JOURNAL OF COMBINATORICS,
INFORMATION & SYSTEM SCIENCES

Vol. 4, No. 1, 76–115 (1979)

Coloring Hypergraphs: A New Approach to Multi-user Source Coding—I

RUDOLF AHLSWEDE

Fakulat f. Mathematik Universität Bielefeld Universitässtr. 1, *D–4800 Bielefeld* 1

INTRODUCTION

In spite of the great activity in multi-user communication theory during the last decade and in spite of the many interesting results which have been obtained by researchers all over the world we feel that progress on the essential and harder problems has been rather slow.

We think that this is due to the fact that too much thought is given to specific coding problems with ever increasing complexity, which can be treated by composition of known techniques, and too little effort has been put in trying to understand the basic principles in the subject or to create new ones. As things stand now a beginner in the field must be frightened away by this huge collection of tricks known only to a few experts.

In order to keep the size of the paper in proportion we limit ourselves to source coding even though there are many connections between channel- and source coding problems in multi-user communication theory. We intend to return to them on another occasion.

We hope to bring some clarity into the subject by introducing two principles: the *coloring* principle and the *covering* principle. Those together with Shannon's idea [2] of the test channel for describing certain coverings and the very elementary properties of typical sequences (see [6]) and the entropy (resp. information) function seem to be sufficient to prove most existing source coding theorems. It is asked too much to check this in all cases, but as a justification for our belief in the power of the present approach we give the solution to several outstanding problems.

In Part I we give the rate regions for the following source coding problems:

I. Gallager's problem of coding arbitrarily varying sources (AVS) with side information at the decoder (last paragraph of Section II of [22]).

II. AVS with partial side information at the decoder (in the spirit of [17], [18]).

III. Arbitrarily varying correlated sources (AVCS) with side information at the decoder.

IV. AVCS, that is a robust version and generalization of the Slepian-Wolf Theorem [15] under a positiveness assumption.

Part II will deal with the 2-helper side information problem, rate-distortion versions of the above problems, and a *robustification technique* of wide applicability, which makes it possible to convert coding theorems for compound ([7], [33], [6]) multi-user sources (and -channels) to those for arbitrarily varying multi-user sources (and -channels).

Whenever in the literature channel theorems are used to solve source coding problems things get overcomplicated and this already indicates that something is not really understood. Our present approach seems more canonical.

In addition to the two most basic principles mentioned we emphasize some ideas which may or may not be applicable in a particular situation, but which help as guides in finding solutions: the idea of *separate encoding* the information and the side information and the idea of *decomposing information* or entropy.

SECTION 1. THE SOURCE CODING PROBLEMS AND THEIR RATE REGION

We give now a description of the source coding problems treated and state the results about their rate regions.

§1. ARBITRARILY VARYING SOURCES WITH SIDE INFORMATION AT THE DECODER—A PROBLEM BY GALLAGER [22]

Similar as in channel coding (see for instance [8], [13], [6]) one can describe sources by more robust models, where the source output is governed by an unknown probability distribution (PD) from a prescribed class of PD's.

An arbitrarily varying (discrete memoryless) source (AVS) is a model for a source whose letter distribution depends on a state which may vary within a certain set $\mathcal{S}$ of states from one time instant to the next in an arbitrary manner. We give now the formal description.

Let $\mathcal{X}$, $\mathcal{S}$ be finite sets, and let $\mathcal{P} = \{p(\cdot \mid s): s \in \mathcal{S}\}$ be a set of PD's on $\mathcal{X}$. For every $s^n = (s_1, \ldots, s_n) \in \mathcal{S}^n = \prod_1^n \mathcal{S}$ define PD's $P(\cdot \mid s^n)$ on $\mathcal{X}^n$ by

$$P(x^n \mid s^n) = \prod_{t=1}^{n} P(x_t \mid s_t),\ x^n = (x_1, \ldots, x_n) \in \mathcal{X}^n. \tag{1.1}$$

Set $\mathcal{P}^n = \{P(\cdot \mid s^n)\, s^n : \in \mathcal{S}^n\}$. We call the sequence $(\mathcal{P}^n)_{n=1}^{\infty}$ an AVS. Instead of specifying the distributions we could equivalently consider $(\{X(s^n): s^n \in \mathcal{S}^n\})_{n=1}^{\infty}$, where $X(s^n)$ has distribution $P(\cdot \mid s^n)$. The difference to a correlated source $(X_t, S_t)_{t=1}^{\infty}$ is that not the joint distribution but only

conditional distributions are specified and that the $\mathcal{S}$-outputs are not governed by a probabilistic law.

The rate region for an AVS (without any side information) is well-known, easily follows from the Carrier Lemma in Section 3, and can be expressed as follows:

Denote by $\bar{\mathcal{S}}$ the set of all *formal* convex combinations of elements from $\mathcal{S}$:

$$\bar{\mathcal{S}} = \{\bar{s}: \bar{s} = \sum_{i=1}^{r} \alpha_i s_i \text{ with } \alpha_i \geqslant 0, \sum_{i=1}^{r} \alpha_i = 1, s_i \in \mathcal{S}, r \in \mathbf{N}\}. \quad (1.2)$$

We can then define the distribution

$$P(\cdot \mid \bar{s}) = \sum_{i=1}^{r} \alpha_i P(\cdot \mid s_i) \quad (1.3)$$

and a RV $X(\bar{s})$ with distribution $P(\cdot \mid \bar{s})$.

