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 Polyanskiy-Poor-Verdú. 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 $M-(F,X)-(F,Y)-W$ is a Markov chain. Letting $N(y|x) = P_{Y|X}(y|x)$, $E(f,x|m) = P_{FX|M}(f,x|m)$, and $D(w|f,y) = P_{W|FY}(w|f,y)$, we have

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

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 that

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

(3)

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 $P_{WYXM}$ depends on $E$ and $D$ only through the distribution $Z$.

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

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

(5)

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

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

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

$$\tilde{Z}(m,w|x,y) := \frac{P_{MW|XY}(m,w|x,y)}{P_{Y|X}(y|x)P_{X}(x)}.$$  

(6)

(7)

(8)

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

(9)

The final equality follows from (1). Note that $P_{X}(x) = \sum_{m} Z_{X|M}(x|m)P_{M}(m)$, 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

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

(10)

It follows that

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

(11)

We will make use of this expression in the next section, and give a quantum generalisation of it in Section V. Note that $\tilde{Z}(m,w|x,y)$ 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

$$Z(x, w|m, y) = E(x|m)D(w|y).$$  \hfill (12)

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

$$\forall w, y, m, m' : \sum_x Z(x, w|m, y) = \sum_x Z(x, w|m', y),$$  \hfill (13)

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

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

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

$$\Pr(M = W) = \sum_{m, x, y} Z(m, m|x, y)N_Y|X(y|x)P_X(x)$$  \hfill (15)

$$= \sum_{m, x, y} \hat{Z}(m, m|x, y)N_Y|X(y|x)P_X(x).$$  \hfill (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

$$Q_{MW}(m, w) = \sum_{x, y} \hat{Z}(m, w|x, y)P_X(x)Q_Y(y)$$  \hfill (17)

$$= P_M(m)\sum_y Z_{W|Y}(w|y)Q_Y(y),$$  \hfill (18)

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

$$\forall Q_Y : \sum_{m, x, y} \hat{Z}(m, m|x, y)P_X(x)Q_Y(y) = 1/k.$$  \hfill (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 $P_{X|Y}$ for some arbitrary $Q_Y$. A hypothesis test is specified by

$$T[x, y] := \Pr(\text{Accept null}|X = x, Y = y).$$  \hfill (20)

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

$$\beta^*_\epsilon(P_{XY}) := \min_{Q_Y} \max_{Q_{X|Y}} \sum_y T[x, y]P_X(x)Q_Y(y)$$  \hfill (21)

subject to

$$\sum_y T[x, y]P_X(x) \geq 1 - \epsilon.$$  \hfill (22)

Let us define for distributions $p$ and $q$,

$$\beta_\epsilon(p||q) := \min_p \left\{ \sum_x T[z|p(z) : \sum_x T[z|p(z) \geq 1 - \epsilon] \right\}.$$  \hfill (23)

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

$$\beta^*_\epsilon(P_{XY}) = \max_{Q_{X|Y}} p_\epsilon(P_{X|Y}||P_XQ_Y).$$  \hfill (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

$$\sum_{x, y} T[x, y]N_Y|X(y|x)P_X(x) = 1 - \epsilon,$$  \hfill (25)

$$\forall Q_Y : \sum_{x, y} T[x, y]Q_Y(y)P_X(x) = 1/k.$$  \hfill (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

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

$$\hat{Z}(m, m|x, y) = \frac{1}{k} \delta_{mw}T[x, y]$$  \hfill (28)

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

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

$$\sum_x \hat{Z}(m, w|x, y)P_X(x) = \frac{1}{k} \delta_{mw} \sum_x T[x, y]P_X(x)$$  \hfill (29)

$$+ \frac{1}{k(k-1)} (1 - \sum_x T[x, y]P_X(x))(1 - \delta_{mw}) = 1/k^2.$$  \hfill (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),

$$\Pr(M = W) = \sum_{m, x, y} \hat{Z}(m, m|x, y)N_Y|X(y|x)P_X(x)$$  \hfill (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: Sup-
pose we have a test T^y which satisfies
\[ \sum_{x \in y} T^y(x, y)N_{y|x}(y|x)P_X(x) \geq 1 - \epsilon, \]  
and
\[ \forall y : \sum_{x \in y} T^y(x, y)Q_Y(y|x)P_X(x) \leq \beta. \] 
The later condition is equivalent to
\[ \forall y : \sum_{x \in y} T^y(x, y)P_X(x) = c_y \leq \beta. \] 
If we let
\[ T^y(x, y) = (1 - \lambda_y)T^y(x, y) + \lambda_y, \] 
where \[ \lambda_y = \frac{\beta - c_y}{1 - c_y}, \] then
\[ \forall y : \sum_{x \in y} T^y(x, y)P_X(x) = \beta, \] 
and since \[ T^y(x, y) \leq T(x, y) \leq 1 \] for all \( x, y \)
\[ \sum_{x \in y} T^y(x, y)N_{y|x}(y|x)P_X(x) \geq 1 - \epsilon. \] 
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
\[ \frac{1}{k} \geq \beta^*(N_{y|x}(y|x)P_X(x)). \] 

