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 $N(y|x) = P_{Y|X}(y|x)$, $E(f,x|m) = P_{F|X|M}(f,x|m)$, and $D(w|f,y) = P_{W|F|Y}(w|f,y)$, we have

$$P_{WYXM}(w,y,x,m) = Z(x,w|m,y)N(y|x)P_{M}(m)$$  \hspace{2cm} (1)

where

$$Z(x,w|m,y) := \sum_{f} D(w|f,y)E(f,x|m).$$  \hspace{2cm} (2)

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

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

$$=: Z_{X|M}(x|m),$$  \hspace{2cm} (3)

in particular

$$\forall x,m,y : \sum_{w} Z(x,w|m,y) = \sum_{f} E(f,x|m).$$  \hspace{2cm} (4)

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

$$P_{W|M}(w|m) = \sum_{x,y} Z(x,w|m,y)N(y|x).$$  \hspace{2cm} (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

$$P_{WYXM}(w,y,x,m) = \hat{Z}(m,w|x,y)P_{XY}(x,y)$$  \hspace{2cm} (6)

where, for $x$ such that $P_X(x) > 0$ we define

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

$$\frac{P_{WYXM}(w,y,x,m)}{P_{Y|X}(y|x)P_X(x)} = \frac{Z(x,w|m,y)P_{M}(m)}{P_X(x)}.$$  \hspace{2cm} (9)

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

$$\hat{Z}(m,w|x,y) := P_{M}(m)P_{W}(w).$$  \hspace{2cm} (10)

It follows that

$$P_{MW}(m,w) = \sum_{m,w} \hat{Z}(m,w|x,y) \sum_{x} N_Y|X(y|x)P_X(x).$$  \hspace{2cm} (11)

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 $NY|X$. 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 $Z$ 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 $NY|X$ is

$$\Pr(M = W) = \sum_{m, x, y} Z(m, m|x, y)NY|X(y|x)P_M(m)$$

(15)

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

(16)

**Remark 1.** Fixing $NY|X$, the success probability (15) is a linear functional of $M$ 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 $RY|X$ where $Y$ and $X$ are independent, i.e. $RY|X(y|x) = Q_Y(y)$ 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 $P_X$. The alternative hypothesis is a composite hypothesis, which states that $X$ and $Y$ are distributed according to $P_XQ_Y$ 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

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

(21)

subject to

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

(22)

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

$$\beta_p(q) := \min_{T} \left\{ \sum_{x, y} T[x, y]q(x) : \sum_{x, y} T[x, y]p(x) \geq 1 - \epsilon \right\}.$$  

(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

$$\beta^*_\epsilon(P_{XY}) = \max_{Q_Y} \beta_p(P_X),$$

(24)

**Proposition 2.** There is a non-signalling code of size $k$, input distribution $P_X$, and error probability $\epsilon$ for channel $NY|X$ if and only if there is a test $T$ with

$$\sum_{x, y} T[x, y]NY|X(y|x)P_X(x) = 1 - \epsilon, \text{ and}$$

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

(25)

(26)

**Proof.** Suppose that we have a non-signalling code of size $k$ which attains error probability $\epsilon$ for channel $NY|X$. 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

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

(28)

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

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

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

(29)

$$+ \frac{1}{k(k-1)}(1 - \delta_{m, w})(1 - T[x, y]) = 1/k^2$$  

(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),

$$\Pr(M = W) = \sum_{m, x, y} \hat{Z}(m, m|x, y)NY|X(y|x)P_X(x)$$

(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 “P₆-
balanced”), and as noted there, we may relax this condition
without changing the minimax type-II error probability: Sup-
pose we have a test \( T^* \) which satisfies
\[
\sum_{xy} T^*[x, y]N_{Y|x}(y|x)P_X(x) \geq 1 - \epsilon, \quad \text{and} \quad \forall y : \sum_{xy} T^*[x, y]Q_Y(y)P_X(x) \leq \beta.
\]
The latter condition is equivalent to
\[
\forall y : \sum_{x} T^*[x, y]P_X(x) = c_y \leq \beta.
\]
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 \( x, y \)
\[
\sum_{xy} T[x, y]N_{Y|x}(y|x)P_X(x) \geq 1 - \epsilon.
\]
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
\[
1/k \geq \beta^*\left(N_{Y|x}(y|x)P_X(x)\right).
\]

