A Converse Coding Theorem for Mismatched Decoding at the Output of Binary-Input Memoryless Channels *

V.B.Balakirsky
The author is from The Data Security Association "Confident",
St.-Petersburg, Russia
E-mails:
vbal@stoic.spb.su
vbal@newton.spb.su
vbal@mathematik.uni-bielefeld.de

Abstract. An upper bound on the maximal transmission rate over binary-input memoryless channels, provided that the decoding decision rule is given, is derived. If the decision rule is equivalent to the maximum likelihood decoding (matched decoding), then the bound coincides with the channel capacity. Otherwise (mismatched decoding), it coincides with a known lower bound.

Key words :
channel capacity, mismatched decoding.

*The work was supported by a Scholarship from the Swedish Institute, Stockholm, Sweden.

1 Introduction

Shannon's coding theorem on the capacity of memoryless channels [1] may be presented as a result of maximization of the transmission rate over all possible block codes and all possible decoding algorithms. An open information theory problem is to generalize this theorem to the case when optimization over decoding algorithms is forbidden [2-10]. More precisely, the decoder calculates the value of some distortion (metric) function for each codeword and makes a decision that the codeword with the smallest distortion was sent. Since the distortion function is not necessarily matched to the channel's characteristics, the maximal transmission rate can be less than the channel capacity. The current state of the mismatched problem and its connections with other open problems of information and coding theory are given in the recent paper by I.Csiszár and P.Narayan [9], and we cannot do it better.

We deal with the binary-input memoryless channels and prove a converse statement to the direct coding theorem [3]. As a result, we obtain that the maximal transmission rate over any binary-input memoryless channel can be found for any distortion function using a single-letter characterization.

The paper is organized as follows. In Section 2 we introduce some notations, which will be used in the analysis. The main result is formulated and discussed in Section 3. In Section 4 we discuss the basic ideas of the proof of the theorem, which lead to the results, called a 'combinatorial approximation lemma' and a 'permutation lemma'. Section 5 is devoted to the proof of the combinatorial approximation lemma, and Section 6 is devoted to the proof of the permutation lemma. Some properties of the mismatched decoding and basic ideas of the proof are illustrated in the Appendix for specific data.

2 Notation

1. The channel's input and output alphabet will be denoted by X and Y, respectively. We assume that $X = \{0,1\}$, but write X instead of $\{0,1\}$ meaning that many of the further considerations can be extended to a general case.

2. The number of symbols $x \in X$ in $\mathbf{x} \in X^n$, the number of symbols $y \in Y$ in $\mathbf{y} \in Y^n$, and the number of pairs of symbols $(x,y) \in X \times Y$

in $(\mathbf{x},\mathbf{y}) \in X^n \times Y^n$ will be denoted by

$$\begin{aligned} n_x(\mathbf{x}) &= \sum_{j=1}^{n} \chi\{\ x_j = x\ \} \\ n_y(\mathbf{y}) &= \sum_{j=1}^{n} \chi\{\ y_j = y\ \} \\ n_{x,y}(\mathbf{x},\mathbf{y}) &= \sum_{j=1}^{n} \chi\{\ x_j = x, y_j = y\ \}. \end{aligned}$$

Hereafter, χ denotes the indicator function of the event in the braces: $\chi\{\beta\} = 1$ if the statement β is true and $\chi\{\beta\} = 0$ otherwise.

3. We introduce the special notation for the empirical probability distributions generated by the numbers $n_x(\mathbf{x})$, $n_y(\mathbf{y})$, and $n_{x,y}(\mathbf{x},\mathbf{y})$, where $(x,y) \in X \times Y$. These distributions will be referred to as the compositions of $\mathbf{x} \in X^n$, $\mathbf{y} \in Y^n$, and $(\mathbf{x},\mathbf{y}) \in X^n \times Y^n$:

$$\begin{aligned} Comp(\ \mathbf{x}\ ) &= \{n_x(\mathbf{x})/n\} \\ Comp(\ \mathbf{y}\ ) &= \{n_y(\mathbf{y})/n\} \\ Comp(\ \mathbf{x},\mathbf{y}\ ) &= \{n_{x,y}(\mathbf{x},\mathbf{y})/n\}. \end{aligned}$$

Furthermore, we introduce the conditional composition of $\mathbf{y} \in Y^n$ given $\mathbf{x} \in X^n$

$$Comp(\ \mathbf{y}|\mathbf{x}\ ) = \{n_{x,y}(\mathbf{x},\mathbf{y})/n_x(\mathbf{x})\}.$$

4. The probability distributions $\{P_x\}$, $\{V_x(y)\}$, and $\{W_x(y)\}$ will be denoted by P, V, and W, respectively.

5. The set of types $\mathcal{P}^n$ on X^n and the set of conditional types $\mathcal{V}_P^n$ on Y^n given $P \in \mathcal{P}^n$ are introduced as

$$\begin{aligned} \mathcal{P}^n &= \{\ P:\ \ n \cdot P_x \ \text{ are integers for all } \ x \in X\ \} \\ \mathcal{V}_P^n &= \{\ V:\ \ n \cdot P_x \cdot V_x(y) \ \text{ are integers for all } \ (x,y) \in X \times Y\ \}, \end{aligned}$$

The set of sequences of type $P \in \mathcal{P}^n$ and the set of sequences of conditional type $V \in \mathcal{V}_P^n$ given $\mathbf{x} \in \mathbf{T}_P^n$ are denoted by

$$\begin{aligned} \mathbf{T}_P^n &= \{\ \mathbf{x} \in X^n:\ \ Comp(\ \mathbf{x}\ ) = P\ \} \\ \mathbf{T}_V^n(\mathbf{x}) &= \{\ \mathbf{y} \in Y^n:\ \ Comp(\ \mathbf{y}|\mathbf{x}\ ) = V\ \}. \end{aligned}$$

6. A code G^n of rate R and length n, consisting of e^{nR} codewords, such that each codeword has the type P, i.e.,

$$Comp(\ \mathbf{x}\ ) = P, \quad \text{for all } \ \mathbf{x} \in G^n$$

will be referred to as a P-composition code and denoted by G_P^n.

7. The distortion (metric) function is introduced as $d = \{d_x(y)\}$, where

$$\begin{aligned} d_0(y) &= 0 \\ 0 \leq \ d_x(y) &\leq d_{max} < \infty, \quad \text{for all } \ x \in X, y \in Y. \end{aligned}$$

An additive extension of this function is defined as

$$d(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^{n} d_{x_j}(y_j).$$

8. We use the following notation for the marginal distribution PV on Y, the entropy function $H(PV)$, the conditional entropy function $H(V|P)$, the mutual information function $I(P, V)$, and the average distortion function in the ensemble $\{\ X \times Y,\ P_x \cdot V_x(y)\ \}$:

$$\begin{aligned} PV(y) &= \sum_x P_x \cdot V_x(y) \\ H(PV) &= -\sum_y PV(y) \cdot \ln PV(y) \\ H(V|P) &= -\sum_{x,y} P_x \cdot V_x(y) \cdot \ln V_x(y) \\ I(P, V) &= H(PV) - H(V|P) \\ \\ d(P, V) &= \sum_{x,y} P_x \cdot V_x(y) \cdot d_x(y). \end{aligned}$$

9. To simplify formalization, we will write

$$|\tilde{V} - V| = \max_{x,y} |\tilde{V}_x(y) - V_x(y)|$$

for any conditional probability distributions $\tilde{V} = \{\tilde{V}_x(y)\}$ and $V = \{V_x(y)\}$.

The notation above is conventional, and is given to make the paper self-contained. The function, introduced below, seems to be new. This function will be used thoughout the paper.

Definition 2.1 (Upsilon Notation): Let $\{\alpha_n\}$ be a given sequence. We introduce a Boolean function $\Upsilon(\{\alpha_n\})$ such that $\Upsilon(\{\alpha_n\})$ = 'TRUE' if and only if (iff) the following statement: "there exist an $\varepsilon > 0$ and $n_0(\varepsilon) < \infty$ such that $\alpha_n > \varepsilon$ for all $n > n_0(\varepsilon)$", is valid. Otherwise, $\Upsilon(\{\alpha_n\})$ = 'FALSE'. For the values of the function Υ we will use the relation $' \leq'$ meaning that

$$\begin{aligned} \text{'FALSE'} &\leq \text{'FALSE'} \\ \text{'FALSE'} &\leq \text{'TRUE'} \\ \text{'TRUE'} &\leq \text{'TRUE'}. \end{aligned}$$

For any given sequences, $\{\alpha_n\}$ and $\{\beta_n\}$, we say that $\{\beta_n\}$ approximates $\{\alpha_n\}$ if

$$\Upsilon(\{\alpha_n\}) \leq \Upsilon(\{\beta_n\})$$

i.e.,

$$\Upsilon(\{\alpha_n\}) = \text{'TRUE'} \implies \Upsilon(\{\beta_n\}) = \text{'TRUE'}$$

and that $\{\alpha_n\}$ approximates $\{\beta_n\}$ if

$$\Upsilon(\{\beta_n\}) \leq \Upsilon(\{\alpha_n\})$$

i.e.,

$$\Upsilon(\{\beta_n\}) = \text{'TRUE'} \implies \Upsilon(\{\alpha_n\}) = \text{'TRUE'}.$$

If both statements are valid, we say that $\{\alpha_n\}$ and $\{\beta_n\}$ approximate each other. In this case we write

$$\Upsilon(\{\alpha_n\}) = \Upsilon(\{\beta_n\})$$

i.e.,

$$\Upsilon(\{\alpha_n\}) = \text{'TRUE'} \iff \Upsilon(\{\beta_n\}) = \text{'TRUE'}.$$

3 Statement of the Problem and a Converse Coding Theorem

Our considerations of the mismatched decoding for a memoryless channel W are based on constructing the references to a memoryless channel V, which can be called a 'test-chanel'. For formal convenience, we introduce the test channel as a real channel, existing between the sender and the receiver, and consider the system model for information transmission given in Fig.1.

Let us suppose that a P-composition block code G_P^n is used to transmit data over two parallel memoryless channels $V = \{V_x(y)\}$ and $W = \{W_x(y')\}$. The conditional probabilities that the decoders $\mathcal{D}$ and $\mathcal{D}'$ receive the vectors $\mathbf{y} = (y_1, ..., y_n) \in Y^n$ and $\mathbf{y}' = (y_1', ..., y_n') \in Y^n$, when a codeword $\mathbf{x} = (x_1, ..., x_n) \in G_P^n$ was sent, are given as

$$\begin{aligned} V(\mathbf{y}|\mathbf{x}) &= \prod_{j=1}^{n} V_{x_j}(y_j) \\ W(\mathbf{y}'|\mathbf{x}) &= \prod_{j=1}^{n} W_{x_j}(y_j'). \end{aligned} \tag{3.1}$$

We suppose that the decoders $\mathcal{D}$ and $\mathcal{D}'$ estimate the transmitted codeword as $\hat{\mathbf{x}} \in G_P^n$ and $\hat{\mathbf{x}}' \in G_P^n$, respectively, using the same partitioning of the space Y^n with respect to the minimal value of the distortion (metric) function d, i.e.,

$$\begin{aligned} d(\hat{\mathbf{x}}, \mathbf{y}) &= \min_{\mathbf{x} \in G_P^n} d(\mathbf{x}, \mathbf{y}) \\ d(\hat{\mathbf{x}}', \mathbf{y}') &= \min_{\mathbf{x}' \in G_P^n} d(\mathbf{x}', \mathbf{y}'). \end{aligned} \tag{3.2}$$

If the minimum is attained for several codewords we assume that the received vector is decoded incorrectly. Then the decoding error probabilities for the codeword $\mathbf{x}$ can be expressed as

$$\begin{aligned} P_d^{(n)}(\mathbf{x}, V) &= \sum_{\mathbf{y}} V(\mathbf{y}|\mathbf{x}) \cdot \chi\{\ d(\mathbf{x}, \mathbf{y}) \geq D(\mathbf{x}, \mathbf{y})\ \} \\ P_d^{(n)}(\mathbf{x}, W) &= \sum_{\mathbf{y}'} W(\mathbf{y}'|\mathbf{x}) \cdot \chi\{\ d(\mathbf{x}, \mathbf{y}') \geq D(\mathbf{x}, \mathbf{y}')\ \}, \end{aligned} \tag{3.3}$$

where

$$\begin{aligned} D(\mathbf{x}, \mathbf{y}) &= \min_{\hat{\mathbf{x}} \in G_P^n \setminus \{\mathbf{x}\}} d(\hat{\mathbf{x}}, \mathbf{y}) \\ D(\mathbf{x}, \mathbf{y}') &= \min_{\hat{\mathbf{x}}' \in G_P^n \setminus \{\mathbf{x}\}} d(\hat{\mathbf{x}}', \mathbf{y}'). \end{aligned} \tag{3.4}$$

Let

$$
\begin{aligned}
P_d^{(n)}(V) &= \max_{\mathbf{x}\in G_P^n} P_d^{(n)}(\mathbf{x}, V) \\
P_d^{(n)}(W) &= \max_{\mathbf{x}\in G_P^n} P_d^{(n)}(\mathbf{x}, W)
\end{aligned}
\tag{3.5}
$$

denote the maximal decoding error probabilities for the code G_P^n, and let $\{P_d^{(n)}(V)\}$ and $\{P_d^{(n)}(W)\}$ denote the sequences of the maximal decoding error probabilities constructed for a sequence of codes $\{G_P^n\}$ having a fixed rate R.

The upsilon notation, introduced in Section 2, allows us to represent the lower bound on the maximal transmission rate for mismatched decoding as a corollary of the result, formulated below.

Theorem 3.1: Let P be a given probability distribution on $X = \{0, 1\}$ and let W be a given binary-input memoryless channel. Let

$$
\begin{aligned}
\Pi_d(P, W) = \{\ V:\ & PV = PW, \\
& d(P, V) \le d(P, W)\ \}.
\end{aligned}
\tag{3.6}
$$

Then

$$
\Upsilon(\{P_d^{(n)}(V)\}) \le \Upsilon(\{P_d^{(n)}(W)\}) \tag{3.7}
$$

for all $V \in \Pi_d(P, W)$.