**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
\[ k \leq \min_{Q_Y} \beta^*(N_{y|x}P_X|Q_Y P_X)^{-1}. \] 

There is a non-signalling code of size \( k \) and error probability \( \epsilon \) for \( N_{y|x} \) if and only if
\[ k \leq \max_{P_Y} \min_{Q_Y} \beta^*(N_{y|x}P_X|Q_Y P_X)^{-1}. \] 

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 \( \mathcal{P} \) of equal dimension \( d \) we de-
finite \( |\Phi^+\rangle_{QQ} := \sum_{j \in d} |j\rangle_{Q} \otimes |j\rangle_{Q} \) and \( \Phi^+_{QQ} = |\Phi^+\rangle|\Phi^+\rangle_{QQ}. \) The vector \( |\Phi^+\rangle_{QQ} \) has the property that for any operator \( L_Q \)
\[ L_Q|\Phi^+\rangle_{QQ} = L_Q^T|\Phi^+\rangle_{QQ}. \] 
where \( L_Q^T \) is the transpose of \( L_Q := id^{Q-\mathcal{P}} L_Q \) in the
computational basis \( id^{Q-\mathcal{P}} = \text{the linear map which takes the} \)
computational basis for operators on \( \mathcal{P} \) to that for \( Q \), i.e.
\( id^{Q-\mathcal{P}} : |i\rangle|j\rangle_{\mathcal{P}} \mapsto |i\rangle|j\rangle_{Q}. \) This fact is sometimes referred
to as the ‘transpose trick’. We also note that \( Tr_Q \Phi^+_{QQ} = \mathbb{1}_Q \)
and \( Tr_Q \Phi^+_{QQ} = 1_Q. \) From this property it follows that, for any
density operator \( \rho_A \), \( \rho_A^{1/2} \Phi_A \rho_A^{1/2} \) is a purification of \( \rho_A \). Let \( \mathcal{H}_A \) and \( \mathcal{H}_B \) be Hilbert spaces of finite dimension. Any linear
\( \mathcal{L}^{\mathcal{B}^{\mathcal{C}}\mathcal{A}} \) from operators on \( \mathcal{H}_A \) to operators on \( \mathcal{H}_B \), has an
operator representation \( \mathcal{L}^{\mathcal{B}^{\mathcal{C}}\mathcal{A}} \). We note that
\[ \mathcal{L}^{\mathcal{B}^{\mathcal{C}}\mathcal{A}} : \mathcal{H}_A \rightarrow Tr_A \mathcal{H}_A \mathcal{L}^{\mathcal{B}^{\mathcal{C}}\mathcal{A}} \mathcal{H}_A. \] 
(This correspondence between linear maps between operators
and operators is known as the ‘Choi-Jamiolkowski isomor-
phism’.) Complete positivity of a map corresponds to its
operator representation being positive semidefinite. \( \mathcal{L}^{\mathcal{B}^{\mathcal{C}}\mathcal{A}} \) is
trace preserving if and only if \( Tr_B \mathcal{L}^{\mathcal{B}^{\mathcal{C}}\mathcal{A}} = \mathbb{1}_A \). A
quantum operation from system \( \mathcal{A} \) to system \( \mathcal{B} \) 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
\[ \rho_{AB} = \mathcal{W}^{\mathcal{B}^{\mathcal{C}}\mathcal{A}} \rho_{A\mathcal{A}} \] 
where \( \rho_{A\mathcal{A}} = \rho_A^{1/2} \mathcal{W}^{\mathcal{B}^{\mathcal{C}}\mathcal{A}} \rho_A^{1/2} \) and \( \mathcal{W}^{\mathcal{B}^{\mathcal{C}}\mathcal{A}} \) is an operation
which we may specify explicitly in terms of its operator
representation: Let \( \rho_A^{1/2} \) denote the generalised inverse
of \( \rho_A^{1/2} \), which is the unique operator such that \( \rho_A^{1/2} \rho_A^{1/2} \) and \( \rho_A \rho_A^{1/2} \) are equal to the orthogonal projection operator, \( \rho_A^{0} \)
on the support of \( \rho_A \). Then, for any state \( \tau_B \), the operation
\[ \mathcal{W}^{\mathcal{B}^{\mathcal{C}}\mathcal{A}} \rho_{A\mathcal{A}} = \rho_A^{1/2} \rho_{AB} \rho_A^{1/2} + (1 - \rho_A^{0}) \otimes \tau_B \] 
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 \( \mathcal{E}^{X \leftrightarrow \mathcal{M}} \). System \( F \) is transferred noiselessly to Bob, while an operation \( \mathcal{A}^{Y \leftrightarrow \mathcal{X}} \) is applied to \( X \) leaving Bob with system \( Y \).

