Good Codes Can Be Produced by a Few Permutations

RUDOLF AHLSWEDE AND GUNTER DUECK, MEMBER, IEEE

Abstract—Our main result is that good codes, even those meeting the random coding bound, can be produced with relatively few (linear in the block length) permutations from a single codeword. This cutdown in complexity may be of practical importance. The motivation for looking at such codes came from Ahlswede's covering lemma, which makes it possible to build correlated source codes from channel codes via permutations. In Appendix I we show that the problem of finding the best error exponents for coding sources with full side information at the decoder, which has received attention in the recent literature, can easily be reduced to the familiar one for the discrete memoryless channel (DMC). Finally, in Appendices II and III we give rather precise double exponentially small bounds on the probabilities that a randomly chosen code will fail to meet the random coding or expurgated bound for the DMC. According to these results, good codes are hard to miss if selected at random. This also explains why good codes of a low complexity (such as those produced by permutations) do exist.

I. Introduction

In [5, part II, sections 5 and 6] Ahlswede suggested as a *program in coding theory to systematically investigate the symmetric group* $\mathfrak{S}_n$ (the group of permutations) acting on the components $\{1,\cdots,n\}$. The immediate use of this group is due to the fact that it leaves probability distributions specifying stationary memoryless multi-user sources and channels invariant. As a justification for his belief in this program he presented a general robustification technique, and he derived Slepian–Wolf's [6] source coding theorem from the coding theorem for the discrete memoryless channel (DMC) via a *covering lemma* (see Appendix I). By this method source codes are built from channel codes.

Here we show that channel codes, which achieve capacity (and even the random coding bound), can also be built up iteratively by producing bigger codes from good smaller codes with suitable permutations $\pi_1,\cdots,\pi_t$, say, which we call *code producers*. In particular, this is possible for subcodes consisting just of one codeword.

To fix ideas, we describe the production first in this case. Suppose we are given a single codeword $x^n=(x_1,\cdots,x_n)$ of length n and the permutations operate on $\{1,\cdots,n\}$. By $\pi_1 x^n$ we mean the n-sequence obtained from x^n by permuting the components of x^n according to π_1, i.e.,

$$\pi_1 x^n = \left(x_{\pi_1 1},\cdots,x_{\pi_1 n}\right).$$

Now we have two codewords x^n and $\pi_1 x^n$. Form now $\pi_2 x^n$ and $\pi_2\circ\pi_1 x^n$, then $\pi_3 x^n$, $\pi_3\circ\pi_1 x^n$, $\pi_3\circ\pi_2 x^n$, $\pi_3\circ\pi_2\circ\pi_1 x^n$, etc. In each step we double the cardinality of our codeword set, if repetitions are counted with multiplicity. In this manner it is possible to construct simply structured codes. Note that in order to give a code book for such a code we have to list t permutations, say, instead of $\exp\{t\}$ codewords.

Finally, we prove right away a somewhat stronger result than just achievability of the random coding bound: the same set of permutations can serve for every positive rate below capacity as follows. If the rate is R, then use the first t' permutations, where t' is minimal with the property $\exp\{t'\}\geq\exp\{nR\}$. Moreover, we also establish universality in the sense of Goppa [18], that is, the same set of permutations can be used for all channels of bounded alphabet sizes. The exact statements are given in the Main Theorem. For ordinary codes Goppa proved universality with respect to the capacities and this result was sharpened by Csiszár, Körner, and Marton [15] to the universal achievability of the random coding bound. Those authors also proved that the expurgated bound can be achieved using a universal set of codewords, and Csiszár and Körner established in [9] the (universal) achievability of both bounds simultaneously. We do not know yet whether those results can be proved for our simply structured codes for we do not even know whether the expurgated bound can be achieved at all. The immediate reason is that expurgation destroys the algebraic structure.

We would like to draw attention to another *problem* of some interest. Generally speaking the idea of building bigger structures from smaller structures is very common in human life (also the reverse process, which is often an unfortunate fact), in science, and, especially, in engineering. It is often wasteful to build a new machine from scratch, if functioning parts are available and could be used. Code producers perform this task for all discrete memoryless channels with properly bounded alphabet sizes and all rates. However, they do so only for fixed block length n. Hence, it may be interesting to try now to build producers from smaller ones, that is to introduce "producers of producers."

Our *main tool* for proving the main theorem is a new kind of maximal code method for abstract bipartite graphs, which was given by Ahlswede in [5, part II, section 4, §3]. The method uses average errors. Other differences from

Manuscript received February 7, 1980; revised April 30, 1981. This work was presented at the IEEE Symposium on Information Theory, Santa Monica, CA, 1981.

The authors are with the Department of Mathematics, University of Bielefeld, D-4800 Bielefeld, West Germany.

0018-9448/82/0500-0430$00.75 ©1982 IEEE

Feinstein's maximal code method [20], which is for maximal errors, are explained in [5]. An important feature of the method is that while finding codewords iteratively, the error probability of any initial code can be linked to the error probability of the extended code.

Moreover, the selection of a codeword at each step can be done at random and the probability of finding good code extensions can be estimated rather precisely. These estimates are used in the Appendices II and III to derive bounds on the probability that a randomly chosen (nonexpurgated or suitably expurgated) code achieves the best known error bounds. They are also used for showing the existence of universal code producers.

In applying the abstract maximal method to "channel graphs" the actual calculations of graphic parameters such as degrees, etc., involve information quantities. These calculations are very similar to those used in the proofs of [15, theorem R and theorem EX] stated in the next section. They also can be found in the forthcoming book [24]. Since it will be widely available soon, we adopt its notation and refer to it for proofs of auxiliary results.

In another paper the first author will give applications of the abstract maximal coding method and of other methods of [5] to other graphs and hypergraphs of genuine information theoretical interest. There the graphic parameters *cannot* be described by *information quantities* and this will, as we hope, convince more people of the use of the abstract approach to information theory developed in [5].

II. Notation and Known Facts

Script capitals $\mathcal{X}, \mathcal{Y}, \cdots$ will denote finite sets. The cardinality of a set $\mathcal{C}$ and of the range of a function f will be denoted by $|\mathcal{C}|$ and $\|f\|$, respectively. The letters P, Q will always stand for probability distributions (PD's) on finite sets, and $X, Y, \cdots$ denote random variables (RV's).

A. Channels, Types, Generated Sequences

A stochastic matrix $W = \{W(y|x)\colon\ y \in \mathcal{Y},\ x \in \mathcal{X}\}$ uniquely defines a DMC 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 $x^n = (x_1,\cdots,x_n) \in \mathcal{X}^n$, $y^n = (y_1,\cdots,y_n) \in \mathcal{Y}^n$, $n = 1,2,3,\cdots$.

We denote by $\mathcal{P}$ the sets of all PD's on $\mathcal{X}$ and by $\mathcal{W}$ (resp. $\mathcal{V}$) the set of all channels with alphabets $\mathcal{X}, \mathcal{Y}$ (resp. $\mathcal{X}, \mathcal{X}$).

For positive integers n we set

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

For any $P \in \mathcal{P}_n$, called *type*, we define the set $\mathcal{W}_n(P) = \{\tilde{W} \in \mathcal{W} \mid \tilde{W}(y|x) \in \{0, 1/(nP(x)), 2/(nP(x)), \cdots, 1\}$ for all $x \in \mathcal{X}$, $y \in \mathcal{Y}\}$. $\mathcal{V}_n(P)$ is defined similarly.

The *type* of a sequence $x^n \in \mathcal{X}^n$ is the distribution $P_{x^n} \in \mathcal{P}_n$ defined by letting $P_{x^n}(x)$ count the relative frequency of the letter x in the n-sequence x^n. The *joint type* of a pair $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ is the distribution $P_{x^n y^n}$ on $\mathcal{X} \times \mathcal{Y}$ defined analogously. For $P \in \mathcal{P}$, the set $\mathcal{T}_P^n$ of all P-typical sequences in $\mathcal{X}^n$ is given by

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

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

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

The set of those sequences is denoted by $\mathcal{T}_{\tilde{W}}^n(x^n)$. Observe that $\mathcal{T}_P^n \neq \phi$ if and only if $P \in \mathcal{P}_n$ and $\mathcal{T}_{\tilde{W}}^n(x^n) \neq \phi$ if and only if $\tilde{W} \in \mathcal{W}_n(P_{x^n})$.

B. Entropy and Information Quantities

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

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

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

$$D(\tilde{P}\|P) = \sum_{x \in \mathcal{X}} \tilde{P}(x) \log \frac{\tilde{P}(x)}{P(x)}$$

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

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

stands for the conditional I-divergence. Finally, for $x^n \in \mathcal{X}^n$, $y^n \in \mathcal{Y}^n$

$$I(x^n \wedge y^n) = \sum_x \sum_y P_{x^n, y^n}(x, y) \log \frac{P_{x^n y^n}(x, y)}{P_{x^n}(x) \cdot P_{y^n}(y)}.$$

C. Elementary Properties of Typical Sequences and Generated Sequences

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

$$|\mathcal{W}_n(P)| \leq (n+1)^{|\mathcal{X}| \cdot |\mathcal{Y}|}, \qquad \text{for } P \in \mathcal{P}_n, \tag{2}$$

$$|\mathcal{V}_n(P)| \leq (n+1)^{|\mathcal{X}| \cdot |\mathcal{X}|}, \qquad \text{for } P \in \mathcal{P}_n, \tag{3}$$

$$|\mathcal{T}_P^n| = \frac{n!}{\prod_{x \in \mathcal{X}} (nP(x))!}, \qquad \text{for } P \in \mathcal{P}_n, \tag{4}$$

$$(n+1)^{-|\mathcal{X}|} \exp\{nH(P)\} \leq |\mathcal{T}_P^n| \leq \exp\{nH(P)\}, \qquad \text{for all } P \in \mathcal{P}_n. \tag{5}$$

For $P \in \mathcal{P}_n$, $\tilde{W} \in \mathcal{W}_n(P)$, $x^n \in \mathcal{T}_P^n$:

$$(n+1)^{-|\mathcal{X}| \cdot |\mathcal{Y}|} \exp\big\{nH(\tilde{W}|P)\big\} \leq |\mathcal{T}_{\tilde{W}}^n(x^n)| \leq \exp\big\{nH(\tilde{W}|P)\big\}. \tag{6}$$

For $P \in \mathcal{P}_n$, $\tilde{P} \in \mathcal{P}$, $x^n \in \mathcal{T}_P^n$:

$$\tilde{P}^n(x^n) = \exp\big\{-n\big(D(P\|\tilde{P}) + H(P)\big)\big\}, \tag{7}$$

where $\tilde{P}^n$ is the n-fold extension of $\tilde{P}$. For $P \in \mathcal{P}$; $\tilde{W}, \tilde{\tilde{W}} \in \mathcal{W}$; $x^n \in \mathcal{T}_P^n$, $y^n \in \mathcal{T}_{\tilde{W}}^n(x^n)$:

$$\tilde{\tilde{W}}^n(y^n \mid x^n) = \exp\left\{-n\left(D\left(\tilde{W} \| \tilde{\tilde{W}} \mid P\right) + H\left(\tilde{W} \mid P\right)\right)\right\}. \tag{8}$$

For $P \in \mathcal{P}_n$, $\tilde{W} \in \mathcal{W}_n(P)$, $y^n \in \mathcal{T}_{P\tilde{W}}^n$:

$$\begin{aligned}(n+1)^{-|\mathcal{X}| \cdot |\mathcal{Y}|} & \exp\left\{n\left(H(P) - I(P, \tilde{W})\right)\right\} \\ & \leq \left| \left\{x^n \in \mathcal{T}_P^n \mid y^n \in \mathcal{T}_{\tilde{W}}^n(x^n)\right\} \right| \\ & \leq \exp\left\{n\left(H(P) - I(P, \tilde{W})\right)\right\},\end{aligned} \tag{9}$$

where $P\tilde{W}$ denotes the PD on $\mathcal{Y}$ given by $P\tilde{W}(y) = \Sigma_x P(x)\tilde{W}(y \mid x)$ for $y \in \mathcal{Y}$.

D. Historical Sketch of the Bounds on the Reliability Function

An (n, N) code $\mathcal{C}$ for the DMC is a system of pairs $\{(u_i, \mathcal{D}_i) \mid i = 1, \cdots, N\}$ with $u_i \in \mathcal{X}^n$ and pairwise disjoint subsets $\mathcal{D}_i \subset \mathcal{Y}^n$ $(i = 1, \cdots, N)$. $\bar{\lambda}(\mathcal{C}, W)$ denotes the average error probability of $\mathcal{C}$, i.e.,

$$\bar{\lambda}(\mathcal{C}, W) = \frac{1}{N} \sum_{i=1}^{N} W^n(\mathcal{D}_i^c \mid u_i),$$

where $\mathcal{D}_i^c = \mathcal{X}^n - \mathcal{D}_i$. $\lambda_{\max}(\mathcal{C}, W) = \max_i W^n(\mathcal{D}_i^c \mid u_i)$ denotes the maximal error of $\mathcal{C}$. $\mathcal{C}$ is called an ML code (maximum likelihood code), if for $i = 1, \cdots, N$ the sets $\mathcal{D}_i$ consist of those n-words $y^n \in \mathcal{Y}^n$ such that

$$\begin{aligned} W^n(y^n \mid u_i) &\geq W^n(y^n \mid u_j), \qquad \text{for all } j \neq i \\ W^n(y^n \mid u_i) &> W^n(y^n \mid u_j), \qquad \text{for all } j < i. \end{aligned}$$

If we define for any rate R

$$\bar{\lambda}(n, R, W) = \min\{\bar{\lambda}(\mathcal{C}, W) \mid \mathcal{C} \text{ is an } (n, N) \text{ code with } N \geq \exp\{nR\}\},$$

then

$$E(R, W) = \limsup_{n \to \infty} -\frac{1}{n} \log \bar{\lambda}(n, R, W)$$

is the familiar reliability function for the DMC W.

