Identification via Channels

RUDOLF AHLSWEDE AND GUNTER DUECK, MEMBER, IEEE

Abstract—**Our main finding is that any object among $N = 2^{2^{nR}}$ (doubly exponentially many!) objects can be *identified* in blocklength n with arbitrarily small error probability via a discrete memoryless channel (DMC), if *randomization* can be used for the encoding procedure. Moreover, we present a novel doubly exponential coding theorem, which determines the optimal R, that is, *the identification capacity* of the DMC as a function of its transmission probability matrix. Surprisingly, this identification capacity is a well-known quantity, namely, Shannon's transmission capacity for the DMC.**

I. Results and Preliminaries

A. Formulation of the Classical Transmission Problem

TO PUT our new coding theorems for identification in a proper perspective, we describe first the analogous classical situation for transmission. A stochastic matrix $W = \{W(y|x)\colon x \in \mathcal{X},\ y \in \mathcal{Y}\}$ uniquely defines a discrete memoryless channel with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and transmission probabilities

$$W^n(y^n|x^n) = \prod_{t=1}^{n} W(y_t|x_t)$$

for n-sequences $x^n = (x_1, \cdots, x_n) \in \mathcal{X}^n$, $y^n = (y_1, \cdots, y_n) \in \mathcal{Y}^n$, $n = 1, 2, 3, \cdots$.

For a set $\mathcal{S}$, $\mathcal{P}(\mathcal{S})$ always stands for the set of probability distributions on $\mathcal{S}$. An (n, M, λ) code for W is a set of pairs $\{(u_i, \mathcal{D}_i)|i = 1, \cdots, M\}$ with the properties

$$u_i \in \mathcal{X}^n,\ \mathcal{D}_i \subset \mathcal{Y}^n, \qquad \text{for } i \in \{1, \cdots, M\} \quad (1.1)$$

$$\mathcal{D}_i \cap \mathcal{D}_j = \varnothing \qquad \text{for } i, j \in \{1, \cdots, M\} \text{ with } i \neq j \quad (1.2)$$

$$W^n(\mathcal{D}_i|u_i) \geq 1 - \lambda, \qquad \text{for } i \in \{1, \cdots, M\}. \quad (1.3)$$

Let $M(n, \lambda)$ be the maximal integer M for which an (n, M, λ) code exists.

Theorem S (Shannon [2]):

$$\lim_{n \to \infty} \frac{1}{n} \log M(n, \lambda) = C \qquad \text{for all } \lambda \in (0,1)$$

where $C = \max_P I(P, W)$.

Here $I(P, W) = H(PW) - H(W|P)$ is the familiar mutual information associated with $P \in \mathcal{P}(\mathcal{X})$. Actually, Shannon proved in [2] only the direct part of the Theorem. The so-called strong converse was proved by Wolfowitz (see [3]).

B. Formulation of the Identification Problem

A (randomized) identification (ID) code $(n, N, \lambda_1, \lambda_2)$ is a family

$$\{(Q(\cdot|i), \mathcal{D}_i)|i = 1, \cdots, N\}$$

of pairs with

$$Q(\cdot|i) \in \mathcal{P}(\mathcal{X}^n),\ D_i \subset \mathcal{Y}^n, \qquad \text{for } i = 1, \cdots, N \quad (1.4)$$

and with errors of the first (resp. second) kind satisfying

$$\sum_{x^n \in \mathcal{X}^n} Q(x^n|i) W^n(\mathcal{D}_i^c|x^n) \leq \lambda_1 \quad (1.5)$$

and

$$\sum_{x^n \in \mathcal{X}^n} Q(x^n|j) W^n(\mathcal{D}_i|x^n) \leq \lambda_2 \quad (1.6)$$

for all $i = 1, \cdots, N$, $j = 1, \cdots, N$ with $j \neq i$.

Of course, we also could have defined deterministic ID codes where the $Q(\cdot|i)$ denote point masses on points $u_i \in \mathcal{X}^n$. However, the study of deterministic ID codes leads only to very poor results (see the Discussion, Section IV). Therefore, we consider only the much more powerful randomized ID codes.

The essential difference between ID codes and classical transmission codes is that no disjointness condition is imposed on the decoding sets $\mathcal{D}_i$. In an ID code, the decoding sets have to be only pairwise significantly different in the sense specified by (1.5) and (1.6).

We now explain how ID codes arise naturally as the appropriate code concept in an identification problem. Assume there is a set $\mathcal{E} = \{e_1, \cdots, e_N\}$ of events (or objects), any one of which may occur. The event is known to the sender of the channel, but unknown to the receiver. On the receiver's side is a set of persons (or devices) $\mathcal{F} = \{F_1, \cdots, F_N\}$ observing the output of the channel. *Person F_i wants to know whether or not event e_i occurred.* The sender can transmit his knowledge of the event over the channel. For this transmission procedure, randomization is allowed, that is, an encoding rule for an event e_i is formally described by a probability distribution (PD) $Q(\cdot|i)$ out of $\mathcal{P}(\mathcal{X}^n)$. Clearly, F_i can choose a decision rule specifying sequences y^n for which s/he assumes that e_i has occurred. This rule is represented by the decoding set $\mathcal{D}_i \subset \mathcal{Y}^n$. Thus one is led to the notion of an ID code as described above. Randomized decision rules on the re-

Manuscript received April 9, 1986; revised February 9, 1988. This paper was presented at the 17th European Meeting of Statisticians, Thessalonini, Greece, August 24–28, 1987, and at the IEEE Workshop, Bellagio, Italy, June 1987.

R. Ahlswede is with the Facultät für Mathematik, Universität Bielefeld, Universitätsstrasse, Postfach 8640, 4800 Bielefeld 1, Germany.

G. Dueck was with the Universität Bielefeld, Bielefeld, Germany. He is now with the IBM Scientific Center Heidelberg, Tiergartenstraße 15, 6900 Heidelberg, Germany.

IEEE Log Number 8825709.

ceiver's side are not considered because they yield only minor improvements in the present coding problem.

The identification problem can also be stated in the following way. Instead of N persons, we can assume that the receiver wants to know whether or not e_j occurred. The parameter j is not known to the sender; that is, the sender does not know what the receiver wants to identify.

At the end of the paper we give examples for which the present model is suitable and discuss its relation to identification problems found in the literature [4], [5].

C. The Double Exponent Coding Theorem

Let $N(n,\lambda)$ be the maximal number N such that an $(n, N, \lambda_1, \lambda_2)$ ID code with $\lambda_1, \lambda_2 \le \lambda$ exists, and let C be Shannon's *transmission* capacity of the DMC W.

Theorem 1 (Coding Theorem and Soft Converse):

a) $\liminf_{n\to\infty} 1/n \log\log N(n,\lambda) \ge C$, for all $\lambda \in (0,1]$.

b) $\limsup_{n\to\infty} 1/n \log\log N(n,2^{-n\epsilon}) \le C$, for all $\epsilon > 0$.

We used the term "soft" converse for statement b) of Theorem 1 because the error probability on the left side is exponentially small. In the usual terminology a "weak" converse would mean

$$\inf_{\lambda\in(0,1)} \limsup_{n\to\infty} \frac{1}{n} \log\log N(n,\lambda) \le C \qquad (1.7)$$

which is a sharper bound than that in b). An even stronger bound

$$\limsup_{n\to\infty} \frac{1}{n} \log\log N(n,\lambda) \le C, \quad \text{for all } \lambda \in \left(0, \tfrac{1}{2}\right) \qquad (1.8)$$

would be called a "strong" converse. Note that (1.8) is not true for $\lambda > 1/2$.

To make this plausible, consider an arbitrary system of encoding PD's together with the following decoding rule: if y^n is received and if it is to be decided whether or not e_i occurred, decide "yes" with probability $1/2$. Clearly, both error probabilities (first and second kind) are below $\lambda > 1/2$.

We do not know whether (1.7) or (1.8) holds. One has to keep in mind that in ordinary coding theory, when dealing with constant λ the codelength grows exponentially. Thus, since now $N(n,\lambda)$ grows doubly exponentially in n, one may have to change the scale of error performance to an exponential decline to get a fair comparison. In any case, a proof of (1.7) (or even (1.8)), if true, would require very delicate estimates.

We derive better estimates than those stated in Theorem 1. For their description we use the notions of information and I-divergence, and some notation from [6] and [1]. The reader not familiar with these is referred to Section I-D. In those sharper estimates, we are concerned with *error exponents*, which can be achieved with a certain (second-order) rate.

The triple (R, E_1, E_2) is called achievable if, for all $\delta > 0$ and $n \ge n(\delta, |\mathcal{X}|, |\mathcal{Y}|)$, an ID code exists for N messages and error probabilities $\lambda_1(n)$, $\lambda_2(n)$ such that

$$\frac{1}{n} \log\log N \ge R - \delta, \ \lambda_i \le \exp\{-n(E_i - \delta)\}, \quad i = 1,2. \qquad (1.9)$$

For achievable triples we have the following result.

Theorem 2: a) If $P \in \mathcal{P}(\mathcal{X})$ satisfies $I(P,W) > R + 2E_2$, then

$$\left(R, \min_{I(P,V)\le R+2E_2} D(V\|W|P), E_2\right)$$

is achievable. b) If $E_1 > 0$ and $R + 2E_2 > C$, then (R, E_1, E_2) is not achievable.

Remarks:

1) Theorem 2-a) clearly implies Theorem 1-a).

2) Theorem 2-b) implies, formally, that

$$\liminf_{n\to\infty} \frac{1}{n} \log\log N(n,2^{-n\epsilon}) \le C.$$

From the proof of Theorem 2-b), however, it will become clear that the same is true for the limes superior.

3) Since $D(V\|W|P)$ is a continuous function with the property $D(V\|W|P) > 0$ if and only if $V \ne W$ we see that the condition $I(P,W) > R + 2E_2$ in Theorem 2-a) implies that

$$\min_{I(P,V)\le R+E_2} D(V\|W|P) > 0.$$

4) Theorem 2 completely characterizes the set of achievable pairs (R, E_2) in the limit $E_1 \to 0$. More precisely,

$$\lim_{E_1\to 0} \{(R,E_2): (R,E_1,E_2) \text{ is achievable}\} = \{(R,E_2): R \le C - 2E_2\}.$$

In the remainder of this section we prepare the reader for the results of Theorem 1 and its proof. The fact that the maximal codelength grows doubly exponentially can more easily be understood for the very special case of a noiseless binary channel. We include a complete proof. We then comment on our proof for the direct part of Theorem 1 for the general DMC, and, finally, on the proof of the converse part.

We start with the construction of n-block ID codes for the binary channel W given by the input alphabet $\mathcal{X} := \{0,1\}$, $\mathcal{Y} := \{0,1\}$, and $W(1|1) = W(0|0) = 1$. We use the standard maximal coding argument.

Let n be the blocklength, and let $\lambda \in (0,1/2)$ be given. Let 2^l be the smallest power of 2, such that

$$\lambda \cdot \log(2^l - 1) > 1 \text{ and } 2^l > 6.$$

Suppose that n is large compared with 2^l. Set

$$M := 2^{n-l}.$$

We define an n-block ID code

$$\{(Q(\cdot|i), \mathcal{D}_i) | i = 1, \cdots, N\}$$

such that $\log\log N$ is close to $n\log 2$. We restrict our

attention to distributions $Q(\cdot|i)$ which are equidistributions on sets $\mathscr{A}_i \subset \mathscr{X}^n$ with cardinality M. Since M equals 2^{n-l}, we therefore consider only equidistributions on relatively large subsets of $\mathscr{X}^n$. Suppose now we have found subsets $\mathscr{A}_1, \cdots, \mathscr{A}_N \subset \mathscr{X}^n$, all of which have cardinality M and such that

$$|\mathscr{A}_i \cap \mathscr{A}_j| < \lambda \cdot M, \qquad \text{for all } i, j \in \{1, \cdots, N\}, i \neq j. \tag{1.10}$$

Then we define

$$Q(\cdot|i) := \text{equidistribution on } \mathscr{A}_i$$
$$\mathscr{D}_i := \mathscr{A}_i, \qquad \text{for } i = 1, \cdots, N.$$

Consider $\{(Q(\cdot|i), \mathscr{D}_i) | i = 1, \cdots, N\}$. We claim that this system is an $(n, N, 0, \lambda)$ ID code. This is true because

$$\sum_{x^n} Q(x^n|i) W^n(\mathscr{D}_i|x^n) = 1$$
$$\sum_{x^n} Q(x^n|j) W^n(\mathscr{D}_i|x^n) = M^{-1}|\mathscr{A}_i \cap \mathscr{A}_j| < \lambda$$

for $i, j \in \{1, \cdots, N\}$, $i \neq j$. Here we used the special nature of W and assumption (1.10).

We have seen now that it suffices to show the existence of a large family $\mathscr{A}_1, \cdots, \mathscr{A}_N$ of sets of cardinality M which satisfies (1.10).

