{Tr}_{\bar{Q}} \Phi^+_{QQ} = \mathbb{I}_{\bar{Q}}$. From this property it follows that, for any
density operator $\rho$, $\rho^{1/2} \Phi_{\bar{Q}} \rho^{1/2}$ is a purification of $\rho$.
Let $\mathcal{H}_{A}$ and $\mathcal{H}_{B}$ be Hilbert spaces of finite dimension. Any linear map $\mathcal{L}_{\bar{A} \rightarrow B}$ from operators on $\mathcal{H}_{A}$ to operators on $\mathcal{H}_{B}$, has an
operator representation $\mathcal{L}_{\bar{A} \rightarrow B} \Phi^+_{\bar{A}}$. We note that
\[ \mathcal{L}_{\bar{A} \rightarrow B} : \kappa_{A} \mapsto \text{Tr}_{\bar{A}} \kappa_{\bar{A}} \mathcal{L}_{\bar{A} \rightarrow B} \Phi^+_{\bar{A}}. \] (41)
(This correspondence between linear maps between operators and operators is known as the “Choi-Jamiolkowski isomorphism.”) Complete positivity of a map corresponds to its operator representation being positive semidefinite. $\mathcal{L}_{\bar{A} \rightarrow B}$ is trace preserving if and only if $\text{Tr}_{\bar{B}} \mathcal{L}_{\bar{A} \rightarrow B} \Phi^+_{\bar{A}} = \mathbb{I}_{\bar{B}}$. A quantum operation from system $A$ to system $B$ is a linear
map from $\mathcal{H}_{A}$ to $\mathcal{H}_{B}$ which is completely positive and trace-preserving.

Given any density operator $\rho_{AB}$ we can write
\[ \rho_{AB} = \mathcal{N}_{\bar{A} \rightarrow B} \rho_{\bar{A}} \] (42)
where $\rho_{\bar{A}} = \rho_{\bar{A}}^{1/2} \Phi_{\bar{A}}^{1/2}$ and $\mathcal{N}_{\bar{A} \rightarrow B}$ is an operation
which we may specify explicitly in terms of its operator representation: Let $\rho_{\bar{A}}^{1/2}$ denote the generalised inverse of $\rho_{\bar{A}}^{1/2}$, which is the unique operator such that $\rho_{\bar{A}} \rho_{\bar{A}}^{1/2}$ and $\rho_{\bar{A}}^{1/2} \rho_{\bar{A}}$ are equal to the orthogonal projection operator, $\rho_{\bar{A}}^{0}$, onto the support of $\rho_{\bar{A}}$. Then, for any state $\tau_{\bar{B}}$, the operation
\[ \mathcal{N}_{\bar{A} \rightarrow B} \Phi_{\bar{A}} = \rho_{\bar{A}}^{1/2} \rho_{\bar{B}} \rho_{\bar{A}}^{1/2} + (1 - \rho_{\bar{A}}^{0}) \otimes \tau_{\bar{B}} \] (43)
satisfies equation (42).

**V. LINEAR TRANSFORMATIONS OF QUANTUM OPERATIONS**

We will now develop the quantum generalisation of the classical results given earlier, starting with Section I.

Alice has some system $M$ to which she applies an operation $\mathcal{E}_{X \rightarrow \mathcal{M}}$. System $F$ is transferred noiselessly to Bob, while an operation $\Lambda^{\mathcal{X} \rightarrow \mathcal{F}}$ is applied to $X$ leaving Bob with system $Y$. Bob applies an operation $\mathcal{D}_{\mathcal{W} \rightarrow \mathcal{Y}}$ to $\mathcal{Y}$, leaving him with system $W$. The overall operation from $M$ to $W$ is
\[ \mathcal{D}_{\mathcal{W} \rightarrow \mathcal{Y}} : \mathcal{I}^{\mathcal{F} \rightarrow \mathcal{M}} \odot \Lambda^{\mathcal{X} \rightarrow \mathcal{F}} \mathcal{E}_{X \rightarrow \mathcal{M}}. \] (44)
Fixing $\mathcal{D}_{\mathcal{W} \rightarrow \mathcal{Y}}$ and $\mathcal{E}_{X \rightarrow \mathcal{M}}$ (44) is a linear function of $\Lambda^{\mathcal{X} \rightarrow \mathcal{F}}$ which maps any operation $\Lambda^{\mathcal{X} \rightarrow \mathcal{F}}$ to an operation. In fact it satisfies a strictly stronger property, which is that if $\Lambda^{\mathcal{X} \rightarrow \mathcal{F}}$ is an operation, then it will be mapped to an operation. As shown in [5], any linear map from operations to operations with this property can be written in the form (44).