Since Shannon discovered the coding theorem for the DMC in his famous paper [1] there has been considerable effort in improving bounds on the error probability for codes of a given rate or, equivalently, on the reliability function $E(R, W)$. Well-known upper bounds on $E(R, W)$ are the sphere packing bound $E_{\text{sp}}(R, W)$ and the straight line bound $E_{\text{sl}}(R, W)$. These bounds were derived by Shannon, Gallager, and Berlekamp [10]. $E_{\text{sp}}(R, W)$ was first established (with an incomplete proof) by Fano [11]. For rates $R > C$ Wolfowitz's strong converse [12] implies

$$\liminf_{n \to \infty} \bar{\lambda}(n, R, W) = 1.$$

For $R > C$ the problem is to evaluate

$$\liminf_{n \to \infty} -\frac{1}{n} \log\left(1 - \bar{\lambda}(n, R, W)\right).$$

Arimoto [13] extended the sphere packing exponent for rates above capacity, and finally Dueck and Körner [14] showed that this exponent is optimal. A partial result in this direction was obtained earlier by Omura [25].

The best known lower bounds for $R < C$ are the random coding bound $E_r(R, W)$, which was derived by Fano and given a simpler proof by Gallager [2], and the expurgated bound $E_{\text{ex}}(P, W)$, which is due to Gallager [2].

Our results here mainly concern those lower bounds. Csiszár, Körner, and Marton [15] have rederived those bounds via types incorporating earlier ideas of Haroutunian [16], Blahut [17], and Goppa [18]. Their approach leads to universal codes. The function $E_r(R, W)$ and to a certain extent also the function $E_{\text{ex}}(R, W)$ appear in the new derivations in a form somewhat more linked to information quantities than the familiar analytic expression [19]. The results of [15] are

Theorem R: For every $R > 0$, $\delta > 0$, $n \geq n_0(|\mathcal{X}|, |\mathcal{Y}|, \delta)$, and every type $P \in \mathcal{P}_n$ there exists an (n, N) code

$$\mathcal{C} = \left\{(u_i, \mathcal{D}_i) \mid i = 1, \cdots, N\right\} \qquad \text{with } u_i \in \mathcal{T}_P^n \text{ and } \frac{1}{n} \log N \geq R - \delta$$

such that

$$\bar{\lambda}(\mathcal{C}, W) \leq \exp\left\{-n\left(E_r(R, P, W) - \delta\right)\right\} \tag{10}$$

for any $W \in \mathcal{W}$, where

$$E_r(R, P, W) = \min_{\tilde{W} \in \mathcal{W}} \left\{D\left(\tilde{W} \| W \mid P\right) + \left[I(P, \tilde{W}) - R\right]^+\right\},$$

and $[t]^+ = \max\{0, t\}$.

Theorem EX: For every $R > 0$, $\delta > 0$, $n \geq n_0(|\mathcal{X}|, |\mathcal{Y}|, \delta)$, and every type $P \in \mathcal{P}_n$ there exist codewords

$$u_1, \cdots, u_N \in \mathcal{T}_P^n \quad \text{with} \quad \frac{1}{n} \log N \geq R - \delta$$

such that for every $W \in \mathcal{W}$ the corresponding ML code

$$\mathcal{C}^W = \left\{\left(u_i, \mathcal{D}_i^W\right) \mid i = 1, \cdots, N\right\}$$

[i.e., the $\mathcal{D}_i^W$ denote the maximum likelihood decoding sets with respect to W] satisfies

$$\bar{\lambda}(\mathcal{C}^W, W) \leq \exp\left\{-n\left(E_{\text{ex}}(R, P, W) - \delta\right)\right\},$$

where

$$E_{\text{ex}}(R, P, W) = \min_{\substack{X, \tilde{X}\ P\text{-distributed} \\ I(X \wedge \tilde{X}) \leq R}} \left\{Ed(X, \tilde{X}) + I(X \wedge \tilde{X}) - R\right\}$$

and $d(x, \tilde{x}) = -\log \Sigma_{y \in \mathcal{Y}} \sqrt{W(y \mid x) \cdot W(y \mid \tilde{x})}$ for $x, \tilde{x} \in \mathcal{X}$. $Ed(\cdot)$ means the expectation of $d(\cdot)$.

Actually, in [9] a unified description of the random coding and expurgated bound was given, but this description will not be used here. Also, Theorem *EX* will be relevant only in the Appendix.

III. The Main Result: Channel Codes Produced by Permutations

Let k, n be positive integers with $k \cdot n \le n!$. We call any subset $\{\pi_1,\cdots,\pi_{n\cdot k}\} \subset \mathcal{S}_n$ a *code producer*. Such a code producer works as follows. Assume we are given a DMC W with alphabet $\mathcal{X}, \mathcal{Y}$, where $|\mathcal{X}|, |\mathcal{Y}| \le 2^k$ and we want to transmit one of 2^m messages $(m \le n \cdot k)$ over this channel using an n-length block code. First we identify the messages with m-sequences in $\{0,1\}^m$, then we choose a proper type $P \in \mathcal{P}_n$ and build the "canonical P-sequence" u_P defined by

$$u_P = \left(x_1,\cdots,x_1, x_2,\cdots,x_2,\cdots,x_{|\mathcal{X}|},\cdots,x_{|\mathcal{X}|}\right) \in \mathcal{T}_P^n,$$

where $\mathcal{X} = \{x_1,\cdots,x_{|\mathcal{X}|}\}$.

Suppose that message $z^m = (z_1,\cdots,z_m) \in \{0,1\}^m$ is to be sent over the channel W. Then the encoder puts

$$\pi_m^{z_m} \circ \pi_{m-1}^{z_{m-1}} \circ \cdots \circ \pi_z^{z_1} \circ id(u_P)$$

into the channel, where $id \in \mathcal{S}_n$ is the identity mapping and $\pi_i^0 = id$, $\pi_i^1 = \pi_i$ for $i = 1,\cdots,n \cdot k$. Thus, the codeword set produced for the given parameters $\mathcal{X}$, $P \in \mathcal{P}_n$, and m is

$$\left\{\pi_m^{z_m} \circ \cdots \circ \pi_z^{z_1} \circ id(u_P) \,|\, z^m = (z_1,\cdots,z_m) \in \{0,1\}^m\right\}.$$

We denote the ML code with respect to the channel W for this codeword set by

$$\mathcal{C}(\pi_1,\cdots,\pi_{n\cdot k}, P, \mathcal{X}, \mathcal{Y}, W, R),$$

where

$$R = \frac{1}{n}\log 2^m.$$

Two sequences $\pi_m^{z_m} \circ \pi_1^{z_1} \circ id(u_P)$, $\pi_m^{z'_m} \circ \cdots \circ \pi_1^{z'_1} id(u_P)$ are considered as different if $z^m \neq z'^m$, even though they may represent the same element of $\mathcal{T}_P^n$. Therefore the cardinalities of the produced codeword sets are always powers of two.

If N is given and we want to produce an n-length block code with N messages $(R = (1/n)\log N)$, then by $\mathcal{C}(\pi_1,\cdots,\pi_{nk}, P, \mathcal{X}, \mathcal{Y}, W, R)$ we mean always the code having 2^m codewords, where 2^m is the smallest power of 2 with $2^m \ge N$.

Main Theorem: Fix a positive integer k and $\delta > 0$. Then for any $n \ge n_0(k,\delta)$ there exists a producer $\{\pi_1,\cdots,\pi_{n\cdot k}\} \subset \mathcal{S}_n$ with the properties

$$\bar{\lambda}(\mathcal{C}(\pi_1,\cdots,\pi_{nk}, P, \mathcal{X}, \mathcal{Y}, W, R), W) \le \exp\{-n(E_r(R,P,W) - \delta)\},$$

for *every* $\mathcal{X}, \mathcal{Y}$ with $|\mathcal{X}|, |\mathcal{Y}| \le 2^k$, for every $P \in \mathcal{P}_n$, for *every* channel W with alphabets $\mathcal{X}$ and $\mathcal{Y}$, and for *every* rate $R > 0$.

The theorem is an immediate consequence of the following lemma.

Main Lemma: Fix alphabets $\mathcal{X}, \mathcal{Y}$ and $\delta > 0$. Then for *every* $n \ge n_0(|\mathcal{X}|, |\mathcal{Y}|, \delta)$, *every* type $P \in \mathcal{P}_n$, and *every* code $\mathcal{C} = \{(u_i, \mathcal{D}_i) \,|\, i = 1,\cdots,N\}$; $u_i \in \mathcal{T}_P^n$ for $i = 1,\cdots,N$; there exists a permutation $\pi \in \mathcal{S}_n$ and suitable decoding sets $\mathcal{E}_1,\cdots,\mathcal{E}_N, \mathcal{E}_{1,\pi},\cdots,\mathcal{E}_{N,\pi}$ such that the enlarged code

$$\mathcal{C}'_\pi = \{(u_1,\mathcal{E}_1),\cdots,(u_N,\mathcal{E}_N),(\pi u_1,\mathcal{E}_{1,\pi}),\cdots,(\pi u_N,\mathcal{E}_{N,\pi})\}$$

satisfies for *every* $W \in \mathcal{W}$

$$\bar{\lambda}(\mathcal{C}'_\pi, W) \le \bar{\lambda}(\mathcal{C}, W) + \exp\{-n(E_r(R,P,W) - \delta)\}, \tag{11}$$

where $R = (1/n)\log N$.

Moreover, for a randomly (according to the uniform distribution on $\mathcal{S}_n$) chosen π (11) holds for all $W \in \mathcal{W}$ with a probability larger than $1 - \exp\{-(\delta/2)\cdot n\}$.

The proof is based on the maximal coding idea of [5]. In its original form codewords are added iteratively to a given code. Here we add permutations iteratively and thus keep doubling the lengths of codes. The reader may find it easier to study first Theorems 2 and 3 in Appendix II, whose proofs use the original form. These theorems are needed for the derivation of double exponential bounds on the probability that a randomly chosen code fails to meet the random coding or expurgated bound for the DMC. They also imply Theorem R and Theorem EX and thus we have an alternative proof of those theorems by maximal coding.

For the proof of the main lemma we need Lemmas 1 and 2 below. They involve quantities which we now define.

Fix $R > 0$, $\delta > 0$, $P \in \mathcal{P}_n$, and let $\{u_1,\cdots,u_N\} \subset \mathcal{T}_P^n$ and $N \le \exp\{nR\}$ be given.

For any pair $\tilde{W}, \tilde{\tilde{W}} \in \mathcal{W}$ we define the function $g_{\tilde{W},\tilde{\tilde{W}}}$ on $\mathcal{X}^n$ by

$$g_{\tilde{W},\tilde{\tilde{W}}}(u) = \left|\mathcal{T}_{\tilde{W}}^n(u) \cap \bigcup_{i=1}^N \mathcal{T}_{\tilde{\tilde{W}}}^n(u_i)\right|, \qquad \text{for } u \in \mathcal{X}^n. \tag{12}$$

$g_{\tilde{W},\tilde{\tilde{W}}}(u)$ measures the size of intersections of sets generated by n and of sets generated by the given system of codewords.

Furthermore, for permutations $\pi \in \mathcal{S}_n$ we define the function $g^*_{\tilde{W},\tilde{\tilde{W}}}$ by

$$g^*_{\tilde{W},\tilde{\tilde{W}}}(\pi) = \sum_{i=1}^N g_{\tilde{W},\tilde{\tilde{W}}}(\pi u_i).$$

Let U be a random variable equidistributed on $\mathcal{T}_P^n$ and let Π be a random variable equidistributed on $\mathcal{S}_n$.

Lemma 1: For every pair $\tilde{W}, \tilde{\tilde{W}} \in \mathcal{W}$

a) $Eg_{\tilde{W},\tilde{\tilde{W}}}(U) \le (n+1)^{|\mathcal{X}|}\exp\{n(H(\tilde{W}|P) - [I(P,\tilde{\tilde{W}}) - R]^+\}$, where $[t]^+ = \max\{0,t\}$.