Proposition 1: Let $\mathscr{X}$ be a finite set and let $\lambda \in (0, 1/2)$ be given. If ϵ is so small that

$$\lambda \log\left(\frac{1}{\epsilon} - 1\right) > 2 \qquad \frac{1}{\epsilon} > 6,$$

then a family $\mathscr{A}_1, \cdots, \mathscr{A}_N$ of subsets of $\mathscr{X}$ exists satisfying

$$|\mathscr{A}_i| = \lfloor \epsilon |Z| \rfloor, \qquad \text{for all } i \in \{1, \cdots, N\},$$
$$|\mathscr{A}_i \cap \mathscr{A}_j| < \lambda \lfloor \epsilon \cdot |\mathscr{X}| \rfloor, \qquad \text{for } i, j \in \{1, \cdots, N\}, \quad i \neq j$$

and

$$N \geq |\mathscr{X}|^{-1} \cdot 2^{\lfloor \epsilon |\mathscr{X}| \rfloor} - 1.$$

Proof: Choose as a starting point an arbitrary $\mathscr{A}_1 \subset \mathscr{X}$, $|\mathscr{A}_1| = \lfloor \epsilon \cdot |\mathscr{X}| \rfloor$. We count how many sets $\mathscr{A} \subset \mathscr{X}$ exist with cardinality $\lfloor \epsilon \cdot |\mathscr{X}| \rfloor$ and

$$|\mathscr{A}_1 \cap \mathscr{A}| \geq \lambda \lfloor \epsilon |\mathscr{X}| \rfloor.$$

We define $M' := \lfloor \epsilon |\mathscr{X}| \rfloor$. The number of those sets $\mathscr{A}$ in question is then

$$\sum_{i = \lceil \lambda \cdot M' \rceil}^{M'} \binom{|\mathscr{X}| - M'}{M' - i} \binom{M'}{i}. \tag{1.11}$$

For $\lambda < 1/2$ and $1/\epsilon > 6$ the first summand in the sum is the maximal one. This is easy to establish. Therefore, the sum in (1.11) can be upper-bounded by

$$M' \cdot \binom{|\mathscr{X}| - M'}{M' - \lceil \lambda M \rceil} \binom{M'}{\lceil \lambda M' \rceil} \leq M' \cdot \binom{|\mathscr{X}|}{M' - \lceil \lambda M' \rceil} \cdot 2^{M'} =: T. \tag{1.12}$$

Hence at most T sets $\mathscr{A}$ of cardinality M' exist such that

$$|\mathscr{A}_1 \cap \mathscr{A}| \geq \lambda \cdot M'.$$

There are $\binom{|\mathscr{X}|}{M'}$ sets of cardinality M'. If $T < \binom{|\mathscr{X}|}{M'}$, then an $\mathscr{A}_2 \subset \mathscr{X}$ exists with $|\mathscr{A}_2| = M'$ and $|\mathscr{A}_1 \cap \mathscr{A}_2| < \lambda \cdot M'$. Furthermore, if $2T < \binom{|\mathscr{X}|}{M'}$, then $\mathscr{A}_3 \subset \mathscr{X}$ exists with $|\mathscr{A}_3| = M'$, $|\mathscr{A}_3 \cap \mathscr{A}_1| < \lambda \cdot M'$, and $|\mathscr{A}_3 \cap \mathscr{A}_2| < \lambda M'$. By repeatedly using this argument, we get the following result.

There are M'-element sets $\mathscr{A}_1, \cdots, \mathscr{A}_{N^*} \subset \mathscr{X}$ such that $|\mathscr{A}_i \cap \mathscr{A}_j| < \lambda \cdot M'$ for $i \neq j$, if

$$N^* \cdot T < \binom{|\mathscr{X}|}{M'}.$$

Hence a family of sets $\mathscr{A}_1, \cdots, \mathscr{A}_N$ exists with

$$N := \left\lfloor \binom{|\mathscr{X}|}{M'} \cdot T^{-1} \right\rfloor - 1. \tag{1.13}$$

Recall that T was defined in (1.12). It is now easy to lower-bound N. By (1.12) and (1.13),

$$N \geq 2^{-M'} \cdot M'^{-1} \cdot \prod_{i=1}^{\lceil \lambda M' \rceil} \frac{|\mathscr{X}| - M' + i}{M' - \lceil \lambda M' \rceil + 1} - 1.$$

Since $M' = \lfloor \epsilon |\mathscr{X}| \rfloor$ and $\lambda \leq 1/2$, for $i \in \{1, \cdots, \lceil \lambda M' \rceil\}$

$$\frac{|\mathscr{X}| - M' + i}{M' - \lceil \lambda M' \rceil + i} \geq \frac{1}{\epsilon} - 1.$$

Hence

$$N + 1 \geq 2^{-M'} \cdot M'^{-1} \cdot \left(\frac{1}{\epsilon} - 1\right)^{\lceil \lambda \cdot M' \rceil}$$
$$\geq 2^{-M'} \cdot \left(\frac{1}{\epsilon} - 1\right)^{M'} \cdot M'^{-1}$$
$$= 2^{M'(\lambda \log((1/\epsilon) - 1) - 1)} \cdot M'^{-1}$$
$$\geq 2^{M'} \cdot |\mathscr{X}|^{-1} = 2^{\lfloor \epsilon \cdot |\mathscr{X}| \rfloor} \cdot |\mathscr{X}|^{-1},$$

where we have used the assumption in the proposition. The proof is complete.

We return to the binary noiseless channel. We apply the result of Proposition 1 to $\{0,1\}^n$ instead of $\mathscr{X}$ and with 2^{-l} instead of ϵ. We conclude that there are at least

$$N := 2^{-n} \cdot 2^{2^{n-l}} - 1$$

sets $\mathscr{A}_1, \cdots, \mathscr{A}_N$ of $\{0,1\}^n$ with cardinality 2^{n-l} such that

$$|\mathscr{A}_i \cap \mathscr{A}_j| < \lambda \cdot 2^{n-l} \qquad \text{for } i \neq j.$$

In other words, we have found an $(n, N, 0, \lambda)$ identification code. Clearly, $(1/n) \log \log N$ is arbitrarily close to $\log 2$, the capacity of the binary noiseless channel, if n grows to infinity.

Thus Theorem 1-a) is proved for the noiseless binary channel. The validity of Theorem 1-b) is easy to see for this channel. Obviously, the number of messages in an identification code cannot exceed the number of possible decoding sets, because all the decoding sets have to be different.

Since all the decoding sets of n-block length codes are subsets of $\{0,1\}^n$, there are at most 2^{2^n} decoding sets in an identification code. We shall see in Section II that the construction in Proposition 1 can be used to construct

good ID codes with the help of an underlying classical transmission code with rate close to capacity.

The converse part of the proof, however, is highly complicated. One has to show that there is no advantage in considering equidistributions on subsets with a cardinality larger than $\exp\{nC\}$.

We originally wanted to show that, starting with a given code with equidistributions on large sets, we can find "smaller" equidistributions on sets with cardinality smaller than $\exp\{nC\}$ such that the resulting decoding sets in the new code are nearly the same as in the originally given code. Then one could conclude that any ID code could have at most as many messages as subsets of $\mathcal{X}^n$ with cardinality smaller than $\exp\{nC\}$, and the proof would be complete.

Unfortunately, we were not able to prove the converse part in this elegant version; we were, however, able to prove it following this basic idea. In the next section (Section I-D) we introduce some notation. This notation is not needed for the proof of Theorem 1-a). The reader not interested in exponential error bounds may proceed directly to Section II-A.

D. Notation and Known Facts

1) Channels, Types, Generated Sequences: We use essentially the notation of [1]. Script capitals $\mathcal{X}, \mathcal{Y}, \cdots$ denote finite sets. The cardinality of a set $\mathcal{A}$ is denoted by $|\mathcal{A}|$. The letters P, Q always stand for probability distributions on finite sets. $X, Y, \cdots$ denote random variables (RV's). The functions "log" and "exp" are understood to be to the base 2. For a stochastic $|\mathcal{X}|\times|\mathcal{Y}|$-matrix W we have already defined the transmission probabilities W^n of a DMC, and we have also introduced $\mathcal{P}(\mathcal{X}^n)$ as the set of PD's on $\mathcal{X}^n$. We abbreviate $\mathcal{P}(\mathcal{X})$ as $\mathcal{P}$. $\mathcal{W}$ denotes the set of all channels V with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. For positive integers n we set

$$\mathcal{P}_n = \{P \in \mathcal{P} | P(x) \in \{0, 1/n, 2/n, \cdots, 1\} \text{ for all } x \in \mathcal{X}\}.$$

For any $P \in \mathcal{P}_n$, called type or n-type, we define the set

$$\mathcal{W}_n(P) = \left\{V \in \mathcal{W} | V(y|x) \in \left\{0, \frac{1}{nP(x)}, \frac{2}{nP(x)}, \cdots, 1\right\}, \quad x \in \mathcal{X},\ y \in \mathcal{Y}\right\}.$$

For $x^n \in \mathcal{X}^n$ we define for every $x \in \mathcal{X}$

$$P_{x^n}(x) = \frac{1}{n} \cdot (\text{number of occurrences of } x \text{ in } x^n).$$

P_{x^n} is a member of $\mathcal{P}_n$ by definition. P_{x^n} is called type of x^n. Similarly, we define the type $P_{x^n y^n}$ for pairs $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$. For $P \in \mathcal{P}$ the set $\mathcal{T}_P^n$ of all P-typical sequences in $\mathcal{X}^n$ is given by

$$\mathcal{T}_P^n = \{x^n | P_{x^n} = P\}.$$

For $V \in \mathcal{W}$, a sequence $y^n \in \mathcal{Y}^n$ is said to be V-generated by x^n if, for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$,

$$P_{x^n y^n}(x, y) = P_{x^n}(x) \cdot V(y|x).$$

The set of those sequences is denoted by $\mathcal{T}_V^n(x^n)$. Notice that $\mathcal{T}_P^n \neq \emptyset$ if and only if $P \in \mathcal{P}_n$ and $\mathcal{T}_V^n(x^n) \neq \emptyset$ if and only if $V \in \mathcal{W}_n(P_{x^n})$. For $P \in \mathcal{P}$, $V \in \mathcal{W}$ we write PV for the PD on $\mathcal{Y}$ given by

$$PV(y) = \sum_x P(x) V(y|x), \qquad y \in \mathcal{Y}.$$

$\mathcal{T}_{PV}^n$ is the set of PV-typical sequences in $\mathcal{Y}^n$.

2) Entropy and Information Quantities: Let X be an RV with values in $\mathcal{X}$ and distribution $P \in \mathcal{P}$, and let Y be an RV with values in $\mathcal{Y}$ such that the joint distribution of (X, Y) on $\mathcal{X} \times \mathcal{Y}$ is given by

$$\Pr(X = x, Y = y) = P(x) \cdot V(y|x), \qquad V \in \mathcal{W}.$$

We write $H(P)$, $H(V|P)$, and $I(P, V)$ for the entropy $H(X)$, the conditional entropy $H(Y|X)$, and the mutual information $I(X \wedge Y)$, respectively. For $P, \tilde{P} \in \mathcal{P}$

$$D(\tilde{P}||P) = \sum_x \tilde{P}(x) \log \frac{\tilde{P}(x)}{P(x)}$$

denotes the I-divergence and for $V, \tilde{V} \in \mathcal{W}$ the quantity

$$D(\tilde{V}||V|P) = \sum_x P(x) D(\tilde{V}(\cdot|x)||V(\cdot|x))$$

stands for the conditional I-divergence.

3) Elementary Properties of Typical Sequences and Generated Sequences:

$$|\mathcal{P}_n| \leq (n+1)^{|\mathcal{X}|} \tag{1.14}$$

$$|\mathcal{W}_n(P)| \leq (n+1)^{|\mathcal{X}| \cdot |\mathcal{Y}|} \quad \text{for every } P \in \mathcal{P}_n \tag{1.15}$$

$$(n+1)^{-|\mathcal{X}|} \cdot \exp\{nH(P)\} \leq |\mathcal{T}_P^n| \leq \exp\{nH(P)\} \quad \text{for } P \in \mathcal{P}_n \tag{1.16}$$

$$|\mathcal{T}_V^n(x^n)| \geq (n+1)^{-|\mathcal{X}| \cdot |\mathcal{Y}|} \cdot \exp\{nH(V|P)\}$$
$$|\mathcal{T}_V^n(x^n) \leq \exp\{nH(V|P)\}$$
$$\text{for } P \in \mathcal{P}_n,\ V \in \mathcal{W}_n(P),\ x^n \in \mathcal{T}_P^n \tag{1.17}$$

$$W^n(y^n|x^n) = \exp\{-n(D(V||W|P) + H(V|P))\}$$
$$\text{for } P \in \mathcal{P}_n,\ V \in \mathcal{W}_n(P),\ x^n \in \mathcal{T}_P^n,\ y^n \in \mathcal{T}_V^n(x^n) \tag{1.18}$$

$$W^n(\mathcal{T}_V^n(x^n)|x^n) \leq \exp\{-nD(V||W|P)\}$$
$$W^n(\mathcal{T}_V^n(x^n)|x^n) \geq (n+1)^{-|X| \cdot |\mathcal{Y}|} \cdot \exp\{-nD(V||W|P)\}$$
$$\text{for } P \in \mathcal{P}_n,\ V \in \mathcal{W}_n(P),\ x^n \in \mathcal{T}_P^n. \tag{1.19}$$

II. The Direct Parts of the Coding Theorems

A. Proof of the Direct Part: Theorem 1-a)

We simply apply Proposition 1 for a classical transmission code for W. Let $\lambda \in (0,1)$ be given, and let $\epsilon > 0$ be so small such that

$$\lambda \log\left(\frac{1}{\epsilon} - 1\right) > 2, \qquad \frac{1}{\epsilon} > 6.$$