Bob applies an operation \( \mathcal{D}^{W \leftrightarrow \mathcal{Y}} \) to \( \mathcal{Y} \), leaving him with system \( W \). The overall operation from \( M \) to \( W \) is
\[ \mathcal{D}^{W \leftrightarrow \mathcal{F}} \otimes \mathcal{A}^{Y \leftrightarrow \mathcal{X}} \mathcal{E}^{X \leftrightarrow \mathcal{M}}. \] 

Fixing \( \mathcal{D}^{W \leftrightarrow \mathcal{F}} \) and \( \mathcal{E}^{X \leftrightarrow \mathcal{M}} \), (44) is a linear function of \( \mathcal{A}^{Y \leftrightarrow \mathcal{X}} \) which maps any operation \( \mathcal{A}^{Y \leftrightarrow \mathcal{X}} \) to an operation. In fact it
satisfies a strictly stronger property, which is that if \( \mathcal{A}^{Y \leftrightarrow \mathcal{X}} \) 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 \( \mathcal{Z}^{XW \leftrightarrow \mathcal{M}} \) via
\[ \mathcal{Z}^{XW \leftrightarrow \mathcal{M}} := \mathcal{D}^{W \leftrightarrow \mathcal{F}} \otimes \mathcal{E}^{X \leftrightarrow \mathcal{M}}. \] 

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
\[ \forall \rho_Y : Tr_W \mathcal{Z}^{XW \leftrightarrow \mathcal{M}} \mathcal{1} \otimes \rho_Y = \mathcal{Z}^{X \leftrightarrow \mathcal{M}}. \]
In particular, $\mathcal{Z}^{X\leftarrow M} = \text{Tr}_X \mathcal{E}^{FX\leftarrow M}$. 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 $\tilde{M}$ have the same dimension as system $M$ and suppose that, initially, Alice has systems $\tilde{M}M$. The ‘transpose trick’ tells us that $\mu_{\tilde{M}} = \text{Tr}_M \mu_{\tilde{MM}}$. Let

$$\omega_{\tilde{M}W} := D^{W\leftarrow Y} \mathcal{E}^{F\leftarrow X} \Phi_{\tilde{X}W}^{+} \otimes \mathcal{N}_{Y\leftarrow X} \Phi_{XY}^{+} \rho_{\tilde{M}W}$$

(48)

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

$$\rho_{\tilde{X}X} := \rho_{X}^{1/2} \Phi_{XX}^{+} \rho_{X}^{1/2}$$

(49)

Note that $\rho_{X} := \text{Tr}_X \rho_{\tilde{X}X} = \text{id} \tilde{X}^{X} \rho_{X}^{1/2} \rho_{X}^{1/2}$. Then

$$\omega_{MW} := D^{W\leftarrow Y} \mathcal{N}_{Y\leftarrow X} \rho_{\tilde{M}FW} = \tilde{\mathcal{E}}^{\tilde{M}W\leftarrow XY} = \mathcal{E}^{M\leftarrow X} \rho_{FX}$$

(50)

where

$$\tilde{\mathcal{E}}^{\tilde{M}W\leftarrow XY} := \tilde{\mathcal{E}}^{M\leftarrow X} \mathcal{E}^{F\leftarrow X}.$$

(51)

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

$$\rho_{X}^{1/2} (\tilde{\mathcal{E}}^{\tilde{M}W\leftarrow XY} \Phi_{XY}^{+} \rho_{\tilde{X}X} \mu_{\tilde{M}}^{1/2}) = \mu_{M}^{1/2} (\mathcal{E}^{M\leftarrow XY} \Phi_{XY}^{+} \rho_{FX}) \mu_{M}^{1/2}.$$

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, $\mu_{M} = \frac{1}{k} \mathbb{1}_{M}$. 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 $M$ and $W$. The POVM element corresponding to this outcome is

$$\Pi_{\tilde{M}W} := \sum_{m} |m\rangle \langle m| \otimes |m\rangle \langle m|,$$

so the success probability of the code is

$$1 - \epsilon = \text{Tr} \Pi_{\tilde{M}W} \tilde{\mathcal{E}}^{\tilde{M}W\leftarrow XY} \mathcal{E}^{Y\leftarrow X} \rho_{\tilde{X}X}.$$ (52)