Furthermore, for any $\delta > 0$, $\xi \ge 0$ and $n \ge n_0(\delta, |\mathcal{X}|, |\mathcal{Y}|)$;

b) $\Pr\{g_{\tilde{W},\tilde{\tilde{W}}}(U) \ge \exp\{n(H(\tilde{W}|P) - [I(P,\tilde{\tilde{W}}) - R - \xi]^+ + (3/4)\delta$ for some $\tilde{W}, \tilde{\tilde{W}} \in \mathcal{W}\} \le \exp\{-n((\delta/2) + \xi)\}$.

Lemma 2: For every pair $\tilde{W}, \tilde{\tilde{W}} \in \mathcal{W}$

a) $Eg_{\tilde{W},\tilde{\tilde{W}}}(\Pi) = N \cdot Eg_{\tilde{W},\tilde{\tilde{W}}}(U)$.

For any $\delta > 0$ and $n \ge n_0(\delta, |\mathcal{X}|, |\mathcal{Y}|)$;

b) $\Pr\{g_{\tilde{W},\tilde{\tilde{W}}}(\Pi) \geq N \cdot \exp\{n(H(\tilde{W} \mid P) - [I(P,\tilde{\tilde{W}}) - R]^{+} + (3/4)\delta)\}$ for some $\tilde{W},\tilde{\tilde{W}} \in \mathcal{W}\} \leq \exp\{-n(\delta/2)\}$.

Proof of Lemma 1: Choose any $\tilde{W},\tilde{\tilde{W}} \in \mathcal{W}$ and note that $g_{\tilde{W},\tilde{\tilde{W}}}$ is zero for sequences in $\mathcal{T}_P^n$ if $\tilde{W} \notin \mathcal{W}_n(P)$ or $\tilde{\tilde{W}} \notin \mathcal{W}_n(P)$. Let $P\tilde{W}$ denote the distribution on $\mathcal{Y}$ given by

$$P\tilde{W}(y) = \sum_x P(x)\tilde{W}(y \mid x), \qquad \text{for } y \in \mathcal{Y}.$$

Note again that $g_{\tilde{W},\tilde{\tilde{W}}}$ is zero for sequences in $\mathcal{T}_P^n$ if $P\tilde{W} \neq P\tilde{\tilde{W}}$. Hence we assume that $\tilde{W},\tilde{\tilde{W}} \in \mathcal{W}_n(P)$ and $P\tilde{W} = P\tilde{\tilde{W}}$.

$$\begin{aligned}
Eg_{\tilde{W},\tilde{\tilde{W}}}(U) &= E\left|\mathcal{T}_{\tilde{W}}^n(U) \cap \bigcup_{i=1}^N \mathcal{T}_{\tilde{\tilde{W}}}^n(u_i)\right| \\
&\leq \sum_{i=1}^N E\left|\mathcal{T}_{\tilde{W}}^n(U) \cap \mathcal{T}_{\tilde{\tilde{W}}}^n(u_i)\right| \\
&= N \cdot E\left|\mathcal{T}_{\tilde{W}}^n(U) \cap \mathcal{T}_{\tilde{\tilde{W}}}^n(u_1)\right| \quad \text{(by symmetry)} \\
&= N \cdot \sum_{y^n \in \mathcal{T}_{\tilde{\tilde{W}}}^n(u_1)} \Pr\left(y^n \in \mathcal{T}_{\tilde{W}}^n(U)\right).
\end{aligned}$$

Since U is equidistributed over $\mathcal{T}_P^n$, we have for every $y^n \in \mathcal{Y}^n$ that

$$\Pr\left(y^n \in \mathcal{T}_{\tilde{W}}^n(U)\right) = \frac{\left|\{x^n \mid x^n \in \mathcal{T}_P^n,\ y^n \in \mathcal{T}_{\tilde{W}}^n(x^n)\}\right|}{|\mathcal{T}_P^n|};$$

therefore, (5) and (9) yield

$$\begin{aligned}
&Eg_{\tilde{W},\tilde{\tilde{W}}}(U) \\
&\leq N \cdot \left|\mathcal{T}_{\tilde{\tilde{W}}}^n(u_1)\right| \\
&\quad \cdot \exp\{n(H(P) - I(P,\tilde{W}) - H(P))\} \cdot (n+1)^{|\mathcal{X}|} \\
&\leq N \cdot \exp\left\{n\left(H(\tilde{\tilde{W}} \mid P) - I(P,\tilde{W})\right)\right\} \cdot (n+1)^{|\mathcal{X}|}. \qquad (13)
\end{aligned}$$

By assumption, $P\tilde{W} = P\tilde{\tilde{W}}$ and thus, $I(P,\tilde{W}) = H(P\tilde{W}) - H(\tilde{W} \mid P)$.

We therefore get from (13)

$$Eg_{\tilde{W},\tilde{\tilde{W}}}(U) \leq N \cdot \exp\left\{n\left(H(\tilde{W} \mid P) - I(P,\tilde{\tilde{W}})\right)\right\} \cdot (n+1)^{|\mathcal{X}|}. \qquad (14)$$

On the other hand, it is obvious from the definition of $g_{\tilde{W},\tilde{\tilde{W}}}$ and from (6) that

$$Eg_{\tilde{W},\tilde{\tilde{W}}}(U) \leq E\left|\mathcal{T}_{\tilde{W}}^n(U)\right| \leq \exp\left\{nH(\tilde{W} \mid P)\right\}. \qquad (15)$$

Since $N \leq \exp\{nR\}$, (14) and (15) imply a). b) follows from a) by applying Chebychev's inequality.

Proof of Lemma 2: Let $\tilde{W},\tilde{\tilde{W}} \in \mathcal{W}$. Then

$$\begin{aligned}
Eg^*_{\tilde{W},\tilde{\tilde{W}}}(\Pi) &= \frac{1}{n!}\sum_{i=1}^N \sum_{\pi \in \mathcal{S}_n} g_{\tilde{W},\tilde{\tilde{W}}}(\pi u_i) \\
&= \frac{1}{n!}\sum_{i=1}^N \sum_{v \in \mathcal{T}_P^n} \left|\{\pi \in \mathcal{S}_n \mid \pi u_i = v\}\right| \cdot g_{\tilde{W},\tilde{\tilde{W}}}(v) \\
&= \frac{1}{n!}\prod_{x \in \mathcal{X}} (nP(x))! \sum_{i=1}^N \sum_{v \in \mathcal{T}_P^n} g_{\tilde{W},\tilde{\tilde{W}}}(v) \qquad (16) \\
&= N \cdot Eg_{\tilde{W},\tilde{\tilde{W}}}(U). \qquad (17)
\end{aligned}$$

Equation (17) follows from (4). Thus part a) of the lemma is proved. Part b) is an application of Chebychev's inequality.

Proof of the Main Lemma: Lemma 2 guarantees the existence of a permutation $\pi \in \mathcal{S}_n$ with

$$g^*_{\tilde{W},\tilde{\tilde{W}}}(\pi),\ g^*_{\tilde{W},\tilde{\tilde{W}}}(\pi^{-1}) \leq N \cdot \exp\left\{n\left(H(\tilde{W} \mid P) - \left[I(P,\tilde{\tilde{W}}) - R\right]^{+} + \frac{3}{4}\delta\right)\right\}$$

for any pair $\tilde{W},\tilde{\tilde{W}} \in \mathcal{W}$. π^{-1} denotes the inverse (18) permutation of π.

Let $\mathcal{C} = \{(u_i,\mathcal{D}_i) \mid i = 1,\cdots,N\}$ be a code for the given codeword set $\{u_1,\cdots,u_N\} \subset \mathcal{T}_P^n$. Define new decoding sets

$$\mathcal{E}_i = \mathcal{D}_i - \{y^n \mid I(\pi u_j \wedge y^n) \geq I(u_i \wedge y^n) \text{ for some } j\}$$

for $i = 1,\cdots,N$ and

$$\mathcal{E}_{i,\pi} = \pi\mathcal{D}_i - \{y^n \mid I(u_j \wedge y^n) > I(\pi u_i \wedge y^n) \text{ for some } j\}$$

for $i = 1,\cdots,N$.

Notice that the sets $\mathcal{E}_1,\cdots,\mathcal{E}_N,\mathcal{E}_{1,\pi},\cdots,\mathcal{E}_{N,\pi}$ are disjoint and set

$$\mathcal{C}' = \{(u_1,\mathcal{E}_1),\cdots,(u_N,\mathcal{E}_N),(\pi u_1,\mathcal{E}_{1,\pi}),\cdots,(\pi u_N,\mathcal{E}_{N,\pi})\}.$$

Now we have for every $W \in \mathcal{W}$

$$\bar{\lambda}(\mathcal{C}',W) = \frac{1}{2N}\left(\sum_{i=1}^N (W^n(\mathcal{D}_i - \mathcal{E}_i \mid u_i) + W^n(\pi\mathcal{D}_i - \mathcal{E}_{i,\pi} \mid \pi u_i)) + 2N\bar{\lambda}(\mathcal{C},W)\right). \qquad (19)$$

First we estimate

$$\begin{aligned}
\sum_{i=1}^N W^n(\mathcal{D}_i - \mathcal{E}_i \mid u_i) &= \sum_{i=1}^N W^n\left(\{y^n \mid I(\pi u_j \wedge y^n) \geq I(u_i \wedge y^n) \text{ for some } j\} \mid u_i\right) = \sum_{\substack{\tilde{W},\tilde{\tilde{W}} \in \mathcal{W}_n(P) \\ I(P,\tilde{W}) \leq I(P,\tilde{\tilde{W}})}} \sum_{i=1}^N W^n\left(\mathcal{T}_{\tilde{W}}^n(u_i)\right. \\
&\quad \left. \cap \bigcup_{j=1}^N \mathcal{T}_{\tilde{\tilde{W}}}^n(\pi u_j) \mid u_i\right) = \sum_{\substack{\tilde{W},\tilde{\tilde{W}} \in \mathcal{W}_n(P) \\ I(P,\tilde{W}) \leq I(P,\tilde{\tilde{W}})}} \sum_{i=1}^N W^n\left(\mathcal{T}_{\tilde{W}}^n(\pi^{-1}u_i) \cap \bigcup_{j=1}^N \mathcal{T}_{\tilde{\tilde{W}}}^n(u_j) \mid \pi^{-1}u_i\right) \\
&= \sum_{\substack{\tilde{W},\tilde{\tilde{W}} \in \mathcal{W}_n(P) \\ I(P,\tilde{W}) \leq I(P,\tilde{\tilde{W}})}} \exp\left\{-n\left(D(\tilde{W} \| W \mid P) + H(\tilde{W} \mid P)\right)\right\} g^*_{\tilde{W},\tilde{\tilde{W}}}(\pi^{-1})
\end{aligned}$$

by (8). Now apply (18) to get

$$\sum_{i=1}^{N} \dot{W}^n(\mathcal{D}_i - \mathcal{E}_i \mid u_i)$$

$$\le \sum_{\substack{\tilde{W}, \tilde{\tilde{W}} \in \mathcal{W}_n(P) \\ I(P, \tilde{W}) \le I(P, \tilde{\tilde{W}})}} N \cdot \exp\Big\{-n\Big(D(\tilde{W} \| W \mid P) + [I(P, \tilde{W}) - R]^+ - \frac{3}{4}\delta\Big)\Big\} \tag{20}$$

$$\le N \exp\{-n(E_r(R, P, W) - \delta)\}, \quad \text{for } n \ge n_0(|\mathcal{X}|, |\mathcal{Y}|, \delta). \tag{21}$$

In (20) we have used $[I(P, \tilde{W}) - R]^+ \le [I(P, \tilde{\tilde{W}}) - R]^+$. In the same way,

$$\sum_{i=1}^{N} W^n(\pi\mathcal{D}_i - \mathcal{E}_{i,\pi} \mid \pi u_i) \le \sum_{\substack{\tilde{W}, \tilde{\tilde{W}} \in \mathcal{W}_n(P) \\ I(P, \tilde{W}) \le I(P, \tilde{\tilde{W}})}} \sum_{i=1}^{N} W^n\Big(\mathcal{T}^n_{\tilde{W}}(\pi u_i) \cap \bigcup_{j=1}^{N} \mathcal{T}^n_{\tilde{\tilde{W}}}(u_j) \mid \pi u_i\Big)$$

$$= \sum_{\substack{\tilde{W}, \tilde{\tilde{W}} \in \mathcal{W}_n(P) \\ I(P, \tilde{W}) \le I(P, \tilde{\tilde{W}})}} \exp\{-n(D(\tilde{W} \| W \mid P) + H(\tilde{W} \mid (P)))\}\, g^*_{\tilde{W}, \tilde{\tilde{W}}}(\pi)$$

$$\le N \exp\{-n(E_r(R, P, W) - \delta)\}, \qquad \text{for } n \ge n_0(|\mathcal{X}|, |\mathcal{Y}|, \delta). \tag{22}$$

The result now follows from (21), (22), and (19).

The second part of the claim of the Main Lemma follows directly from Lemma 2 to b) and the argument given in this proof.