By Theorem S there exists for any large n an n-length

block code,

$$\mathscr{C} = \{(u_i, \mathscr{E}_i) | i = 1, \cdots, M\}$$

with maximal error bounded by λ and

$$M \geq 2^{n(\mathscr{C}-\epsilon)}.$$

Let $\mathscr{Z} := \{u_1, \cdots, u_N\}$. By Proposition 1 a family of subsets $\mathscr{A}_1, \cdots, \mathscr{A}_N$ of $\mathscr{Z}$ exists satisfying

$$|\mathscr{A}_i| = \lfloor \epsilon |\mathscr{Z}| \rfloor = \lfloor \epsilon \cdot M \rfloor \quad \text{for } i \in \{1, \cdots, N\} \tag{2.1}$$

$$|\mathscr{A}_i \cap \mathscr{A}_j| < \lambda \lfloor \epsilon \cdot |\mathscr{Z}| \rfloor \quad \text{for all } i, j \in \{1, \cdots, N\}, i \neq j, \tag{2.2}$$

$$N \geq |\mathscr{Z}|^{-1} \cdot 2^{\lfloor \epsilon |\mathscr{Z}| \rfloor}. \tag{2.3}$$

From the sets $\mathscr{A}_1, \cdots, \mathscr{A}_N$ we construct an ID code in the following simple manner. Define for $i \in \{1, \cdots, N\}$

$$Q(\cdot|i) := \text{equidistribution on } \mathscr{A}_i$$

and

$$\mathscr{D}_i := \bigcup_{u_k \in \mathscr{A}_i} \mathscr{E}_k.$$

Form the ID code

$$\{(Q(\cdot|i), \mathscr{D}_i) | i = 1, \cdots, N\}.$$

We look at the errors of first and second kind of this ID code (recall (1.5) and (1.6)). Let $i \in \{1, \cdots, N\}$, and let $u_k \in \mathscr{A}_i$. Then

$$W^n(\mathscr{D}_i^c | u_k) \leq W^n(\mathscr{E}_k^c | u_k) \leq \lambda$$

because of $\mathscr{E}_k \subset \mathscr{D}_i$ and because $\mathscr{E}_k$ is the decoding set for u_k in the transmission code $\mathscr{C}$. Hence

$$\sum_{x^n \in \mathscr{X}^n} Q(x^n|i) W^n(\mathscr{D}_i^c | x^n) = \sum_{u_k \in \mathscr{A}_k} \frac{1}{|\mathscr{A}_k|} \cdot W^n(\mathscr{D}_i^c | u_k) \leq \lambda.$$

On the other hand, for a $j \in \{1, \cdots, N\}$, $j \neq i$,

$$\begin{aligned} &\sum_{x^n \in \mathscr{X}^n} Q(x^n|j) W^n(\mathscr{D}_i | x^n) \\ &= \sum_{u_l \in \mathscr{A}_j} \frac{1}{|\mathscr{A}_j|} \cdot W^n(\mathscr{D}_i | u) \\ &= \frac{1}{|\mathscr{A}_j|} \left(\sum_{u_l \in \mathscr{A}_j \cap \mathscr{A}_i} W^n(\mathscr{D}_i | u) + \sum_{u_l \notin \mathscr{A}_j \cap \mathscr{A}_i} W^n(\mathscr{D}_i | u_l) \right) \\ &\leq \frac{1}{|\mathscr{A}_j|} \cdot \left(|\mathscr{A}_j \cap \mathscr{A}_i| + \sum_{u_l \notin \mathscr{A}_j \cap \mathscr{A}_i} W^n(\mathscr{D}_i | u_l) \right). \end{aligned}$$

If $u_l \in \mathscr{A}_i$, then $\mathscr{E}_l \cap \mathscr{D}_i = \emptyset$. Hence for such u_l the relation $\mathscr{D}_i \subset \mathscr{E}_l^c$ holds. This observation together with (2.2) yields

$$\sum_{x^n \in \mathscr{X}^n} Q(x^n|j) W^n(\mathscr{D}_i | x^n) \leq 2\lambda.$$

Equation (2.3) finally gives

$$\begin{aligned} N &\geq |M|^{-1} \cdot 2^{\lfloor \epsilon |\mathscr{Z}| \rfloor} \\ &\geq |\mathscr{X}|^{-n} \cdot 2^{\lfloor \epsilon 2^{n(C-\epsilon)} \rfloor}. \end{aligned}$$

In summary, $\{(Q(\cdot|i), \mathscr{D}_i) | i = 1, \cdots, N\}$ is an $(n, N, \lambda, 2\lambda)$ ID code with $(1/n) \log \log N$ close to $C - \epsilon$. Since λ and ϵ could be chosen arbitrarily small, Theorem 1 is proved.

Remark: Observe that for the proof we needed only Proposition 1 (which is just Gilbert's bound for constant weight sequences) and a given code for the channel W. Thus we can conclude that Theorem 1-a) holds in fact for all channels having a capacity. It is not necessary to assume that W is discrete or memoryless.

B. Proof of the Direct Part: Theorem 2-a)

Of course, one could easily derive exponential error bounds with the construction in the preceding section. The difference would be instead of a code with maximal error λ one would start with a code having exponentially small error probability. Furthermore, one would choose $2^{-n\xi}$ instead of ϵ.

However, Theorem 2-a) gives a stronger result than the one obtainable by this simple method. Theorem 2-a) gives, in the sense expressed in Remark 4, a best possible error exponent. The principal idea is random selection of ID codes rather than a maximal coding idea which led to Proposition 1. The key step is the application of Proposition 2 which we will present soon. Its proof is rather technical, so we give here only the short proof of Theorem 2-a) assuming that Proposition 2 holds. The proof of Proposition 2 can be found in the Appendix.

Let $P \in \mathscr{P}_n$. We consider here only ID codes of a special structure. Every message i is encoded by the uniform distribution on a family $\mathscr{U}_i$ of members of $\mathscr{T}_P^n$ satisfying $|\mathscr{U}_i| = M$ for all $i = 1, \cdots, N = \lfloor 2^{2^{nR}} \rfloor$.

Let $(R, \mathscr{E}_2)$ be given. We assign to $\mathscr{U}_i$ a decoding set $\mathscr{D}_i = \mathscr{D}(\mathscr{U}_i)$ defined by

$$\mathscr{D}(\mathscr{U}_i) = \bigcup_{u \in \mathscr{U}_i} \mathscr{F}_u \tag{2.4}$$

where

$$\mathscr{F}_u = \bigcup_{V:\, I(P,V) > R + 2E_2} \mathscr{T}_V^n(u). \tag{2.5}$$

First notice that

$$\begin{aligned} W^n(\mathscr{F}_u^c | u) &\leq \sum_{V:\, I(P,V) \leq R + 2E_2} W^n(\mathscr{T}_V^n(u) | u) \\ &\leq \sum_{V:\, I(P,V) \leq R + 2E_2} \exp\{-nD(V\|W|P)\} \\ &\leq (n+1)^{|\mathscr{X}| \cdot |\mathscr{Y}|} \\ &\quad \cdot \exp\left\{ -n \min_{V:\, I(P,V) \leq R + 2E_2} D(V\|W|P) \right\} \end{aligned}$$

by (1.19) and (1.15) and thus

$$\frac{1}{M}\sum_{u\in\mathscr{U}_i} W^n(\mathscr{D}_i^c|u) \le (n+1)^{|\mathscr{X}|\cdot|\mathscr{Y}|} \cdot \exp\Big\{-n \min_{V:\, I(P,V)\le R+2E_2} D(V\|W|P)\Big\}.$$

This means that regardless of the choice of $\mathscr{U}_i$ the error exponent of the first kind

$$E_1 := \min_{V:\, I(P,V)\le R+2E_2} D(V\|W|P)$$

is achievable. We now specify the sets $\mathscr{U}_i$ to achieve (R, E_2). We choose

$$M = \lfloor \exp\{n(R+E_2)\}\rfloor \tag{2.6}$$

and define the $\mathscr{U}_i$ by random selection as follows.

Let U_{ij}, $i=1,\cdots,N$; $j=1,\cdots,M$, be independent random variables, all uniformly distributed over $\mathscr{T}_P^n$. Define the random families

$$\bar{\mathscr{U}}_i = \{U_{i1},\cdots,U_{iM}\}, \qquad i=1,\cdots,N.$$

Every realization of the U_{ij} gives rise to an ID code as described before.

We want to show that a large fraction of these randomly selected ID codes has an error exponent of the second kind at least $E_2-\delta$ ($\delta>0$ arbitrarily small, n sufficiently large for δ). In fact, we can get this result (and therefore Theorem 2-a)) by the following result for two messages.

Let $\mathscr{U}_1$ be any subset of $\mathscr{T}_P^n$ of cardinality M and let $\bar{\mathscr{U}}_2$ be as described above. We consider the probability P^* that for $\gamma>0$:

$$\frac{1}{M}\sum_{u\in\mathscr{U}_1} W^n\big(\mathscr{D}(\bar{\mathscr{U}}_2)|u\big) \le \exp\{-n(E_2-3\gamma)\} \tag{2.7}$$

and

$$\frac{1}{M}\sum_{u\in\bar{\mathscr{U}}_2} W^n\big(\mathscr{D}(\mathscr{U}_1)|u\big) \le \exp\{-n(E_2-3\gamma)\}. \tag{2.8}$$

Then we have the following proposition.

Proposition 2: For any $\gamma>0$ and $n\ge n(\gamma)$,

$$P^* \ge 1-(n+1)\exp\{-(n\gamma-2)\cdot\exp\{nR\}\}.$$

A proof is given in the Appendix. We show here that Theorem 2-a) follows from Proposition 2. Imagine that the random selection is performed iteratively and that the realizations $\mathscr{U}_1,\cdots,\mathscr{U}_t$ have the desired error performances.

Then, by Proposition 2, with the choice $\gamma=\delta/3$, $(\mathscr{U}_1,\cdots,\mathscr{U}_t,\bar{\mathscr{U}}_{t+1})$ has the desired error performances with probability exceeding $1-t(1-P^*)$. Therefore, there exists a sufficiently good realization $\mathscr{U}_{t+1}$ of $\bar{\mathscr{U}}_{t+1}$, if

$$1-t(1-P^*)>0.$$

Since by Proposition 2 even $1-N(1-P^*)>0$ for large n, it is possible to find a realization $(\mathscr{U}_1,\cdots,\mathscr{U}_N)$ of $(\bar{\mathscr{U}}_1,\cdots,\bar{\mathscr{U}}_N)$ with the desired error performances.

The proof of Proposition 2 follows the large deviations approach of [3] in the improved form of [1]. Specifically, we shall need a very useful lemma on large deviations which we state at the end of the section. The proof can be found in the Appendix.

Lemma LD (Generalized Chebyshev Inequality, Bernstein's Trick, Chernov's Bound): Let $\Psi_1,\cdots,\Psi_M$ be independent identically distributed RV's with values in $\{0,1\}$. Suppose that the expectation $\mathbf{E}\Psi_1$ of Ψ_1 satisfies $\mathbf{E}\Psi_1\le\mu<\lambda\le 1$; then

$$\Pr\Big(\sum_{i=1}^{M}\Psi_i > M\cdot\lambda\Big) \le \exp\{-M\cdot D(\lambda\|\mu)\},$$

where $D(\lambda\|\mu)$ denotes the I-divergence between $(\lambda,1-\lambda)$ and $(\mu,1-\mu)$.

III. The Converses

A. Rearranging ID Codes

In the following analysis it is essential that the error probabilities λ_i be exponentially small (or at least of the order $O(n^{-c})$ for sufficiently large c), that is

$$\lambda_i \le 2^{-\epsilon_i n} \quad \text{with } \epsilon_i>0 \quad \text{for } i=1,2. \tag{3.1}$$

Starting with an $(n,N,\lambda_1,\lambda_2)$ ID code $\{(Q(\cdot|i),\mathscr{D}_i)|i=1,\cdots,N\}$, we proceed in several steps to construct a modified code whose structure is described in Proposition 3 below. This modified code will have essentially the same length and error performances. The intermediate results in each step are put into the form of lemmas.

Step 1: From general $Q(\cdot|i)$ to uniform distributions on sets $\mathscr{U}_i\subset\mathscr{X}^n$ $(i=1,\cdots,N)$.