We define a bipartite operation $\mathcal{E}_{X \rightarrow \mathcal{M}}$ via
\[ \mathcal{E}_{X \rightarrow \mathcal{M}} := \mathcal{D}_{\mathcal{W} \rightarrow \mathcal{Y}} \mathcal{E}_{X \rightarrow \mathcal{M}}. \] (45)
This operation completely determines the map from operations to operations discussed above (see [12]). Evidently this operation is implemented by local operations and one-way quantum communication from Alice to Bob. Any operation of this form is non-signalling from Bob to Alice [1], in the sense that
\[ \forall \rho_{\bar{Y}} : \text{Tr}_{\bar{W}} \mathcal{E}_{X \rightarrow \mathcal{M}} \rho_{\bar{Y}} = \mathcal{E}_{X \rightarrow \mathcal{M}}. \] (46)
In particular, \( \mathcal{Z}^{X\leftarrow M} = \text{Tr}_X \mathcal{E} F_{X\leftarrow M} \). Conversely, any bipartite operation which satisfies (46) can be implemented by local operations and quantum communication from Alice to Bob [3]. That is, it can be written in the form (45).

Let \( \hat{M} \) have the same dimension as system \( M \) and suppose that, initially, Alice has systems \( MM \). The 'transpose trick' tells us that \( \mu_{MM} = \text{Tr}_M \mu_{MM} \). Let \( \hat{M} \) be the state of the system MM in the state

\[
\mu_{\hat{M}M} := \mu_M^{1/2} \phi_{\hat{M}M}^+ \phi_{\hat{M}M}^{1/2} = \mu_M^{1/2} \phi_{MM}^+ \phi_{MM}^{1/2} \tag{47}
\]

where \( \mu_M := \text{Tr}_M \mu_{MM} \). The 'transpose trick' tells us that \( \mu_M^{1/2} = \text{id} \mu_{MM}^{1/2} \). Let

\[
\omega_{\hat{M}MM} := D^{W\leftarrow Y} \text{id} F \otimes \mathcal{N}^{Y\leftarrow X} \mathcal{E} F_{X\leftarrow M} \mu_{\hat{M}M} \tag{48}
\]

After Alice applies \( \mathcal{E} \), the system MM is in the state \( \rho_{\hat{M}MM} = \mu_{\hat{M}M}^{1/2} (\mathcal{E} F_{X\leftarrow M} \phi_{\hat{M}MM}^+ \phi_{\hat{M}MM}^{1/2} \mu_{MM}^{1/2} \). The solution is illustrated in the top half of the figure. Let \( \hat{E} \) be an operation (see previous section) such that \( \rho_{\hat{E}MM} = \mathcal{E} \hat{E} \rho_{\hat{E}XX} \) where (for the remainder of this article) \( \rho_{\hat{E}XX} \) is defined to be the state

\[
\rho_{\hat{E}XX} := \rho_X^{1/2} \phi_{\hat{E}XX}^+ \phi_{\hat{E}XX}^{1/2}. \tag{49}
\]

Note that \( \rho_X := \text{Tr}_X \rho_{\hat{E}XX} = \text{id} \phi_{\hat{E}XX} F \rho_{\hat{E}XX} \). Then

\[
\omega_{\hat{E}MM} = D^{W\leftarrow Y} \mathcal{N}^{Y\leftarrow X} \rho_{\hat{E}MM} = \mathcal{Z}^{W\leftarrow Y} \mathcal{N}^{Y\leftarrow X} \rho_{\hat{E}XX} \tag{50}
\]

where

\[
\mathcal{Z}^{W\leftarrow Y} := D^{W\leftarrow Y} \mathcal{E} \mathcal{N}^{Y\leftarrow X}. \tag{51}
\]

Note the analogy between the expression (50) for the final state of \( \hat{M}W \) and the expression (11) for the joint distribution of \( M \) and \( W \). In terms of the operator representations of \( \hat{Z} \) and \( Z \), we have

\[
\frac{1}{\sqrt{2}} (\mathcal{Z}^{W\leftarrow X} \rho_{\hat{E}XX}^+ \phi_{\hat{E}XX}) \rho_X^{1/2} = \mu_{\hat{M}M}^{1/2} (\mathcal{Z}^{W\leftarrow X} \rho_{\hat{E}XX}^+ \phi_{\hat{E}XX}) \rho_X^{1/2}. \tag{52}
\]