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

$$\mathcal{Z}^{X\leftarrow MW} = \mathcal{E}^{X\leftarrow M} \otimes \mathcal{D}^{W\leftarrow Y}$$

(53)

where $\mathcal{E}^{X\leftarrow M}$ and $\mathcal{D}^{X\leftarrow M}$ are the 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 $\tilde{\mathcal{E}}^{\tilde{M}W\leftarrow XY}$ this condition is

$$\tilde{\mathcal{E}}^{\tilde{M}W\leftarrow XY} \rho_{\tilde{X}} \otimes 1_{Y} = \mu_{\tilde{M}} \otimes \mathcal{E}^{W\leftarrow Y},$$

(54)

and given any operation $\tilde{Z}$ which satisfies this condition the corresponding $\tilde{Z}$ 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 normalisation 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

$$\forall \sigma_{Y} : \text{Tr} \Pi_{\tilde{M}W} \tilde{\mathcal{E}}^{\tilde{M}W\leftarrow XY} \rho_{\tilde{X}} \otimes \sigma_{Y} = 1/k.$$ (55)
VII. QUANTUM HYPOTHESIS TESTING CONVERSE

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

\[
\beta_{\epsilon}^{|p_{xy}|} := \min_{0 \leq T_{XY} \leq 1} \max_{\sigma_{Y}} \text{Tr} T_{XY} \rho_{X} \otimes \sigma_{Y}
\]

subject to \( \text{Tr} T_{XY} \rho_{X} \geq 1 - \epsilon \).

(56)

For any two states \( \rho_{0} \) and \( \rho_{1} \) of the same system we define

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

(57)

By von Neumann’s minimax theorem

\[
\beta_{\epsilon}^{|p_{xy}|} (\rho_{XY}) = \max_{\sigma_{Y}} \beta(\rho_{XY} || \rho_{X} \otimes \sigma_{Y}).
\]

(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 \( \rho_{X} \) and error probability \( \epsilon \) for operation \( \mathcal{N}_{X}^{\hat{X}} \) if and only if there is a quantum hypothesis test \( T_{XY} \) satisfying

\[
\text{Tr} T_{XY} \mathcal{N}_{X}^{\hat{X}} \rho_{XY} = 1 - \epsilon, \text{ and}
\]

\[
\forall \sigma_{Y} : T_{XY} \rho_{XY} \otimes \sigma_{Y} = 1/k.
\]

(59)

(60)

where \( \rho_{XY} = \rho_{X}^{1/2} \hat{X}^{\rho_{X}} \rho_{X}^{1/2} \).

Proof. First the converse part: Suppose that there is a non-signalling code \( \mathcal{C} \) with properties stated in (5). Consider the test obtained by applying the operation \( \mathcal{Z}_{\hat{X}W}^{\hat{X}Y} \) to system \( \hat{X}Y \), measuring both \( \hat{M} \) and \( \hat{W} \) 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 \( T_{XY} \) be a test satisfying (59) and (60), and let

\[
\mathcal{Z}_{\hat{X}W}^{\hat{X}Y} : A_{\hat{X}} \mapsto \frac{1}{k} \Pi_{\hat{X}W} T_{XY} A_{\hat{X}} \]

\[
+ \frac{1}{k(k-1)} \left( \mathbb{I} - \Pi_{\hat{X}W} \right) \text{Tr}(1 - T_{XY}). A_{\hat{X}}
\]

(61)

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

**Corollary 6.** If there is a non-signalling code of size \( k \) and average input state \( \rho_{X} \) and error probability \( \epsilon \) for \( \mathcal{N}_{X}^{\hat{X}} \) then

\[
k \leq \min_{\sigma_{Y}} \beta(\mathcal{N}_{X}^{\hat{X}} \rho_{XY} || \rho_{X} \otimes \sigma_{Y})^{-1}.
\]

(62)

If there is a non-signalling code of size \( k \) and error probability \( \epsilon \) for \( \mathcal{N}_{X}^{\hat{X}} \) then

\[
k \leq \max_{\rho_{X}} \min_{\sigma_{Y}} \beta(\mathcal{N}_{X}^{\hat{X}} \rho_{XY} || \rho_{X} \otimes \sigma_{Y})^{-1}.
\]

(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 \( Tr_{XY} \rho_{X} \otimes \sigma_{Y} = \beta \) for all \( \beta \) 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**