Corollary 3.2: If

$$
R > C_d(P, W), \tag{3.8}
$$

where

$$
C_d(P, W) = \min_{V\in\Pi_d(P,W)} I(P, V) \tag{3.9}
$$

and the set $\Pi_d(P, W)$ is defined in (3.6), then

$$
\Upsilon(\{P_d^{(n)}(W)\}) = \text{'TRUE'} \tag{3.10}
$$

for all P-composition codes G_P^n, i.e., there exist $\varepsilon_d(P, W) > 0$ and $n_0(\varepsilon, d, P, W) < \infty$ such that

$$
P_d^{(n)}(W) \ge \varepsilon_d(P, W) \ \text{ for all } \ n > n_0(\varepsilon, d, P, W)
$$

and all P-composition codes G_P^n.

Proof: For a sequence of P-composition codes $\{G_P^n\}$, we use a converse statement to the coding theorem for the channel V, which minimizes the mutual information at the right hand side of (3.9), and conclude that, if (3.8) is valid, then

$$\Upsilon(\{P_d^{(n)}(V)\}) = \text{'TRUE'}. \tag{3.11}$$

Combining (3.7) and (3.11), we obtain (3.10). Q.E.D.

Corollary 3.3: If

$$R > C_d(W) \tag{3.12}$$

where

$$C_d(W) = \max_P C_d(P, W) \tag{3.13}$$

and the function $C_d(P, W)$ is defined in (3.9), then

$$\Upsilon(\{P_d^{(n)}(W)\}) = \text{'TRUE'} \tag{3.14}$$

for all codes G^n, i.e., there exist $\varepsilon_d(W) > 0$ and $n_0(\varepsilon, d, W) < \infty$ such that

$$P_d^{(n)}(W) \geq \varepsilon_d(W), \text{ for all } n > n_0(\varepsilon, d, W) \tag{3.15}$$

and all codes G^n.

Proof: There exists P such that any sequence of codes $\{G^n\}$, having a fixed code rate, contains a subsequence of P-composition codes $\{G_P^n\}$, having asymptotically the same rate [2]. Thus, (3.14) follows from (3.10). Q.E.D.

Corollary 3.4: Let $PW_0 = \{PW_0(y)\}$ be a given probability distribution on Y and let $\alpha \leq 0$ be a given constant. If

$$V_x(y) = f_x \cdot \varphi(y) \cdot e^{\alpha d_x(y)} \tag{3.16}$$

where $\{f_x\}$ and $\{\varphi(y)\}$ are chosen in such a way that

$$\sum_y V_x(y) = 1, \quad \text{for all } x \in X \tag{3.17}$$

and

$$\sum_x P_x \cdot V_x(y) = PW_0(y) \text{ for all } y \in Y \tag{3.18}$$

then the maximal transmission rate, when a P-composition code is used, is not greater than

$$C_d(P,W) = H(PW_0) - H(V|P)$$

for all binary-input channels W such that

$$\begin{aligned} PW &= PW_0 \\ d(P,W) &\geq d(P,V). \end{aligned} \tag{3.19}$$

Proof: If the marginal distribution on Y is fixed by PW_0, then the minimization of the mutual information function at the right hand side of (3.9) is equivalent to the maximization of the conditional entrropy function $H(V|P)$ under the linear restrictions (3.17) and (3.18). This function is convex up, and we obtain (3.16) as a result of optimization using Lagrange multipliers. The restrictions (3.19) coincide with the restrictions at the right hand side of (3.6) when we consider them as restrictions on W given PW and V. Q.E.D.

Corollary 3.5: For any P-composition code G_P^n, decoding, which minimizes the distortion function d, is equivalent to the maximum likelihood decoding for the channel V, given in (3.16), and the maximal transmission rate is not greater than $I(P,V)$.

Proof: Using (3.1) and (3.16) we write :

$$\ln V(\mathbf{y}|\mathbf{x}) = n \cdot \sum_x P_x \cdot f_x + \sum_{j=1}^{n} \varphi(y_j) + \alpha \cdot d(\mathbf{x},\mathbf{y}).$$

Since $\alpha < 0$, maximization of the function $\ln V(\mathbf{y}|\mathbf{x})$ over all codewords is equivalent to the minimization of the distortion function $d(\mathbf{x},\mathbf{y})$. Q.E.D.

Discussion: Let V be a probability distribution minimizing the mutual information at the right hand side of (3.9). We are interested in the case

$$I(P,V) < R < I(P,W)$$

because, otherwise, the maximum-likelihood decoding does not achieve an arbitrarily small error probability. Let

$$\mathbf{A}_d^n(\mathbf{x}) = \{\ \mathbf{y} :\ d(\mathbf{x},\mathbf{y}) \leq n \cdot d(P,W)\ \}.$$

Then $\mathbf{T}_V^n(\mathbf{x}), \mathbf{T}_W^n(\mathbf{x}) \subseteq \mathbf{A}_d^n(\mathbf{x})$. Roughly speaking, a vector $\mathbf{y} \in \mathbf{T}_W^n(\mathbf{x})$ will be realized as a result of transmission of $\mathbf{x}$ over the channel W. Since $R < I(P,W)$, we can choose a code G_P^n, for which "almost" all the vectors belonging to $\mathbf{T}_W^n(\mathbf{x})$ do not coincide with the vectors belonging to $\mathbf{T}_W^n(\mathbf{x}')$, $\mathbf{x}' \in G_P^n \backslash \{\mathbf{x}\}$. However, to analyze the decoding error probability in our case it is necessary to estimate the number of d-"bad" points for $\mathbf{x}$, i.e., the size of the intersection of $\mathbf{T}_W^n(\mathbf{x})$ and $\mathbf{A}_d^n(\mathbf{x}')$, $\mathbf{x}' \in G_P^n \backslash \{\mathbf{x}\}$. As it is well-known [2],

$$\begin{aligned} \ln |\mathbf{T}_W^n(\mathbf{x})|/n &\simeq H(W|P) \\ \ln |\mathbf{T}_V^n(\mathbf{x})|/n &\simeq H(V|P) \end{aligned}$$

and because V minimizes the mutual information at the right side of (3.9), it maximizes the conditional entropy function $H(V|P)$. Therefore, most of the vectors belonging to $\mathbf{A}_d^n(\mathbf{x})$ have the conditional type V. Since $R > I(P,V)$, the size of the intersection of $\mathbf{T}_V^n(\mathbf{x})$ and the union of $\mathbf{T}_V^n(\mathbf{x}')$, $\mathbf{x}' \in G_P^n \backslash \{\mathbf{x}\}$ is asymptotically the same as $|\mathbf{T}_V^n(\mathbf{x})|$ (note that we use only a weak converse statement for the channel V and lower-bound this size as $\varepsilon \cdot |\mathbf{T}_V^n(\mathbf{x})|$, where $\varepsilon > 0$ does not depend on n). In fact, the main result of the paper is the statement that, for binary-input channels, this condition is sufficient to show that the size of the intersection between $\mathbf{T}_W^n(\mathbf{x})$ and the union of $\mathbf{A}_d^n(\mathbf{x}')$, $\mathbf{x}' \in G_P^n \backslash \{\mathbf{x}\}$ is asymptotically the same as $|\mathbf{T}_W^n(\mathbf{x})|$.

4 Basic Ideas of the Proof of the Converse Coding Theorem

4.1 Combinatorial Approximation of Memoryless Channels

Let us consider the transmission of a codeword $\mathbf{x} \in G_P^n$ over a memoryless channel V. Suppose that V is a type on Y^n given $P \in \mathcal{P}^n$, i.e., $V \in \mathcal{V}_P^n$. Then we may write :

$$Comp(\mathbf{y}|\mathbf{x}) \approx V \text{ with the high probability}$$

meaning that, with the high probability, the conditional compositions of the received sequences given $\mathbf{x}$ are close to the conditional type V. The exact formulation of this statement is shown below.

Lemma 4.1 ([2, Section 2.1]): Let

$$\mathcal{N}(P,V) = \{ \ n : \ P \in \mathcal{P}^n \text{ and } V \in \mathcal{V}_P^n \ \}$$

be the set of lengths such that a given probability distribution P is a type on X^n, and a given conditional probability distribution V is a conditional type on Y^n for that P. For any increasing sequence $\{n\}$, whose elements belongs to $\mathcal{N}(P,V)$, there exist sequences $\{\delta_n\}$ such that

$$\delta_n \to 0, \quad \delta_n \sqrt{n} \to \infty, \quad \text{as} \quad n \to \infty$$

and, for any $\mathbf{x} \in \mathbf{T}_P^n$,

$$\sum_{\mathbf{y}} V(\mathbf{y}|\mathbf{x}) \cdot \chi\{ \ |Comp(\mathbf{y}|\mathbf{x}) - V| < \delta_n \ \} \geq 1 - \gamma_n \tag{4.1}$$

where $\{\gamma_n\}$ is a sequence, depending on $\{\delta_n\}$, such that $\gamma_n \to 0$, as $n \to \infty$.

Convention 4.2: We assume that $P_x > 0$ for all $x \in X$, and that $n \in \mathcal{N}(P,V)$. In later considerations, when we deal with two channels, V and W, we assume that

$$n \in \mathcal{N}(P,V) \bigcap \mathcal{N}(P,W).$$

The sequence $\{\delta_n\}$, satisfying the conditions of Lemma 4.1, is assumed to be given.

Let

$$[V] = \{ \ \tilde{V} \in \mathcal{V}_P^n : \ |\tilde{V} - V| < \delta_n \ \}. \tag{4.2}$$

Then the inequality (4.1) can be rewritten as

$$V(\mathbf{T}_{[V]}^n(\mathbf{x})|\mathbf{x}) \geq 1 - \gamma_n \tag{4.3}$$

for any $\mathbf{x} \in \mathbf{T}_P^n$, where

$$\mathbf{T}_{[V]}^n(\mathbf{x}) = \bigcup_{\tilde{V} \in [V]} \mathbf{T}_{\tilde{V}}^n(\mathbf{x}). \tag{4.4}$$

Note also that

$$d(\mathbf{x},\mathbf{y})/n = d(P,V), \quad \text{for all } \ \mathbf{y} \in \mathbf{T}_V^n(\mathbf{x}) \tag{4.5}$$

and

$$|d(\mathbf{x},\mathbf{y})/n - d(P,V)| < \delta_n \cdot d_{max}, \quad \text{for all } \mathbf{y} \in \mathbf{T}^n_{[V]}(\mathbf{x}). \tag{4.6}$$

The conditional probability to receive $\mathbf{y}$ depends only on $Comp(\mathbf{y}|\mathbf{x})$, and, for all $\tilde{V} \in \mathcal{V}^n_P$, we may write

$$V(\mathbf{y}|\mathbf{x}) = p(\tilde{V}|V) \cdot \frac{1}{|\mathbf{T}^n_{\tilde{V}}(\mathbf{x})|} \ \text{ if } \ Comp(\mathbf{y}|\mathbf{x}) = \tilde{V} \tag{4.7}$$

where

$$p(\tilde{V}|V) = |\mathbf{T}^n_{\tilde{V}}(\mathbf{x})| \cdot \exp\{\ n \cdot \sum_{x,y} P_x \cdot \tilde{V}_x(y) \cdot \ln V_x(y)\ \}. \tag{4.8}$$

Therefore, information transmission over a memoryless channel V can be represented as a choice of the conditional composition in accordance with the distribution $p(\tilde{V}|V)$, $\tilde{V} \in \mathcal{V}^n_P$, and a choice of a particular received sequence in accordance with the uniform distribution on $\mathbf{T}^n_{\tilde{V}}(\mathbf{x})$. The step, which we call 'a combinatorial approximation', consists in the substitution of different probabilities, $p'(\tilde{V}|V)$, $\tilde{V} \in \mathcal{V}^n_P$, for $p(\tilde{V}|V)$, $\tilde{V} \in \mathcal{V}^n_P$, into the expression at the right hand side of (4.7). Then we obtain a different channel, which may have memory. For example, we can assign $p'(\tilde{V}|V)$ as a uniform distribution on the set $[V]$ and, because of (4.3), expect that the decoding error probability for such a channel is approximately the same as for V. However, we will use a less tight approximation and assign $p'(\tilde{V}|V)$ as the indicator function of the event $\tilde{V} = V$. The definition below formalizes this step, and the statement of Lemma 4.4 shows that this approximation can be used for our purposes.

The considerations above are also valid for the channel W, and we continue the parallel definitions of Section 3.

Definition 4.3: The channels $V^n = \{V^n(\mathbf{y}|\mathbf{x})\}$ and $W^n = \{W^n(\mathbf{y}'|\mathbf{x})\}$, where

$$\begin{aligned} V^n(\mathbf{y}|\mathbf{x}) &= \begin{cases} |\mathbf{T}^n_V(\mathbf{x})|^{-1}, & \text{if } \mathbf{y} \in \mathbf{T}^n_V(\mathbf{x}) \\ 0, & \text{otherwise} \end{cases} \\ W^n(\mathbf{y}'|\mathbf{x}) &= \begin{cases} |\mathbf{T}^n_W(\mathbf{x})|^{-1}, & \text{if } \mathbf{y}' \in \mathbf{T}^n_W(\mathbf{x}) \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{4.9}$$

will be referred to as combinatorial channels V^n and W^n. For these channels we define the decoding error probabilities for a codeword $\mathbf{x} \in G^n_P$:

$$P^{(n)}_d(\mathbf{x}, V^n) \ = \ \sum_{\mathbf{y}} V^n(\mathbf{y}|\mathbf{x}) \cdot \chi\{\ d(\mathbf{x},\mathbf{y}) \geq D(\mathbf{x},\mathbf{y})\ \} \tag{4.10}$$

$$P_d^{(n)}(\mathbf{x}, W^n) \quad = \quad \sum_{\mathbf{y}'} W^n(\mathbf{y}'|\mathbf{x}) \cdot \chi\{\ d(\mathbf{x}, \mathbf{y}') \geq D(\mathbf{x}, \mathbf{y}')\ \}$$

the maximal decoding error probabilities :

$$\begin{aligned} P_d^{(n)}(V^n) &= \max_{\mathbf{x} \in G_P^n} P_d^{(n)}(\mathbf{x}, V^n) \\ P_d^{(n)}(W^n) &= \max_{\mathbf{x} \in G_P^n} P_d^{(n)}(\mathbf{x}, W^n) \end{aligned} \tag{4.11}$$

and the sequences $\{P_d^{(n)}(V^n)\}$, $\{P_d^{(n)}(W^n)\}$ in the same way as for the channels V and W in (3.3)-(3.5).