VI. QUANTUM NON-SIGNALLING CODES

We can view block coding of classical (or quantum) information over a quantum channel as a special case of the scenario described in the previous section. In this case \( M \) and \( W \) are of the same dimension, \( k \) (which we call the size of the code). If (as in the classical case) we are concerned with the transmission of a uniformly distributed classical message, then \( M \) stores a uniformly distributed classical message in the computational basis. That is, \( \mu_M = \mathbb{I}_M/k \). If \( M \) is measured in the computational basis then we obtain a copy of the message that was sent. The probability of successful transmission is, therefore, the probability of obtaining equal results computational basis measurements are performed on \( \hat{M} \) and \( W \). The POVM element corresponding to this outcome is

\[
\Pi_{\hat{M}W} := \sum_m |m\rangle \langle m|_{\hat{M}} \otimes |m\rangle \langle m|_W,
\]

so the success probability of the code is

\[
1 - \epsilon = \text{Tr} \Pi_{\hat{M}W} \mathcal{Z}^{W\leftarrow Y} \mathcal{N}^{Y\leftarrow X} \rho_{\hat{E}XX}. \tag{53}
\]

In a conventional code, there is no auxiliary forward communication and the bipartite operation is of the form

\[
\mathcal{Z}^{X\leftarrow W} = \mathcal{E}_{X\leftarrow M} \otimes \mathcal{D}^{W\leftarrow Y} \tag{54}
\]

where \( \mathcal{E}_{X\leftarrow M} \) and \( \mathcal{D}^{W\leftarrow Y} \) are encoding and decoding operations. The bipartite operation for such codes is not only non-signalling from Bob to Alice, but also from Alice to Bob. We call any forward-assisted quantum code whose bipartite operation is non-signalling in both directions a quantum non-signalling code [12]. In terms of the operation \( \mathcal{Z}^{X\leftarrow W} \mathcal{N}^{Y\leftarrow X} \) this condition is

\[
\mathcal{Z}^{W\leftarrow X} \rho_{\hat{E}XX} \otimes 1_Y = \mu_M \otimes \mathcal{Z}^{W\leftarrow X}. \tag{55}
\]

and given any operation \( \hat{Z} \) which satisfies this condition the corresponding \( Z \) is non-signalling from Bob to Alice.

Remark 4. In terms of the operator representation of \( Z \), the success probability is a linear functional, the non-signalling and normalising constraints on \( Z \) are affine, while the complete positivity of \( Z \) is equivalent to the operator representation being positive semidefinite. Therefore, maximising the success probability over non-signalling quantum codes is a semidefinite program (see [12]).

The quantum analog of a channel for which \( Y \) and \( X \) are independent is for the operation \( \mathcal{N}^{Y\leftarrow X} \) to have the form \( \mathcal{N}^{Y\leftarrow X} = \sigma_Y \text{Tr}_X \). As one would expect, the success probability of a quantum non-signalling code of size \( k \) for any such channel is simply \( 1/k \), that is

\[
\forall \sigma_Y : \text{Tr} \Pi_{\hat{M}W} \mathcal{Z}^{W\leftarrow X} \rho_{\hat{E}XX} \otimes \sigma_Y = 1/k. \tag{56}
\]
VII. QUANTUM HYPOTHESIS TESTING CONVERSE

Consider the quantum hypothesis testing problem where the null hypothesis is that the state of $\tilde{X}\tilde{Y}$ is $\rho_{\tilde{X}\tilde{Y}}$ and the (composite) alternative hypothesis is that that state of $\tilde{X}\tilde{Y}$ is of the form $\rho_{\tilde{X}} \otimes \sigma_{\tilde{Y}}$ where $\rho_{\tilde{X}} = Tr_{\tilde{Y}}\rho_{\tilde{X}\tilde{Y}}$ and $\sigma_{\tilde{Y}}$ is any state. We can specific a quantum hypothesis test by giving the POVM element $T_{\tilde{X}\tilde{Y}}$ corresponding to acceptance of the null hypothesis.

\[
\beta^*_{\rho_{\tilde{X}\tilde{Y}}} := \min_{0 \leq T_{\tilde{X}\tilde{Y}} \leq 1} \max_{\sigma_{\tilde{Y}}} Tr T_{\tilde{X}\tilde{Y}} \rho_{\tilde{X}} \otimes \sigma_{\tilde{Y}} \tag{56}
\]
subject to $Tr T_{\tilde{X}\tilde{Y}} \rho_{\tilde{X}\tilde{Y}} \geq 1 - \epsilon$. \tag{57}

For any two states $\rho_0$ and $\rho_1$ of the same system we define

\[
\beta_{\epsilon}(\rho_0 || \rho_1) := \min \{ Tr T \rho_1 : Tr T \rho_0 \geq 1 - \epsilon, 0 \leq T \leq 1 \}.
\]

By von Neumann’s minimax theorem

\[
\beta^*_{\rho_{\tilde{X}\tilde{Y}}} = \max_{\sigma_{\tilde{Y}}} \beta_{\epsilon}(\rho_{\tilde{X}} || \rho_{\tilde{X}} \otimes \sigma_{\tilde{Y}}) \tag{58}
\]

We now give the quantum generalisation of Proposition 2.