Lemma 4: For every $(n,N,\lambda_1,\lambda_2)$ ID code $\{(Q(\cdot|i),\mathscr{D}_i)|i=1,\cdots,N\}$ and every $\epsilon>0$, sets $\mathscr{U}_i\subset\mathscr{X}$ exist for $i=1,\cdots,N$ such that the ID code $\{(Q'(\cdot|i),\mathscr{D}_i)|i=1,\cdots,N\}$ formed with the uniform distributions $Q'(\cdot|i)$ on $\mathscr{U}_i$ has error probabilities λ_i' satisfying

$$\lambda_i' \le \lambda_i\big(\epsilon^{-1}\log|\mathscr{X}|+1\big)\cdot 2^{n\epsilon}\cdot(1-2^{-n\epsilon})^{-1}.$$

Proof: For $i\in\{1,\cdots,N\}$ and $k:=\lfloor\epsilon^{-1}\log|\mathscr{X}|+1\rfloor$, define

$$\mathscr{X}^n(l,i) = \{x^n|2^{-l n\epsilon} < Q(x^n|i) \le 2^{-(l-1)n\epsilon}\} \tag{3.2}$$

for $l=1,\cdots,k$. Clearly, this definition implies

$$Q\Big(\mathscr{X}^n - \bigcup_{l=1}^{k}\mathscr{X}^n(l,i)|i\Big) \le 2^{-kn\epsilon}|\mathscr{X}^n| \le 2^{-n\epsilon}.$$

Now choose $l^*=l(i)$ such that

$$Q(\mathscr{X}^n(l^*,i)|i) \ge Q(\mathscr{X}^n(l,i)|i) \qquad \text{for } l=1,\cdots,k.$$

Then, for $\mathscr{U}_i := \mathscr{X}^n(l^*,i)$, we have

$$Q(\mathscr{U}_i|i) \ge k^{-1}(1-2^{-n\epsilon}). \tag{3.3}$$

Define Q' by

$$Q'(x^n|i) = \begin{cases} |\mathscr{U}_i|^{-1}, & \text{for } x^n\in\mathscr{U}_i \\ 0, & \text{otherwise} \end{cases}. \tag{3.4}$$

From (3.4) we conclude that for $x^n\in\mathscr{U}_i$

$$Q'(x^n|i) = |\mathscr{U}_i|^{-1} \le |\mathscr{U}_i|^{-1}k(1-2^{n\epsilon})^{-1}Q(\mathscr{U}_i|i). \tag{3.5}$$

By the definition of $\mathcal{U}_i$ we have for $x^n, \bar{x}^n \in \mathcal{U}_i$

$$Q(\bar{x}^n|i) \le 2^{n\epsilon} Q(x^n|i).$$

Thus $|\mathcal{U}_i|^{-1} Q(\mathcal{U}_i|i) \le 2^{n\epsilon} Q(x^n|i)$, and from (3.5) we derive for $x^n \in \mathcal{U}_i$

$$Q'(x^n|i) \le k(1-2^{-n\epsilon})^{-1} 2^{n\epsilon} Q(x^n|i). \tag{3.6}$$

Therefore, for any $\mathcal{D} \subset \mathcal{Y}^n$ and $i = 1, \cdots, N$

$$\sum_{x^n} Q'(x^n|i) W^n(\mathcal{D}|x^n) \le k \cdot (1-2^{-n\epsilon})^{-1} 2^{n\epsilon} \cdot \sum_{x^n} Q(x^n|i) W^n(\mathcal{D}|x^n). \tag{3.7}$$

Choosing $\mathcal{D}_i^c$ (resp. $\mathcal{D}_j$) for $\mathcal{D}$ in (3.7) one gets the desired bounds on λ_1' (resp. λ_2').

Step 2: Towards locally good codes.

We call a code $\{(Q(\cdot|i), \mathcal{D}_i) | i = 1, \cdots, N\}$ locally good if, for all $i = 1, \cdots, N$,

$$W^n(\mathcal{D}_i|x^n) \ge 1 - \lambda_1 \qquad \text{for all } x^n \text{ with } Q(x^n|i) > 0. \tag{3.8}$$

Lemma 5: To every $(n, N, \lambda_1, \lambda_2)$ ID code $\{(Q(\cdot|i), \mathcal{D}_i) | i = 1, \cdots, N\}$ we can assign a locally good code $\{(Q^*(\cdot|i), \mathcal{D}_i) | i = 1, \cdots, N\}$ with error performances $\lambda_i^* \le 2\lambda_i$ $(i = 1, 2)$. Moreover, if $Q(\cdot|i)$ is a uniform distribution, then $Q^*(\cdot|i)$ can also be chosen to be uniform.

Proof: Consider $\mathcal{U}_i^* = \{x^n | W^n(\mathcal{D}_i|x^n) \ge 1 - 2\lambda_1\}$ and define

$$Q^*(x^n|i) = \begin{cases} Q(x^n|i) \cdot Q(\mathcal{U}_i^*|i)^{-1}, & \text{for } x^n \in \mathcal{U}_i^* \\ 0, & \text{otherwise} \end{cases}. \tag{3.9}$$

Clearly,

$$\sum_{x^n} Q^*(x^n|i) W^n(\mathcal{D}_i|x^n) \ge 1 - 2\lambda_1.$$

The definition of $\mathcal{U}_i^*$ and the code properties imply that $Q(\mathcal{U}_i^*|i) \ge 1/2$. Therefore,

$$\sum_{x^n} Q^*(x^n|i) W^n(\mathcal{D}_j|x^n) \le \sum_{x^n} Q(x^n|i) Q(\mathcal{U}_i^*|i)^{-1} W^n(\mathcal{D}_j|x^n) \le 2 \sum_{x^n} Q(x^n|i) W^n(\mathcal{D}_j|x^n) \le 2\lambda_2 \qquad \text{for } j \ne i.$$

Step 3: Reduction to sequences of one type.

Lemma 6: Suppose that $\{(Q(\cdot|i), \mathcal{D}_i) | i = 1, \cdots, N\}$ is a locally good $(n, N, \lambda_1, \lambda_2)$ ID code with $Q(\cdot|i)$ the uniform distribution on $\mathcal{U}_i \subset \mathcal{X}^n$. Then types $P_i \in \mathcal{P}_n$ exist such that, for the uniform distribution $Q'(\cdot|i)$ on $\mathcal{U}_i \cap \mathcal{T}_{P_i}^n$, the system $\{(Q'(\cdot|i), \mathcal{D}_i)\colon i = 1, \cdots, N\}$ is an $(n, N, (n+1)^{|\mathcal{X}|}\lambda_1, (n+1)^{|\mathcal{X}|}\lambda_2)$ ID code.

Proof: Since $|\mathcal{P}_n| \le (n+1)^{|\mathcal{X}|}$, for any i there is a $P_i \in \mathcal{P}_n$ with

$$|\mathcal{U}_i \cap \mathcal{T}_{P_i}^n| \ge (n+1)^{-|\mathcal{X}|} |\mathcal{U}_i|.$$

Therefore, for any $\mathcal{D} \subset \mathcal{Y}^n$

$$\sum_{x^n} Q'(x^n|i) W^n(\mathcal{D}|x^n) = \sum_{x^n \in \mathcal{U}_i \cap \mathcal{T}_{P_i}^n} Q(x^n|i) \cdot |\mathcal{U}_i| \cdot |\mathcal{T}_{P_i}^n \cap \mathcal{U}_i|^{-1} W^n(\mathcal{D}|x^n) \le (n+1)^{|\mathcal{X}|} \sum_{x^n} Q(x^n|i) \cdot W^n(\mathcal{D}|x^n).$$

This implies the error bounds we claimed in the lemma. Finally, notice that among the sets

$$\mathcal{N}(P, m) = \{i | i = 1, \cdots, N;\ P_i = P;\ |\mathcal{U}_i \cap \mathcal{T}_P^n| = m\}$$

for $P \in \mathcal{P}_n$ there is one set of cardinality at least

$$N \cdot |\mathcal{X}|^{-n} (n+1)^{-|\mathcal{X}|}.$$

This fact and the previous lemmas imply the following Proposition.

Proposition 3: For every $(n, N, \lambda_1, \lambda_2)$ ID code $\{(Q(\cdot|i), \mathcal{D}_i) | i = 1, \cdots, N\}$ there exists a locally good $(n, N, \lambda_1', \lambda_2')$ ID code $\{(Q'(\cdot|i), \mathcal{D}_i) | i = 1, \cdots, N\}$ with the following properties:

a) the $Q'(\cdot|i)$ are uniform distributions on sets $\mathcal{U}_i' \subset \mathcal{T}_P^n$ for some fixed $P \in \mathcal{P}_n$;
b) $|\mathcal{U}_i'| = M$ for $i = 1, \cdots, N'$;
c) $N' \ge N \cdot |\mathcal{X}|^{-n} \cdot (n+1)^{-|\mathcal{X}|}$;
d) $\lambda_i' \le (\epsilon^{-1} \log |\mathcal{X}| + 1) \cdot 2^{n\epsilon} \cdot (1 - 2^{-n\epsilon})^{-1} \cdot 2 \cdot (n+1)^{|\mathcal{X}|} \cdot \lambda_i$ for $i = 1, 2$ and all $\epsilon > 0$.

This proposition shows that it suffices to analyze locally good ID codes with properties a), b), c), and exponentially small error probabilities. We refer to them as *canonical* ID codes $(n, N, M, P, \lambda_1, \lambda_2)$. They are specified by a set of pairs $\{(\mathcal{U}_i, \mathcal{D}_i) | i = 1, \cdots, N\}$ with the required parameters.

B. Basic Observations on the Structure of ID Codes

1) The Key Idea: By Proposition 3 it suffices to consider for the proof of Theorem 1-b) and 2-b) canonical ID codes $\{(\mathcal{U}_i, \mathcal{D}_i) | i = 1, \cdots, N\}$ with parameters $(n, N, M, P, \lambda_1, \lambda_2)$. If $\lambda_1 < 1/2$ and $\lambda_2 < 1/2$, it is clear from the error bound conditions on the ID code that

a) the $\mathcal{U}_i$ are *distinct.*

Suppose that for any $\delta > 0$ and all sufficiently large n our codes satisfy

b) $M \le \exp\{(I(P, W) + \delta)n\} \le \exp\{n(C + \delta)\}$.

Then, of course, N cannot be larger than the set of different subsets of $\mathcal{X}^n$ which have cardinality M. Thus, in case b), $N \le \binom{|\mathcal{X}|^n}{M} \le \exp\{n \log |\mathcal{X}| \exp\{n(C + \delta)\}\}$, and

$$\frac{1}{n} \log \log N \le C + \delta + \frac{\log n + \log \log |\mathcal{X}|}{n}.$$

This completes the proof of Theorem 1-b).

Now recall that in the proof of the direct Theorems 1-a) and 2-a) we used for $\mathcal{U}_i$ the set of codewords of an ordinary code. Here b) holds. It seems therefore natural to

try to "eliminate some unnecessary randomization in the encoding" by selecting suitable subsets $\mathcal{U}_i' \subset \mathcal{U}_i$, all of a cardinality M', such that $\{\mathcal{U}_i' | i = 1, \cdots, N\}$ satisfies a) and b). We presently do not know whether this can be done in such a way that the ID code $\{(\mathcal{U}_i', \mathcal{D}_i) | i = 1, \cdots, N\}$ has good error performances λ_1', λ_2'. However, this is not needed in our proof. We actually have to show the existence of the "representatives" $\mathcal{R}_i$ only for an essential proportion of the message set $\{i | i = 1, \cdots, N\}$. Still the calculations are involved, mainly because we work in a space which is larger than the one usually considered in information theory. Furthermore, the structural consequences of bounds for the error have to be extracted. The main observation in this respect is presented next.

2) A Global Property of List Sizes: Let $\{(Q(\cdot|i), \mathcal{D}_i) | i = 1, \cdots, N\}$ be an $(n, N, \lambda_1, \lambda_2)$ ID code. For $y^n \in \mathcal{Y}^n$ we define the list size

$$L(y^n) = \sum_{i=1}^{N} l_{\mathcal{D}_i}(y^n) \tag{3.10}$$

where $l_{\mathcal{D}}$ denotes the indicator function of $\mathcal{D}$. We look for an estimate of those list sizes.

If a $y^n \in \mathcal{Y}^n$ is contained in too many $\mathcal{D}_i$, this may cause an error event of the second kind in many messages if y^n is received. In other words, the error probability of the second kind cannot be very low if the numbers $L(y^n)$ are large. We quantify this observation as follows.

Let J be an RV that takes values in $\{1, \cdots, N\}$ with probabilities $1/N$. Using J we define an "output variable" Y^n by the distribution:

$$\begin{aligned} \Pr(Y^n = y^n) &= \sum_{i=1}^{N} \Pr(J = i) \sum_{x^n} Q(x^n|i) W^n(y^n|x^n) \\ &= \frac{1}{N} \sum_{i=1}^{N} \sum_{x^n} Q(x^n|i) W^n(y^n|x^n). \end{aligned} \tag{3.11}$$

Proposition 4: For any $(n, N, \lambda_1, \lambda_2)$ ID code $\{(Q(\cdot|i), D_i) | i = 1, \cdots, N\}$, the expected list size, $\mathbb{E}L(Y^n)$ can be bounded as follows:

$$\mathbb{E}L(Y^n) \le (N-1)\lambda_2 + 1. \tag{3.12}$$

Proof:

$$\begin{aligned} \mathbb{E}L(Y^n) &= \sum_{y^n} \frac{1}{N} \sum_{i=1}^{N} \sum_{x^n} Q(x^n|i) W^n(y^n|x^n) \sum_{j=1}^{N} l_{\mathcal{D}_j}(y^n) \\ &= \sum_{y^n} \frac{1}{N} \sum_{i=1}^{N} \sum_{x^n} Q(x^n|i) W^n(y^n|x^n) l_{\mathcal{D}_i}(y^n) \\ &\quad + \sum_{y^n} \frac{1}{N} \sum_{i=1}^{N} \sum_{x^n} Q(x^n|i) W^n(y^n|x^n) \sum_{j \ne i} l_{\mathcal{D}_j}(y^n) \\ &\le (1-\lambda_1) + \frac{1}{N} \sum_{i=1}^{N} \sum_{j \ne i} \sum_{x^n} Q(x^n|i) W^n(\mathcal{D}_j|x^n) \\ &\le (1-\lambda_1) + (N-1)\lambda_2 \le (N-1)\lambda_2 + 1. \end{aligned} \tag{3.13}$$

3) Local Properties of List Sizes: Our next results (Proposition 5) involving list sizes concern a local property, that is, a property for a fixed message i in a canonical $(n, N, M, P, \lambda_1, \lambda_2)$ ID code $\{(\mathcal{U}_i, \mathcal{D}_i) | i = 1, \cdots, N\}$. Here, only the error of the first kind $\lambda_1 \le \exp\{-\epsilon_1 n\}$ is of interest. The results are expressed in terms of hypergraphs $\mathcal{H}_i$ which we associate with the single messages i. We then apply the covering lemma [2, lemma 3, part II], to these hypergraphs to find the representatives mentioned in Section II-B1).