**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
\[
k \leq \min_{Q_Y} \beta(N_{Y|x}P_X||Q_YP_X)^{-1}.
\]
There is a non-signalling code of size \( k \) and error probability \( \epsilon \) for \( N_{Y|x} \) if and only if
\[
k \leq \max_{P_Y} \min_{Q_Y} \beta(N_{Y|x}P_X||Q_YP_X)^{-1}.
\]
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 \( \bar{Q} \) of equal dimension \( d \) we define \(|\Phi^+\rangle_{QQ} := \sum_{0 \leq j < d} |j\rangle_Q \otimes |j\rangle_Q \) 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 \( L_Q := id_{Q^{e\rightarrow Q}}L_Q \) in the computational basis \((id_{Q^{e\rightarrow Q}} \) is the linear map which takes the computational basis for operators on \( Q \) to that for \( Q, \) i.e. \( id_{Q^{e\rightarrow Q}} : |i\rangle|j\rangle_Q \mapsto |i\rangle|j\rangle_Q \). This fact is sometimes referred
to as the ‘transpose trick’. We also note that \( Tr_Q\Phi^+_{QQ} = I_Q \) and \( Tr_Q\Phi^+_{QQ} = I_Q \). 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 \( H_A \) and \( H_B \) be Hilbert spaces of finite dimension. Any linear map \( L^{B\leftarrow A} \) from operators on \( H_A \) to operators on \( H_B \), has an
operator representation \( L^{B\leftarrow A}\Phi_A^+ \). We note that
\[
L^{B\leftarrow A} : \kappa_A \mapsto Tr_A\kappa_A^TL^{B\leftarrow A}\Phi_A^+.
\]
(41)

(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. \( L^{B\leftarrow A} \) is
trace preserving if and only if \( Tr_BL^{B\leftarrow A}\Phi_A^+ = 1_A \). A quantum operation from system \( A \) to system \( B \) is a linear
map from \( H_A \) to \( H_B \) which is completely positive and trace-

preserving.

Given any density operator \( \rho_{AB} \) we can write
\[
\rho_{AB} = \Psi^{A\leftarrow B}\rho_{AA} \quad (42)
\]
where \( \rho_{AA} = \rho_A^{1/2}\Phi_A^+\rho_A^{1/2} \) and \( \Psi^{A\leftarrow B} \) is an operation
which we may specify explicitly in terms of its operator representation: Let \( \rho_A^{1/2} \) denote the generalised inverse
of \( \rho_A \), 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, \( \rho_A^0 \), onto the support of \( \rho_A \). Then, for any state \( \tau_B \), the operation
\[
\Psi^{B\leftarrow A}\Phi_A^+ = \rho_A^{1/2}\rho_B\rho_A^{1/2} + (1 - \rho_A^0) \otimes \tau_B
\]
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 \( E^{Fx\leftarrow M} \). System \( F \) is transferred noiselessly to Bob, while an operation \( \Lambda_{X'\leftarrow X} \) is applied to \( X \) leaving Bob with system \( Y \). Bob applies an operation \( D^{We\leftarrow FY} \) to \( FY \), leaving him with system \( W \). The overall operation from \( M \) to \( W \) is
\[
D^{We\leftarrow FY}id^{F\leftarrow F} \otimes \Lambda_{X'\leftarrow X}E^{Fx\leftarrow M}.
\]
(44)

Fixing \( D^{We\leftarrow FY} \) and \( E^{Fx\leftarrow M} \), (44) is a linear function of \( \Lambda_{X'\leftarrow X} \) which maps any operation \( \Lambda_{X'\leftarrow X} \) to an operation. In fact it
satisfies a strictly stronger property, which is that if \( \Lambda_{X'\leftarrow X'} \) 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 \( Z^{XW\leftarrow MY} \) via
\[
Z^{XW\leftarrow MY} := D^{We\leftarrow FY}E^{Fx\leftarrow M}.
\]
(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
\[
\forall \rho_Y : Tr_WZ^{XW\leftarrow MY}1_M \otimes \rho_Y = Z^{X\leftarrow M}.
\]
(46)
In particular, $Z_{X\leftarrow M} = \text{Tr}_X \mathcal{E} F^X_{\leftarrow M}$. 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 $\hat{M} M$. The 'transpose trick' tells us that $\mu^{T}_{\hat{M}} = \text{id}^{\hat{M} \leftarrow M} \mu_{\hat{M} M}$. Let

$$\omega^{\hat{M} M} := D^{\hat{M} \leftarrow X} \text{id}^{\hat{X} \leftarrow X} \cdot \mathcal{E} F_{\leftarrow X} M^{\hat{M} \leftarrow M} \mu^{\hat{M} M}_{\hat{M} M}.$$ 

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

$$\rho_{\hat{M} F X} := \rho_{\hat{X} \leftarrow X}^{1/2} \rho_{\hat{X}}^{1/2}. \tag{49}$$