Lemma 4.4 (Combinatorial Approximation Lemma): The sequences of the maximal decoding error probabilities for the memoryless channels V and W, defined in (3.1), and for the combinatorial channels V^n and W^n, defined in (4.9), approximate each other, i.e.,

$$\begin{aligned} \Upsilon(\{P_d^{(n)}(V)\}) &= \Upsilon(\{P_d^{(n)}(V^n)\}) \\ \Upsilon(\{P_d^{(n)}(W)\}) &= \Upsilon(\{P_d^{(n)}(W^n)\}). \end{aligned} \tag{4.12}$$

4.2 Permutation Lemma

The results of the previous subsection are valid for any memoryless channel. In particular, they are valid for a channel V, satisfying the restrictions

$$\begin{aligned} V &\in \mathcal{V}_P^n \\ I(P, V) &\leq I(P, W) \\ PV &= PW \\ d(P, V) &\leq d(P, W). \end{aligned} \tag{4.13}$$

The idea of the proof of the theorem is to connect the maximal decoding error probabilities for combinatorial channels V^n and W^n in such a way that $\{P_d^{(n)}(W^n)\}$ approximates $\{P_d^{(n)}(V^n)\}$, i.e.,

$$\Upsilon(\{P_d^{(n)}(V^n)\}) \leq \Upsilon(\{P_d^{(n)}(W^n)\}).$$

Then, based on the combinatorial approximation lemma, we can use this inequality in the middle part of a logical chain whose first part is the statement

$$\Upsilon(\{P_d^{(n)}(V)\}) = \Upsilon(\{P_d^{(n)}(V^n)\})$$

and the last part is the statement

$$\Upsilon(\{P_d^{(n)}(W^n)\}) = \Upsilon(\{P_d^{(n)}(W)\}).$$

As a result, we obtain (3.7) and prove the theorem.

Lemma 4.5 (Permutation Lemma): Let the probability distributions $P \in \mathcal{P}^n$ and $W \in \mathcal{V}_P^n$ be given, and let G_P^n be a P-composition binary code. Let V be some probability distribution satisfying the restrictions (4.13). Then the sequence of the maximal decoding error probability for the combinatorial channel W^n approximates the sequence of the maximal decoding error probability for the combinatorial channel V^n, i.e.,

$$\Upsilon(\{P_d^{(n)}(V^n)\}) \leq \Upsilon(\{P_d^{(n)}(W^n)\}). \tag{4.14}$$

5 Proof of the Combinatorial Approximation Lemma

The distortion (metric) function was introduced in Section 2 in such a way that, for binary-input channels,

$$\begin{aligned} & d_0(y) &= 0 \\ 0 \leq\ & d_1(y) &\leq d_{max} < \infty, \quad \text{for all } y \in Y. \end{aligned}$$

Without loss of generality, we assume that

$$d_1(0) = 0. \tag{5.1}$$

Proposition 5.1: Let $\underline{V}^n = \{\underline{V}^n(\mathbf{y}|\mathbf{x})\}$ and $\overline{V}^n = \{\overline{V}^n(\mathbf{y}|\mathbf{x})\}$ be the combinatorial channels constructed for the memoryless channels $\underline{V} = \{\underline{V}_x(y)\}$ and $\overline{V} = \{\overline{V}_x(y)\}$ such that

$$\underline{V}_0(y) = V_0(y) = \overline{V}_0(y), \quad \text{for all } y \in Y \tag{5.2}$$

and

$$\begin{aligned} \underline{V}_1(y) &= \begin{cases} V_1(y) + (|Y|-1)\delta_n, & \text{if } y = 0 \\ V_1(y) - \delta_n, & \text{if } y \neq 0 \end{cases} \qquad (5.3) \\ \overline{V}_1(y) &= \begin{cases} V_1(y) - (|Y|-1)\delta_n, & \text{if } y = 0 \\ V_1(y) + \delta_n, & \text{if } y \neq 0. \end{cases} \end{aligned}$$

Then the following statements :

$$\Upsilon(\{P_d^{(n)}(\underline{V}^n)\}) \leq \Upsilon(\{P_d^{(n)}(V)\}) \leq \Upsilon(\{P_d^{(n)}(\overline{V}^n)\}) \tag{5.4}$$

are valid, where

$$\begin{aligned} P_d^{(n)}(\underline{V}^n) &= \max_{\mathbf{x}\in G_P^n} P_d^{(n)}(\mathbf{x}, \underline{V}^n) \\ P_d^{(n)}(\overline{V}^n) &= \max_{\mathbf{x}\in G_P^n} P_d^{(n)}(\mathbf{x}, \overline{V}^n) \end{aligned}$$

and

$$\begin{aligned} P_d^{(n)}(\mathbf{x}, \underline{V}^n) &= \sum_{\mathbf{y}} \underline{V}^n(\mathbf{y}|\mathbf{x}) \cdot \chi\{\, d(\mathbf{x},\mathbf{y}) \geq D(\mathbf{x},\mathbf{y}) \,\} \\ P_d^{(n)}(\mathbf{x}, \overline{V}^n) &= \sum_{\mathbf{y}} \overline{V}^n(\mathbf{y}|\mathbf{x}) \cdot \chi\{\, d(\mathbf{x},\mathbf{y}) \geq D(\mathbf{x},\mathbf{y}) \,\}. \end{aligned}$$

Proof of the Combinatorial Approximation Lemma Based on Proposition 5.1: Let us note that Proposition 5.1 also gives the inequalities :

$$\begin{aligned} \Upsilon(\{P_d^{(n)}(V')\}) &\leq \Upsilon(\{P_d^{(n)}(\underline{V}^n)\}) \\ \Upsilon(\{P_d^{(n)}(\overline{V}^n)\}) &\leq \Upsilon(\{P_d^{(n)}(V'')\}) \end{aligned} \tag{5.5}$$

where

$$V_0'(y) = V_0(y) = V_0''(y), \quad \text{for all} \;\; y \in Y$$

and

$$\begin{aligned} V_x'(y) &= \begin{cases} V_1(y) + 2(|Y|-1)\delta_n, & \text{if } \; y = 0 \\ V_1(y) - 2\delta_n, & \text{if } \; y \neq 0 \end{cases} \\ V_x''(y) &= \begin{cases} V_1(y) - 2(|Y|-1)\delta_n, & \text{if } \; y = 0 \\ V_1(y) + 2\delta_n, & \text{if } \; y \neq 0. \end{cases} \end{aligned}$$

Using the converse statement to the coding theorem for the channels V', V, and V'' we obtain :

$$\Upsilon(\{P_d^{(n)}(V')\}) = \Upsilon(\{P_d^{(n)}(V)\}) = \Upsilon(\{P_d^{(n)}(V'')\}). \tag{5.6}$$

Hence, (5.4)-(5.6) lead to the statement

$$\Upsilon(\{P_d^{(n)}(\underline{V}^n)\}) = \Upsilon(\{P_d^{(n)}(V)\}) = \Upsilon(\{P_d^{(n)}(\overline{V}^n)\}). \tag{5.7}$$

However, as it is easy to see

$$\Upsilon(\{P_d^{(n)}(\underline{V}^n)\}) \le \Upsilon(\{P_d^{(n)}(V^n)\}) \le \Upsilon(\{P_d^{(n)}(\overline{V}^n)\}) \tag{5.8}$$

and combining (5.7), (5.8) we complete the proof of the combinatorial approximation lemma.

Proof of Proposition 5.1: The idea of the proof is to construct references to the vectors $\underline{\mathbf{y}} \in \mathbf{T}^n_{\underline{V}}(\mathbf{x})$ and $\overline{\mathbf{y}} \in \mathbf{T}^n_{\overline{V}}(\mathbf{x})$ located from the received vector $\mathbf{y} \in \mathbf{T}^n_{[V]}(\mathbf{x})$ at the minimal Hamming distance.

Let us introduce the sets

$$\begin{aligned}
\underline{\mathbf{S}}^n_V(\mathbf{x},\mathbf{y}) &= \{\, \underline{\mathbf{y}} \in \mathbf{T}^n_{\underline{V}}(\mathbf{x}) :\ d_H(\mathbf{y},\underline{\mathbf{y}}) = \min_{\mathbf{y}' \in \mathbf{T}^n_{\underline{V}}(\mathbf{x})} d_H(\mathbf{y},\mathbf{y}') \,\} \\
\overline{\mathbf{S}}^n_V(\mathbf{x},\mathbf{y}) &= \{\, \overline{\mathbf{y}} \in \mathbf{T}^n_{\overline{V}}(\mathbf{x}) :\ d_H(\mathbf{y},\overline{\mathbf{y}}) = \min_{\mathbf{y}' \in \mathbf{T}^n_{\overline{V}}(\mathbf{x})} d_H(\mathbf{y},\mathbf{y}') \,\}.
\end{aligned}$$

Then, using (5.1), we note that

$$\begin{aligned}
&\chi\{\, d(\mathbf{x},\underline{\mathbf{y}}) \ge D(\mathbf{x},\underline{\mathbf{y}}) \,\} \\
&\qquad \le \chi\{\, d(\mathbf{x},\mathbf{y}) \ge D(\mathbf{x},\mathbf{y}) \,\} \le \\
&\qquad\qquad \chi\{\, d(\mathbf{x},\overline{\mathbf{y}}) \ge D(\mathbf{x},\overline{\mathbf{y}}) \,\}
\end{aligned} \tag{5.9}$$

for all $\underline{\mathbf{y}} \in \underline{\mathbf{S}}^n_V(\mathbf{x},\mathbf{y})$ and $\overline{\mathbf{y}} \in \overline{\mathbf{S}}^n_V(\mathbf{x},\mathbf{y})$. Let us also introduce the uniform distributions on the sets $\underline{\mathbf{S}}^n_V(\mathbf{x},\mathbf{y})$ and $\overline{\mathbf{S}}^n_V(\mathbf{x},\mathbf{y})$:

$$\begin{aligned}
\underline{R}^n_V(\underline{\mathbf{y}}|\mathbf{x},\mathbf{y}) &= \begin{cases} |\underline{\mathbf{S}}^n_V(\mathbf{x},\mathbf{y})|^{-1}, & \text{if } \underline{\mathbf{y}} \in \underline{\mathbf{S}}^n_V(\mathbf{x},\mathbf{y}) \\ 0, & \text{otherwise} \end{cases} \\
\overline{R}^n_V(\overline{\mathbf{y}}|\mathbf{x},\mathbf{y}) &= \begin{cases} |\overline{\mathbf{S}}^n_V(\mathbf{x},\mathbf{y})|^{-1}, & \text{if } \overline{\mathbf{y}} \in \overline{\mathbf{S}}^n_V(\mathbf{x},\mathbf{y}) \\ 0, & \text{otherwise.} \end{cases}
\end{aligned} \tag{5.10}$$

Note that

$$\sum_{\mathbf{y} \in \mathbf{T}^n_{[V]}(\mathbf{x})} V(\mathbf{y}|\mathbf{x}) \cdot \underline{R}^n_V(\underline{\mathbf{y}}|\mathbf{x},\mathbf{y}) = \underline{V}^n(\underline{\mathbf{y}}|\mathbf{x}). \tag{5.11}$$

To prove (5.11), we use the symmetry properties of the sets $\mathbf{T}^n_{[V]}(\mathbf{x})$ and $\mathbf{T}^n_{\underline{V}}(\mathbf{x})$ and conclude that the sum at the left hand side of (5.11) is the same for all $\underline{\mathbf{y}} \in \mathbf{T}^n_{\underline{V}}(\mathbf{x})$ and that this sum is equal to zero if $\underline{\mathbf{y}} \notin \mathbf{T}^n_{\underline{V}}(\mathbf{x})$. Therefore, this sum gives the uniform distribution on $\mathbf{T}^n_{\underline{V}}(\mathbf{x})$, which is $\underline{V}^n(\underline{\mathbf{y}}|\mathbf{x})$. Using (5.9)-(5.11), we write :

$$\begin{aligned}
P_d^{(n)}(\mathbf{x}, V) &\geq \sum_{\mathbf{y}\in\mathbf{T}^n_{[V]}(\mathbf{x})} V(\mathbf{y}|\mathbf{x}) \cdot \chi\{\ d(\mathbf{x},\mathbf{y}) \geq D(\mathbf{x},\mathbf{y})\ \} \qquad (5.12)\\
&= \sum_{\mathbf{y}\in\mathbf{T}^n_{[V]}(\mathbf{x})} V(\mathbf{y}|\mathbf{x}) \cdot \sum_{\underline{\mathbf{y}}} \underline{R}^n_V(\underline{\mathbf{y}}|\mathbf{x},\mathbf{y}) \cdot \chi\{\ d(\mathbf{x},\mathbf{y}) \geq D(\mathbf{x},\mathbf{y})\ \}\\
&\geq \sum_{\mathbf{y}\in\mathbf{T}^n_{[V]}(\mathbf{x})} V(\mathbf{y}|\mathbf{x}) \cdot \sum_{\underline{\mathbf{y}}} \underline{R}^n_V(\underline{\mathbf{y}}|\mathbf{x},\mathbf{y}) \cdot \chi\{\ d(\mathbf{x},\underline{\mathbf{y}}) \geq D(\mathbf{x},\underline{\mathbf{y}})\ \}\\
&= \sum_{\underline{\mathbf{y}}} \underline{V}^n(\underline{\mathbf{y}}|\mathbf{x}) \cdot \chi\{\ d(\mathbf{x},\underline{\mathbf{y}}) \geq D(\mathbf{x},\underline{\mathbf{y}})\ \}\\
&= P_d^{(n)}(\mathbf{x}, \underline{V}^n).
\end{aligned}$$

Similar considerations lead to the inequality :

$$P_d^{(n)}(\mathbf{x}, V) \leq \gamma_n + P_d^{(n)}(\mathbf{x}, \overline{V}^n). \qquad (5.13)$$

Since the inequalities (5.12) and (5.13) are valid for all $\mathbf{x} \in G^n_P$, we obtain

$$P_d^{(n)}(\underline{V}^n) \leq P_d^{(n)}(V) \leq \gamma_n + P_d^{(n)}(\overline{V}^n)$$

and prove (5.5). Q.E.D.

6 Proof of the Permutation Lemma

6.1 Basic Ideas of the Proof

The proof of the permutation lemma can be presented as a result of several sequential steps, which are described in Subsections 6.2-6.6.