If the given canonical ID code is built from P-sequences $x^n \in \mathcal{T}_P^n$, then define $V \in \mathcal{W}_n(P)$ to be a channel such that

$$W^n(\mathcal{T}_V^n(x^n)|x^n) = \max_{V' \in \mathcal{W}} W^n(\mathcal{T}_{V'}^n(x^n)|x^n) \tag{3.14}$$

for $x^n \in \mathcal{T}_P^n$. Clearly, if W itself is contained in $\mathcal{W}_n(P)$, then $V = W$. Otherwise, V is the best approximation of W in $\mathcal{W}_n(P)$. V satisfies

$$W^n(\mathcal{T}_V^n(x^n)|x^n) \ge (n+1)^{-|\mathcal{X}|\cdot|\mathcal{Y}|} \text{ for } x^n \in \mathcal{T}_P^n \tag{3.15}$$

and, if any $\delta > 0$ is given, it also satisfies

$$|I(P, W) - I(P, V)| \le \delta \tag{3.16}$$

for every $n \ge n_0(\delta)$.

For this $V \in \mathcal{W}$ we define the hypergraph

$$\mathcal{H}_i := \left(\mathcal{D}_i, (\mathcal{T}_V^n(u) \cap \mathcal{D}_i)_{u \in \mathcal{U}_i}\right).$$

For a $\tau \in (0,1)$, which we later choose sufficiently small, we partition the set $\mathcal{D}_i$ into the sets

$$\left\{\mathcal{D}_i(l) | 0 \le l \le \left\lceil |\mathcal{X}| \tau^{-1} \right\rceil\right\}$$

according to the vertex degrees or local list sizes

$$d_i(y^n) := \left|\{u \in \mathcal{U}_i | y^n \in \mathcal{T}_V^n(u) \cap \mathcal{D}_i\}\right| \tag{3.17}$$

for $y^n \in \mathcal{D}_i$ as follows:

$$\mathcal{D}_i(l) := \left\{y^n \in \mathcal{D}_i | 2^{(l-1)n\tau} \le d_i(y^n) < 2^{ln\tau}\right\} \tag{3.18}$$

for $l = 1, 2, \cdots, \lceil |\mathcal{X}| \tau^{-1} \rceil$.

Since $d_i(y^n) \le |\mathcal{U}_i| \le |\mathcal{X}|^n$, we have

$$d_i(y^n) \le \exp\{n \log |\mathcal{X}|\} \le \exp\left\{\left\lceil |\mathcal{X}| \cdot \tau^{-1} \right\rceil n\tau\right\}.$$

This, together with the above definitions, implies that

$$\mathcal{D}_i(0) := \{y^n \in \mathcal{D}_i | d_i(y^n) = 0\} = \mathcal{D}_i - \bigcup_{l \ge 1} \mathcal{D}_i(l).$$

For any $i \in \{1, \cdots, N\}$, choose $l_i \in \{0, \cdots, \lceil |\mathcal{X}|/\tau \rceil\}$ such that

$$\frac{1}{|\mathcal{U}_i|} \sum_{u \in \mathcal{U}_i} W^n(\mathcal{D}_i(l_i)|u) = \max_{l} \frac{1}{|\mathcal{U}_i|} \sum_{u \in \mathcal{U}_i} W^n(\mathcal{D}_i(l)|u). \tag{3.19}$$

Then, by this definition and (3.15),

$$\begin{aligned} &\frac{1}{|\mathcal{U}_i|} \sum_{u \in \mathcal{U}_i} W^n(\mathcal{D}_i(l_i) \cap \mathcal{T}_V^n(u)|u) \\ &\qquad \ge (n+1)^{-|\mathcal{X}|\cdot|\mathcal{Y}|} \cdot \left\lceil |\mathcal{X}| \tau^{-1} \right\rceil^{-1} - \lambda_1. \end{aligned} \tag{3.20}$$

For convenience, we make a little change from the given code $\{(\mathcal{U}_i, \mathcal{D}_i) | i = 1, \cdots, N\}$ to a new system $\{(\mathcal{U}_i, \mathcal{D}_i^*) | i = 1, \cdots, N^*\}$ which is simpler and contains all information necessary for the proof of the theorem.

Since $l_i \in \{1,\cdots,\lceil|\mathcal{X}|/\tau\rceil\}$ and $|\mathcal{D}_i(l_i)| \in \{1,\cdots,|\mathcal{Y}|^n\}$, we conclude that there are numbers l^*, T^* such that

$$N^* := \left|\{i | l_i = l^*, |\mathcal{D}_i(l_i)| = T^*\}\right| \tag{3.21}$$

is lower-bounded by $N|\mathcal{Y}|^{-n}\lceil|\mathcal{X}|\tau^{-1}\rceil^{-1}$. For those i for which $l_i = l^*$ and $|\mathcal{D}_i(l_i)| = T^*$, we set

$$\mathcal{D}_i^* := \mathcal{D}_i(l_i).$$

After relabeling, we can assume that, for the indices $i = 1,\cdots,N^*$, the conditions $l_i = l^*$ and $|\mathcal{D}_i(l_i)| = T^*$ are satisfied.

In what follows we study $\{(\mathcal{U}_i, \mathcal{D}_i^*) | i = 1,\cdots,N^*\}$ instead of the ID code given originally. Observe that by (3.21) the limitation to only N^* messages does not affect the second order rate. The $\mathcal{D}_i^*$ are rather small in comparison with the $\mathcal{D}_i$. However, we shall see that they are substantial enough parts of the $\mathcal{D}_i$ for the purposes of our proof.

We now consider the new hypergraphs

$$\mathcal{H}_i^* := \left(\mathcal{D}_i^*, \left(\mathcal{T}_V^n(u) \cap \mathcal{D}_i^*\right)_{u \in \mathcal{U}_i}\right), \qquad i = 1,\cdots,N^*$$

with vertex degrees (or local list sizes) $d_i^*(y^n)$ defined as in (3.17). We have our next proposition.

Proposition 5: For $\epsilon_1, \delta, \tau \in (0,1)$ the constructed hypergraphs $\mathcal{H}_i^*$ above ($i = 1,\cdots,N^*$) have the following properties

a) $|\mathcal{D}_i^*| = T^*$;
b) $2^{(l^*-1)n\tau} \le d_i^*(y^n) < 2^{l^*n\tau}$ for all $y^n \in \mathcal{D}_i^*$;
c) $N \ge N^* \ge N \cdot |\mathcal{Y}|^{-n} \cdot \lceil|\mathcal{X}|\tau^{-1}\rceil^{-1}$;
d) $1/N \sum_{u \in \mathcal{U}_i} W^n(\mathcal{D}_i^* \cap \mathcal{T}_V^n(u)|u) \ge (n+1)^{-|\mathcal{X}|\cdot|\mathcal{Y}|} \cdot \lceil|\mathcal{X}|\tau^{-1}\rceil^{-1} - \lambda_1$; and
e) $|I(P,W) - I(P,V)| \le \delta$ if n is sufficiently large.

4) Choosing Representatives via Balanced Hypergraph Covering: We consider the system $\{(\mathcal{U}_i, \mathcal{D}_i^*) | i = 1,\cdots,N^*\}$ and the hypergraphs $\mathcal{H}_i^*$, $i = 1,\cdots,N^*$. With the help of a covering lemma from [2] we choose representation subsets $\mathcal{R}_i \subset \mathcal{U}_i \subset \mathcal{T}_P^n$.

In Sections III-C to III-E we shall show that the sets $\mathcal{R}_i$ we define here have a cardinality which is essentially bounded by $\lambda_2 \exp\{nI(P,V)\}$ (Section III-C). Furthermore, we show in Section III-E that the sets $\mathcal{R}_i$ are not just distinct but, rather differ greatly. In particular, we demonstrate that the $\mathcal{R}_i$ cannot have large subsets in common. This strong distinctness property immediately leads to a proof of Theorem 2-b). For a proof of Theorem 1-b) it would be sufficient to prove just the normal distinctness of the sets $\mathcal{R}_i$.

We start the formal arguments for the definition of the $\mathcal{R}_i$. Following [2] we call $\mathcal{C} = \{E_1,\cdots,E_k\} \subset \mathcal{E}$ a c-balanced covering of a hypergraph $(\mathcal{V}, \mathcal{E})$ if

$$\bigcup_{j=1}^{k} E_j = \mathcal{V}$$

and

$$\left|\{E \in \mathcal{C} | v \in E\}\right| \le c \qquad \text{for all } v \in \mathcal{V}. \tag{3.22}$$

Covering Lemma 3 [2, p. II, p. 250]: A hypergraph with

$$\alpha := \max_{v \in \mathcal{V}} \deg(v), \quad \min_{v \in \mathcal{V}} \deg(v) =: \beta > 0$$

has a c-balanced covering with exactly k edges, if

a) $k \ge |\mathcal{E}|\beta^{-1}(\log|\mathcal{V}| + 1)$
b) $c \le k \le c|\mathcal{E}|\alpha^{-1}$, and
c) $\exp\{(h(ck^{-1}) + \log(\alpha|\mathcal{E}|^{-1}))k + \log|\mathcal{V}|\} < 1/2$.

If a) holds and $c > k$, the result is again true (already by Covering Lemma 1 in [2, pt. I]). Note that $h(\cdot)$ denotes the binary entropy.

We now apply this Lemma to the hypergraphs $\mathcal{H}_i^*$. Recall that $|\mathcal{U}_i| = M \ge 2$ for every $i = 1,\cdots,N^*$.

Corollary 1: Every hypergraph $\mathcal{H}_i^*$, $i = 1,\cdots,N^*$, has a c-balanced covering

$$\mathcal{C}_i := \left\{\mathcal{T}_V^n(u) \cap \mathcal{D}_i^* | u \in \mathcal{R}_i\right\}$$

with the properties:

a) $|\mathcal{R}_i| = \lfloor M \cdot 2^{-(l^*-2)n\tau}\rfloor$, and
b) $c \le 2^{4n\tau}$, if n is sufficiently large.

Proof: The parameters of $\mathcal{H}_i^*$ are

$$|\mathcal{E}| = M \qquad |\mathcal{V}| = |\mathcal{D}_i^*| = T^* \le |\mathcal{Y}|^n$$

and

$$2^{(l^*-1)n\tau} \le \beta \le \alpha \le 2^{l^*n\tau}.$$

We choose $k = \lfloor M 2^{-(l^*-2)n\tau}\rfloor$ and $c = \lfloor 2^{4n\tau}\rfloor$.

1) $|\mathcal{E}| \cdot \beta^{-1}(\log|\mathcal{V}| + 1) \le M \cdot 2^{-(l^*-1)\tau n} \cdot 2n \log|\mathcal{Y}| \le k$, for $n \ge n(\tau)$.

2) $c|\mathcal{E}| \cdot \alpha^{-1} \ge \lfloor 2^{4n\tau}\rfloor \cdot M \cdot 2^{-nl^*\tau} \ge k$ for $n \ge n(\tau)$.

3) In case $c > k$ we are done. Otherwise, $k \ge c$ and $c \cdot k^{-1}$. Hence $h(ck^{-1})$ is defined and $0 \le h(ck^{-1}) \le 1$.

Furthermore, $k \ge c$ implies

$$\lfloor 2^{4n\tau}\rfloor \le M \cdot 2^{-(l^*-2)n\tau}$$

and therefore

$$M \ge 2(l^*+1)n\tau \qquad \text{for large } n.$$

This gives us

$$\log(\alpha|\mathcal{E}|^{-1}) \le \log(2^{l^*n\tau}M^{-1}) \le \log 2^{-n\tau} = -n\tau.$$

Therefore, the left side of c) in Covering Lemma 3 is here upper-bounded by (again use $k \ge c$): $\exp\{1 - n\tau)\lfloor 2^{4n\tau}\rfloor + n\log|\mathcal{Y}|\}$ for large n, which converges to zero if n tends to infinity. This completes the proof of Corollary 1.

We next establish an important upper bound on $|\mathcal{R}_i|$ in terms of $T^* = |\mathcal{D}_i^*|$. We start with a simple observation. By (1.18) we have for any $y^n \in \mathcal{T}_V^n(x^n)$, $x^n \in \mathcal{T}_P^n$:

$$W^n(y^n|x^n) \le \exp\{-nH(V|P)\}.$$

With the abbreviation

$$\gamma_n = (n+1)^{-|\mathcal{X}|\cdot|\mathcal{Y}|}\left\lceil|\mathcal{X}|\cdot\tau^{-1}\right\rceil^{-1} - \lambda_1$$

we can rewrite (3.20) as

$$M \le \gamma_n^{-1} \cdot \sum_{u \in \mathcal{U}_i} W^n(\mathcal{D}_i^* \cap \mathcal{T}_V^n(u)|u)$$

and then by (1.18),

$$M \le \gamma_n^{-1} \cdot \exp\{-nH(V|P)\} \cdot \sum_{u \in \mathcal{U}_i} |\mathcal{D}_i^* \cap \mathcal{T}_i^n(u)|$$
$$\le \gamma_n^{-1} \cdot \exp\{-nH(V|P)\} \cdot 2^{n\tau l^*} |\mathcal{D}_i^*|$$
$$\le 2^{n\tau} \cdot \exp\{-nH(V|P)\} \cdot 2^{n\tau l^*} |\mathcal{D}_i^*|, \qquad \text{for } n \ge n(\tau). \tag{3.23}$$

Thus we can use (3.23) and the property a) of Corollary 1 to obtain the following.