Proof of the Main Theorem: The Main Lemma states that if one chooses the permutation π randomly according to the equidistribution on $\mathcal{S}_n$, then the probability is at most $\exp\{-(\delta/2)n\}$ that (11) cannot be fulfilled. Since the number of types in $\mathcal{P}_n$ the number of different alphabets $\mathcal{X}, \mathcal{Y}$ with $|\mathcal{X}|, |\mathcal{Y}| \le 2^k$ is "exponentially small" it is clear that one can obtain the Main Theorem immediately from the Main Lemma.

Appendix I

Correlated Source Codes Produced by Permutations from Ordinary Channel Codes

Gallager [7] and Košelev [8] have derived a "random coding" error exponent for discrete memoryless correlated sources (DMCS's) $\{(X_t, Y_t)\}_{t=1}^{\infty}$ in case the decoder is informed about the outputs of one of the sources. Csiszár and Körner [9] recently improved those results by establishing what they considered to be the counterpart of the expurgated bound in source coding. Our results below confirm this view. In [5, part II, section 8], it is shown that their result can also be derived via a hypergraph coloring lemma, which slightly generalizes [23, lemma 4]. In [5, section 6, part II], Ahlswede showed that the Slepian–Wolf source coding theorem can easily be derived from the coding theorem for the discrete memoryless channel via the following lemma.

Covering Lemma: Fix n and $P \in \mathcal{P}_n$ and let $\mathcal{A} \subset \mathcal{T}^n_P$. Then there exist permutations $\pi_1, \cdots, \pi_k \in \mathcal{S}_n$ such that

$$\bigcup_{i=1}^{k} \pi_i \mathcal{A} = \mathcal{T}^n_P,$$

if $k > |\mathcal{A}|^{-1} \cdot |\mathcal{T}^n_P| \log |\mathcal{T}^n_P|$. (Here $\pi\mathcal{A} = \{\pi x^n \mid x^n \in \mathcal{A}\}$ and $\pi x^n = (x_{\pi 1}, \cdots, x_{\pi n})$ for $x^n = (x_1, \cdots, x_n) \in \mathcal{X}^n$.)

Here we first show that by this approach every upper bound on the error probability for the DMC yields immediately an upper bound on the error probability for the DMCS, with one source known to the decoder. This way the random coding bound is transformed into the bound derived by Gallager and Košelev, and the expurgated bound is transformed into the bound formed by Csiszár and Körner.

Next we show that, conversely, every lower bound on the error probabilities for the DMC yields also a lower bound on the error probabilities for the DMCS. *Theorem 1 below shows the intimate connection between the source and channel reliability functions. We now give the exact statements.*

For the DMCS $\{(X_t, Y_t)\}_{t=1}^{\infty}$ we consider the communication situation "source coding with (full) side information," that is, an encoder observes the X-source and he has the task to encode this source reliably for a decoder who can observe the Y-source. An n-length block code (f, F) for this problem consists of an encoding function f: $\mathcal{X}^n \to \mathcal{Z}$, where $\mathcal{Z}$ is the range of f, and of a decoding function F: $\mathcal{Z} \times \mathcal{Y}^n \to \mathcal{X}^n$. If $x^n \in \mathcal{X}^n$ is observed by the encoder, he gives $f(x^n)$ to the decoder. Having observed the side information y^n the decoder votes for $F(f(x^n), y^n) \in \mathcal{X}^n$ as being the output of the X-source. The (average) error probability of this code $\lambda(f, F)$ is given by

$$\lambda(f, F) = \sum_{x^n \in \mathcal{X}^n} \sum_{y^n \in \mathcal{Y}^n} Q^n(x^n, y^n)\eta(x^n, F(f(x^n), y^n)),$$

where

$$Q^n(x^n, y^n) = \prod_{t=1}^{n} Q(x_t, y_t), Q(x, y) = \Pr\{X = x, Y = y\},$$

and

$$\eta(x^n, x'^n) = \begin{cases} 1, & \text{for } x^n \ne x'^n \\ 0, & \text{for } x^n = x'^n. \end{cases}$$

For $R > 0$ define $\lambda(n, R) = \min \lambda(f, F)$, where the minimum is taken over all n-length block codes (f, F) satisfying $\|f\| \le$

$\exp\{nR\}$. We are interested in the reliability curve

$$e(R) = -\limsup_{n\to\infty} \frac{1}{n}\log \lambda(n, R)$$

for any rate $R > H(X \mid Y)$.

The joint distribution $Q(x, y)$ of (X, Y) induces a channel W, given by

$$W(y \mid x) := \Pr(Y = y \mid X = x), \qquad \text{for } x \in \mathcal{X},\ y \in \mathcal{Y}.$$

For any type $P \in \mathcal{P}_n$ define

$$\bar{\lambda}(n, R, P, W) = \min \bar{\lambda}(\mathcal{C}, W),$$

where the minimum is taken over all n-length block codes for W with codewords from $\mathcal{T}_P^n$ and rate at least R. We denote the distribution of X by Q_1. We establish the following connection between $\lambda(n, R)$ and the numbers $\bar{\lambda}(n, R, P, W)$.

Theorem 1: For any $\delta > 0$ and $n \geq n_0(|\mathcal{X}|, |\mathcal{Y}|, \delta)$,

a) $$-\frac{1}{n}\log \lambda(n, R+\delta) \geq \min_{P\in\mathcal{P}_n}\left[D(P\|Q_1) - \frac{1}{n}\log \bar{\lambda}(n, H(P) - R, P, W)\right] - \delta,$$

b) $$-\frac{1}{n}\log \lambda(n, R) \leq \min_{P\in\mathcal{P}_n}\left[D(P\|Q_1) - \frac{1}{n}\log \bar{\lambda}(n, H(P) - R - \delta, P, W)\right] + \delta.$$

In order to get estimates on $e(R)$ we can therefore use the familiar estimates on $\bar{\lambda}(n, H(P) - R, P, W)$ and thus obtain the following corollary.

Corollary:

$$e(R) \geq \min_{P\in\mathcal{P}}\left[D(P\|Q_1) + E_r(H(P) - R, P, W)\right], \tag{23}$$

$$e(R) \geq \min_{P\in\mathcal{P}}\left[D(P\|Q_1) + E_{\text{ex}}(H(P) - R, P, W)\right], \tag{24}$$

$$e(R) \leq \min_{P\in\mathcal{P}}\left[D(P\|Q_1) + E_{\text{sp}}(H(P) - R, P, W)\right], \tag{25}$$

where

$$E_{\text{sp}}(R, P, W) = \min_{\substack{\tilde{W}\in\mathcal{W} \\ I(P,\tilde{W})\geq R}} D(\tilde{W}\|W \mid P).$$

Remark 2: Equations (23) and (25) were obtained in a different form via Chernov bounds by Gallager [7] and Košelev [8]. Equation (24) was proved by Csiszár and Körner [9]. In the present form (23) can be found in [9] and (25) in [24].

VI. Proof of Theorem 1

a) Fix $R > 0$, $\delta > 0$, and $n \geq n_0(|\mathcal{X}|, |\mathcal{Y}|, \delta)$, $P \in \mathcal{P}_n$. Recall the definition of $\bar{\lambda}(n, R, P, W)$ and note that any (n, N) code $\mathcal{C} = \{(u_i, \mathcal{D}_i) \mid i = 1, \cdots, N\}$ for W contains at least $N/2$ codewords u_i such that $W^n(\mathcal{D}_i^c \mid u_i) \leq 2\bar{\lambda}(\mathcal{C}, W)$.

We conclude that for any fixed $P \in \mathcal{P}_n$ there is an (n, N_p) code $\mathcal{C}_P = \{(u_i^P, \mathcal{D}_i^P) \mid i = 1, \cdots, N_P\}$ for the induced channel W such that

$$\mathcal{U}_P = \left\{u_1^P, \cdots, u_{N_P}^P\right\} \subset \mathcal{T}_P^n$$

and

$$N_P \geq \tfrac{1}{2}\exp\{n(H(P) - R)\}, \tag{26}$$

and

$$\lambda_{\max}(P, W) \leq 2\bar{\lambda}(n, H(P) - R, P, W). \tag{27}$$

(It is important here to have a good *maximal* error code.) From these "best" channel codes, constructed for every $P \in \mathcal{P}_n$, we form a source code as follows.

By the covering lemma there exist permutations $\pi_1^P, \cdots, \pi_{k_P}^P \in \mathcal{S}_n$ such that

$$\bigcup_{i=1}^{k_P} \pi_i^P \mathcal{U}_P = \mathcal{T}_P^n, \tag{28}$$

and

$$k_P = \left\lceil N_P^{-1} \cdot |\mathcal{T}_P^n| \log |\mathcal{T}_P^n| \right\rceil. \tag{29}$$

For every $P \in \mathcal{P}_n$ we partition the set $\mathcal{T}_P^n$ into the sets

$$\mathcal{A}_{i,P} = \pi_i^P \mathcal{U}_P - \bigcup_{j=1}^{i-1} \pi_j^P \mathcal{U}_P.$$

We now define an n-length block code (f, F): for every $x^n \in \mathcal{X}^n$ set

$$f(x^n) = (i, P_{x^n}), \qquad \text{if } x^n \in \mathcal{A}_{i, P_{x^n}}, \tag{30}$$

and for every $P \in \mathcal{P}_n$, $i \in \{1, \cdots, k_P\}$, and $y^n \in \mathcal{Y}^n$ set

$$F(i, P, y^n) = \pi_i^P u_j^P, \qquad \text{if } y^n \in \pi_i^P \mathcal{D}_j^P. \tag{31}$$

Next we compute the rate and the error probability of the source code (f, F).

$$\begin{aligned}\|f\| &\leq |\mathcal{P}_n| \cdot \max_{P\in\mathcal{P}_n} k_P \\ &\leq (n+1)^{|\mathcal{X}|} \cdot \max_{P\in\mathcal{P}_n} \big[2\exp\{-n(H(P) - R) + nH(P)\} \\ &\qquad \cdot n \cdot H(P) + 1\big] \\ &\leq \exp\{n(R+\delta)\}\end{aligned}$$

for $n \geq n_0(|\mathcal{X}|, |\mathcal{Y}|, \delta)$, where the steps are justified by (26), (29), (1), and (5). Further,

$$\begin{aligned}\lambda(f, F) &= \sum_{x^n, y^n} Q^n(x^n, y^n) \cdot \eta(x^n, F(f(x^n), y^n)) \\ &= \sum_{P\in\mathcal{P}_n} \sum_{i=1}^{k_P} \sum_{x^n\in\mathcal{A}_{i,P}} \sum_{y^n} Q_1(x^n) \cdot W^n(y^n \mid x^n) \\ &\qquad \cdot \eta(x^n, F(f(x^n), y^n)) \\ &= \sum_{P\in\mathcal{P}_n} \exp\{-n(D(P\|Q_1) + H(P))\} \\ &\qquad \cdot \sum_{i=1}^{k_P} \sum_{\pi_i^P u_j^P \in \mathcal{A}_{i,P}} W^n\left(\left(\pi_i^P \mathcal{D}_j^P\right)^c \mid \pi_i^P u_j^P\right) \\ &\leq 2\sum_{P\in\mathcal{P}_n} \exp\{-n(D(P\|Q_1) + H(P))\} \\ &\qquad \cdot |\mathcal{T}_P^n| \cdot \bar{\lambda}(n, H(P) - R, P, W)\end{aligned}$$

(by (27) and (7)).