We introduce a combinatorial broadcast channel $\{F^n(\mathbf{y},\mathbf{y}'|\mathbf{x})\}$, which has a codeword $\mathbf{x} \in G^n_P$ at the input and the elements of the sets $\mathbf{T}^n_V(\mathbf{x})$ and $\mathbf{T}^n_W(\mathbf{x})$ at the output. The probabilities $F^n(\mathbf{y},\mathbf{y}'|\mathbf{x})$ are assigned in such a way that

$$\begin{aligned}
\sum_{\mathbf{y}'} F^n(\mathbf{y},\mathbf{y}'|\mathbf{x}) &= V^n(\mathbf{y}|\mathbf{x}) \qquad (6.1)\\
\sum_{\mathbf{y}} F^n(\mathbf{y},\mathbf{y}'|\mathbf{x}) &= W^n(\mathbf{y}'|\mathbf{x})
\end{aligned}$$

and

$$F^n(\mathbf{y}, \mathbf{y}'|\mathbf{x}) > 0 \iff d_H(\mathbf{y}, \mathbf{y}') = k \tag{6.2}$$

where

$$k = \min_{\mathbf{y}' \in \mathbf{T}_W^n(\mathbf{x})} d_H(\mathbf{y}, \mathbf{y}'), \quad \mathbf{y} \in \mathbf{T}_V^n(\mathbf{x}). \tag{6.3}$$

If $R > I(P, V)$, then the decoder $\mathcal{D}$, which receives $\mathbf{y}$, cannot realize a reliable decoding procedure, i.e., with the high probability, there exists an incorrect codeword $\hat{\mathbf{x}}$ such that

$$d(\hat{\mathbf{x}}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{y}). \tag{6.4}$$

Using (6.1) and the inequality $d(P, V) \leq d(P, W)$ we conclude that

$$d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{y}') \tag{6.5}$$

i.e., the conditions for the decoder $\mathcal{D}'$, which receives $\mathbf{y}'$, are worse than the conditions for the decoder $\mathcal{D}$ from the point of view of the correct codeword $\mathbf{x}$. Our intention is to prove that the conditions for $\mathcal{D}'$, as a rule, are not worse than the conditions for $\mathcal{D}$ from the point of view of the incorrect codeword $\hat{\mathbf{x}}$, i.e., as a rule,

$$d(\hat{\mathbf{x}}, \mathbf{y}') \leq d(\hat{\mathbf{x}}, \mathbf{y}). \tag{6.6}$$

Then using (6.4)-(6.6) we conclude that, as a rule, $d(\hat{\mathbf{x}}, \mathbf{y}') \leq d(\mathbf{x}, \mathbf{y}')$, and $\mathcal{D}'$ cannot do better than $\mathcal{D}$.

The result is named a 'permutation lemma' since $\mathbf{y} \in \mathbf{T}_V^n(\mathbf{x})$ and $\mathbf{y}' \in \mathbf{T}_W^n(\mathbf{x})$ can be obtained one from the other by permutations of the components. This statement follows from the equation $PV = PW$, which is very important for our considerations. The minimal number of pairwise permutations transforming $\mathbf{y}$ to some element of the set $\mathbf{T}_W^n(\mathbf{x})$ is equal to $k/2$, where k is defined in (6.3). In Subsection 6.2 we describe the structure of these permutations and note that there is a complementary property : for any given $x \in X$, the set Y can be splitted into two disjoint subsets, Y_x^+ and Y_x^-, such that

$$x_j = x, \; y_j \neq y_j' \implies y_j \in Y_x^-, \; y_j' \in Y_x^+ \text{ or } y_j \in Y_x^+, \; y_j' \in Y_x^-. \tag{6.7}$$

On the base of this fact we represent F^n as a result of combination of four conditional distributions : $Q^n, U^{n-k}, \Delta V^k$, and ΔW^k in Subsection 6.3. The distribution U^{n-k} is used to assign the coinciding symbols in $\mathbf{y}$ and $\mathbf{y}'$, and we interprete it as a common element in V^n and W^n. The distributions ΔV^k and ΔW^k are used to assign the noncoinciding symbols in $\mathbf{y}$ and $\mathbf{y}'$, and we interprete them as individual contributions of V^n and W^n. A binary vector $\mathbf{z}$ of length n and the Hamming weight k, which determines the positions j such that $y_j \neq y'_j$, is assigned in accordance with the distribution Q^n. In the end of Subsection 6.3 we represent the transformation of $\mathbf{y} \in \mathbf{T}^n_V(\mathbf{x})$ to $\mathbf{y}' \in \mathbf{T}^n_W(\mathbf{x})$ as interchanging of the individual contributions of V^n and W^n.

Since $d(P,V) \leq d(P,W)$, all the transformations of $\mathbf{y} \in \mathbf{T}^n_V(\mathbf{x})$ to $\mathbf{y}' \in \mathbf{T}^n_W(\mathbf{x})$ do not decrease the value of the distortion function for the codeword $\mathbf{x}$. If $\hat{\mathbf{x}} \in G^n_P \backslash \{\mathbf{x}\}$ is an incorrect codeword, then a particular transformation either increases or decreases the value of the distortion function for $\hat{\mathbf{x}}$. The result depends on the distributions of the components $y_j \neq y'_j$ located at positions where $x_j \neq \hat{x}_j$. In Subsection 6.4 we show that there are distributions which definitely do not increase this value. Therefore, we conclude that $d(\mathbf{x},\mathbf{y}') \geq d(\hat{\mathbf{x}},\mathbf{y}')$ if $d(\mathbf{x},\mathbf{y}) \geq d(\hat{\mathbf{x}},\mathbf{y})$ and call these transformations of $\mathbf{y}$ to $\mathbf{y}'$ as permutations conserving relations between distortions.

The last step of the proof is to show that the decoder $\mathcal{D}$ can restrict the code G^n_P in such a way that the distribution on the components, where $\mathbf{y}$ is transformed to $\mathbf{y}'$, is fixed. On the other hand, the restriction of the code keeps the converse statement to the coding theorem for the channel V. These considerations are given in Subsection 6.5.

A very short formal proof of the permutation lemma is given in Subsection 6.6.

We recommend reading Subsections 6.2-6.5 simultaneously with the material given in the Appendix, where the main steps of the proof are illustrated numerically.

6.2 The Structure of Minimal Permutations Between $\mathbf{T}^n_V(\mathbf{x})$ and $\mathbf{T}^n_W(\mathbf{x})$

Let

$$\begin{aligned} Y_x^+ &= \{\ y \in Y : V_x(y) > W_x(y)\ \} \\ Y_x^- &= \{\ y \in Y : V_x(y) < W_x(y)\ \} \end{aligned} \tag{6.8}$$

$$
\begin{aligned}
k_x(y) &= \begin{cases} nP_x \cdot (V_x(y) - W_x(y)), & \text{if } y \in Y_x^+ \\ 0, & \text{if } y \in Y_x^- \end{cases} \\
k(y) &= \sum_x k_x(y) \\
k_x &= \sum_y k_x(y).
\end{aligned}
$$

For binary-input channels, the equation $PV = PW$ means that

$$P_0 \cdot V_0(y) + P_1 \cdot V_1(y) = P_0 \cdot W_0(y) + P_1 \cdot W_1(y)$$

for all $y \in Y$. Therefore, as it is easy to see,

$$Y_0^- = Y_1^+, \quad Y_0^+ = Y_1^-$$

and

$$k = \sum_y k(y) = \sum_x k_x$$

where the parameter k is defined in (6.3).

Let us suppose that we want to get a vector $\mathbf{y}' \in \mathbf{T}_W^n(\mathbf{x})$ from $\mathbf{y} \in \mathbf{T}_V^n(\mathbf{x})$ using a minimal number of permutations of the components (these permutations are referred to as minimal permutations). If $y \in Y_0^+$, then $V_0(y) > W_0(y)$, $V_1(y) < W_1(y)$, and

$$nP_0 \cdot V_0(y) - nP_0 \cdot W_0(y) = nP_1 \cdot W_1(y) - nP_1 \cdot V_1(y)$$

because of the condition $PV = PW$. The similar statement is valid for all $y \in Y_1^+$. Therefore, we should select $k_0(y)$ indices j such that $(x_j, y_j) = (0, y)$ for all $y \in Y_0^+$ and $k_1(y)$ indices j such that $(x_j, y_j) = (1, y)$ for all $y \in Y_1^+$. Then k_0 and k_1 indices j such that $x_j = 0$ and $x_j = 1$ will be selected, and $k_0 = k_1$. The components of the vector $\mathbf{y}$ located at k_0 positions where $x_j = 0$ should be replaced with the components located at k_1 positions where $x_j = 1$ and *vice versa*. As a result of this procedure, we obtain a vector $\mathbf{y}' \in \mathbf{T}_W^n(\mathbf{x})$ with the following properties:

$$
\begin{aligned}
n_{xyy}(\mathbf{x}, \mathbf{y}, \mathbf{y}') &= nP_x \cdot \min\{V_x(y), W_x(y)\} \\
\sum_{y' \neq y} n_{xyy'}(\mathbf{x}, \mathbf{y}, \mathbf{y}') &= \begin{cases} nP_x \cdot (V_x(y) - W_x(y)) & \text{if } y \in Y_x^+ \\ nP_x \cdot (W_x(y) - V_x(y)) & \text{if } y \in Y_x^- \end{cases}
\end{aligned}
$$

where $n_{xyy'}(\mathbf{x}, \mathbf{y}, \mathbf{y}')$ is the number of indices j such that $x_j = x$, $y_j = y$, and $y_j' = y'$.

6.3 Decomposition of the Distributions V^n and W^n

Convention 6.1 (Z-Convention): For any binary vector $\mathbf{z} = (z_1, ..., z_n)$, we introduce the sets

$$Z^* = \{ j : z_j = 0 \}, \quad Z = \{ j : z_j = 1 \}$$

and associate $\mathbf{z}$ with the pair (Z^*, Z). For any vector $\mathbf{u}$, we write : $\mathbf{u} = (\mathbf{u}_{Z^*}, \mathbf{u}_Z)$, where $\mathbf{u}_{Z^*}$ is the vector composed of the components $u_j, j \in Z^*$, and $\mathbf{u}_Z$ is the vector composed of the components $u_j, j \in Z$ (in further considerations we substitute $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{y}'$ for $\mathbf{u}$).

Let us define the conditional probability distribution $Q = \{Q_x(z)\}$ on $\{0,1\}$ in such a way that

$$Q_x(z) = \begin{cases} (nP_x - k_x)/nP_x, & \text{if } z = 0 \\ k_x/nP_x, & \text{if } z = 1 \end{cases}$$

for all $x \in X$. Since Q is as a type on $\{0,1\}^n$ given $\mathbf{x} \in \mathbf{T}_P^n$, we can also refer to the set

$$\mathbf{T}_Q^n(\mathbf{x}) = \{ \mathbf{z} \in \{0,1\}^n : Comp(\mathbf{x}, \mathbf{z}) = \{ P_x \cdot Q_x(z) \} \}$$

and define a combinatorial channel $Q^n = \{Q^n(\mathbf{z}|\mathbf{x})\}$ as the uniform distribution on $\mathbf{T}_Q^n(\mathbf{x})$, i.e.,

$$Q^n(\mathbf{z}|\mathbf{x}) = \begin{cases} |\mathbf{T}_Q^n(\mathbf{x})|^{-1}, & \text{if } \mathbf{z} \in \mathbf{T}_Q^n(\mathbf{x}) \\ 0, & \text{otherwise.} \end{cases}$$

We also define the conditional probability distributions $U = \{U_x(y)\}$, $\Delta V = \{\Delta V_x(y)\}$, and $\Delta W = \{\Delta W_x(y)\}$ in such a way that

$$\begin{aligned} U_x(y) &= \frac{nP_x \cdot \min\{V_x(y), W_x(y)\}}{nP_x - k_x} \quad (6.9) \\ \Delta V_x(y) &= \frac{nP_x V_x(y) - (nP_x - k_x) \cdot U_x(y)}{k_x} \\ \Delta W_x(y) &= \frac{nP_x W_x(y) - (nP_x - k_x) \cdot U_x(y)}{k_x} \end{aligned}$$

for all $x \in X,\ \ y \in Y$. Given $\mathbf{x} \in \mathbf{T}_P^n$ and $\mathbf{z} \in \mathbf{T}_Q^n(\mathbf{x})$, we refer to the sets

$$
\begin{aligned}
\mathbf{T}_U^{n-k}(\mathbf{x}_{Z^*}) &= \{\ \mathbf{y}_{Z^*} \in Y^{n-k} : \\
&\qquad Comp(\mathbf{x}_{Z^*}, \mathbf{y}_{Z^*}) = \{\ (nP_x - k_x) \cdot U_x(y)/(n-k)\ \}\ \} \\
\mathbf{T}_{\Delta V}^k(\mathbf{x}_Z) &= \{\ \mathbf{y}_Z \in Y^k :\ Comp(\mathbf{x}_Z, \mathbf{y}_Z) = \{\ k_x \cdot \Delta V_x(y)/k\ \}\ \} \\
\mathbf{T}_{\Delta W}^k(\mathbf{x}_Z) &= \{\ \mathbf{y}'_Z \in Y^k :\ Comp(\mathbf{x}_Z, \mathbf{y}'_Z) = \{\ k_x \cdot \Delta W_x(y)/k\ \}\ \}
\end{aligned}
$$

where we used the definition of $Comp$, given in Section 2 as the set of ratios of the number of entries (x, y) in a particular pair of vectors and the length of these vectors.