Corollary 2:

$$|\mathcal{R}_i| \le 2^{3n\tau} \cdot T^* \exp\{-nH(V|P)\}, \qquad \text{for } n \ge n(\tau).$$

C. An Upper Bound on T^*

In the preceding section we defined representatives $\mathcal{R}_i$ for the messages i. We recall briefly that we first want to upper-bound the cardinality of the $\mathcal{R}_i$. Then we want to show that the $\mathcal{R}_i$ differ greatly. We shall conclude that there are not too many $\mathcal{R}_i$.

The preceding section ended with an upper bound on $|\mathcal{R}_i|$. For a final result, however, we still need an upper bound on T^*. This bound is provided here.

Proposition 6:

a) $T^* \le \exp\{nH(PV)\} \cdot (\lambda_2 + 1/N^*) 2^{5\tau n}$

b) $|\mathcal{R}_i| \le \exp\{nI(P,V)\} \cdot (\lambda_2 + 1/N^*) 2^{8\tau n}$

if $n \ge n(\tau)$ is sufficiently large.

Proof: In Proposition 4 we proved that, for λ_2 small, the $\mathcal{D}_i^*$ cannot overlap too much. This result is used here to prove that the $\mathcal{D}_i^*$ cannot be too large.

First, we need a simple observation. As before, we consider the system

$$\{(\mathcal{U}_i, \mathcal{D}_i^*) | i = 1, \cdots, N^*\}$$

instead of the original system $\{(\mathcal{U}_i, \mathcal{D}_i) | i = 1, \cdots, N\}$.

We want to apply Proposition 4 to the new system. We remark here that this presents no problems because the sets $\mathcal{D}_i^*$ are subsets of the original sets. The reason is that, if one reduces the decoding sets of an ID code in size then the error probability of the second kind cannot increase.

We consider the list sizes

$$L^*(y^n) = |\{i | i \in \{1, \cdots, N^*\},\ y^n \in \mathcal{D}_i^*\}|. \tag{3.24}$$

Let J^* be uniformly distributed over $\{1, \cdots, N^*\}$ and Y^{*n} be defined in analogy to (3.11), for the new system $\{(\mathcal{U}_i, \mathcal{D}_i^*) | i = 1, \cdots, N^*\}$.

Proposition 4 yields:

$$N^* \lambda_2 + 1 \ge \sum_{y^n} \Pr(Y^{*n} = y^n) L^*(y^n)$$
$$\ge \sum_{y^n \in \mathcal{T}_{PV}^n} \Pr(Y^{*n} = y^n) L^*(y^n). \tag{3.25}$$

Since the $\mathcal{D}_i^*$'s are contained in $\mathcal{T}_{PV}^n$,

$$\sum_{y^n \in \mathcal{T}_{PV}^n} L^*(y^n) \ge \sum_{i=1}^{N^*} |\mathcal{D}_i^*| = N^* \cdot T^*. \tag{3.26}$$

For $y^n \in \mathcal{T}_{PV}^n$ we have by (3.15) and (1.17)

$$\Pr(Y^n = y^n) = \frac{1}{N^*} \sum_{i=1}^{N^*} \sum_{u \in \mathcal{U}_i} \frac{1}{M} W^n(y^n|u)$$
$$\ge \frac{1}{N^*} \sum_{i:\, y^n \in \mathcal{D}_i^*} \sum_{\substack{u:\, y^n \in \mathcal{T}_i^n(u) \\ u \in \mathcal{U}_i}} \frac{1}{M} W^n(y^n|u)$$
$$\ge \frac{1}{N^*} \cdot \frac{1}{M} \cdot \exp\{-nH(V|P)\}$$
$$\cdot 2^{(l^*-1)n\tau} L^*(y^n) (n+1)^{-|\mathcal{X}| \cdot |\mathcal{Y}|}.$$

Substituting this into (3.25) we get

$$N^* \lambda_2 + 1 \ge \frac{1}{N^* M} \cdot \exp\{-nH(V|P)\} \cdot (n+1)^{-|\mathcal{X}| \cdot |\mathcal{Y}|}$$
$$\cdot 2^{(l^*-1)n\tau} \sum_{y^n \in \mathcal{T}_{PV}^n} (L^*(y^n))^2. \tag{3.27}$$

The last sum can be written in the form

$$|\mathcal{T}_{PV}^n| \cdot \sum_{y^n \in \mathcal{T}_{PV}^n} \frac{L^*(y^n)^2}{|\mathcal{T}_{PV}^n|}$$

which by convexity of the square function exceeds

$$|\mathcal{T}_{PV}^n| \cdot \left(\frac{1}{|\mathcal{T}_{PV}^n|} \sum L^*(y^n) \right)^2.$$

By (3.26), this is lower-bounded by

$$|\mathcal{T}_{PV}^n|^{-1} \cdot N^{*2} \cdot T^{*2}.$$

Therefore, (3.27) yields

$$N^* \lambda_2 + 1 \ge \frac{1}{M} \cdot N^* \cdot T^{*2} \cdot |\mathcal{T}_{PV}^n|^{-1} \cdot \exp\{-nH(V|P)\}$$
$$\cdot 2^{(l^*-1)n\tau} (n+1)^{-|\mathcal{X}| \cdot |\mathcal{Y}|}. \tag{3.28}$$

We now estimate one term only on the right side namely T^* with the aid of the inequality in Corollary 2:

$$N^* \lambda_2 + 1 \ge \frac{1}{M} \cdot N^* \cdot T^* \cdot |\mathcal{T}_{PV}^n|^{-1} \cdot 2^{-3n\tau} \cdot 2^{(l^*-1)n\tau}$$
$$\cdot (n+1)^{-|\mathcal{X}| \cdot |\mathcal{Y}|} \cdot |\mathcal{R}_i|. \tag{3.29}$$

By (1.16) we have $|\mathcal{T}_{PV}^n| \le \exp\{nH(PV)\}$. Thus, by Corollary 1-a),

$$N^* \lambda_2 + 1 \ge N^* \cdot T^* \exp\{-nH(PV)\} \cdot 2^{-4n\tau}$$
$$\cdot (n+1)^{-|\mathcal{X}| \cdot |\mathcal{Y}|}. \tag{3.30}$$

We see that for large n we get part a) of Proposition 6. Part b) follows from part a) and Corollary 2, if n is sufficiently large.

Remark: The bounds on T^* (resp. $|\mathcal{R}_i|$) we derived here are not needed for a proof of Theorem 1-b). For this theorem, the trivial bound

$$T^* \le |\mathcal{T}_{PV}^n| \le \exp\{nH(PV)\}$$

would suffice for a proof. Thus in this case Proposition 4 is not used at all.

D. A Probabilistic Property of the Hypergraph $\mathcal{H}_i^*$

For a complete proof of our Converse Theorem 2 it remains to show a strong distinctness of the representatives $\mathcal{R}_i$. The technical Lemma 7 of this section provides a tool for this final step. We show here that our c-balanced hypergraph $\mathcal{H}_i^* = (\mathcal{D}_i^*, (\mathcal{T}_V^n(u) \cap \mathcal{D}_i^*)_{u \in \mathcal{R}_i})$ is in a certain sense also probabilistically balanced. Define

$$\mathcal{A}_i = \left\{ y^n \in \mathcal{D}_i^* \,\middle|\, \sum_{u \in \mathcal{R}_i} W^n(y^n|u) \le 2^{11\tau n} \exp\{-nH(V|P)\} \right\}.$$

Lemma 7:

$$\frac{1}{|\mathcal{R}_i|} \cdot \sum_{u \in \mathcal{R}_i} W^n(\mathcal{A}_i \cap \mathcal{T}_V^n(u)|u) \ge 2^{-4\tau n}$$

if n is sufficiently large.

Proof: Let

$$a_i := \sum_{u \in \mathcal{R}_i} W^n(\mathcal{A}_i \cap \mathcal{T}_V^n(u)|u)$$

$$b_i := \sum_{u \in \mathcal{R}_i} W^n((\mathcal{D}_i^* - \mathcal{A}_i) \cap \mathcal{T}_V^n(u)|u).$$

First, note that by (1.18)

$$a_i \le \sum_{y^n \in \mathcal{A}_i} \exp\{-nH(V|P)\} \cdot |\{u \in \mathcal{R}_i | y^n \in \mathcal{T}_V^n(u)\}| \tag{3.31}$$

and

$$b_i \le \sum_{y^n \in \mathcal{D}_i^* - \mathcal{A}_i} \exp\{-nH(V|P)\} |\{u \in \mathcal{R}_i | y^n \in \mathcal{T}_V^n(u)\}|. \tag{3.32}$$

On the other hand, by the definition of the representatives $\mathcal{R}_i$, the sets $\mathcal{T}_V^n(u)$, $u \in \mathcal{R}_i$, cover $\mathcal{D}_i^*$. Therefore, (3.15) and (1.17) give

$$a_i + b_i \ge (n+1)^{-|\mathcal{X}| \cdot |\mathcal{Y}|} \cdot \exp\{-nH(V|P)\} \cdot |\mathcal{D}_i^*|.$$

Continuing with Corollary 2 we get

$$a_i + b_i \ge (n+1)^{-|\mathcal{X}| \cdot |\mathcal{Y}|} \cdot 2^{-3n\tau} |\mathcal{R}_i| \tag{3.33}$$

for sufficiently large n. This is used to obtain

$$\sum_{u \in \mathcal{R}_i} W^n(\mathcal{D}_i^*|u) \le |\mathcal{R}_i| \le (n+1)^{|\mathcal{X}| \cdot |\mathcal{Y}|} \cdot 2^{3n\tau} \cdot (a_i + b_i). \tag{3.34}$$

We now derive a lower bound for the left side of (3.34). If $y^n \in \mathcal{D}_i^*$, then apparently

$$\sum_{u \in \mathcal{R}_i} W^n(y^n|u) \ge \exp\{-nH(V|P)\} \cdot \{u \in \mathcal{R}_i | y^n \in \mathcal{T}_V^n(u)\} | \cdot (n+1)^{-|\mathcal{X}| \cdot |\mathcal{Y}|}. \tag{3.35}$$

By Corollary 1, $|\{u \in \mathcal{R}_i | y^n \in \mathcal{T}_V^n(u)\}| \le 2^{4n\tau}$ for $y^n \in \mathcal{D}_i^*$. If $y^n \in \mathcal{D}_i^* - \mathcal{A}_i$ this, together with the definition of $\mathcal{A}_i$, gives:

$$\sum_{u \in \mathcal{R}_i} W^n(y^n|u) \ge 2^{7\tau n} \cdot (n+1)^{-|\mathcal{X}| \cdot |\mathcal{Y}|} \cdot \exp\{-nH(V|P)\} \cdot |\{u \in \mathcal{R}_i | y^n \in \mathcal{T}_V^n(u)\}|. \tag{3.36}$$

Hence, by (3.31), (3.32), (3.35), and (3.36)

$$\sum_{u \in \mathcal{R}_i} W^n(\mathcal{D}_i^*|u) \ge (n+1)^{-|\mathcal{X}| \cdot |\mathcal{Y}|} \cdot a_i + (n+1)^{-|\mathcal{X}| \cdot |\mathcal{Y}|} \cdot 2^{7n\tau} b_i. \tag{3.37}$$

From (3.34) and (3.37) we finally get

$$a_i + 2^{7n\tau} b_i \le (n+1)^{2|\mathcal{X}| \cdot |\mathcal{Y}|} \cdot 2^{3n\tau} (a_i + b_i).$$

Since $a_i \le |\mathcal{R}_i|$ and $a_i \ge 0$,

$$\left(2^{7n\tau} - (n+1)^{2|\mathcal{X}| \cdot |\mathcal{Y}|} \cdot 2^{3n\tau}\right) b_i \le (n+1)^{2|\mathcal{X}| \cdot |\mathcal{Y}|} \cdot 2^{3n\tau} |\mathcal{R}_i|. \tag{3.38}$$

This gives an upper bound on b_i. We conclude with (3.33)

$$a_i \ge (n+1)^{-|\mathcal{X}| \cdot |\mathcal{Y}|} \cdot 2^{-3n\tau} |\mathcal{R}_i| - b_i. \tag{3.39}$$

One need only substitute (3.38) into (3.39) to get the claim of the Lemma for sufficiently large n.

Corollary 3: At least $\frac{1}{2} 2^{-4n\tau} |\mathcal{R}_i|$ members $u \in \mathcal{R}_i$ satisfy

$$W^n(\mathcal{T}_V^n(u) \cap \mathcal{A}_i|u) \ge 2^{-5n\tau}$$

if n is sufficiently large.

Proof: Let $\mathcal{B}_i = \{u \in \mathcal{R}_i | W^n(\mathcal{T}_V^n(u) \cap \mathcal{A}_i|u) \ge 2^{-5n\tau}\}$. Lemma 7 yields

$$|\mathcal{R}_i| \cdot 2^{-4\tau n} \le \sum_{u \in \mathcal{R}_i} W^n(\mathcal{T}_V^n(u) \cap \mathcal{A}_i|u) \le |\mathcal{B}_i| + 2^{-5n\tau} |\mathcal{R}_i|.$$

Hence

$$|\mathcal{B}_i| \ge |\mathcal{R}_i| \cdot (2^{-4n\tau} - 2^{-5n\tau}) \ge \frac{1}{2} |\mathcal{R}_i| \cdot 2^{-4n\tau}, \quad \text{for } n \text{ large.}$$

E. A Strong Distinctness Property of the Representatives $\mathcal{R}_i$

For every $i \in \{1, \cdots, N\}$ choose $\frac{1}{2} 2^{-4n\tau} |\mathcal{R}_i| \cdot (2^{-n\epsilon_2} \cdot 2^{23n\tau})$ elements $u \in \mathcal{R}_i$ with

$$W^n(\mathcal{T}_V^n(u) \cap \mathcal{A}_i|u) \ge 2^{-5n\tau}. \tag{3.40}$$

This is possible by Corollary 3, if τ is small. Let $\mathcal{R}_i^*$ be the subset of these chosen elements.