Hence, by (5) and (1)

$$\begin{aligned}\lambda(f, F) &\leq 2\sum_{P\in\mathcal{P}_n} \exp\{-nD(P\|Q_1)\} \cdot \bar{\lambda}(n, H(P) - R, P, W) \\ &\leq 2(n+1)^{|\mathcal{X}|} \max_{P\in\mathcal{P}_n} \big[\exp\{-nD(P\|Q_1)\} \\ &\qquad \cdot \bar{\lambda}(n, H(P) - R, P, W)\big],\end{aligned}$$

and Theorem 1a) follows.

b) Let any code (f, F) of block length n be given. Let $\mathcal{Z} = \{z_1, \cdots, z_{\|f\|}\}$ be the range of f. For any $z \in \mathcal{Z}$ and every type

$P \in \mathcal{P}_n$ define

$$\mathcal{A}_{P,z} = \{x^n \mid P_{x^n} = P, f(x^n) = z\}.$$

For fixed $P \in \mathcal{P}_n$, $z \in \mathcal{Z}$, $x^n \in \mathcal{A}_{P,z}$ define

$$\mathcal{D}_{x^n,P,z} = \{y^n \mid F(z, y^n) = x^n\}.$$

We now consider for any $P \in \mathcal{P}_n$, $z \in \mathcal{Z}$ the system

$$\mathcal{C}_{P,z} = \{(x^n, \mathcal{D}_{x^n,P,z}) \mid x^n \in \mathcal{A}_{P,z}\} \tag{32}$$

as a code for the induced channel W. Clearly, $|\mathcal{T}_P^n|/2$ sequences in $\mathcal{T}_P^n$ are contained in sets $\mathcal{A}_{P,z}$ satisfying

$$|\mathcal{A}_{P,z}| \geq \tfrac{1}{2}|\mathcal{T}_P^n| \cdot \|f\|^{-1}. \tag{33}$$

For any $P \in \mathcal{P}_n$ let $\mathcal{Z}'(P)$ be the set of those elements in $\mathcal{Z}$ which satisfy (33). We analyze now the relation between $\lambda(f, F)$ and the error probabilities of the codes in (32). We get

$$\begin{aligned}\lambda(f, F) &= \sum_{x^n}\sum_{y^n} Q^n(x^n, y^n)\eta(x^n, F(f(x^n), y^n))\\ &= \sum_{P\in\mathcal{P}_n}\sum_{z\in\mathcal{Z}}\sum_{x^n\in\mathcal{A}_{P,z}}\sum_{y^n} Q_1^n(x^n)\\ &\quad \cdot W^n(y^n \mid x^n)\eta(x^n, F(f(x^n), y^n))\\ &= \sum_{P\in\mathcal{P}_n} \exp\{-n(D(P\|Q_1) + H(P))\}\\ &\quad \cdot \sum_{z\in\mathcal{Z}}\sum_{x^n\in\mathcal{A}_{P,z}} W^n(\mathcal{D}^c_{x^n,P,z} \mid x^n)\\ &\geq \max_{P\in\mathcal{P}_n} \exp\{-n(D(P\|Q_1) + H(P))\}\\ &\quad \cdot \sum_{z\in Z'(P)}\sum_{x^n\in\mathcal{A}_{P,z}} W^n(\mathcal{D}^c_{x^n,P,z} \mid x^n),\end{aligned}$$

where we have applied (7). Furthermore,

$$\begin{aligned}\sum_{x^n\in\mathcal{A}_{P,z}} W^n(\mathcal{D}^c_{x^n,P,z} \mid x^n) &= |\mathcal{A}_{P,z}|\bar\lambda(\mathcal{C}_{P,z}, W)\\ &\geq |\mathcal{A}_{P,z}|\bar\lambda\left(n, \frac{1}{n}\log|\mathcal{A}_{P,z}|, P, W\right).\end{aligned}$$

Now use (33) to obtain

$$\sum_{z\in Z'(P)} |\mathcal{A}_{P,z}| \geq \tfrac{1}{2}|\mathcal{T}_P^n|$$

and continue again by using (33) and (5) to obtain

$$\begin{aligned}\lambda(f, F) &\geq \max_{P\in\mathcal{P}_n} \exp\{-n(D(P\|Q_1) + H(P))\}\\ &\quad \cdot \frac{1}{2}|\mathcal{T}_P^n|\bar\lambda\left(n, \frac{1}{n}\log\left(\frac{1}{2}\cdot|\mathcal{T}_P^n|\cdot\|f\|^{-1}\right), P, W\right)\\ &\geq \max_{P\in\mathcal{P}_n} \exp\{-nD(P\|Q_2)\\ &\quad + \log\bar\lambda(n, H(P) - R - \delta, P, W) - n\delta\}\end{aligned}$$

for $n \geq n_0(|\mathcal{X}|, |\mathcal{Y}|, \delta)$.

Theorem 1 is proved.

Appendix II

An Iterative Code Construction Achieving the Random Coding and the Expurgated Bound

Theorem 2: For any $R > 0$, $\delta > 0$, $n \geq n_0(|\mathcal{X}|, |\mathcal{Y}|, \delta)$ and every type $P \in \mathcal{P}_n$ the following is true.

a) Let $\mathcal{C} = \{(u_i, \mathcal{D}_i) \mid i = 1,\cdots,N\}$ be an (n, N) code such that $(1/n)\log N \leq R$ and $u_i \in \mathcal{T}_P^n$ for $i = 1,\cdots,N$. Then there exist an n-sequence $u_{N+1} \in \mathcal{T}_P^n$ and proper decoding sets $\mathcal{E}_1,\cdots,\mathcal{E}_{N+1}$ such that the enlarged $(n, N+1)$ code

$$\mathcal{C}' = \{(u_i, \mathcal{E}_i) \mid i = 1,\cdots,N+1\}$$

satisfies for any channel $W \in \mathcal{W}$ the inequality

$$\begin{aligned}\bar\lambda(\mathcal{C}', W) \leq \frac{1}{N+1}\big(N\cdot\bar\lambda(\mathcal{C}, W)\\ +2\exp\{-n(E_r(R, P, W) - \delta)\}\big).\end{aligned} \tag{34}$$

In particular, if $\bar\lambda(\mathcal{C}, W)$ is less than 2 exp $\{-n(E_r(R, P, W) - \delta)\}$, then also $\bar\lambda(\mathcal{C}', W)$ is smaller than this quantity.

b) Furthermore, if we prolong the (n, N) code $\mathcal{C}$ to $\mathcal{C}'$ by choosing u_{N+1} at random according to the equidistribution on $\mathcal{T}_P^n$, then the probability of selecting an u_{N+1} for which

$$\begin{aligned}\bar\lambda(\mathcal{C}', W) \leq \frac{1}{N+1}\big(N\cdot\bar\lambda(\mathcal{C}, W)\\ +2\exp\{-n(E_r(R+\xi, P, W) - \delta)\}\big)\end{aligned} \tag{35}$$

holds for any $W \in \mathcal{W}$ is larger than

$$1 - \exp\left\{-n\left(\frac{\delta}{2} + \xi\right)\right\}.$$

Theorem 3: For any $R > 0$, $n \geq n_0(|\mathcal{X}|, |\mathcal{Y}|, \delta)$ and every type $P \in \mathcal{P}_n$ the following is true.

a) Let $u_1,\cdots,u_N \in \mathcal{T}_P^n$ arbitrary n-sequences, $N \leq \exp\{nR\}$. For every $W \in \mathcal{W}$ let $\mathcal{C}^W = \{(u_i, \mathcal{D}_i^W) \mid i = 1,\cdots,N\}$ be the ML code with respect to W to the codewords $u_1,\cdots,u_N$. Then there exists an n-sequence $u_{N+1} \in \mathcal{T}_P^n$ such that for every $W \in \mathcal{W}$ the ML code with respect to W satisfies

$$\begin{aligned}\bar\lambda(\mathcal{C}'^W, W) \leq (1/(N+1))\big(N\cdot\bar\lambda(\mathcal{C}^W, W)\\ +2\exp\{-nE_{\mathrm{ex}}(R+\delta, P, W)\}\big).\end{aligned} \tag{36}$$

Again, if $\bar\lambda(\mathcal{C}^W, W)$ is less than 2 $\exp\{-nE_{\mathrm{ex}}(R + \delta, P, W)\}$, then also $\bar\lambda(\mathcal{C}'^W, W)$ is smaller than this quantity.

b) If the additional codeword u_{N+1} is chosen according to the equidistribution on $\mathcal{T}_P^n$, then the probability that (36) can be fulfilled is larger than $1 - \exp\{-(\delta/2)n\}$.

Remark 2: Since for $N = 1$ $\bar\lambda(\mathcal{C}, W) \leq 2$ $\exp\{-n(E_r(R, P, W) - \delta)\}$ and $\bar\lambda(\mathcal{C}^W, W) \leq 2$ $\exp\{-nE_{\mathrm{ex}}(R + \delta, P, W) - \delta)\}$ are obviously achievable, Theorem R (resp. Theorem EX) are immediate consequences of Theorem 2 (Theorem 3).

Proof of Theorem 2: Suppose we are given $\delta > 0$ and an (n, N) code

$$\mathcal{C} = \{(u_i, \mathcal{D}_i) \mid i = 1,\cdots,N\},$$

where $(1/n)\log N \leq R$ and $u_i \in \mathcal{T}_P^n$ for $i = 1,\cdots,N$. By Lemma 1a) there exists a $u_{N+1} \in \mathcal{T}_P^n$ such that

$$g_{\tilde W, \tilde{\tilde W}}(u_{N+1}) \leq \exp\left\{n\left(H(\tilde W \mid P) - [I(P, \tilde{\tilde W}) - R]^+ + \frac{3}{4}\delta\right)\right\} \tag{37}$$

holds for any pair $\tilde W, \tilde{\tilde W} \in \mathcal{W}$. We show that with such a choice of u_{N+1} (34) can be fulfilled, so that Theorem 2a) will follow. It is clear that then Theorem 2b) follows directly from this proof and from Lemma 1b).

First we define new decoding sets

$$\mathcal{E}_i = \mathcal{D}_i - \{y^n \mid I(u_{N+1} \wedge y^n) > I(u_i \wedge y^n)\}, \qquad \text{for } i \in \{1,\cdots,N\}$$

and

$$\mathcal{E}_{N+1} = \{y^n \mid I(u_{N+1} \wedge y^n) > I(u_i \wedge y^n), \qquad \text{for all } i \in \{1,\cdots,N\}\}.$$

Obviously, the $\mathcal{E}_i$ are disjoint subsets of $\mathcal{Y}^n$. Set $\mathcal{C}' = \{(u_i, \mathcal{E}_i) \mid i = 1,\cdots,N+1\}$. For these codes, $\mathcal{C}$ and $\mathcal{C}'$, we show (34). We estimate $\bar{\lambda}(C', W)$ for any $W \in \mathcal{W}$. Now

$$\begin{aligned}\bar{\lambda}(\mathcal{C}', W) &= \frac{1}{N+1} \sum_{i=1}^{N+1} W^n(\mathcal{E}_i^c \mid u_i) \\ &= \frac{1}{N+1}\left(\sum_{i=1}^{N} \left(W^n(\mathcal{D}_i^c \mid u_i) + W^n(\mathcal{D}_i - \mathcal{E}_i \mid u_i)\right) + W^n(\mathcal{E}_{N+1}^c \mid u_{N+1})\right) \\ &= \frac{1}{N+1}\left(N \cdot \bar{\lambda}(\mathcal{C}, W) + \sum_{i=1}^{N} W^n(\mathcal{D}_i - \mathcal{E}_i \mid u_i) + W^n(\mathcal{E}_{N+1}^c \mid u_{N+1})\right). \end{aligned} \tag{38}$$

First we bound the error probability of u_{N+1} from above.

$$\begin{aligned} W^n(\mathcal{E}_{N+1}^c \mid u_{N+1}) &= W^n(\{y^n \mid I(u_{N+1} \wedge y^n) \le I(u_i \wedge y^n), \text{ for some } 1 \le i \le N\} \mid u_{N+1}) \\ &= \sum_{\substack{\tilde{W}, \tilde{\tilde{W}} \in \mathcal{W}_n(P) \\ I(P,\tilde{W}) \le I(P,\tilde{\tilde{W}})}} W^n\left(\mathcal{T}^n_{\tilde{W}}(u_{N+1}) \cap \bigcup_{i=1}^{N} \mathcal{T}^n_{\tilde{\tilde{W}}}(u_i) \mid u_{N+1}\right) \\ &= \sum_{\substack{\tilde{W}, \tilde{\tilde{W}} \in \mathcal{W}_n(P) \\ I(P,\tilde{W}) \le I(P,\tilde{\tilde{W}})}} g_{\tilde{W},\tilde{\tilde{W}}}(u_{n+1}) \cdot \exp\left\{-n\left(D(\tilde{W} \| W \mid P) + H(\tilde{W} \mid P)\right)\right\} \end{aligned}$$

by (8) and the definition of $g_{\tilde{W},\tilde{\tilde{W}}}$. Observing that $I(P,\tilde{W}) \le I(P,\tilde{\tilde{W}})$ implies $[I(P,\tilde{W}) - R]^+ \le [I(P,\tilde{\tilde{W}}) - R]^+$ we obtain with (37)

$$\begin{aligned} W^n(\mathcal{E}_{N+1}^c \mid u_{N+1}) &\le \sum_{\substack{\tilde{W}, \tilde{\tilde{W}} \in \mathcal{W}_n(P) \\ I(P,\tilde{W}) \le I(P,\tilde{\tilde{W}})}} \exp\left\{-n\left(D(\tilde{W} \| W \mid P) + [I(P,\tilde{W}) - R]^+ - \frac{3}{4}\delta\right)\right\} \\ &\le |\mathcal{W}_n(P)|^2 \cdot \max_{\tilde{W} \in \mathcal{W}} \exp\left\{-n\left(D(\tilde{W} \| W \mid P) + [I(P,\tilde{W}) - R]^+ - \frac{3}{4}\delta\right)\right\} \qquad (39) \\ &\le \exp\{-n(E_r(R,P,W) - \delta)\} \qquad (40) \end{aligned}$$

for $n \ge n_0(\delta, |\mathcal{X}|, |\mathcal{Y}|)$, because of (2). Further

$$\begin{aligned} \sum_{i=1}^{N} W^n(\mathcal{D}_i - \mathcal{E}_i \mid u_i) &= \sum_{i=1}^{N} W^n(\{y^n \in \mathcal{D}_i \mid I(u_{N+1} \wedge y^n) > I(u_i \wedge y^n)\} \mid u_i) \\ &= \sum_{\substack{\tilde{W}, \tilde{\tilde{W}} \in \mathcal{W}_n(P) \\ I(P,\tilde{W}) < I(P,\tilde{\tilde{W}})}} \sum_{i=1}^{N} W^n\left(\mathcal{D}_i \cap \mathcal{T}^n_{\tilde{W}}(u_i) \cap \mathcal{T}^n_{\tilde{\tilde{W}}}(u_{N+1}) \mid u_i\right). \end{aligned} \tag{41}$$