Note that $\rho_{\hat{X}} := \text{Tr}_X \rho_{\hat{M} F X} = \text{id}^{\hat{X} \leftarrow X} \rho_{\hat{X}}$. Then

$$\omega^{\hat{M} F X} := D^{\hat{M} \leftarrow X} \mathcal{E} F_{\leftarrow X} \rho_{\hat{M} F X} = \hat{Z}^{\hat{M} \leftarrow X} \mathcal{E} F_{\leftarrow X} \rho_{\hat{M} F X}.$$ 

where

$$\hat{Z}^{\hat{M} \leftarrow X} := D^{\hat{M} \leftarrow X} \mathcal{E} F_{\leftarrow X}.$$ 

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

$$\rho_{\hat{X}}^{1/2} (\hat{Z}^{\hat{M} \leftarrow X} \mathcal{E} F_{\leftarrow X} \rho_{\hat{X}}) \rho_{\hat{X}}^{1/2} = \mu^{1/2}_{\hat{M}} (Z_{W \leftarrow M} \mathcal{F}_{\leftarrow X} \rho_{\hat{X}}) \mu^{1/2}_{\hat{M}}.$$ 

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, $\mu_{\hat{M}} = \text{id}^{M} / k$. 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

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

so the success probability of the code is

$$1 - \epsilon = \text{Tr} \Pi_{M W} \hat{Z}^{\hat{M} \leftarrow X} \mathcal{E} F_{\leftarrow X} \rho_{\hat{M} F X}. \tag{52}$$

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

$$Z_{W \leftarrow M} = \mathcal{E} F_{\leftarrow X} M \otimes D^{W \leftarrow Y} \tag{53}$$

where $\mathcal{E} F_{\leftarrow X}$ and $D^{W \leftarrow Y}$ 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 $\hat{Z}^{\hat{M} \leftarrow X}$ this condition is

$$\hat{Z}^{\hat{M} \leftarrow X} \rho_{\hat{X}} \otimes \mathcal{E} F_{\leftarrow X} \rho_{\hat{X}}.$$ 

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 $N_{Y \leftarrow X}$ to have the form $N_{Y \leftarrow X} = \sigma_{Y} \text{Tr}_{X}$. 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

$$\forall \sigma_{Y} : \text{Tr} \Pi_{M W} \hat{Z}^{\hat{M} \leftarrow X} \rho_{\hat{X}} \otimes \mathcal{E} F_{\leftarrow X} \rho_{\hat{X}} = 1/k. \tag{55}$$
VII. QUANTUM HYPOTHESIS TESTING CONVERSE

Consider the quantum hypothesis testing problem where the null hypothesis is that the state of $XY$ is $\rho_{XY}$ and the (composite) alternative hypothesis is that that state of $XY$ 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

$$\beta^*_\epsilon (\rho_{XY}) := \min_{0 \leq T_{XY} \leq 1} \max_{\sigma_Y} \frac{1}{2} Tr T_{XY} \rho_X \otimes \sigma_Y$$

subject to $Tr T_{XY} \rho_{XY} \geq 1 - \epsilon$. \hfill (56)

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

$$\beta(\rho_0||\rho_1) := \min \{ Tr T \rho_1 : Tr T \rho_0 \geq 1 - \epsilon, 0 \leq T \leq 1 \}.$$\hfill (57)

By von Neumann’s minimax theorem

$$\beta^*_\epsilon (\rho_{XY}) = \max_{\sigma_Y} \beta (\rho_X \| \rho_X \otimes \sigma_Y).$$

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 $N^{\epsilon - X}$ if and only if there is a quantum hypothesis test $T_{XY}$ satisfying

$$Tr T_{XY} N^{\epsilon - X} \rho_{XY} = 1 - \epsilon,$$ \hfill (59)

$$\forall \sigma_Y : Tr T_{XY} \rho_X \otimes \sigma_Y = 1/k$$ \hfill (60)

where $\rho_{XY} = \rho_X^{1/2} \Phi^+_{XX} \rho_X^{1/2}$.

**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 $Z^{\epsilon_{X} - X}$ to system $XY$, 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 $T_{XY}$ be a test satisfying (59) and and (60), and let

$$Z^{\epsilon_{X} - X} : A_X \mapsto \frac{1}{k} \Pi_{\epsilon_{X} - X} Tr T_{XY} A_X$$

$$+ \frac{1}{k(k - 1)} \left( \Pi_{\epsilon_{X} - X} - \Pi_{\epsilon_{X} - X} \right) Tr (1 - T_{XY}) A_X$$ \hfill (61)

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

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

$$k \leq \min_{\sigma_Y} \beta (N^{\epsilon - X} \rho_X \| \rho_X \otimes \sigma_Y)^{-1}.\hfill (62)$$

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

$$k \leq \max_{\rho_X} \min_{\sigma_Y} \beta (N^{\epsilon - X} \rho_X \| \rho_X \otimes \sigma_Y)^{-1}.\hfill (63)$$

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 $Tr T_{XY} \rho_X \otimes \sigma_Y = \beta$ for all $\sigma_Y$ 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**