Proposition 7: The sets $\mathcal{R}_i^*$ are distinct, if τ is small in comparison with ϵ_1, ϵ_2 and if n is sufficiently large.

Remark: Note that we prove here a much stronger property than just the distinctness of the $\mathcal{R}_i$ themselves. In Section III-D we proved that many $u \in \mathcal{R}_i$ satisfy (3.40). In light of these results, we see that the definition of the sets $\mathcal{R}_i^*$ means essentially: "Take *any* subset of $2^{-n\epsilon_2} |\mathcal{R}_i|$ elements $u \in \mathcal{R}_i$ and call it $\mathcal{R}_i^*$." The fact that we restrict our attention only to those $u \in \mathcal{R}_i$ satisfying (3.40) is only

a technical necessity. Therefore, Proposition 6 says essentially: The sets $\mathscr{R}_i$ are very much different in the sense that they cannot contain common subsets of size $2^{-n\epsilon_2}|\mathscr{R}_i|$. We now give the proof of Proposition 7 along with some easy arguments to complete the proof of Theorem 2-b).

Proof: Suppose $i \neq j$ and $\mathscr{R}_i^* = \mathscr{R}_j^*$.

We lower-bound the error probability of the second kind

$$\frac{1}{M}\sum_{u\in\mathscr{U}_i} W^n(\mathscr{D}_j|u),$$

which we assumed is less than or equal to λ_2. We will be led to a contradiction.

Suppose that $y^n \in \mathscr{A}_i$. Then

$$\begin{aligned}
&\frac{1}{M}\sum_{u\in\mathscr{U}_i} W^n(y^n|u)\\
&\quad\geq \frac{1}{M}\sum_{\substack{u\in\mathscr{U}_i\\ y^n\in\mathscr{T}_V^n(u)}} W^n(y^n|u)\\
&\quad\geq \frac{1}{M}(n+1)^{-|\mathscr{X}|\cdot|\mathscr{Y}|}\cdot\exp\{-nH(V|P)\}\cdot 2^{(l^*-1)n\tau}\\
&\quad\geq \frac{1}{|\mathscr{R}_i|}(n+1)^{-|\mathscr{X}|\cdot|\mathscr{Y}|}\cdot 2^{-n\tau}\exp\{-nH(V|P)\}
\end{aligned}$$

because of Corollary 1. Since $y^n \in \mathscr{A}_i$,

$$\begin{aligned}
&\frac{1}{M}\sum_{u\in\mathscr{U}_i} W^n(y^n|u)\\
&\quad\geq \frac{1}{|\mathscr{R}_i|}(n+1)^{-|\mathscr{X}|\cdot|\mathscr{Y}|}\cdot 2^{-12n\tau}\cdot\sum_{u\in\mathscr{R}_i} W^n(y^n|u)\\
&\quad\geq \frac{1}{|\mathscr{R}_i|}(n+1)^{-|\mathscr{X}|\cdot|\mathscr{Y}|}\cdot 2^{-12n\tau}\cdot\sum_{u\in\mathscr{R}_i^*} W^n(y^n|u). \qquad (3.41)
\end{aligned}$$

Recall now that $\mathscr{R}_i^* = \mathscr{R}_j^*$ and that the original code was assumed to be locally good, i.e., that

$$W^n(\mathscr{D}_j^c|u) \leq \lambda_1 \qquad \text{for all } u\in\mathscr{U}_j.$$

Since $\mathscr{R}_i^* = \mathscr{R}_j^*$ we also have $\mathscr{R}_i^* \subset \mathscr{U}_j$. Thus we can continue with (3.41).

$$\begin{aligned}
&\frac{1}{M}\sum_{u\in\mathscr{U}_i} W^n(\mathscr{D}_j|u)\\
&\quad\geq \frac{1}{M}\sum_{u\in\mathscr{U}_i} W^n(\mathscr{D}_j\cap\mathscr{A}_i|u)\\
&\quad\geq \frac{1}{|\mathscr{R}_i|}(n+1)^{-|\mathscr{X}|\cdot|\mathscr{Y}|}\cdot 2^{-12n\tau}\sum_{u\in\mathscr{R}_i^*} W^n(\mathscr{D}_j\cap\mathscr{A}_i|u)\\
&\quad\geq \frac{1}{|\mathscr{R}_i|}(n+1)^{-|\mathscr{X}|\cdot|\mathscr{Y}|}\\
&\qquad\cdot 2^{-12n\tau}\left(\sum_{u\in\mathscr{R}_i^*} W^n(\mathscr{A}_i|u) - |\mathscr{R}_i^*|\cdot\lambda_1\right)\\
&\quad\geq \frac{|\mathscr{R}_i^*|}{|\mathscr{R}_i|}(n+1)^{-|\mathscr{X}|\cdot|\mathscr{Y}|}\cdot 2^{-12n\tau}\left(2^{-5n\tau}-\lambda_1\right) \qquad (3.42)
\end{aligned}$$

where we used (3.40) in the last step. The definition of $|\mathscr{R}_i^*|$ gives

$$\frac{1}{M}\sum_{u\in\mathscr{U}_i} W^n(\mathscr{D}_j|u) > 2^{-n\epsilon_2} \geq \lambda_2 \qquad (3.43)$$

if τ is small in comparison with ϵ_1, ϵ_2 and if n is sufficiently large. Proposition 7 is proved.

Final Arguments for the Proof of Theorem 2-b: From Proposition 7 we see that the map

$$i \to \mathscr{R}_i^*, \qquad i\in\{1,\cdots,N^*\}$$

is injective. Thus N^* is bounded by the number of subsets of $\mathscr{X}^n$ with cardinality $|\mathscr{R}_i^*|$. We conclude:

$$N^* \leq \binom{|\mathscr{X}|^n}{|\mathscr{R}_i^*|} \leq |\mathscr{X}|^{n\cdot|\mathscr{R}_i^*|} \leq |\mathscr{X}|^{n\cdot(1/2)\cdot 2^{-19n\tau}\cdot 2^{-n\epsilon_2}\cdot|\mathscr{R}_i|}. \qquad (3.44)$$

From (3.21) we see that $\log\log N^*$ and the original parameter $\log\log N$ do not differ too much. In Proposition 6 we gave an estimate for $|\mathscr{R}_i|$.

We bounded $(1/n)\log|\mathscr{R}_i|$ essentially by $I(P,V)-\epsilon_2$. Hence (3.44) bounds $1/n\log\log N^*$ essentially by $I(P,V)-2\epsilon_2$, which was to be proved.

IV. Discussion

In all coding problems previously studied in information theory, the maximal codelengths grow only exponentially in blocklength. Therefore, our double exponent coding theorem is the first of its kind. The identification problem solved seems to be a natural one. In our judgment it enlarges the basis of information theory, which in Shannon's foundation was restricted to the transmission problem. The success of Shannon's theory relies on the fact that the semantic aspect of information was excluded, but the identification problem also has its place in a presemantic theory. It therefore is satisfying to see that this meaningful question finds an answer in a smooth mathematical theory. Moreover, the result is quite sophisticated from the mathematical point of view. Of course, we expect that simpler proofs will be found; we do not expect, however, that a very simple proof of the converse will be given soon.

A few historical remarks seem in order. In 1970 one of the authors presented a manuscript entitled "A New Information Theory: Information Transfer at Rates Above Shannon's Capacity" to the late Jack Wolfowitz. Within 24 hours Wolfowitz responded with a letter entitled, "New Information Theory for Those who Don't Know the Old." He was absolutely right, because the calculation of the error probability for a random encoding procedure used only two-codeword error probabilities and had completely ignored the union bound. Nonetheless, somehow information was conveyed, and in another letter two days later, Wolfowitz wrote "The result is perhaps completely useless, but I like it!"

At the Information Theory Workshop at Gränna, Sweden, during a discussion on Yao's two-way communication

complexity (see [4]). Ephremides drew attention to a recent unpublished work of Ja'Ja' (see [5]). Immediately, the bell rang. The ancient result had a proper interpretation in the context of identification. The observation of Ja'Ja' is that, for the binary symmetric channel with crossover probability $\epsilon \neq 1/2$, one can identify at a rate arbitrarily close to 1. This is immediately clear, if one uses Gilbert's bound for the Hamming distance $d = \delta \cdot n(\delta \to 0)$ and Hamming spheres of radius equal to $(\epsilon + \eta)n(\epsilon < 1/n, \eta \ll \epsilon)$ as decoding sets.

One can apply the same idea to the general DMC to get a (non-randomized) identification capacity equal to $\log_2$ of the number of distinct row vectors in W. The unsatisfactory aspect of this result is that the actual values of the positive entries in W do not matter.

Our idea to use randomization in the encoding therefore is fruitful in two respects: it leads to much better performance and also eliminates the shortcoming mentioned. Since $\sum_{x^n} Q(x^n|i)W(\mathcal{D}_i|x^n) \geq 1-\lambda$ implies the existence of an u_i with $W(\mathcal{D}_i|u_i) \geq 1-\lambda$, the effect of randomization is on the error of the second kind. For the transmission problem on the DMC, it does not help at all!

It must also be emphasized that even for noiseless channels our result is of interest. Suppose that one out of N possible events occurred. Shannon was concerned with the question, "Which event occurred?" The question asked in identification is "Did event i occur?" Here i could be any member of $\{1,2,\cdots,N\}$. There are many situations in which the answer to this question is of interest.

Example 1: Let $S_1,\cdots,S_N$ be sailors on a ship, and let sailor S_i be associated with lady L_i. In a stormy night one sailor, say S_j, drowns in the ocean. One could now broadcast his name to the radio stations of the country from which all sailors are known to come, hoping that the lady L_j listens to the news, so that she hears about the tragic event. However, this takes $\lceil \log_2 N \rceil$ bits and the news is (primarily) of interest only to one lady. If we now permit a certain error probability, which is not much of a prize in an imperfect (as the tragedy shows) world, then by our result $O(\log\log N)$ bits suffice!

Example 2: In many countries the winning m-digit state lottery number is made public on radio and television. Again, by tolerating a certain error probability, this number could be replaced by a properly produced random number of $O(\log m)$ digits and still every winner and every loser would be informed correctly with probability close to 1. Also one could modify the lottery so that the chance errors become part of the lottery.

These examples show that there is a need for explicit *constructions* of ID codes. If such codes achieve positive second-order rates, then they are already much better than the naive error-free identification codes.

There is a multitude of other problems which can now be studied. Almost every known coding theorem concerning the transmission problem can be reconsidered in the context of identification. Also, new phenomena arise. We are preparing two papers entitled "Identification in the Presence of Feedback" and "Identification for Multi-Way Channels" to expand the discussion.

APPENDIX

Proof of Lemma LD

If $T \geq 0$ is a random variable, then by Chebyshev's inequality,

$$\Pr(T \geq \delta) \leq \frac{\mathbf{E}T}{\delta}$$

for $\delta > 0$. If $\alpha > 0$ then

$$\Pr(T \geq \delta) = \Pr(2^{\alpha T} \geq 2^{\alpha\delta}).$$

We apply the above inequality to the right side.

$$\Pr(T \geq \delta) \leq 2^{-\alpha\delta} \cdot \mathbf{E}2^{\alpha T}.$$

This gives for the RV's $\Psi_1,\cdots,\Psi_M$

$$\Pr\left(\sum_{j=1}^{M} \Psi_j > M\cdot\lambda\right) \leq 2^{-\alpha\lambda M}\cdot\mathbf{E}\prod_{j=i}^{M} 2^{\alpha\Psi_j} = 2^{-\alpha\lambda M}\cdot\left(\mathbf{E}2^{\alpha\Psi_1}\right)^M$$

for any $\alpha > 0$, because the Ψ_j are independent and identically distributed.

$$\mathbf{E}2^{\alpha\Psi_1} = \Pr(\Psi_1 = 0)\cdot 2^0 + \Pr(\Psi_1 = 1)\cdot 2^\alpha = 1 - \mathbf{E}\Psi_1 + 2^\alpha\cdot\mathbf{E}\Psi_1.$$

Since $\alpha > 0$, we have $2^\alpha > 1$. Hence $\mathbf{E}2^{\alpha\Psi_1} \leq 1 - \mu + 2^\alpha\cdot\mu$, because $\mathbf{E}\Psi_1 \leq \mu$ by assumption. Now choose

$$\alpha' := -\log\left(\frac{1-\lambda}{\lambda}\cdot\frac{\mu}{1-\mu}\right).$$

Since $\mu < \lambda$, also $(1-\lambda)/\lambda < (1-\mu)/\mu$; therefore, $\alpha' > 0$. We get easily

$$\left(\mathbf{E}2^{\alpha'\Psi_1}\right)^M = \left(1 - \mu + \mu\cdot\frac{\lambda}{1-\lambda}\cdot\frac{1-\mu}{\mu}\right)^M$$
$$= 2^{-M(\log(1-\lambda) - \log(1-\mu))}$$

and also

$$2^{-\alpha'\Psi\cdot M} = 2^{-M(-\lambda\log((1-\lambda)/\lambda\cdot\mu/(1-\mu)))}.$$

Note now that the product of the last two right sides expressions equals $2^{-MD(\lambda\|\mu)}$.