By (8),

$$\begin{aligned} W^n\left(\mathcal{D}_i \cap \mathcal{T}^n_{\tilde{W}}(u_i) \cap \mathcal{T}^n_{\tilde{\tilde{W}}}(u_{N+1}) \mid u_i\right) &= \exp\left\{-n\left(D(\tilde{W} \| W \mid P) + H(\tilde{W} \mid P)\right)\right\} \\ &\quad \cdot \left|\mathcal{D}_i \cap \mathcal{T}^n_{\tilde{W}}(u_i) \cap \mathcal{T}^n_{\tilde{\tilde{W}}}(u_{N+1})\right|. \end{aligned} \tag{42}$$

Since the sets $\mathcal{D}_i$ are disjoint we get

$$\begin{aligned} \sum_{i=1}^{N} \left|\mathcal{D}_i \cap \mathcal{T}^n_{\tilde{W}}(u_i) \cap \mathcal{T}^n_{\tilde{\tilde{W}}}(u_{N+1})\right| &\le \left|\mathcal{T}^n_{\tilde{\tilde{W}}}(u_{N+1}) \cap \bigcup_{i=1}^{N} \mathcal{T}^n_{\tilde{W}}(u_i)\right| = g_{\tilde{\tilde{W}},\tilde{W}}(u_{N+1}). \end{aligned} \tag{43}$$

Combining (41), (42), and (43) we obtain as before with (38)

$$\begin{aligned} \sum_{i=1}^{N} W^n(\mathcal{D}_i - \mathcal{E}_i \mid u_i) &\le \sum_{\substack{\tilde{W}, \tilde{\tilde{W}} \in \mathcal{W}_n(P) \\ I(P,\tilde{W}) \le I(P,\tilde{\tilde{W}}),\, P\tilde{W} = P\tilde{\tilde{W}}}} \exp\left\{-n\left(D(\tilde{W} \| W \mid P) + H(\tilde{W} \mid P)\right)\right\} \\ &\quad \cdot \exp\left\{n\left(H(\tilde{\tilde{W}} \mid P) - [I(P,\tilde{W}) - R]^+ + \frac{3}{4}\delta\right)\right\}. \end{aligned}$$

Since $I(P,\tilde{W}) \le I(P,\tilde{\tilde{W}})$ and $P\tilde{W} = P\tilde{\tilde{W}}$ (by assumption) imply $H(\tilde{\tilde{W}} \mid P) \le H(\tilde{W} \mid P)$, we conclude (as previously for (39) and (40)) that

$$\sum_{i=1}^{N} W^n(\mathcal{D}_i - \mathcal{E}_i \mid u_i) \le \exp\{-n(E_r(R,P,W) - \delta)\}, \tag{44}$$

for $n \ge n_0(|\mathcal{X}|, |\mathcal{Y}|, \delta)$.

Theorem 2 is proved.

For the proof of Theorem 3 we shall need an auxiliary result, which is an analogy to Lemma 1 for the expurgated bound.

Fix $R > 0$, $\delta > 0$, $P \in \mathcal{P}_n$, and let $\{u_1,\cdots,u_N\} \subset \mathcal{T}_P^n$, $N \le \exp\{nR\}$ be given.

For any $V \in \mathcal{V}$ we define the function f_V on $\mathcal{X}^n$ by

$$f_V(u) = |\{i \mid u \in \mathcal{T}_V^n(u_i)\}|, \qquad \text{for } u \in \mathcal{X}^n. \tag{45}$$

$f_V(u)$ measures the V-relationship of u to the given code word system $\{u_1,\cdots,u_N\}$. Note that $f_V(u) = 0$ if $V \notin \mathcal{V}_n(P)$, because in this case $\mathcal{T}_V^n(u_i) = \varnothing$ for all $i = 1,\cdots,N$.

Lemma 3: Let U be a random variable equidistributed in $\mathcal{T}_P^n$. Then for any $V \in \mathcal{V}$

a) $Ef_V(U) \le (n+1)^{|\mathcal{X}|} \cdot \exp\{n(R - I(P,V))\}$,

b) $\Pr\{f_V(U) \ge \exp\{n(R - I(P,V) + 3/4\delta)\}$, for some $V \in \mathcal{V}\} \le \exp\{-n(\delta/2)\}$.

Proof:

$$\begin{aligned}
&Ef_V(U)\\
&\quad= \sum_{u\in\mathcal{T}_P^n} \Pr(U=u)f_V(u)\\
&\quad\le \exp\{-nH(P)\}\cdot(n+1)^{|\mathcal{X}|}\sum_{u\in\mathcal{T}_P^n} f_V(u)\\
&\quad= \exp\{-nH(P)\}\cdot(n+1)^{|\mathcal{X}|}\sum_{i=1}^{N}\left|\{u\in\mathcal{T}_P^n \mid u\in\mathcal{T}_V^n(u_i)\}\right| \qquad (46)\\
&\quad= \exp\{-nH(P)\}\cdot(n+1)^{|\mathcal{X}|}\sum_{i=1}^{N}\left|\mathcal{T}_V^n(u_i)\right|\\
&\quad\le N\cdot\exp\{n(H(V\mid P)-H(P))\}\cdot(n+1)^{|\mathcal{X}|} \qquad (47)\\
&\quad\le (n+1)^{|\mathcal{X}|}\exp\{n(R+H(V\mid P)-H(P))\}.
\end{aligned}$$

The first inequality follows from (5) and the fact that U is equidistributed. (46) is obtained by counting and (47) is a consequence of (6).

Now let PV be the distribution on $\mathcal{X}$ given by

$$PV(\tilde{x}) = \sum_{x} P(x)V(\tilde{x}\mid x), \qquad \text{for } \tilde{x}\in\mathcal{X}.$$

Then from the definition of f_V it is clear that for $u\in\mathcal{T}_P^n$ $f_V(u)=0$ if $PV\neq P$. Therefore, we can assume that $PV=P$. Then, however, $H(V\mid P)-H(P) = H(V\mid P)-H(PV) = -I(P,V)$. Hence, in any case

$$Ef_V(U)\le(n+1)^{|\mathcal{X}|}\exp\{n(R-I(P,V))\}.$$

Part b) follows by Chebychev's inequality.

Proof of Theorem 3: Let $\delta, R, P\in\mathcal{P}_n$, $u_1,\cdots,u_N\in\mathcal{T}_P^n$ be given. Then by Lemma 3 there exists a u_{N+1} satisfying

$$f_V(u_{N+1})\le(n+1)^{|\mathcal{X}|}\exp\{n(R-I(P,V))\} \qquad (48)$$

for any $V\in\mathcal{V}$.

For any $W\in\mathcal{W}$ we consider the ML codes $\mathcal{C}^W=\{(u_i,\mathcal{D}_i^W)\mid i=1,\cdots,N\}$ and $\mathcal{C}'^W=\{(u_i,\mathcal{E}_i^W)\mid i=1,\cdots,N+1\}$. We estimate for every $W\in\mathcal{W}$

$$\bar{\lambda}(\mathcal{C}'^W,W)=\frac{1}{N+1}\sum_{i=1}^{N+1}W^n\left((\mathcal{E}_i^W)^c\mid u_i\right). \qquad (49)$$

First we bound the error probability for u_{N+1}.

$$\begin{aligned}
&W^n\left((\mathcal{E}_{N+1}^W)^c\mid u_{N+1}\right)\\
&\quad\le\sum_{i=1}^{N}\ \sum_{y^n:\,W^n(y^n|u_i)>W^n(y^n|u_{N+1})} W^n(y^n\mid u_{N+1}) \qquad (50)\\
&\quad\le\sum_{i=1}^{N}\sum_{y^n\in\mathcal{Y}^n}\sqrt{W^n(y^n\mid u_i)\cdot W^n(y^n\mid u_{N+1})}.
\end{aligned}$$

Now recall the definition of the function d in Theorem EX and observe that

$$\sum_{y^n\in\mathcal{Y}^n}\sqrt{W^n(y^n\mid u_i)W^n(y^n\mid u_{N+1})}=\exp\{-nEd(X,\tilde{X})\}, \qquad (51)$$

where $\tilde{X}, X$ are random variables on $\mathcal{X}$ of joint distribution $P_{u_i,\,u_{N+1}}$.

Now we count how often every sum of the form (51) occurs in (50). We use (48). Note that in (48) $f_V(u_{N+1})$ is a positive integer so that $f_V(u_{N+1})=0$, if $R+(3/4)\delta<I(P,V)$. Hence, we get

$$\begin{aligned}
&W^n\left((\mathcal{E}_{N+1}^W)^c\mid u_{N+1}\right)\\
&\quad\le|\mathcal{V}_n|\cdot\exp\left\{-n\cdot\min_{\substack{I(X\wedge\tilde{X})\le R+(3/4)\delta\\ X,\tilde{X}\ P\text{-distributed}}}\left[Ed(X,\tilde{X})+I(X\wedge\tilde{X})-R-\frac{3}{4}\delta\right]\right\}\\
&\quad\le(n+1)^{|\mathcal{X}|\cdot|\mathcal{Y}|}\exp\left\{-nE_{\mathrm{ex}}\left(R+\frac{3}{4}\delta,P,W\right)\right\}\\
&\quad\le\exp\{-nE_{\mathrm{ex}}(R+\delta,P,W)\}, \qquad (52)
\end{aligned}$$

for $n\ge n_0(|\mathcal{X}|,|\mathcal{Y}|,\delta)$. Since the code $\mathcal{C}'^W$ is an enlarged version of $\mathcal{C}^W$ and since both $\mathcal{C}'^W$ and $\mathcal{C}^W$ are ML codes, obviously

$$\mathcal{E}_i^W\subset\mathcal{D}_i^W \quad\text{resp.}\quad (\mathcal{E}_i^W)^c\supset(\mathcal{D}_i^W)^c, \qquad \text{for } i=1,\cdots,N.$$

Therefore we can write for $i=1,\cdots,N$

$$W^n\left((\mathcal{E}_i^W)^c\mid u_i\right)=W^n\left((\mathcal{D}_i^W)^c\mid u_i\right)+W^n\left(\mathcal{D}_i^W-E_i^W\mid u_i\right), \qquad (53)$$

where

$$\mathcal{D}_i^W-\mathcal{E}_i^W=\{y^n\in\mathcal{D}_i^W\mid W^n(y^n\mid u_{N+1})>W^n(y^n\mid u_i)\}$$

is a subset of $\mathcal{E}_{N+1}^W$.

Using (48), by the same arguments as above, we get the estimates

$$\begin{aligned}
&\sum_{i=1}^{N}W^n\left(\mathcal{D}_i^W-\mathcal{E}_i^W\mid u_i\right)\\
&\quad=\sum_{i=1}^{N}\ \sum_{y^n\in\mathcal{D}_i^W-\mathcal{E}_i^W}W^n(y^n\mid u_i) \qquad (54)\\
&\quad\le\sum_{i=1}^{N}\sum_{y^n\in\mathcal{Y}^n}\sqrt{W^n(y^n\mid u_{N+1})W^n(y^n\mid u_i)}\\
&\quad\le\exp\{-nE_{\mathrm{ex}}(R+\delta,P,W)\}
\end{aligned}$$

for $n\ge n_0(|\mathcal{X}|,|\mathcal{Y}|,\delta)$.

Summarizing we obtain by (49) and (52)–(54)

$$\begin{aligned}
\bar{\lambda}(\mathcal{C}'^W,W)&\le\frac{1}{N+1}\left(\sum_{i=1}^{N}W^n\left((\mathcal{D}_i^W)^c\mid u_i\right)+2\exp\{-nE_{\mathrm{ex}}(R+\delta,P,W)\}\right)\\
&=\frac{1}{N+1}\left(N\cdot\bar{\lambda}(\mathcal{C}^W,W)+2\exp\{-nE_{\mathrm{ex}}(R+\delta,P,W)\}\right)
\end{aligned}$$

for $n\ge n_0(|\mathcal{X}|,|\mathcal{Y}|,\delta)$. Theorem 3a) is proved. Part b) follows directly from this proof and Lemma 3b).

APPENDIX III

Good Codes Are Highly Probable

In the standard Shannon random coding method [22] one derives bounds on the *expected* average error probability and then concludes that at least one code must be as good as the ensemble average. For high rates this leads to asymptotically optimal results ($E_r(R, W) = E_{sp}(R, W)$ for rates near capacity, see [19]) and therefore in this case "most" codes in the ensemble must be close to the optimum. In the study of complex channel systems such as arbitrarily varying channels ([4]) it is necessary to have estimates on the proportion of codes in the ensemble which are good. Also, if random selection is of any practical use, one would like to have bounds on the probability with which a good code can be found. First steps in this direction were taken by Dobrushin and Stambler in [26], and independently in [3] and [4]. The papers [26] and [3] consider the average and the paper [4] the maximal error probability.