Let us introduce the combinatorial channels

$$
\begin{aligned}
U^{n-k} &= \{U^{n-k}(\mathbf{y}_{Z^*}|\mathbf{x}_{Z^*})\} \\
\Delta V^k &= \{\Delta V^k(\mathbf{y}_Z|\mathbf{x}_Z)\},
\end{aligned}
$$

and

$$
\Delta W^k = \{\Delta W^k(\mathbf{y}_Z|\mathbf{x}_Z)\}
$$

as the uniform distributions on the sets $\mathbf{T}_U^{n-k}(\mathbf{x}_{Z^*})$, $\Delta V^k(\mathbf{y}_Z|\mathbf{x}_Z)$, and $\Delta W^k(\mathbf{y}'_Z|\mathbf{x}_Z)$, i.e.,

$$
\begin{aligned}
U^{n-k}(\mathbf{y}_{Z^*}|\mathbf{x}_{Z^*}) &= \begin{cases} |\mathbf{T}_U^{n-k}(\mathbf{x}_{Z^*})|^{-1}, & \text{if } \mathbf{y}_{Z^*} \in \mathbf{T}_U^{n-k}(\mathbf{x}_{Z^*}) \\ 0, & \text{otherwise} \end{cases} \\
\Delta V^k(\mathbf{y}_Z|\mathbf{x}_Z) &= \begin{cases} |\mathbf{T}_{\Delta V}^k(\mathbf{x}_Z)|^{-1}, & \text{if } \mathbf{y}_Z \in \mathbf{T}_{\Delta V}^k(\mathbf{x}_Z) \\ 0, & \text{otherwise} \end{cases} \\
\Delta W^k(\mathbf{y}'_Z|\mathbf{x}_Z) &= \begin{cases} |\mathbf{T}_{\Delta W}^k(\mathbf{x}_Z)|^{-1}, & \text{if } \mathbf{y}'_Z \in \mathbf{T}_{\Delta W}^k(\mathbf{x}_Z) \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}
$$

Then we can represent the process of generating the vectors $\mathbf{y} \in \mathbf{T}_V^n(\mathbf{x})$ and $\mathbf{y}' \in \mathbf{T}_W^n(\mathbf{x})$, located at the Hamming distance k, as a transmission of $\mathbf{x}$ over a combinatorial broadcast channel (Fig.2), defined by the probabilities

$$
\begin{aligned}
F^n(\mathbf{y}, \mathbf{y}'|\mathbf{x}) &= \sum_{\mathbf{z}} Q^n(\mathbf{z}|\mathbf{x}) \cdot f^{n-k}(\mathbf{y}_{Z^*}, \mathbf{y}'_{Z^*}|\mathbf{x}) \cdot \\
&\qquad \cdot \Delta V^k(\mathbf{y}_Z|\mathbf{x}_Z) \cdot \Delta W^k(\mathbf{y}'_Z|\mathbf{x}_Z)
\end{aligned}
$$

where

$$f^{n-k}(\mathbf{y}_{Z^*},\mathbf{y}'_{Z^*}|\mathbf{x}) = U^{n-k}(\mathbf{y}_{Z^*}|\mathbf{x}_{Z^*}) \cdot \chi\{\ \mathbf{y}'_{Z^*} = \mathbf{y}_{Z^*}\ \}.$$

These probabilities satisfy (6.1), (6.2), and the triples

$$(\ \mathbf{y}_{Z^*} = \mathbf{y}'_{Z^*}, \mathbf{y}_Z, \mathbf{y}'_Z\ ) \in (\ \mathbf{T}^{n-k}_U(\mathbf{x}_{Z^*}), \mathbf{T}^k_{\Delta V}(\mathbf{x}_Z), \mathbf{T}^k_{\Delta W}(\mathbf{x}_Z)\ ) \qquad (6.10)$$

are generated at the output of the channel. The decoder $\mathcal{D}$ receives $(\mathbf{y}_{Z^*}, \mathbf{y}_Z)$, while the decoder $\mathcal{D}'$ receives $(\mathbf{y}_{Z^*}, \mathbf{y}'_Z)$ (note that $\mathbf{z}$ is unknown for $\mathcal{D}$ and $\mathcal{D}'$). From the point of view of $\mathcal{D}$, transmission of the codeword $\mathbf{x}$ over the combinatorial channel V^n can be represented as a process consisting of two steps :

1) the channel assigns a binary vector $\mathbf{z} \in \mathbf{T}^n_Q(\mathbf{x})$;

2) the channel distributes $n-k$ components of the vector $\mathbf{y}$ corresponding to the components 0 of the vector $\mathbf{z}$ in accordance with the distribution U^{n-k} and k components of the vector $\mathbf{y}$ corresponding to the components 1 of the vector $\mathbf{z}$ in accordance with the distribution ΔV^k.

From the point of view of $\mathcal{D}'$, transmission of the codeword $\mathbf{x}$ over the combinatorial channel W^n can be described in the same way, but the distribution ΔW^k should be used at the second step instead of ΔV^k.

Note that the distributions ΔW and ΔV have the complementary property :

$$\begin{aligned} \Delta V_0(y) &= \Delta W_1(y) \\ \Delta W_0(y) &= \Delta V_1(y) \end{aligned} \qquad (6.11)$$

for all $y \in Y$. Note also that the average distortion functions $d(P,V)$ and $d(P,W)$ can be written as

$$\begin{aligned} d(P,V) &= \frac{nP_1 - k_1}{n} \cdot \sum_y U_1(y) \cdot d_1(y) + \frac{k_1}{n} \cdot \sum_y \Delta V_1(y) \cdot d_1(y) \\ d(P,W) &= \frac{nP_1 - k_1}{n} \cdot \sum_y U_1(y) \cdot d_1(y) + \frac{k_1}{n} \cdot \sum_y \Delta W_1(y) \cdot d_1(y) \end{aligned}$$

where we used the equations $d_0(y) = 0$ for all $y \in Y$. Therefore, the restriction

$$d(P,V) \leq d(P,W)$$

can be represented as the inequality :

$$\sum_y (\Delta V_1(y) - \Delta W_1(y)) \cdot d_1(y) \geq 0. \qquad (6.12)$$

6.4 Permutations Conserving Relations Between Distortions for Binary-Input Channels

In this subsection we show the idea for a simplified case and then extend the considerations to a more general case.

Let $\mathbf{z} \in \mathbf{T}_Q^n(\mathbf{x})$, $\mathbf{y} \in \mathbf{T}_V^n(\mathbf{x})$, and $\mathbf{y}' \in \mathbf{T}_W^n(\mathbf{x})$ be given. Suppose also that $\hat{\mathbf{x}} \in G_P^n \backslash \{\mathbf{x}\}$ is fixed. Let

$$\begin{aligned} k_{x\hat{x}} &= \sum_{j:z_j=1} \chi\{ \, (x_j, \hat{x}_j) = (x, \hat{x}) \, \} \\ k_{x\hat{x}}(y) &= \sum_{j:z_j=1} \chi\{ \, (x_j, \hat{x}_j, y_j) = (x, \hat{x}, y) \, \} \\ k_{x\hat{x}}(y') &= \sum_{j:z_j=1} \chi\{ \, (x_j, \hat{x}_j, y'_j) = (x, \hat{x}, y') \, \} \end{aligned} \tag{6.13}$$

denote the number of entries $(x, \hat{x})$ in the pair $(\mathbf{x}_Z, \hat{\mathbf{x}}_Z)$, the number of entries $(x, \hat{x}, y)$ in the triple $(\mathbf{x}_Z, \hat{\mathbf{x}}_Z, \mathbf{y}_Z)$, and the number of entries $(x, \hat{x}, y')$ in the triple $(\mathbf{x}_Z, \hat{\mathbf{x}}_Z, \mathbf{y}'_Z)$, respectively.

Proposition 6.2: Let $\mathbf{y} \in \mathbf{T}_V^n(\mathbf{x})$ and $\mathbf{y}' \in \mathbf{T}_W^n(\mathbf{x})$ be connected by (6.10). If $\hat{\mathbf{x}} \in G_P^n$ and $\mathbf{z} \in \mathbf{T}_Q^n(\mathbf{x})$ are chosen in such a way that

$$\begin{aligned} k_{x\hat{x}}(y) &= k_{x\hat{x}} \cdot \Delta V_x(y), \quad \text{for all } \; y \in Y \\ k_{x\hat{x}}(y') &= k_{x\hat{x}} \cdot \Delta W_x(y'), \quad \text{for all } \; y' \in Y \end{aligned} \tag{6.14}$$

then

$$d(\hat{\mathbf{x}}, \mathbf{y}) \le d(\mathbf{x}, \mathbf{y}) \implies d(\hat{\mathbf{x}}, \mathbf{y}') \le d(\mathbf{x}, \mathbf{y}').$$

Proof: Let

$$\delta = (d(\mathbf{x}, \mathbf{y}') - d(\hat{\mathbf{x}}, \mathbf{y}')) - (d(\mathbf{x}, \mathbf{y}) - d(\hat{\mathbf{x}}, \mathbf{y})). \tag{6.15}$$

Since

$$\left.\begin{matrix} d(\mathbf{x}, \mathbf{y}) \\ d(\mathbf{x}, \mathbf{y}') \end{matrix}\right\} = d(\mathbf{x}_{Z^*}, \mathbf{y}_{Z^*}) + \left\{\begin{matrix} d(\mathbf{x}_Z, \mathbf{y}_Z) \\ d(\mathbf{x}_Z, \mathbf{y}'_Z) \end{matrix}\right.$$

$$\left.\begin{matrix} d(\hat{\mathbf{x}}, \mathbf{y}) \\ d(\hat{\mathbf{x}}, \mathbf{y}') \end{matrix}\right\} = d(\hat{\mathbf{x}}_{Z^*}, \mathbf{y}_{Z^*}) + \left\{\begin{matrix} d(\hat{\mathbf{x}}_Z, \mathbf{y}_Z) \\ d(\hat{\mathbf{x}}_Z, \mathbf{y}'_Z) \end{matrix}\right.$$

we change the order of summands in (6.15) and write

$$\delta = (d(\mathbf{x}_Z, \mathbf{y}'_Z) - d(\mathbf{x}_Z, \mathbf{y}_Z)) - (d(\hat{\mathbf{x}}_Z, \mathbf{y}'_Z) - d(\hat{\mathbf{x}}_Z, \mathbf{y}_Z)).$$

Thus

$$
\begin{aligned}
\delta &= \left(\sum_{y'} k_{10}(y') \cdot d_1(y') - \sum_{y} k_{10}(y) \cdot d_1(y)\right) - \qquad (6.16) \\
&- \left(\sum_{y'} k_{01}(y') \cdot d_1(y') - \sum_{y} k_{01}(y) \cdot d_1(y)\right) \\
&= k_{10} \cdot \sum_{y} (\Delta W_1(y) - \Delta V_1(y)) \cdot d_1(y) - k_{01} \cdot \sum_{y} (\Delta W_0(y) - \Delta V_0(y)) \cdot d_1(y) \\
&= (k_{10} + k_{01}) \cdot \sum_{y} (\Delta W_1(y) - \Delta V_1(y)) \cdot d_1(y) \geq 0.
\end{aligned}
$$

The first equation in (6.16) follows from the note that we can consider only noncoinciding components of $\mathbf{x}$ and $\hat{\mathbf{x}}$. The second equation follows from (6.14). Then we have used (6.11) and (6.12). Hence,

$$d(\mathbf{x}, \mathbf{y}') - d(\hat{\mathbf{x}}, \mathbf{y}') = d(\mathbf{x}, \mathbf{y}) - d(\hat{\mathbf{x}}, \mathbf{y}) + \delta \geq 0.$$

Q.E.D.

The statement below generalizes the considerations. The proof is omitted since it is similar to the proof of Proposition 6.2.

Proposition 6.3: Let $\mathbf{y} \in \mathbf{T}_V^n(\mathbf{x})$ and $\mathbf{y}' \in \mathbf{T}_W^n(\mathbf{x})$ be connected by (6.10). If $\hat{\mathbf{x}} \in G_P^n$ and $\mathbf{z} \in \mathbf{T}_Q^n(\mathbf{x})$ are chosen in such a way that

$$
\begin{aligned}
|k_{x\hat{x}}(y) - k_{x\hat{x}} \cdot \Delta V_x(y)| &\leq \alpha_n \cdot k_{x\hat{x}} \text{ for all } (x, \hat{x}, y) \in X \times X \times Y \\
|k_{x\hat{x}}(y') - k_{x\hat{x}} \cdot \Delta W_x(y')| &\leq \alpha_n \cdot k_{x\hat{x}} \text{ for all } (x, \hat{x}, y') \in X \times X \times Y
\end{aligned} \qquad (6.17)
$$

where $k_{x\hat{x}}$, $k_{x\hat{x}}(y)$, and $k_{x\hat{x}}(y')$ are defined in (6.13), and

$$\sum_{y} (\Delta W_1(y) - \Delta V_1(y)) \cdot d_1(y) \geq 2\alpha_n d_{max} \qquad (6.18)$$

then

$$d(\hat{\mathbf{x}}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{y}) \implies d(\hat{\mathbf{x}}, \mathbf{y}') \leq d(\mathbf{x}, \mathbf{y}').$$

6.5 Restrictions of the Code G_P^n

In this subsection, we are dealing with the probabilistic ensemble

$$\{\ XYZ,\ P_x \cdot V_x(y) \cdot \hat{Q}(z|x,y)\ \}$$

where $Z = \{0,1\}$,

$$\hat{Q}_{x,y}(z) \;=\; \begin{cases} 1 - k_x(y)/(nP_x \cdot V_x(y)), & \text{if } z = 0 \\ k_x(y)/(nP_x \cdot V_x(y)), & \text{if } z = 1 \end{cases} \tag{6.19}$$

and the parameters $k_x(y)$, $x \in X$, $y \in Y$, are defined in (6.8). We consider the distribution $\hat{Q} = \{\hat{Q}_{x,y}(z)\}$ as a conditional type on $\{0,1\}^n$ given $\mathbf{y} \in \mathbf{T}_V^n(\mathbf{x})$ and $\mathbf{x} \in \mathbf{T}_P^n$ and introduce the set

$$\begin{aligned} \mathbf{T}_{\hat{Q}}^n(\mathbf{x},\mathbf{y}) \;=\; \{\, \mathbf{z} \in \{0,1\}^n :\ & n_{xyz}(\mathbf{x},\mathbf{y},\mathbf{z}) = nP_x \cdot V_x(y) \cdot \hat{Q}_{x,y}(z) \\ & \text{for all } (x,y,z) \in X \times Y \times Z \,\} \end{aligned}$$

where $n_{xyz}(\mathbf{x},\mathbf{y},\mathbf{z})$ denotes the number of entries (x,y,z) in the triple $(\mathbf{x},\mathbf{y},\mathbf{z})$. Let the uniform distributions on $\mathbf{T}_{\hat{Q}}^n(\mathbf{x},\mathbf{y})$ be given as

$$\hat{Q}^n(\mathbf{z}|\mathbf{x},\mathbf{y}) \;=\; \begin{cases} |\mathbf{T}_{\hat{Q}}^n(\mathbf{x},\mathbf{y})|^{-1}, & \text{if } \mathbf{z} \in \mathbf{T}_{\hat{Q}}^n(\mathbf{x},\mathbf{y}) \\ 0, & \text{otherwise.} \end{cases}$$

Let us fix $\hat{\mathbf{x}} \in G_P^n$, $\mathbf{y} \in \mathbf{T}_V^n(\mathbf{x})$, and $\mathbf{z} \in \mathbf{T}_{\hat{Q}}^n(\mathbf{x},\mathbf{y})$. In Subsection 6.2 we partitioned the output alphabet Y in two disjoint subsets, Y_0^+ and Y_1^+ (see (6.8)), and we can say either $y \in Y_0^+$ or $y \in Y_1^+$ for any $y \in Y$. Let $x(y) \in \{0,1\}$ be defined in such a way that

$$x(y) = x \iff y \in Y_x^+ \tag{6.20}$$

and let

$$k_{x\hat{x}}(y) = \sum_{j:z_j=1} \chi\{\ (\hat{x}_j, y_j) = (\hat{x}, y)\ \},\ \ x = x(y). \tag{6.21}$$

Besides, let

$$k_{x\hat{x}} = \sum_{y \in Y_x^+} \sum_{j:z_j=1} \chi\{\ (\hat{x}_j, y_j) = (\hat{x}, y)\ \}. \tag{6.22}$$

If $k_{x\hat{x}} > 0$, then $\{k_{x\hat{x}}(y)/k_{x\hat{x}}, y \in Y_x^+\}$ are the probability distributions on Y_x^+ for all pairs $(x,\hat{x}) \in X \times X$. In the definition below, we construct a subcode of G_P^n, consisting of codewords, for which these distributions are close to $\{\Delta V_x(y), y \in Y\}$ for all $\hat{x} \in X$.