**Proposition 5.** There is a quantum non-signalling code of size $k$ with input state $\rho_{\tilde{X}}$ and error probability $\epsilon$ for operation $N^{\tilde{X} \to \tilde{Y}}$ if and only if there is a quantum hypothesis test $T_{\tilde{X}\tilde{Y}}$ satisfying

\[
Tr T_{\tilde{X}\tilde{Y}} N^{\tilde{X} \to \tilde{Y}} \rho_{\tilde{X}} = 1 - \epsilon, \quad \forall \sigma_{\tilde{Y}} \colon Tr T_{\tilde{X}\tilde{Y}} \rho_{\tilde{X}} \otimes \sigma_{\tilde{Y}} = 1/k
\]

where $\rho_{\tilde{X}} = \rho_{\tilde{X}}^{1/2} \Phi^+_{\tilde{X}\tilde{Y}} \rho_{\tilde{X}}^{1/2}$.

Proof. First the converse part: Suppose that there is a non-signalling code $Z$ with properties stated in (5). Consider the test obtained by applying the operation $Z^{\tilde{X} \to \tilde{Y}}$ to system $\tilde{X}\tilde{Y}$, measuring both $M$ and $W$ in their computational bases, and accepting (the null hypothesis) when the two results are equal. By (52) and (55) this test has the required properties.

For the direct part, let $T_{\tilde{X}\tilde{Y}}$ be a test satisfying (59) and (60), and let

\[
\begin{aligned}
Z^{\tilde{X} \to \tilde{Y}} : A_{\tilde{X}} \mapsto & \frac{1}{k} \sum_{W} T_{\tilde{X}\tilde{Y}} A_{\tilde{X}}
+ \frac{1}{k(k-1)} \left( I - \sum_{W} T_{\tilde{X}\tilde{Y}} \right) (I - T_{\tilde{X}\tilde{Y}}) A_{\tilde{X}}
\end{aligned}
\]

where $\Pi_{\tilde{X}W} := \sum_{m} |m\rangle \langle m| \otimes |m\rangle \langle m|$. It is easy to check that this is non-signalling from Bob to Alice, and the property (60) ensures that this $Z^{\tilde{X} \to \tilde{Y}}$ satisfies (54). That is, if the desired error probability follows from (52), (59) and $\Pi_{\tilde{X}W}(1 - \Pi_{\tilde{X}W}) = 0$. \qquad \Box

**Corollary 6.** If there is a non-signalling code of size $k$ and average input state $\rho_{\tilde{X}}$ and error probability $\epsilon$ for $N^{\tilde{X} \to \tilde{Y}}$ then

\[
k \leq \min_{\sigma_{\tilde{Y}}} \beta_{\epsilon}(N^{\tilde{X} \to \tilde{Y}} \rho_{\tilde{X}} || \rho_{\tilde{X}} \otimes \sigma_{\tilde{Y}})^{-1}. \tag{62}
\]

If there is a non-signalling code of size $k$ and error probability $\epsilon$ for $N^{\tilde{X} \to \tilde{Y}}$ then

\[
k \leq \max_{\rho_{\tilde{X}}} \min_{\sigma_{\tilde{Y}}} \beta_{\epsilon}(N^{\tilde{X} \to \tilde{Y}} \rho_{\tilde{X}} || \rho_{\tilde{X}} \otimes \sigma_{\tilde{Y}})^{-1}. \tag{63}
\]

This converse applies to entanglement-assisted codes because they are non-signalling. For memoryless channels, analysing the large block length limit of the upper bound on rate that it gives recovers (see [11]) the known, single-letter formula for the entanglement-assisted classical capacity of a quantum channel [2].

As noted in [11], if we are dealing with codes of the form (53), then the hypothesis test constructed in the direct part of (5) can be implemented by local measurements and classical post-processing of the results (to compare the outcomes). This means that we can obtain a better converse for such codes by restricting the optimisation over hypothesis tests to those which can be implemented in this way. In [11] it was shown that if we restrict to those which can implemented by local operations and one-way classical communication from Alice to Bob then the converse obtained is equivalent to the one obtained in [8].

In Corollary 6 we do not have a quantum analog of Theorem 3 because the implication is only one way. If we could show that one can restrict to quantum tests satisfying $Tr T_{\tilde{X}\tilde{Y}} \rho_{\tilde{X}} \otimes \sigma_{\tilde{Y}} = \beta_{\epsilon}$ for all $\sigma_{\tilde{Y}}$ without changing the minimax type-II error probability then we could add the other direction of implication to Corollary 6. Whether this is true is open at the time of writing.

**ACKNOWLEDGMENT**

My thanks to Debbie Leung and Andreas Winter for many useful discussions on this topic.

**REFERENCES**