Here we show considerably more. Whereas in all those papers the error probability was kept constant we allow here λ to meet the random coding bound and still show that for a random selection the probability of not meeting those bounds is double exponentially small. Moreover, we obtain estimates in the double exponential function.

We first state the result. Theorem 4 estimates the probability that randomly selected and expurgated codes are "good." Theorem 5 gives a result for nonexpurgated codes. In order to formulate Theorem 4 we have to introduce some notation concerning the expurgation of a code.

Let $n, \delta > 0$, and $P \in \mathcal{P}_n$ be given. $U_1, \cdots, U_N$ are independent random variables equidistributed on $\mathcal{T}_P^n$, $N = \exp\{nR\}$. For outcomes $u_1, \cdots, u_N \in \mathcal{T}_P^n$ of $U_1, \cdots, U_N$ we define the functions $F(u_1, \cdots, u_N)$ and $G(u_1, \cdots, u_N)$ by

1) $F(u_1, \cdots, u_N) = 1$ if there exist $u_{j_1}, \cdots, u_{j_M} \in \{u_1, \cdots, u_N\}$ and suitable decoding sets $\mathcal{D}_{j_1}, \cdots, \mathcal{D}_{j_M} \subset \mathcal{Y}^n$ such that $M \geq N/2$ and for $\mathcal{C} = \{(u_{j_k}, \mathcal{D}_{j_k}) \mid k = 1, \cdots, M\}$
$$\bar{\lambda}(\mathcal{C}, W) \leq 2 \exp\{-n(E_r(R, P, W) - \delta)\}, \quad (55)$$
for every $W \in \mathcal{W}$. $F(u_1, \cdots, u_N) = 0$ otherwise. Similarly,

2) $G(u_1, \cdots, u_N) = 1$, if there exist $u_{j_1}, \cdots, u_{j_M} \in \{u_1, \cdots, u_N\}$ such that $M \geq N/2$ and such that for every $W \in \mathcal{W}$ the corresponding ML code
$$\mathcal{C}^W = \left\{\left(u_{j_k}, \mathcal{D}_{jk}^W\right) \mid k = 1, \cdots, M\right\}$$
satisfies
$$\bar{\lambda}(\mathcal{C}^W, W) \leq 2 \exp\{-nE_{ex}(R + \delta, P, W)\}.$$
$G(u_1, \cdots, u_N) = 0$ otherwise.

Theorem 4: In the notation above for $n \geq n_0(|\mathcal{X}|, |\mathcal{Y}|, \delta)$
$$\Pr(F = 0) \leq \exp\{-(n \cdot \delta/4 - \log 2) \exp\{nR\}\}, \quad (56)$$
$$\Pr(G = 0) \leq \exp\{-(n \cdot \delta/4 - \log 2) \exp\{nR\}\}, \quad (57)$$
that is, the procedures fail to achieve the random coding bounds (resp. expurgated bounds) uniformly for every $W \in \mathcal{W}$ with double exponentially small error probabilities. Moreover the exponent R is optimal.

By somewhat more refined calculations we obtain the next theorem.

Theorem 5: For any $\delta > 0$, $R > 0$, $n \geq n_0(|\mathcal{X}|, |\mathcal{Y}|, \delta)$, and $P \in \mathcal{P}_n$ the following is true.

Let $U_1, \cdots, U_N$ be independent random variables equidistributed on $\mathcal{T}_P^n$ and for any W let $\mathcal{C}^W(U_1, \cdots, U_N)$ be the ML code for the codewords $U_1, \cdots, U_N$. Then
$$\Pr\big(\bar{\lambda}(\mathcal{C}^W(U_1, \cdots, U_N), W) \geq 2 \exp\{-n(E_r(R, P, W) - 2\delta)\}\big) \leq \exp\{-\exp\{n(R - E_r(R, P, W))\}\}$$
for all $W \in \mathcal{W}$.

Remark 3: This result shows that for $R > E_r(R, P, W)$ codes achieving the random coding bound can hardly be missed by random selection. Notice that for $R < E_r(R, P, W)$ the probability to select a code with P-typical codewords *not* achieving the random coding bound is larger than the probability that in a selected code there are two equal codewords. Since the latter probability is at least exponentially small, for $R < E_r(R, P, W)$ we cannot get any double exponential estimate.

As a new *problem* in the area of error bounds we propose to find the exact exponent for all rates $R > E_r(R, P, W)$.

Proof of Theorem 4: Fix $\delta > 0$, $R > 0$. Let $n \geq n_0(|\mathcal{X}|, |\mathcal{Y}|, \delta)$ such that Theorems 2, 3 hold.

Let $U_1, \cdots, U_N$ be independent random variables equidistributed on $\mathcal{T}_P^n$, $N = \exp\{nR\}$.

Consider the following "expurgated codes": Set $\mathcal{C}_{ex}(U_1) = \{(U_1, \mathcal{D}_{U_1})\}$ with the decoding set $\mathcal{D}_{U_1} = \mathcal{Y}^n$. Clearly, $\bar{\lambda}(\mathcal{C}_{ex}(U_1), W) = 0$ for every $W \in \mathcal{W}$. For $i = 2, \cdots, N$ we define the codes $\mathcal{C}_{ex}(U_1, \cdots, U_i)$ by extending $\mathcal{C}_{ex}(U_1, \cdots, U_{i-1})$. Suppose $i \geq 2$ and assume that $\mathcal{C}_{ex}(U_1, \cdots, U_{i-1}) = \{(U_{j1}, \mathcal{D}_{j1}), \cdots, (U_{jk}, \mathcal{D}_{jk})\}$ with k codewords $U_{j1}, \cdots, U_{jk} \in \{U_1, \cdots, U_{i-1}\}$ has been defined. Then we prolong this code by the codeword U_i to the new code
$$\mathcal{C}_{ex}(U_1, \cdots, U_{i-1} \mid U_i) = \left\{\left(U_{j1}, \mathcal{E}_{j1}\right), \cdots, \left(U_{jk}, \mathcal{E}_{jk}\right), \left(U_i, \mathcal{E}_i\right)\right\},$$
where, for $l = 1, \cdots, k$, $\mathcal{E}_{jl} = \mathcal{D}_{jl} - \{y^n \mid I(U_i \wedge y^n) > I(U_{jl} \wedge y^n)\}$ and where
$$\mathcal{E}_i = \left\{y^n \mid I(U_i \wedge y^n) > I\left(U_{jl} \wedge y^n\right), \text{for all } l = 1, \cdots, k\right\}.$$
If for all $W \in \mathcal{W}$
$$\bar{\lambda}\left(\mathcal{C}_{ex}(U_1, \cdots, U_{i-1} \mid U_i), W\right) \leq 2 \exp\left\{-n\left(E_r(R, P, W) - \delta\right)\right\},$$
then we define
$$\mathcal{C}_{ex}(U_1, \cdots, U_i) = \mathcal{C}_{ex}(U_1, \cdots, U_{i-1} \mid U_i).$$
If this is not the case we set
$$\mathcal{C}_{ex}(U_1, \cdots, U_i) = \mathcal{C}_{ex}(U_1, \cdots, U_{i-1}).$$
In this way we gave a formal definition of the expurgation of a given code with codewords $U_1, \cdots, U_N$.

Now let S_i be a random variable on $\{0, 1\}$ such that $S_i = 0$ if and only if $\mathcal{C}_{ex}(U_1, \cdots, U_i) \neq \mathcal{C}_{ex}(U_1, \cdots, U_{i-1})$, that is, $S_i = 0$ if and only if the codeword U_i was not expurgated. We observe
$$\Pr(F = 0) \leq \Pr\left(\sum_{i=1}^{N} S_i \geq \frac{N}{2}\right), \quad (59)$$
and
$$\Pr(S_i = 1 \mid S_{i-1} = s_{i-1}, \cdots, S_1 = s_1) \leq \exp\left\{-n\frac{\delta}{2}\right\} \quad (60)$$
for any values $s_{i-1}, \cdots, s_1 \in \{0, 1\}$. Equation (59) follows from the definition of the functions F and $S_1, \cdots, S_N$. Equation (60) is a direct application of Theorem 2b). Hence, we only have to

estimate $\Pr(\Sigma_{i=1}^N S_i \ge N/2)$. This can be done by using Bernstein's trick.

For any $\alpha > 0$

$$\Pr\left(\sum_{i=1}^{N} S_i \ge \frac{N}{2}\right) \le \exp\left\{-\alpha \cdot \frac{N}{2}\right\} \cdot E \prod_{i=1}^{N} \exp\{\alpha S_i\}.$$

Now apply (46) to estimate the expected value on the right-hand side. Thus we obtain

$$\Pr\left(\sum_{i=1}^{N} S_i \ge \frac{N}{2}\right) \le \exp\left\{-\alpha \frac{N}{2}\right\} \cdot \left[\exp\left\{-n\frac{\delta}{2}\right\} \cdot \exp\{\alpha\} + \left(1 - \exp\left\{-n\frac{\delta}{2}\right\}\right)\right]^N.$$

Choose in particular

$$\alpha = \log \frac{1 - \exp\left\{-n\frac{\delta}{2}\right\}}{\exp\left\{-n\frac{\delta}{2}\right\}},$$

which is positive for $n \ge n_0(\delta)$. Then,

$$\Pr\left(\sum_{i=1}^{N} S_i \ge N/2\right) \le \exp\{-D((1/2) \| \exp\{-n(\delta/2)\}) \cdot N\},$$

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

We can estimate this divergence:

$$D\left(\frac{1}{2} \| \exp\left\{-n\frac{\delta}{2}\right\}\right) = -\log 2 - \frac{1}{2} \log \exp\left\{-n\frac{\delta}{2}\right\} - \frac{1}{2} \log\left(1 - \exp\left\{-n\frac{\delta}{2}\right\}\right) \ge -\log 2 + n \cdot \frac{\delta}{4}.$$

Thus, $\Pr(F = 0) \le \exp\{-(n(\delta/4) - \log 2) \cdot \exp\{nR\}\}$. This proves the first part of Theorem 4. The proof of the second part is completely analogous.

We have to show that the exponent R is best possible. For this, choose any codeword $u \in \mathcal{T}_P^n$, $P \in \mathcal{P}_n$. Define $\mathcal{C}$ as a code with N codewords $u_1, \cdots, u_N$; $u_i = u$ for all $i = 1, \cdots, N$. We make two observations: $\mathcal{C}$ is a "bad" code, even if one expurgates $\mathcal{C}$. On the other hand, the probability to choose $\mathcal{C}$ at random is of the order $\exp\{-O(n)\exp\{nR\}\}$.

Proof of Theorem 5: Fix $\delta > 0$ and $n \ge n_0(|\mathcal{X}|, |\mathcal{Y}|, \delta)$ such that Theorems 2 and 3 hold and choose $N = \exp\{nR\}$.

Let $U_1, \cdots, U_n$ be independent random variables equidistributed on $\mathcal{T}_P^n$ and let $W \in \mathcal{W}$. We consider the ML codes $\mathcal{C}(U_1, \cdots, U_k)$, $k = 1, \cdots, N$, that is codes with codeword set $\{U_1, \cdots, U_k\}$ and maximum likelihood decoding with respect to the given channel W. We define the random variables $T_1, \cdots, T_N$ on $[0, 1]$ as follows: $T_1 = \bar{\lambda}(\mathcal{C}(U_1), W) = 0$, and for $k = 1, \cdots, N - 1$ the random variable T_{k+1} is defined by

$$\bar{\lambda}(\mathcal{C}(U_1, \cdots, U_{k+1}), W) = \frac{1}{k+1}\left(k \cdot \bar{\lambda}(\mathcal{C}(U_1, \cdots, U_k), W) + T_{k+1}\right).$$

Observe that with this definition

$$\bar{\lambda}(\mathcal{C}(U_1, \cdots, U_k), W) = \frac{1}{k} \sum_{i=1}^{k} T_i$$

for any $k = 1, \cdots, N$.

Using this notation Theorem 2 says that for any $\xi \ge 0$ and for any values $t_1, \cdots, t_k$ of the random variables $T_1, \cdots, T_k$ we have

$$\Pr\left(T_{k+1} > 2\exp\{-n(E_r(R + \xi, P, W) - \delta)\} \mid T_1 = t_1, \cdots, T_k = t_k\right) \le \exp\left\{-n\left(\frac{\delta}{2} + \xi\right)\right\}, \qquad k = 1, \cdots, N - 1. \tag{61}$$

For any $\xi \ge 0$ we define random variables $S_{i,\xi}$, $i = 1, \cdots, N$, on $\{0, 1\}$ such that $S_{i,\xi} = 1$ if and only if

$$T_i > 2\exp\{-n(E_r(R + \xi, P, W) - \delta)\}.$$

Thus, $\Sigma_{i=1}^N S_{i,\xi}$ counts the number of T_i of a certain size. Note that $|T_i| \le 1$ and $E_r(R + |\mathcal{X}|, P, W) = 0$, since $E_r(C, P, W) = 0$, where C is the capacity of W. We express the error probability of the code $\mathcal{C}(U_1, \cdots, U_N)$ with the help of the "counting variables" $\Sigma_{i=1}^N S_{i,\xi}$. Let m be a positive integer, $1/m < \delta/2$. Then

$$\bar{\lambda}(\mathcal{C}(U_1, \cdots, U_N), W) = \frac{1}{N} \sum_{i=1}^{N} T_i \le \frac{1}{N} \sum_{j=1}^{m \cdot |\mathcal{X}|} \left(\sum_{i=1}^{N} S_{i,j/m}\right) \cdot 2\exp\left\{-n\left(E_r\left(R + \frac{j+1}{m}, P, W\right) - \delta\right)\right\}. \tag{62}$$

Here we have counted those T_i which lie in intervals of the form

$$\left[2\exp\left\{-n\left(E_r\left(R + \frac{j}{m}, P, W\right) - \delta\right)\right\}, \cdot 2\exp\left\{-n\left(E_r\left(R + \frac{j+1}{m}, P, W\right) - \delta\right)\right\}\right].$$

Therefore, $\bar{\lambda}(\mathcal{C}(U_1, \cdots, U_N), W)$ becomes large, if the expressions $\Sigma_{i=1}^N S_{i,j/m}$ become large.