Definition 6.4: The code $G_P^n(\mathbf{y}_Z)$ will be referred to as a ΔV-restricted code given $\mathbf{y} \in \mathbf{T}_V^n(\mathbf{x})$ and $\mathbf{z} \in \mathbf{T}_{\hat{Q}}^n(\mathbf{x},\mathbf{y})$ if it consists of the codewords $\hat{\mathbf{x}} \in G_P^n$ such that

$$|k_{x\hat{x}}(y) - k_{x\hat{x}} \cdot \Delta V_x(y)| \le \alpha_n \cdot k_{x\hat{x}} \ \text{ for all } \ (x,\hat{x},y) \in X \times X \times Y \tag{6.23}$$

where the parameters $k_{x\hat{x}}(y)$ and $k_{x\hat{x}}$ are defined in (6.21), (6.22).

Proposition 6.5: Let

$$\hat{P}_d^{(n)}(\mathbf{x}, V^n) \ = \ \sum_{\mathbf{y},\mathbf{z}} V^n(\mathbf{y}|\mathbf{x}) \cdot \hat{Q}^n(\mathbf{z}|\mathbf{x},\mathbf{y}) \cdot \chi\{\ d(\mathbf{x},\mathbf{y}) \ge D(\mathbf{x},\mathbf{y}|\mathbf{z})\ \}$$

where

$$D(\mathbf{x},\mathbf{y}|\mathbf{z}) \ = \ \min_{\hat{\mathbf{x}} \in G_P^n(\mathbf{y}_Z)\setminus\{\mathbf{x}\}} d(\hat{\mathbf{x}},\mathbf{y}). \tag{6.24}$$

Then there exist sequences $\{\alpha_n\}$ such that

$$\alpha_n \to 0, \quad \alpha_n \sqrt{n} \to \infty, \quad \text{as} \ \ n \to \infty \tag{6.25}$$

and that the sequences $\{P_d^{(n)}(V^n)\}$ and $\{\hat{P}_d^{(n)}(V^n)\}$ approximate each other, i.e.,

$$\Upsilon(\{P_d^{(n)}(V^n)\}) = \Upsilon(\{\hat{P}_d^{(n)}(V^n)\}). \tag{6.26}$$

Proof: First of all, we write :

$$\Upsilon(\{\hat{P}_d^{(n)}(V^n)\}) \le \Upsilon(\{P_d^{(n)}(V^n)\}) \tag{6.27}$$

since $\mathbf{x} \in G_P^n(\mathbf{y}_Z)$ for all $\mathbf{y} \in \mathbf{T}_V^n(\mathbf{x})$ and $\mathbf{z} \in \mathbf{T}_{\hat{Q}}^n(\mathbf{x},\mathbf{y})$.

Using (6.8), (6.19), and (6.20) we note that $\hat{Q}_{xy}(1) > 0$ only if $x = x(y)$. Therefore, we can introduce the following conditional probability distribution on $\{0,1\}$:

$$\tilde{Q}_y(z) = \hat{Q}_{x,y}(z), \ \text{ where } \ x = x(y).$$

The distribution $\tilde{Q} = \{\tilde{Q}_y(z)\}$ can be also considered as a conditional type on $\{0,1\}^n$ given $\mathbf{y} \in \mathbf{T}_{PV}^n$, i.e., we refer to the set

$$\mathbf{T}_{\tilde{Q}}^n(\mathbf{y}) \ = \ \{\ \mathbf{z} \in \{0,1\}^n :\ Comp(\mathbf{y},\mathbf{z}) = \{PV(y) \cdot \tilde{Q}_y(z)\}\ \}.$$

Let the uniform distributions on $\mathbf{T}^n_{\tilde{Q}}(\mathbf{y})$ be given as

$$\tilde{Q}^n(\mathbf{z}|\mathbf{y}) \;=\; \begin{cases} |\mathbf{T}^n_{\tilde{Q}}(\mathbf{y})|^{-1}, & \text{if } \mathbf{z} \in \mathbf{T}^n_{\tilde{Q}}(\mathbf{y}) \\ 0, & \text{otherwise.} \end{cases}$$

Let us fix ΔV-restricted codes for all $\mathbf{z} \in \mathbf{T}^n_{\tilde{Q}}(\mathbf{y})$ in the same way as for the vectors $\mathbf{z} \in \mathbf{T}^n_{\hat{Q}}(\mathbf{x},\mathbf{y})$ and define the decoding error probability for the codeword $\mathbf{x}$ as

$$\tilde{P}_d^{(n)}(\mathbf{x}, V^n) \;=\; \sum_{\mathbf{y},\mathbf{z}} V^n(\mathbf{y}|\mathbf{x}) \cdot \tilde{Q}^n(\mathbf{z}|\mathbf{y}) \cdot \chi\{\, d(\mathbf{x},\mathbf{y}) \geq D(\mathbf{x},\mathbf{y}|\mathbf{z}) \,\}$$

where the function $D(\mathbf{x},\mathbf{y}|\mathbf{z})$ is defined in (6.24). Note that

$$\begin{aligned} \tilde{P}_d^{(n)}(\mathbf{x}, V^n) \;=\; & \sum_{\mathbf{y},\mathbf{z}} V^n(\mathbf{y}|\mathbf{x}) \cdot \tilde{Q}^n(\mathbf{z}|\mathbf{y}) \cdot \chi\{\, \mathbf{x} \notin G^n_P(\mathbf{y}_Z) \,\} \\ + & \sum_{\mathbf{y},\mathbf{z}} V^n(\mathbf{y}|\mathbf{x}) \cdot \tilde{Q}^n(\mathbf{z}|\mathbf{y}) \cdot \chi\{\, \mathbf{x} \in G^n_P(\mathbf{y}_Z) \,\} \cdot \\ & \qquad \cdot \chi\{\, d(\mathbf{x},\mathbf{y}) \geq D(\mathbf{x},\mathbf{y}|\mathbf{z}) \,\} \end{aligned} \tag{6.28}$$

and

$$\begin{aligned} \hat{P}_d^{(n)}(\mathbf{x}, V^n) \;\geq\; & \sum_{\mathbf{y},\mathbf{z}} V^n(\mathbf{y}|\mathbf{x}) \cdot \tilde{Q}^n(\mathbf{z}|\mathbf{y}) \cdot \chi\{\, \mathbf{x} \in G^n_P(\mathbf{y}_Z) \,\} \cdot \\ & \qquad \cdot \chi\{\, d(\mathbf{x},\mathbf{y}) \geq D(\mathbf{x},\mathbf{y}|\mathbf{z}) \,\} \end{aligned} \tag{6.29}$$

since the vectors $\mathbf{z} \in \mathbf{T}^n_{\tilde{Q}}(\mathbf{y})$ put stronger restrictions on the incorrect codewords included into the collection $G^n_P(\mathbf{y}_Z)$ for all $\mathbf{y} \in \mathbf{T}^n_V(\mathbf{x})$ compared to the vectors $\mathbf{z} \in \mathbf{T}^n_{\hat{Q}}(\mathbf{x},\mathbf{y})$. However, using the same arguments as in Lemma 4.1, we conclude that there exist sequences $\{\alpha_n\}$ satisfying (6.25) such that

$$\sum_{\mathbf{y}} V^n(\mathbf{y}|\mathbf{x}) \cdot \tilde{Q}^n(\mathbf{z}|\mathbf{y}) \cdot \chi\{\, \mathbf{x} \in G^n_P(\mathbf{y}_Z) \,\} \geq 1 - \beta_n \tag{6.30}$$

where $\{\beta_n\}$ is a sequence, depending on $\{\alpha_n\}$, such that $\beta_n \to 0$, as $n \to \infty$. Thus, combining (6.28)-(6.30), we conclude that

$$\Upsilon(\{\tilde{P}_d^{(n)}(V^n)\}) \leq \Upsilon(\{\hat{P}_d^{(n)}(V^n)\}). \tag{6.31}$$

The condition $R > I(P,V)$ does not give an opportunity to partition the space consisting of all possible pairs $(\mathbf{y},\mathbf{z})$, where $\mathbf{y} \in \mathbf{T}^n_{PV}$ and $\mathbf{z} \in \mathbf{T}^n_{\hat{Q}}(\mathbf{y})$. Therefore,

$$\Upsilon(\{P_d^{(n)}(V^n)\}) \leq \Upsilon(\{\tilde{P}_d^{(n)}(V^n)\}). \tag{6.32}$$

As a result of (6.31) and (6.32), we have

$$\Upsilon(\{P_d^{(n)}(V^n)\}) \leq \Upsilon(\{\hat{P}_d^{(n)}(V^n)\}) \tag{6.33}$$

and complete the proof using (6.27) and (6.33). Q.E.D.

Note that $\mathbf{x} \in G^n_P(\mathbf{y}_Z)$ and that if $\hat{\mathbf{x}} \in G^n_P(\mathbf{y}_Z)$, then the first group of inequalities in (6.17) is satisfied.

The similar considerations are valid for the decoder $\mathcal{D}'$, which receives a vector $\mathbf{y}' \in \mathbf{T}^n_W(\mathbf{x})$. To avoid more complex notations, we distinguish between the parameters of $\mathcal{D}$ and $\mathcal{D}'$ writing y' and $\mathbf{y}'$ instead of y and $\mathbf{y}$, respectively. Let

$$\hat{Q}_{x,y'}(z) = \begin{cases} 1 - k_x(y')/(nP_x \cdot W_x(y')), & \text{if } z = 0 \\ k_x(y')/(nP_x \cdot W_x(y')), & \text{if } z = 1 \end{cases}$$

$$\mathbf{T}^n_{\hat{Q}}(\mathbf{x},\mathbf{y}') = \{\, \mathbf{z} \in \{0,1\}^n : \ n_{xy'z}(\mathbf{x},\mathbf{y}',\mathbf{z}) = nP_x \cdot W_x(y') \cdot \hat{Q}_{x,y'}(z) \text{ for all } (x,y',z) \in X \times Y \times Z \,\}$$

$$\hat{Q}^n(\mathbf{z}|\mathbf{x},\mathbf{y}') = \begin{cases} |\mathbf{T}^n_{\hat{Q}}(\mathbf{x},\mathbf{y}')|^{-1}, & \text{if } \mathbf{z} \in \mathbf{T}^n_{\hat{Q}}(\mathbf{x},\mathbf{y}') \\ 0, & \text{otherwise.} \end{cases}$$

Definition 6.6: The code $G^n_P(\mathbf{y}'_Z)$ will be referred to as a ΔW-restricted code given $\mathbf{y}' \in \mathbf{T}^n_{PW}$ and $\mathbf{z} \in \mathbf{T}^n_{\hat{Q}}(\mathbf{x},\mathbf{y}')$ if it consists of the codewords $\hat{\mathbf{x}} \in G^n_P$ such that

$$|k_{x\hat{x}}(y') - k_{x\hat{x}} \cdot \Delta W_x(y')| \leq \alpha_n \cdot k_{x\hat{x}} \ \text{ for all } (x,\hat{x},y') \in X \times X \times Y \tag{6.34}$$

where

$$k_{x\hat{x}}(y') = \sum_{j: z_j = 1} \chi\{\, (\hat{x}_j, y'_j) = (\hat{x}, y') \,\}, \quad x = x(y')$$

$$k_{x\hat{x}} = \sum_{y' \in Y_x^-} \sum_{j: z_j = 1} \chi\{\, (\hat{x}_j, y'_j) = (\hat{x}, y') \,\}$$

and

$$x(y') = x \iff y' \in Y_x^- .$$

Proposition 6.7: Let

$$\hat{P}_d^{(n)}(\mathbf{x}, W^n) = \sum_{\mathbf{y}',\mathbf{z}} W^n(\mathbf{y}'|\mathbf{x}) \cdot \hat{Q}^n(\mathbf{z}|\mathbf{x},\mathbf{y}') \cdot \chi\{ \, d(\mathbf{x},\mathbf{y}') \geq D(\mathbf{x},\mathbf{y}'|\mathbf{z}) \, \}$$

where

$$D(\mathbf{x},\mathbf{y}'|\mathbf{z}) = \min_{\hat{\mathbf{x}} \in G_P^n(\mathbf{y}'_Z)\backslash\{\mathbf{x}\}} d(\hat{\mathbf{x}},\mathbf{y}').$$

Then there exist sequences $\{\alpha_n\}$ such that

$$\alpha_n \to 0, \quad \alpha_n \sqrt{n} \to \infty, \quad \text{as} \quad n \to \infty \tag{6.35}$$

and that the sequence $\{\hat{P}_d^{(n)}(W^n)\}$ approximates the sequence $\{\hat{P}_d^{(n)}(V^n)\}$, i.e.,

$$\Upsilon(\{\hat{P}_d^{(n)}(V^n)\}) \leq \Upsilon(\{\hat{P}_d^{(n)}(W^n)\}). \tag{6.36}$$