Proof of Proposition 2

The following lemma (lemma 2) is [1, lemma 1, p. 433].

Lemma 2: Let U be uniformly distributed on $\mathcal{T}_P^n$, $P \in \mathcal{P}_n$. Let $\mathcal{U}$ be a subset of $\mathcal{T}_P^n$, $|\mathcal{U}| = \exp\{nR^*\}$. Define for any $V, V' \in \mathcal{W}$ and $u^* \in \mathcal{X}^n$:

$$g_{V,V'}(u^*) = \left|\bigcup_{u\in\mathcal{U}} \mathcal{T}_{V'}^n(u) \cap \mathcal{T}_V^n(u^*)\right|. \tag{A.1}$$

Then

a) $\mathbf{E}g_{V,V'}(U) \leq (n+1)^{|\mathcal{X}|}\cdot\exp\{n(H(V|P) - [I(P,V') - R^*]^+)\}$;

b) for all $\eta > 0, \xi \geq 0$, and $n \geq n(\eta, |\mathcal{X}|, |\mathcal{Y}|)$,

$$\Pr\left(g_{V,V'}(U) \exp\left\{n\left(H(V|P) - [I(P,V') - R^*]^+ + \xi + 2\eta\right)\right\}\right.$$
$$\left.\text{for one pair } V, V' \in \mathcal{W}\right) \leq \exp\{-n(\eta + \xi)\}.$$

The significance of the functions $g_{V,V'}$ lies in the fact that they can be used in deriving upper bounds on the error probabilities of the second kind. Indeed we have the following.

Lemma 1: Suppose that, for every $V, V' \in \mathcal{W}$, $u^* \in \mathcal{T}_P^n$ satisfies

$$g_{V,V'}(u^*) \leq \exp\left\{n\left(H(V|P)\right.\right.$$
$$\left.\left. - [I(P,V') - R - E_2]^+ + 2\eta + \xi\right)\right\}; \tag{A.2}$$

then

$$W^n(\mathcal{D}(\mathcal{U})\cap\mathcal{F}_{u^*}|u^*)\le(n+1)^{2|\mathcal{X}|\cdot|\mathcal{Y}|}\exp\{-n(E_2-2\eta-\xi)\} \tag{A.3}$$

$$\sum_{u\in\mathcal{U}}W^n(\mathcal{F}_{u^*}|u)\le(n+1)^{2|\mathcal{X}|\cdot|\mathcal{Y}|}\exp\{-n(E_2-2\eta-\xi)\}. \tag{A.4}$$

Proof: Notice that

$$\begin{aligned}W^n(\mathcal{D}(U)\cap\mathcal{F}_{u^*}|u^*)&\le\sum_{V:\,I(P,V)\ge R+2E_2}\\&\cdot\sum_{V':\,I(P,V')\ge R+2E_2}W^n\Big(\mathcal{T}_V^n(u^*)\cap\bigcup_{u\in\mathcal{U}}\mathcal{T}_{V'}^n(u)|u^*\Big)\\&\le(n+1)^{2|\mathcal{X}||\mathcal{Y}|}\max_{\substack{V:\,I(P,V)\ge R+2E_2\\V':\,I(P,V')\ge R+2E_2}}g_{V,V'}(u^*)\\&\cdot\exp\{-n(D(V\|W|P)+H(V|P))\}\end{aligned}$$

by (1.15), (1.18), and (A.1). Furthermore, by (A.2) and $D(V\|W|P)\ge 0$

$$\begin{aligned}g_{V,V'}(u^*)\exp\{-n(D(V\|W|P)+H(V|P))\}&\le\exp\{-n([I(P,V')-R-E_2]^+-2\eta-\xi)\}\\&\le\exp\{-n(E_2-2\eta-\xi)\},\end{aligned}$$

if $I(P,V')\ge R+2E_2$.

Substitution of this bound in the previous bound gives (A.3). We show now (A.4). Clearly,

$$\sum_{u\in\mathcal{U}}W^n(\mathcal{F}_{u^*}|u)\le\sum_{u\in\mathcal{U}}\sum_{V}\sum_{V':\,I(P,V')\ge R+2E_2}W^n(\mathcal{T}_{V'}^n(u^*)\cap\mathcal{T}_V^n(u)|u).$$

Obviously, $\mathcal{T}_{V'}^n(u^*)\cap\mathcal{T}_V^n(u)\ne\emptyset$ implies $PV=PV'$. Now use (1.15), (A.1), (A.2), and $D(V\|W|P)\ge 0$ to obtain the upper bound

$$\sum_{\substack{V':\,I(P,V')\ge R+2E_2\\V:\,PV=PV'}}\exp\{-n(H(V|P)-H(V'|P)+[I(P,V)-R-E_2]^+-2\eta-\epsilon)\}.$$

It remains to be shown that

$$H(V|P)-H(V'|P)+[I(P,V)-R-E_2]^+\ge E_2.$$

If $I(P,V)\le R+E_2$, then because of $PV=PV'$

$$H(V|P)\ge H(PV')-R-E_2.$$

Since $I(P,V')\ge R+2E_2$,

$$\begin{aligned}H(V|P)-H(V'|P)+[I(P,V)-R-E_2]^+&\le I(P,V')-R-E_2\ge R+2E_2-R-E_2=E_2.\end{aligned}$$

On the other hand, if $I(P,V)\ge R+E_2$

$$\begin{aligned}&H(V|P)-H(V'|P)+I(P,V)-R-E_2\\&\quad=H(PV)-H(V'|P)-R-E_2=I(P,V')-R-E_2\\&\quad\ge R+2E_2-R-E_2=E_2.\end{aligned}$$

Lemma 1 says something about the error contribution of a u^* satisfying (2.2) if taken as a member of a $\mathcal{U}_2$, say, and if $\mathcal{U}_1=\mathcal{U}$ is already specified. Our last auxiliary result concerns large deviations. We keep $\eta>0$ fixed in (A.2) and prove

Lemma 3: Let $U_{21},U_{22},\cdots,U_{2M}$ be defined as above. For any $\xi\ge 0$ we define

$$S_{j\xi}(U_{2j})=\begin{cases}0, & \text{if } U_{2j} \text{ equals a } u^* \text{ satisfying (2.2)}\\ 1, & \text{otherwise}\end{cases}$$

for $j=1,\cdots,M$. Then for every $\xi\in[0,E_2]$ and $n\ge n(\eta,E_2)$:

$$\Pr\Big\{\sum_{j=1}^{M}S_{j\xi}>\exp\{-n\xi\}\cdot M\Big\}\le\exp\{-\exp\{nR\}\cdot(n\eta-2)\}. \tag{A.5}$$

Proof: For fixed ξ the $S_{1\xi},\cdots,S_{M\xi}$ are independent identically distributed random variables with values in $\{0,1\}$. Lemma 2-b) gives

$$\mathbf{E}S_{1\xi}\le\exp\{-n(\eta+\xi)\}.$$

We apply Lemma LD and get

$$\Pr\Big\{\sum_{j=1}^{M}S_{j\xi}>M\cdot\exp\{-n\xi\}\Big\}\le\exp\{-M\cdot D(2^{-n\xi}\|2^{-n(\xi+\eta)})\}. \tag{A.6}$$

We have to estimate the divergence.

$$\begin{aligned}D(2^{-n\xi}\|2^{-n(\xi+\eta)})&=2^{-n\xi}\log(2^{-n\xi}\cdot 2^{n(\xi+\eta)})\\&\quad+(1-2^{-n\xi})\log\frac{1-2^{-n\xi}}{1-2^{-n(\xi+\eta)}}\\&\ge 2^{-n\xi}\cdot\eta n+(1-2^{-n\xi})\log(1-2^{-n\xi})\\&\ge 2^{-n\xi}\cdot\eta n+\log(1-2^{-n\xi}),\end{aligned}$$

because $t\log t\le 0$ for $t\in[0,1]$.

For small $t>0$ one can estimate

$$\log(1-t)\ge -2t.$$

Therefore,

$$D(2^{-n\xi}\|2^{-n(\xi+\eta)})\ge 2^{-n\xi}\cdot\eta\cdot n-2\cdot 2^{-n\xi}=(n\eta-2)\cdot 2^{-n\xi}, \tag{A.7}$$

if n is large enough. Since $M=\lceil\exp\{n(R+E_2)\}\rceil$ (see (2.6)) and $\xi\in[0,E_2]$, (A.6) and (A.7) imply (A.5).

From Lemmas 1–3 to Proposition 1: We apply Lemmas 1–3 with $\eta=\gamma$. Choose in Lemma 3

$$\xi_k=\frac{E_2\cdot k}{n},\qquad k=0,\cdots,n$$

to obtain

$$\Pr\Big\{\sum_{j=1}^{M}S_{j\xi_k}>M\cdot\exp\{-n\xi_k\}\quad\text{for some } k\in\{0,\cdots,n\}\Big\}\le(n+1)\exp\{-\exp\{nR\}\cdot(n\gamma-2)\}. \tag{A.8}$$

It remains to be shown that (2.7) and (2.8) hold if

$$\sum_{j=1}^{M}S_{j\xi_k}\le M\cdot\exp\{-n\xi_k\}\qquad\text{for all } k\in\{0,\cdots,n\}. \tag{A.9}$$

Suppose now that (A.9) holds. Choose $j\in\{1,\cdots,M\}$. Note that if $S_{j\xi}=0$ for a $\xi\ge 0$, then also $S_{j\xi'}=0$ for every $\xi'\ge\xi$. Similarly, if $S_{j\xi'}=1$ for a $\xi\ge 0$, then also $S_{j\xi'}=0$ for every $0\le\xi'\le\xi$. By the definition of S there is a minimal ξ such that $S_{j\xi}=0$.

Choose $k\in\{0,\cdots,n\}$. The number of $j\in\{1,\cdots,M\}$ such that the corresponding minimal ξ is contained in the interval $(k/n,(k+1)/n)$ is upper-bounded by $\sum_{j=0}^{M}S_{j\xi_k}$ (because of the

monotonicity property discussed above). For these j, of course, $S_{j\xi_{k+1}} = 0$ holds, and *a fortiori*, (A.4) holds with ξ_{k+1}. On the other hand, the number of $j \in \{1, \cdots, M\}$ such that the corresponding minimal ξ is larger than $E_2 (= \xi_n)$ is upper-bounded by $\sum_{j=1}^{M} S_{j\xi_n}$.

With these arguments we get the following estimate:

$$\frac{1}{M} \sum_{u \in \mathcal{U}_1} W^n(\mathcal{D}(\mathcal{U}_2)|u)$$

$$\leq \frac{1}{M} \sum_{j=1}^{M} \left(\sum_{u \in \mathcal{U}_1} W^n(\mathcal{F}_{U_{2j}}|u) \right)$$

$$\leq \frac{1}{M} \sum_{k=0}^{n} \left(\sum_{j=1}^{M} S_{j\xi_k} \right) \cdot (n+1)^{2|\mathcal{X}| \cdot |\mathcal{Y}|}$$

$$\cdot \exp\{-n(E_2 - 2\gamma - \xi_{k+1})\} + \frac{1}{M} \sum_{j=1}^{M} S_{j\xi_n}.$$

Since $\xi_k - \xi_{k+1} = -1/n$ we can continue with (A.9)

$$\frac{1}{M} \sum_{u \in \mathcal{U}_1} W^n(\mathcal{D}(\mathcal{U}_2)|u)$$

$$\leq \left(\sum_{k=0}^{n} (n+1)^{2|\mathcal{X}| \cdot |\mathcal{Y}|} \cdot \exp\{-n(E_2 - 2\gamma - \xi_{k+1} + \xi_k)\} \right)$$

$$+ \exp\{-nE_2\} \leq \exp\{-n(E_2 - 3\gamma)\}$$

for $n \geq n(\gamma, E_2)$. Using (A.3) instead of (A.4) in the foregoing derivation one obtains by the very same arguments

$$\frac{1}{M} \sum_{u \in \bar{\mathcal{U}}_2} W^n(\mathcal{D}(\mathcal{U}_1)|u) \leq \frac{1}{M} \sum_{j=1}^{M} W^n\left(\bigcup_{u \in \mathcal{U}_1} \mathcal{F}_u \cap \mathcal{F}_{U_{2j}} | U_{2j} \right)$$

$$\leq \cdots \leq \exp\{-n(E_2 - 3\gamma)\}$$

for $n \geq n(\gamma, E_2)$.

References

[1] R. Ahlswede and G. Dueck, "Good codes can be produced by a few permutations," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 3, pp. 430–443, 1982.

[2] R. Ahlswede, "Coloring hypergraphs: A new approach to multi-user source coding," *J. Comb. Inf. Syst. Sci.*, Pt. I, vol. 1, pp. 76–115, 1979; Pt. II, vol. 5, pp. 220–268, 1980.

[3] ——, "A method of coding and an application to arbitrarily varying channels," *J. Comb. Inf. Syst. Sci.*, vol. 5, no. 1, pp. 10–35, 1980.

[4] A. C. Yao, "Some complexity questions related to distributive computing," in *Proc. 11th Annu. Symp. Theory of Computing*, Atlanta, GA, 1979, pp. 209–213.

[5] J. Ja'Ja', "Identification is easier than decoding," preprint, 1985.

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