We show that for any $\xi \ge 0$

$$\Pr\left\{\sum_{i=1}^{N} S_{i,\xi} \ge \exp\{n(R - (E_r(R, P, W) - E_r(R + \xi, P, W)))\}\right\} \le \exp\left\{-\left(n\frac{\delta}{2} - 2\right) \cdot \exp\{n(R - (E_r(R, P, W) - E_r(R + \xi, p, W)))\}\right\}. \tag{63}$$

Again we use Bernstein's trick. Abbreviate $\tau = 1 - \exp\{-n(E_r(R, P, W) - E_r(R + \xi, P, W))\}$. Then for any $\alpha > 0$:

$$\Pr\left\{\sum_{i=1}^{N} S_{i,\xi} \ge N \cdot (1 - \tau)\right\} \le \exp\{-\alpha \cdot N(1 - \tau)\} \cdot E \prod_{i=1}^{N} \exp\{\alpha S_{i,\xi}\}. \tag{64}$$

In order to estimate the expectation on the right-hand side it is necessary to have estimates on conditional probabilities of the $S_{i,\xi}$. Now observe that from the definition of the $S_{i,\xi}$ and because of (61) we have for any $\xi \ge 0$ and for any values $s_1, \cdots, s_{i-1} \in \{0, 1\}$

$$\Pr(S_{i,\xi} = 1 \mid S_{1,\xi} = s_1, \cdots, S_{i-1,\xi} = s_{i-1}) \le \exp\left\{-n\left(\frac{\delta}{2} + \xi\right)\right\}. \tag{65}$$

We get from (64) and (65)

$$\Pr\left\{\sum_{i=1}^{N} S_{i,\xi} \geq N(1-\tau)\right\} \leq \exp\{-\alpha N(1-\tau)\}\left[\exp\left\{-n\left(\frac{\delta}{2}+\xi\right)+\alpha\right\} + 1 - \exp\left\{-n\left(\frac{\delta}{2}+\xi\right)\right\}\right]. \tag{66}$$

Set

$$\alpha = \log\left(\frac{1-\exp\left\{-n\left(\frac{\delta}{2}+\xi\right)\right\}}{\exp\left\{-n\left(\frac{\delta}{2}+\xi\right)\right\}} \cdot \frac{1-\tau}{\tau}\right).$$

Since $E_r(R,P,W) - E_r(R+\xi,P,W) \leq \xi$ for all $\xi \geq 0$, the number α is positive for $n \geq n_0(\delta)$. We obtain from (66) with this choice of α:

$$\Pr\left\{\sum_{i=1}^{N} S_{i,\xi} \geq N\cdot(1-\tau)\right\} \leq \exp\left\{-D\left(1-\tau \,\|\exp\left\{-n\left(\frac{\delta}{2}+\xi\right)\right\}\right)\cdot N\right\},$$

where

$$D\left(1-\tau\,\|\exp\left\{-n\left(\frac{\delta}{2}+\xi\right)\right\}\right) \geq \tau\log\tau + (1-\tau)\log(1-\tau) + n\left(\frac{\delta}{2}+\xi\right)\cdot(1-\tau).$$

From the fact that $\log(1-x) \geq -2x$ for small positive x we conclude that $\log\tau \geq -2(1-\tau)$ for n sufficiently large.

Hence, for large n,

$$\begin{aligned} D\left(\tau\,\|\exp\left\{-n\left(\frac{\delta}{2}+\xi\right)\right\}\right) &\geq \left[n\left(\frac{\delta}{2}+\xi-(E_r(R,P,W)-E_r(R+\xi,P,W))\right)\right]\cdot(1-\tau)-2(1-\tau) \\ &\geq \left(n\frac{\delta}{2}-2\right)(1-\tau) \qquad (67) \\ &= \left(n\frac{\delta}{2}-2\right)\cdot\exp\{-n(E_r(R,P,W)-E_r(R+\xi,P,W))\}, \end{aligned}$$

where (63) is true, because $E_r(R,P,W) - E_r(R+\xi,P,W) \leq \xi$. Equation (63) is proved. Finally, we have to show that (63) and (62) imply Theorem 5.

From (63) we conclude first that for all $\xi \geq 0$

$$\Pr\left\{\sum_{i=1}^{N} S_{i,\xi} \geq \exp\{n(R-(E_r(R,P,W)-E_r(R+\xi,P,W)))\}\right\} \leq \exp\left\{-\left(n\frac{\delta}{2}-2\right)\cdot\exp\{n(R-E_r(R,P,W))\}\right\}, \quad \text{if } n \text{ is large.} \tag{68}$$

Suppose now that

$$\sum_{i=1}^{N} S_{i,j/m} \leq \exp\left\{n\left(R-\left(E_r(R,P,W)-E_r\left(R+\frac{j+1}{m},P,W\right)\right)\right)\right\}, \qquad j=1,\cdots,m\cdot|\mathcal{X}|.$$

Then we can continue with (62):

$$\begin{aligned} &\bar{\lambda}(\mathcal{C}(U_1,\cdots,U_N),W) \\ &\leq 2\sum_{j=1}^{m\cdot|\mathcal{X}|}\exp\left\{-n\left(E_r(R,P,W)-E_r\left(R+\frac{j}{m},P,W\right)\right) - n\left(E_r\left(R+\frac{j+1}{m},P,W\right)-\delta\right)\right\} \\ &\leq 2\cdot m\cdot|\mathcal{X}|\cdot\exp\left\{-n\left(E_r(R,P,W)+\frac{1}{m}+\delta\right)\right\} \\ &\leq 2\cdot\exp\{-n(E_r(R,P,W)-2\delta)\}, \end{aligned} \tag{69}$$

for n sufficiently large. Now (69), (68), and (62) yield

$$\Pr\{\bar{\lambda}(\mathcal{C}(U_1,\cdots,U_N),W) \geq 2\exp\{-n(E_r(R,P,W)-2\delta)\}\} \leq \exp\left\{-\left(n\frac{\delta}{2}-2\right)\exp\{n(R-E_r(R,P,W))\}\right\}$$

Theorem 5 is proved.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 632–656, 1948.

[2] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 3–18, 1965.

[3] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Z. Wahrscheinlichkeitstheorie verwandte Gebiete*, vol. 44, pp. 159–175, 1978.

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

[5] ——, "Coloring hypergraphs: A new approach to multi-user source coding," Part I, *J. of Comb., Inform., and System Sciences*, vol. 1, pp. 76–115, 1979, Part II, vol. 5, no. 3, pp. 220–268, 1980.

[6] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, 1973.

[7] R. G. Gallager, "Source coding with side information and universal coding," presented at the *IEEE Intern. Symp. Inform. Theory*, Ronneby, Sweden, 1976, preprint.

[8] V. N. Koševev, "On a problem of separate coding of two dependent sources," *Probl. Peredach. Inform.*, vol. 13, pp. 26–32, 1977.

[9] 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.

[10] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels I–II," *Inform. Contr.*, vol. 10, pp. 65–103 and pp. 522–552, 1967.

[11] R. M. Fano, *Transmission of Information, a Statistical Theory of Communication*. New York: Wiley, 1961.

[12] J. Wolfowitz, "The coding of messages subject to chance errors," *Illinois J. Math.*, vol. 1, pp. 591–606, 1957.

[13] S. Arimoto, "On the converse to the coding theorem for the discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 357–359, May 1973.

[14] G. Dueck and J. Körner, "Reliability function of a discrete memoryless channel at rates above capacity," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 82–85, Jan. 1979.

[15] I. Csiszár, J. Körner, and K. Marton, "A new look at the error exponent of a discrete memoryless channel," presented at the *IEEE Intern. Symp. Inform. Theory*, Ithaca, NY, 1977, preprint.

[16] A. Haroutunian, "Estimates of the error exponent for the semi-continuous memoryless channel (in Russian)," *Probl. Peredach. Inform.*, vol. 4, pp. 37–48, 1968.

[17] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 405–417, 1974.

[18] V. D. Goppa, "Nonprobabilistic mutual information without memory," *Prob. Contr. Inform. Theory*, vol. 4, pp. 97–102, 1975.

[19] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[20] A. Feinstein, "A new basic theorem of information theory," *IRE Trans. Inform. Theory*, vol. 4, pp. 2–22, 1954.

[21] R. E. Blahut, "Composition bounds for channel block codes," *IEEE Trans. Inform. Theory*, IT-23, pp. 656–674, 1977.

[22] C. E. Shannon, "Certain results in coding theory for noisy channels," *Inform. Contr.*, vol. 1, pp. 6–25, 1957.

[23] R. Ahlswede, "Channel capacities for list codes," *J. Appl. Prob.*, vol. 10, pp. 824–836, 1973.

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

[25] J. K. Omura, "A lower bounding method for channel and source coding probabilities," *Inform. Contr.*, vol. 27, pp. 148–177, 1975.

[26] R. L. Dobrushin and S. Z. Stambler, "Coding theorems for classes of arbitrarily varying discrete memoryless channels (in Russian)," *Probl. Peredach. Inform.*, vol. 11, pp. 3–22, 1975.

Alphabet-Constrained Data Compression

JERRY D. GIBSON, MEMBER, IEEE, AND THOMAS R. FISCHER, MEMBER, IEEE

Abstract—**The optimal data compression problem is posed in terms of an alphabet constraint rather than an entropy constraint. Solving the optimal alphabet-constrained data compression problem yields explicit source encoder/decoder *designs*, which is in sharp contrast to other approaches. The alphabet-constrained approach is shown to have the additional advantages that (1) classical waveform encoding schemes, such as pulse code modulation (PCM), differential pulse code modulation (DPCM), and delta modnlation (DM), as well as rate distortion theory motivated tree/trellis coders fit within this theory; (2) the concept of preposterior analysis in data compression is introduced, yielding a rich, new class of coders; and (3) it provides a conceptual framework for the design of joint source/channel coders for noisy channel applications. Examples are presented of single-path differential encoding, delayed (or tree) encoding, preposterior analysis, and source coding over noisy channels.**

I. Introduction

THE GOAL of data compression is to process source information, such as a voice or video signal, to obtain the simplest possible representation of the source with an acceptable loss in quality. Phrases sometimes used as synonyms for data compression are "source coding with a fidelity criterion" and "redundancy reduction," although the latter is somewhat restricted in scope [1, pp. 8–9]. As noted by Gray and Davisson [2], [1, pp. 21–25], work in data compression historically has followed one of two substantially different approaches, each approach having been pursued by an entirely different group of researchers. One group of workers has relied principally on intuition and experience to design data compression systems for specific sources such as speech or images [3, pp. 1–3], [26]. One of the principal contributions to this approach was the invention of the delta modulator by Cutler in 1952 [4]; this differential encoding structure is prevalent in data compression systems today. Another group of researchers has taken the rate distortion theory approach that has its origin in a paper published by Shannon in 1959 [5]. The emphasis of the rate distortion theory based approach has, in past years, been on deriving the optimum performance theoretically attainable (OPTA) for given source, fidelity criterion, and coder assumptions [1], [6]. The proofs of OPTA theorems sometimes provide constructive procedures for designing source codes. For example, the familiar random coding argument indicates that for sufficiently long block lengths, a nearly optimal code can be found by selecting a code at random from a particular code ensemble. Unfortunately, such codes have no discernible structure, and, hence, the required code book grows exponentially with block length. Thus these codes are not instrumentable [6, p. 199]. Much effort has been expended recently on obtaining suboptimal coder designs that are motivated by rate distortion theory considerations [7]–[16], [36]. However, OPTA, not source coder designs, has been the principal output of the rate distortion theory approach.

Manuscript received December 10, 1980; revised May 8, 1981. This work was supported in part by the National Science Foundation under Grant ENG77-28527. This work was presented at the 1981 Conference on Information Sciences and Systems, Johns Hopkins University, Baltimore, MD, March 25–27, 1981.

The authors are with the Telecommunication and Control Systems Laboratory, Department of Electrical Engineering, Texas A & M University, College Station, TX 77843.