Proof: Let us suppose that $d(\mathbf{x},\mathbf{y}) \geq D(\mathbf{x},\mathbf{y}|\mathbf{z})$ for some $\mathbf{y} \in \mathbf{T}_V^n(\mathbf{x})$ and $\mathbf{z} \in \mathbf{T}_{\hat{Q}}^n(\mathbf{x},\mathbf{y})$. Then there exists an incorrect codeword $\hat{\mathbf{x}}$ such that $\hat{\mathbf{x}} \in G_P^n(\mathbf{y}_Z)$ and

$$d(\hat{\mathbf{x}},\mathbf{y}) \leq d(\mathbf{x},\mathbf{y}). \tag{6.37}$$

Let the parameters $k_{x\hat{x}}$, defined in (6.22), correspond to this codeword. We generate $\mathbf{y}' \in \mathbf{T}_W^n(\mathbf{x})$ such that (6.10) is valid, i.e., we set $\mathbf{y}'_{Z^*} = \mathbf{y}_{Z^*}$ and distribute the components of the vector $\mathbf{y}'_Z \in \mathbf{T}_{\Delta W}^k(\mathbf{x}_Z)$. It means that we distribute k_0 symbols $y \in Y_0^-$ on k_{00} and k_{01} positions j where $x_j = 0$, and k_1 symbols $y \in Y_1^-$ on k_{10} and k_{11} positions j where $x_j = 1$. However, these positions are fixed and, typically, we obtain the same conditional distribution on k_{x0} and k_{x1} positions as a result of this procedure (note that the main difficulty of the previous analysis was the point that there are exponentially many incorrect codewords, and we could not fix one of them). Therefore, we need to assign α_n in such a way that the distributions, which do not satisfy (6.34), will be classified as large deviations. We refer to the arguments leading to the Delta-convention again [2, Convention 2.11] and conclude that we can take a sequence $\{\alpha_n\}$, satisfying (6.35). Then we note that the conditions (6.23) and (6.34) coincide with (6.17) for the codeword $\hat{\mathbf{x}}$ satisfying (6.37). We suppose that

$$d(P, V) \leq d(P, W) - 2\alpha_n d_{max} \tag{6.38}$$

and use the result of Proposition 6.3. The formal steps are given below.

$$\begin{aligned}
\hat{P}_d^{(n)}(\mathbf{x}, V^n) &= \sum_{\mathbf{y},\mathbf{z}} V^n(\mathbf{y}|\mathbf{x}) \cdot \hat{Q}^n(\mathbf{z}|\mathbf{x},\mathbf{y}) \cdot \\
&\quad \cdot \chi\{\ \exists \hat{\mathbf{x}} :\ \hat{\mathbf{x}} \in G_P^n(\mathbf{y}_Z) \ \text{ and } \ d(\mathbf{x},\mathbf{y}) \geq d(\hat{\mathbf{x}},\mathbf{y})\ \} \\
&\to \sum_{\mathbf{y},\mathbf{z}} V^n(\mathbf{y}|\mathbf{x}) \cdot \hat{Q}^n(\mathbf{z}|\mathbf{x},\mathbf{y}) \cdot \sum_{\mathbf{y}'_Z} \Delta W^k(\mathbf{y}'_Z) \cdot \\
&\quad \cdot \chi\{\ \exists \hat{\mathbf{x}} :\ \hat{\mathbf{x}} \in G_P^n(\mathbf{y}_Z),\ \hat{\mathbf{x}} \in G_P^n(\mathbf{y}'_Z), \ \text{ and } \ d(\mathbf{x},\mathbf{y}) \geq d(\hat{\mathbf{x}},\mathbf{y})\ \} \\
&\leq \sum_{\mathbf{y},\mathbf{z}} V^n(\mathbf{y}|\mathbf{x}) \cdot \hat{Q}^n(\mathbf{z}|\mathbf{x},\mathbf{y}) \cdot \sum_{\mathbf{y}'_Z} \Delta W^k(\mathbf{y}'_Z) \cdot \\
&\quad \cdot \chi\{\ \exists \hat{\mathbf{x}} :\ \hat{\mathbf{x}} \in G_P^n(\mathbf{y}'_Z) \ \text{ and } \ d(\mathbf{x},\mathbf{y}') \geq d(\hat{\mathbf{x}},\mathbf{y}')\ \} \\
&= \sum_{\mathbf{y},\mathbf{z}} V^n(\mathbf{y}|\mathbf{x}) \cdot \hat{Q}^n(\mathbf{z}|\mathbf{x},\mathbf{y}) \cdot \sum_{\mathbf{y}'_Z} \Delta W^k(\mathbf{y}'_Z) \cdot \chi\{\ d(\mathbf{x},\mathbf{y}') \geq D(\mathbf{x},\mathbf{y}'|\mathbf{z})\ \} \\
&= \sum_{\mathbf{y}',\mathbf{z}} W^n(\mathbf{y}'|\mathbf{x}) \cdot \hat{Q}^n(\mathbf{z}|\mathbf{x},\mathbf{y}') \cdot \chi\{\ d(\mathbf{x},\mathbf{y}') \geq D(\mathbf{x},\mathbf{y}'|\mathbf{z})\ \} \\
&= \hat{P}_d^{(n)}(\mathbf{x}, W^n)
\end{aligned}$$

where we also used the equations

$$\begin{aligned}
V^n(\mathbf{y}|\mathbf{x}) \cdot \hat{Q}^n(\mathbf{z}|\mathbf{x},\mathbf{y}) \cdot \Delta W^k(\mathbf{y}'_Z) &= Q^n(\mathbf{z}|\mathbf{x}) \cdot U^{n-k}(\mathbf{y}_{Z^*}) \cdot \Delta V^k(\mathbf{y}_Z) \cdot \Delta W^k(\mathbf{y}'_Z) \\
&= W^n(\mathbf{y}'|\mathbf{x}) \cdot \hat{Q}^n(\mathbf{z}|\mathbf{x},\mathbf{y}') \cdot \Delta V^k(\mathbf{y}_Z).
\end{aligned}$$

Finally, we note that the sequence $\{\alpha_n\}$ vanishes with n and we can set $\alpha_n = 0$ in the equation (6.38), when we formulate the result using the upsilon function. Q.E.D.

6.6 Formal Proof of the Permutation Lemma

The upsilon notation allows us to write all the proof of the permutation lemma in one line :

$$\Upsilon(\{P_d^{(n)}(V^n)\}) = \Upsilon(\{\hat{P}_d^{(n)}(V^n)\}) \le \Upsilon(\{\hat{P}_d^{(n)}(W^n)\}) = \Upsilon(\{P_d^{(n)}(W^n)\})$$

where we have used (6.26) and (6.36), i.e., (4.14) is valid. The proof of the theorem is obtained when we also write : $\Upsilon(\{P_d^{(n)}(V)\}) =$ as the leftmost term of this line, and $= \Upsilon(\{P_d^{(n)}(W)\})$ as the rightmost term.

7 Acknowledgement

The author wishes to thank Prof. R.Johannesson for the possibility to work at the mismatched problem in his Department and for the help in preparing the manuscript. The author is also greatful to Prof. R.Ahlswede, Prof. N.Cai, Prof. I.Csiszár, Prof. G.Kaplan, Prof. A.Lapidoth, Prof. N.Merhav, Prof. P.Narayan, and Prof. S.Shamai (Shitz) for their interest to the results and fruitful discussions.

Appendix
An Example of Mismatched Decoding

Let

$$Y = \{0, 1, 2, 3\}, \quad P_0 = P_1 = 1/2.$$

The marginal probabilities $\{PW(y)\}$ and the values of the distortion function $\{d_x(y)\}$ are defined in Table 1 by the vector **PW** and the matrix **d**, respectively.

Note that the minimization of the distortion function d, when a fixed P-composition code is used, is equivalent to the maximum likelihood decoding for the memoryless channel defined by the following matrix of transition probabilities :

$$\mathbf{W}_0 = \begin{bmatrix} 1/2 & 1/4 & 1/8 & 1/8 \\ 1/8 & 1/8 & 1/4 & 1/2 \end{bmatrix}.$$

Really

$$-\log_2 \mathbf{W}_0 = \begin{bmatrix} 1 & 2 & 3 & 3 \\ 3 & 3 & 2 & 1 \end{bmatrix}$$

and the minimization of the distortion function

$$d'_x(y) = -\log_2 W_{0,x}(y) \tag{A.1}$$

is equivalent to the minimization of the function

$$d_x(y) = c \cdot (d'_x(y) + f_x + \varphi(y)) \tag{A.2}$$

for any $c > 0$, $\{f_x\}$ and $\{\varphi(y)\}$. If

$$\begin{aligned} & c = 1 \\ & f_0 = 0, \quad f_1 = 2 \\ & \varphi(y) = \log_2 W_{0,0}(y), \quad \text{for } y = 0, 1, 2, 3 \end{aligned}$$

then the values of the distortion function, defined by (A.1), (A.2), coincide with the values given in the matrix **d**.

The matrix **W** given in Table 1 was obtained by changing the first two columns of $\mathbf{W}_0$. It does not affect the marginal probability distribution on Y, but increases the average distortion

$$\begin{aligned} d(P, W_0) &= 0.5 \cdot (4 \cdot 1/8 + 3 \cdot 1/8 + 1 \cdot 1/4 + 0 \cdot 1/2) = 0.5625 \\ d(P, W) &= 0.5 \cdot (4 \cdot 1/4 + 3 \cdot 0/8 + 1 \cdot 1/4 + 0 \cdot 1/2) = 0.625. \end{aligned}$$

Therefore, $V = W_0$ satisfies the restrictions

$$\begin{aligned} PV &= PW \\ d(P,V) &\leq d(P,W) \end{aligned}$$

and belongs to the minimization domain in (3.9).

The transition probabilities $V = \{V_x(y)\}$, given in Table 1 by the matrix $\mathbf{V}$, satisfy the same restrictions, since

$$\begin{aligned} &PV(0) = PV(3) = 0.5 \cdot (0.474 + 0.151) = 0.3125 \\ &PV(1) = PV(2) = 0.5 \cdot (0.240 + 0.135) = 0.1875 \\ &d(P,V) = 0.5 \cdot (4 \cdot 0.151 + 3 \cdot 0.135 + 1 \cdot 0.240 + 0 \cdot 0.474) = 0.6245. \end{aligned}$$

However

$$H(V|P) = 1.807 > H(W_0|P) = 1.213.$$

Thus in comparison with W_0, the distribution V gives stronger restrictions on the maximal transmission rate expressed as follows :

$$C_d(P,W) = H(PW) - H(V|P).$$

It is easy to see that maximum likelihood decoding for the memoryless channel V is also equivalent to the minimization of the distortion function $d = \{d_x(y)\}$. Really

$$-\ln \mathbf{V} = \begin{bmatrix} 1.077 & 2.059 & 2.889 & 2.727 \\ 2.727 & 2.889 & 2.059 & 1.077 \end{bmatrix}.$$

and if

$$\begin{aligned} &c = 1.21 \\ &f_0 = 0, \quad f_1 = 1.65, \\ &\varphi(y) = \ln V_0(y) \quad \text{for } y = 0,1,2,3 \end{aligned}$$

then

$$d_x(y) = c \cdot (-\ln V_x(y) + f_x + \varphi(y)).$$

Let us suppose that $n = 2000$ and that $\mathbf{x} = 0^{1000}1^{1000}$ is the transmitted codeword (we denote by s^i the sequence consisting of i repeatitions of the symbol s). Then

$$\mathbf{y} = 0^{474}1^{240}2^{135}3^{151}0^{151}1^{135}2^{240}3^{474} \in \mathbf{T}_V^n(\mathbf{x})$$

and all the other vectors of the set $\mathbf{T}_V^n(\mathbf{x})$ are obtained by all permutations of the components y_j, $j \in J_0$, and all permutations of the components y_j, $j \in J_1$, where we denote the set of the first 1000 positions by J_0 and the set of the last 1000 positions by J_1. If we want to get a vector $\mathbf{y}' \in \mathbf{T}_W^n(\mathbf{x})$ from $\mathbf{y}$ using the minimal number of pairwise permutations of the components, we have to select 99, 10, and 26 positions $j \in J_0$ where $y_j = 0$, 2, and 3, respectively, and 135 positions $j \in J_1$ where $y_j = 1$. The corresponding components should be interchanged. This procedure is illustrated in Table 2.

We can represent the transmission of the codeword $\mathbf{x}$ over the channel V^n as follows :

1) the channel assigns 135 positions $j \in J_0$ and 135 positions $j \in J_1$;

2) the channel distributes 99, 10, and 26 symbols 0, 2, and 3 on the selected positions of J_0 and generates 135 symbols 1 on the selected positions of J_1;

3) the channel distributes $474 - 99, 240, 135 - 10$, and $151 - 26$ symbols 0, 1, 2, and 3 on the remaining 865 positions of J_0 and 151, 240, and 471 symbols 0, 2, and 3 on the remaining 865 positions of J_1.

The distributions of the step 2) are given by $\mathbf{\Delta V}$, and the distributions of the step 3) are given by $\mathbf{U}$ (Table 1). The vectors $\mathbf{y}' \in \mathbf{T}_W^n(\mathbf{x})$, located from $\mathbf{y}$ at the Hamming distance $k = 270$, will be obatined if we substitute (J_1, J_0) for (J_0, J_1) to this procedure. Then the distribution of the step 2) will given by $\mathbf{\Delta W}$.

We are interested in the case

$$e^{nR} > \frac{2000!}{625! \cdot 385! \cdot 385! \cdot 625!} \cdot \left(\frac{1000!}{474! \cdot 135! \cdot 240! \cdot 151!} \right)^{-2}$$

and suppose that the decoder $\mathcal{D}$, having received one of

$$|\mathbf{T}_V^n(\mathbf{x})| = \left(\frac{1000!}{474! \cdot 135! \cdot 240! \cdot 151!}\right)^2$$

vectors $\mathbf{y} \in \mathbf{T}_V^n(\mathbf{x})$, can assign one of

$$|\mathbf{T}_{\hat{Q}}^n(\mathbf{x},\mathbf{y})| = \binom{474}{99} \cdot \binom{135}{135} \cdot \binom{240}{10} \cdot \binom{151}{26}$$

binary vectors $\mathbf{z} \in \mathbf{T}_{\hat{Q}}^n(\mathbf{x},\mathbf{y})$ of the Hamming weight 270 such that there are 99, 135, 10, and 26 positions j containing $z_j = 1$ and $(x_j, y_j) =$ (0,0), (1,1), (0,2), and (0,3), respectively. Let $\hat{\mathbf{x}} \in G_P^n \backslash \{\mathbf{x}\}$ denote an incorrect codeword, and $k_{00}(y)$ and $k_{01}(y)$ denote the number of indices j such that $(\hat{x}_j, y_j) = (0, y)$ and $(\hat{x}_j, y_j) = (1, y)$, respectively, for $y = 0, 2, 3$. Furthermore, let

$$\begin{aligned} k_{00} &= k_{00}(0) + k_{00}(2) + k_{00}(3) \\ k_{01} &= k_{01}(0) + k_{01}(2) + k_{01}(3). \end{aligned}$$

Note that we write the first index 0 because, in our case, $Y_0^+ = \{0, 2, 3\}$ and $x(y) = 0$ for $y = 0$, 2, and 3. Then the codeword $\hat{\mathbf{x}}$ belongs to a collection, which was called a ΔV-restricted code $G_P^n(\mathbf{y}_Z)$, if

$$\begin{aligned} |k_{00}(0) - k_{00} \cdot 99/135| &< \alpha_n \cdot k_{00} \\ |k_{00}(2) - k_{00} \cdot 10/135| &< \alpha_n \cdot k_{00} \\ |k_{00}(3) - k_{00} \cdot 26/135| &< \alpha_n \cdot k_{00} \end{aligned}$$

and

$$\begin{aligned} |k_{01}(0) - k_{01} \cdot 99/135| &< \alpha_n \cdot k_{01} \\ |k_{01}(2) - k_{01} \cdot 10/135| &< \alpha_n \cdot k_{01} \\ |k_{01}(3) - k_{01} \cdot 26/135| &< \alpha_n \cdot k_{01}. \end{aligned}$$

We examine the characteristics of the decoding when the decoder selects the codeword with the minimal distortion, which belongs to $G_P^n(\mathbf{y}_Z)$, and takes the average value of the decoding error probability for the codeword $\mathbf{x}$ over all possible vectors $\mathbf{z} \in \mathbf{T}_{\hat{Q}}^n(\mathbf{x},\mathbf{y})$. Note that, in this procedure, the decoder has some helping information since the vectors $\mathbf{z}$ are assigned depending on

$\mathbf{x}$. However, such a restriction of the code G_P^n is equivalent to the procedure when the decoder restricts the code using uniformly distributed vectors $\tilde{\mathbf{z}}$ having the Hamming weight

$$99 \cdot 625/474 + 135 \cdot 385/135 + 10 \cdot 385/240 + 26 \cdot 151/625 \approx 639$$

such that there are $99 \cdot 625/474$, $135 \cdot 385/135$, $10 \cdot 385/240$, and $26 \cdot 151/625$ positions j containing $\tilde{z}_j = 1$ and $y_j = 0$, 1, 2, and 3, respectively. An assignement of the vector $\tilde{\mathbf{z}}$ does not depend on the transmitted codeword, and the possibility of a reliable decoding procedure with the 'helping' information will lead to a possibility of an improvement of the behaviour, which is based on the received vector $\mathbf{y}$.

The similar considerations are also valid for the decoder $\mathcal{D}'$, which receives a vector $\mathbf{y}' \in \mathbf{T}_W^n(\mathbf{x})$ and 'constructs' ΔW-restricted codes. The decoding error probability for $\mathcal{D}'$ is not less than for $\mathcal{D}$ when they use restricted codes, and this note gives a necessary connection between the behaviour of $\mathcal{D}$ and $\mathcal{D}'$, which allows us to lower-bound the decoding error probability for $\mathcal{D}'$.

Table 1: The matrices of transition probabilities, which determine the characteristics of mismatched decoding for the channel W with respect to the distortion function d.

$y =$	0	1	2	3
$\mathbf{PW} =$	$(1/2 + 1/8)/2 =$ $= 0.3125$	$(1/4 + 1/8)/2 =$ $= 0.1875$	$(1/8 + 1/4)/2 =$ $= 0.1875$	$(1/8 + 1/2)/2 =$ $= 0.3125$
$\mathbf{d} =$	0 4	0 3	0 1	0 0
$\mathbf{W}_0 =$	1/2 1/8	1/4 1/8	1/8 1/4	1/8 1/2
$\mathbf{W} =$	1/2 - 1/8 = = 0.375 1/8 + 1/8 = = 0.250	1/4 + 1/8 = = 0.375 1/8 - 1/8 = = 0.000	1/8 = = 0.125 1/4 = = 0.250	1/8 = = 0.125 1/2 = = 0.500
$\mathbf{V} =$	0.474 0.151	0.240 0.135	0.135 0.240	0.151 0.474
$\mathbf{U} =$	0.375/ 0.865 = = 0.433 0.151/ 0.865 = = 0.175	0.240/ 0.865 = = 0.277 0.000/ 0.865 = = 0.000	0.125/ 0.865 = = 0.145 0.240/ 0.865 = = 0.277	0.125/ 0.865 = = 0.145 0.474/ 0.865 = = 0.548
$\mathbf{\Delta V} =$	0.099/ 0.135 = = 0.733 0.000/ 0.135 = = 0.000	0.000/ 0.135 = = 0.000 0.135/ 0.135 = = 1.000	0.010/ 0.135 = = 0.074 0.000/ 0.135 = = 0.000	0.026/ 0.135 = = 0.193 0.000/ 0.135 = = 0.000
$\mathbf{\Delta W} =$	0.000/ 0.135 = = 0.000 0.099/ 0.135 = = 0.733	0.135/ 0.135 = = 1.000 0.000/ 0.135 = = 0.000	0.000/ 0.135 = = 0.000 0.010/ 0.135 = = 0.074	0.000/ 0.135 = = 0.000 0.026/ 0.135 = = 0.193

Table 2: The structure of a vector $\mathbf{y} \in \mathbf{T}_V^{2000}(\mathbf{x})$, which is transformed into a vector $\mathbf{y}' \in \mathbf{T}_W^{2000}(\mathbf{x})$; the probability distributions V and W are given in Table 1 and $\mathbf{x} = 0^{1000}1^{1000}$.

	$x = 0$	$x = 1$
$Y_x^+ =$	$\{0, 2, 3\}$	$\{1\}$
$Y_x^- =$	$\{1\}$	$\{0, 2, 3\}$
$n_{x0}(\mathbf{x}, \mathbf{y}) =$	$375 + 99$	$151 + 0$
$n_{x1}(\mathbf{x}, \mathbf{y}) =$	$240 + 0$	$0 + 135$
$n_{x2}(\mathbf{x}, \mathbf{y}) =$	$125 + 10$	$240 + 0$
$n_{x3}(\mathbf{x}, \mathbf{y}) =$	$125 + 26$	$474 + 0$
$k_x(0) =$	99	0
$k_x(1) =$	0	135
$k_x(2) =$	10	0
$k_x(3) =$	26	0
$k_x =$	135	135
$k(0) =$	99	
$k(1) =$	135	
$k(2) =$	10	
$k(3) =$	26	
$k =$	270	

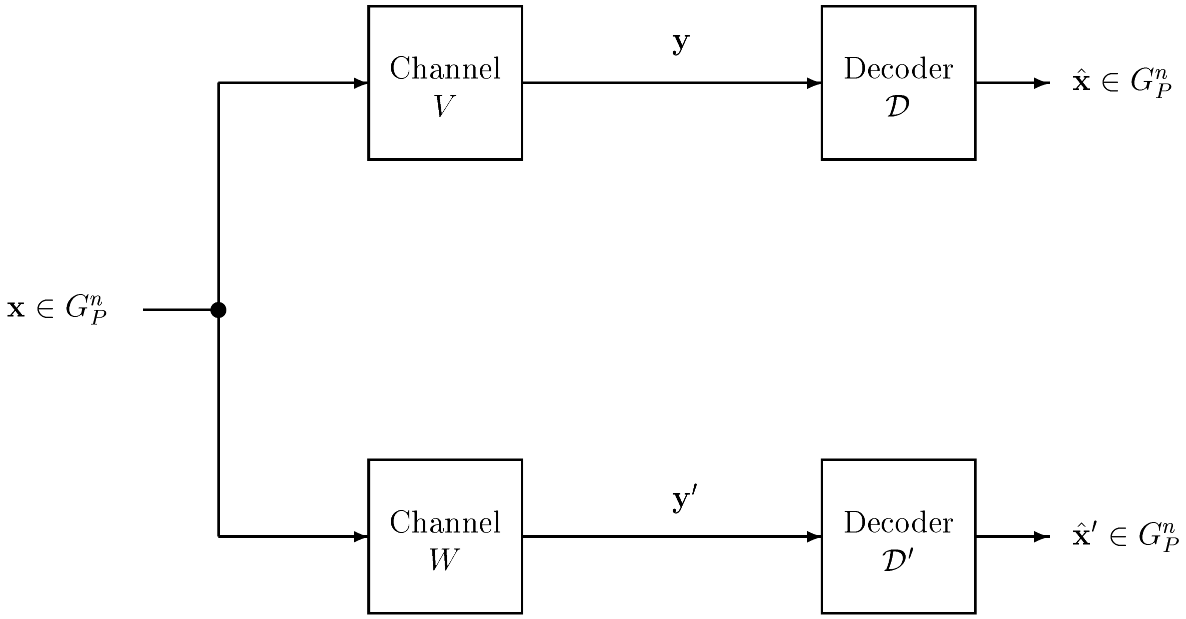

Figure 1: A system model of information transmission for mismatched decoding.

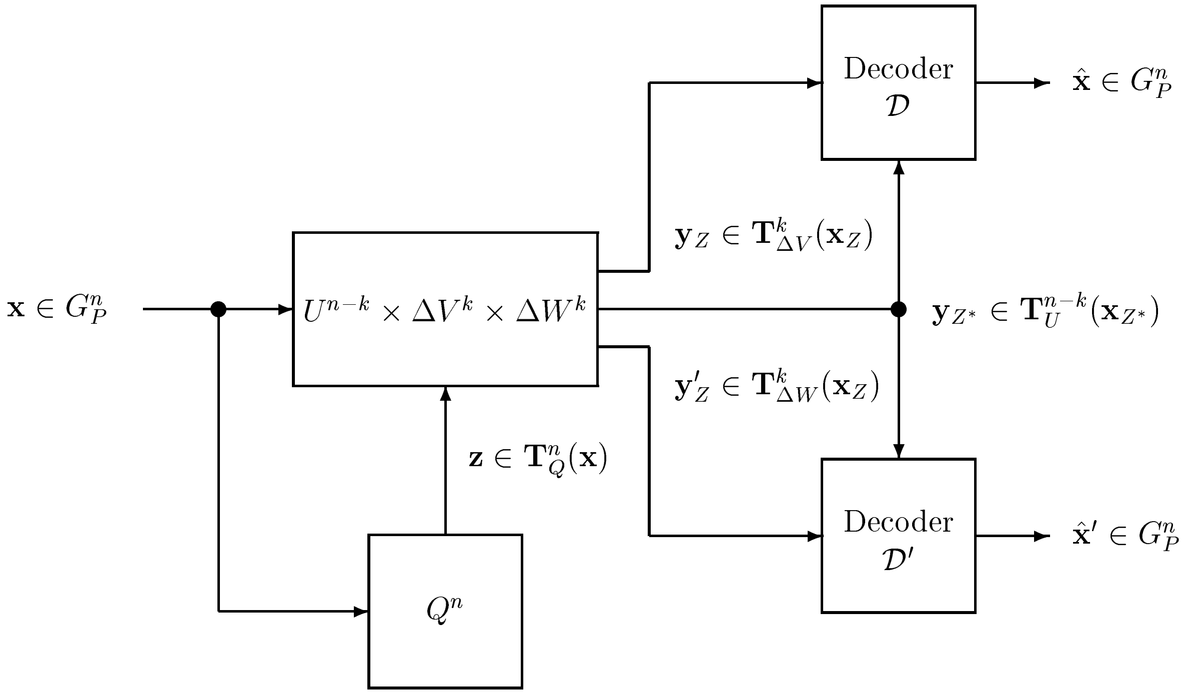

Figure 2: A broadcast model for mismatched decoding.

References

[1] C.E.Shannon, "A mathematical theory of communication," *Bell Syst. Techn. J.,* vol.27, pp.379-423 and 623-656, July and Oct. 1948.

[2] I.Csiszár and J.Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic Press, 1981.

[3] J.Y.N.Hui, Fundamental issues of multiple accessing. Ph.D dissertation, MIT, 1983.

[4] I.Csiszár and J.Körner, "Graph decomposition : A new key to coding theorems," *IEEE Trans. Inform. Theory,* vol.IT-27, pp.5-12, Jan. 1981.

[5] V.B.Balakirsky, "Coding theorems for discrete memoryless channels with given decision rule," *Lecture Notes on Computer Science,* No.573, Proceedings of 1st French-Soviet Workshop on Algebraic Coding, July 1991, pp.142-150.

[6] A.Lapidoth, "Information rates for mismatched decoders," in *Proc. 2-nd Winter Meeting on Coding and Information Theory* (Essen, Germany, Dec. 1993), p.12-15.

[7] I.Csiszár and P.Narayan, "Channel capacity for a given decoding rule," in *Proc. IEEE Int.Symp. on Information Theory* (Trondheim, Norway, June-July 1994), p.378.

[8] A.Lapidoth, "Mismatched decoding and the multiple access channel," in *Proc. IEEE Int.Symp. on Information Theory* (Trondheim, Norway, June-July 1994), p.382.

[9] I.Csiszár and P.Narayan, "Channel capacity for a given decoding rule," *IEEE Trans. Inform. Theory.* vol.IT-41, pp.35-43, Jan. 1995.

[10] N.Merhav, G.Kaplan, A.Lapitoth, and S.Shamai(Shitz), "On information rates for mismatched decoders," *IEEE Trans. Inform. Theory,* vol.IT-40, pp.1953-1967, Nov. 1994.