In this terminology the rate region $\mathcal{R}$ is given by (1.4)

$$\mathcal{R} = \{R: R \geqslant \max_{\bar{s} \in \bar{\mathcal{S}}} H(X(\bar{s}))\}.$$

Gallager considered in [22] the case in which the $\mathcal{S}$-outputs are known exactly to the decoder. He asked to determine the smallest rate at which the $\mathcal{X}$-outputs can be coded with an arbitrarily small error probability uniformly in s^n. The problem seems to be not adaptable to standard techniques and furnishes an interesting example for the power of our coloring techniques. We show that the optimal rate equals $H^* = \max_{s \in \mathcal{S}} H(X(s))$, which is in general smaller than $\max_{\bar{s} \in \bar{\mathcal{S}}} H(X(\bar{s}))$ and therefore the side information helps. As a little exercise the reader may verify that in case the side information is available to the encoder (and only to the encoder) the optimal rate is the same as without side information. Obviously, if both, the encoder and the decoder, have the side information, then the optimal rate equals again H^*.

We give now the formal description of the problem. Let f_n be a mapping of $\mathcal{X}^n$ into some finite set, binary strings of length $\log_2 \|f_n\|$ for instance (As in [17] we use again the notation $\|f\|$ for the cardinality of the range of function f). F_n is a mapping of the cartesian product of the range of f_n with $\mathcal{S}^n$ into $\mathcal{X}^n$. We refer to f_n (resp. F_n) as encoding (resp. decoding) function. The pair (f_n, F_n) is called a code.

The block error probability of the oode is defined by

$$e(f_n, F_n) = \max_{s^n \in \mathcal{S}^n} \text{Prob}\{F_n(f_n(X(s^n)), s^n) \neq X(s^n)\}. \quad (1.5)$$

A non-negative number R is called an *achievable* rate, if for any $\gamma > 0$, $0 < \lambda \leqslant 1$, there exists an $n_0(\lambda, \gamma)$ such that for all $n \geqslant n_0(\lambda, \gamma)$ there exists

a code (f_n, F_n) such that

$$e(f_n, F_n) \leqslant \lambda \tag{1.6}$$

and

$$\log \|f_n\| \leqslant (R + \gamma)n. \tag{1.7}$$

The infimum over all achievable rates is called the *optimal* rate R_D.

THEOREM 1. *For the AVS with side information at the decoder the optimal rate R_D is given by the formula*

$$R_D = \max_{s \in \mathcal{S}} H(X(s)).$$

§2 AVS WITH PARTIAL SIDE INFORMATION AT THE DECODER

In the preceding paragraph we dealt with the situation in which the decoder has *exact* knowledge of the $\mathcal{S}$-outputs. Now we consider a more general problem. Let us imagine that in addition to the $\mathcal{X}$-encoder there is another person (or device), the $\mathcal{S}$-encoder, who observes the $\mathcal{S}$-outputs. He is able to inform the decoder about those outputs at a prescribed rate R_2. Clearly, if $R_2 \geqslant \log |\mathcal{S}|$, then we are back in the old situation of a completely informed decoder. Let us give now the formal description of the coding problem and the result.

f_n (resp. g_n) is a mapping of $\mathcal{X}^n$ (resp. $\mathcal{S}^n$) into a finite set. They are the encoding functions. The decoding function F_n maps the cartesian product of the ranges of f_n and g_n into $\mathcal{X}^n$. The error probability of the code (f_n, g_n, F_n) is defined by

$$e(f_n, g_n, F_n) = \max_{s^n \in \mathcal{S}^n} \text{Prob}\,\{F_n(f_n(X(s^n)), g_n(s^n)) \neq X(s^n)\}. \tag{1.8}$$

A pair of non-negative real numbers (R_1, R_2) is called an achievable pair of rates, if for any $\gamma > 0$, $0 < \lambda \leqslant 1$, there exists an $n_0(\lambda, \gamma)$ such that for all $n \geqslant n_0(\lambda, \gamma)$ there is a code (f_n, g_n, F_n) with

$$e(f_n, g_n, F_n) \leqslant \lambda \tag{1.9}$$

and

$$\log \|f_n\| \leqslant (R_1 + \gamma)n, \tag{1.10}$$

$$\log \|g_n\| \leqslant (R_2 + \gamma)n.$$

The region of all achievable pairs of rates is denoted by R_{DP}. For the presentation of the results it is convenient to adapt the following notation:

$\mathcal{P}(\mathcal{S})$ is the set of all PD's on $\mathcal{S}$. For $p \in \mathcal{P}(\mathcal{S})$, S_p is a random variable (RV) with values in $\mathcal{S}$ and distribution p. X_p takes values in $\mathcal{X}$ and is distributed according to

$$\text{Prob}\,(X_p = x) = \sum_{s \in \mathcal{S}} \text{Prob}\,(X(s) = x)\,\text{Prob}\,(S_p = s). \tag{1.11}$$

U_p stands for a RV with (U_p, S_p, X_p) forming a Markov chain:

$$U_p \to S_p \to X_p.$$

THEOREM 2. *For the AVS with partial side information at the decoder the rate region R_{DP} equals*

$$\mathcal{R}^{**} = V\{(R_1, R_2) : R_1 \geqslant \sup_{p \in \mathcal{P}(\mathcal{S})} H(X_p \mid U_p),\ R_2 \geqslant \sup_{p \in \mathcal{P}(\mathcal{S})} I(S_p \wedge U_p)\},$$

where the union is taken over all sets of Markov chains $(U_p \to S_p \to X_p)_{p \in \mathcal{P}(\mathcal{S})}$. It suffices to use U_p's with $\|U_p\| \leqslant |\mathcal{S}| + 2$.

§3 ARBITRARILY VARYING CORRELATED SOURCES (AVCS) WITH SIDE INFORMATION AT THE DECODER

We consider here (discrete memoryless) arbitrarily varying correlated sources (AVCS), which can be described as follows.

Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{S}$ be finite sets, and let $\{P(\cdot, \cdot \mid s) : s \in \mathcal{S}\}$ be a set of PD's on $\mathcal{X} \times \mathcal{Y}$. For every $s^n = (s_1, \ldots, s_n) \in \mathcal{S}^n = \prod_1^n \mathcal{S}$ define PD's $P(\cdot, \cdot \mid s^n)$ on $\mathcal{X}^n \times \mathcal{Y}^n$ by

$$P(x^n, y^n \mid s^n) = \prod_{t=1}^{n} P(x_t, y_t \mid s_t) \text{ for } (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n. \tag{1.12}$$

We call the sequence $(\{P(\cdot, \cdot \mid s^n) : s^n \in \mathcal{S}^n\})_{n=1}^{\infty}$ an AVCS. Instead of specifying the distributions we could equivalently consider $(\{(X(s^n), Y(s^n)) : s^n \in \mathcal{S}^n\})_{n=1}^{\infty}$ where $(X(s^n), Y(s^n))$ is a pair of RV's with values in $\mathcal{X}^n \times \mathcal{Y}^n$ and distribution $P(\cdot, \cdot \mid s^n)$. In case $|\mathcal{S}| = 1$ one gets the standard correlated source (DMCS) considered by Slepian and Wolf ([15]) and in case $|\mathcal{Y}| = 1$ one gets the AVS of §1. For encoding functions f_n (resp. g_n) defined on $\mathcal{X}^n$ (resp. $\mathcal{Y}^n$) a decoding function F_n shall be a mapping of the cartesian product of the ranges of f_n, g_n and of $\mathcal{S}^n$ into $\mathcal{X}^n \times \mathcal{Y}^n$. The error probability of (f_n, g_n, F_n) is defined by

$$e(f_n, g_n, F_n) = \max_{s^n \in \mathcal{S}^n} \text{Prob}\,\{F_n(f_n(X(s^n)), g_n(Y(s^n)), s^n) \neq (X(s^n), Y(s^n))\}. \tag{1.13}$$

Achievable pairs of rates and the rate region, denoted now by $\mathcal{R}_D^{\mathcal{A}}$, are defined in the usual way.

THEOREM 3. *The rate region $\mathcal{R}_D^{\mathcal{A}}$ for the AVCS $\mathcal{A}$ with side information at the decoder can be characterized as follows:*

$$\mathcal{R}_D^{\mathcal{A}} = \{(R_1, R_2) : R_1, R_2 \text{ satisfy (a), (b), (c)}\}$$

where

(a) $$R_1 \geqslant \max_{s \in \mathcal{S}} H(X(s) \mid Y(s))$$

(b) $$R_2 \geqslant \max_{s \in \mathcal{S}} H(Y(s) \mid X(s))$$

(c) $$R_1 + R_2 \geqslant \max_{s \in \mathcal{S}} H(X(s), Y(s)).$$

If we denote by $\mathcal{R}(s)$ the rate region of the DMCS $((X_t(s), Y_t(s))_{t=1}^{\infty}$, then the Theorem says that

$$\mathcal{R}_D^{\mathcal{A}} = \bigcap_{s \in \mathcal{S}} \mathcal{R}(s).$$

We therefore have the

COROLLARY. *For the AVCS $\mathcal{A}$ with side information at the decoder and* **additional** *side information at one or both encoders the rate region equals (again) $\mathcal{R}_D^{\mathcal{A}}$.*

The cases in which one or both encoders have side information about the states are still to be investigated. More generally one could study the case in which the encoders and the decoder have different *partial* side information in the *following sense*: There are 3 partitions $\Omega_1 = \{A_1, \ldots, A_{k_1}\}$, $\Omega_2 = \{B_1, \ldots, B_{k_2}\}$, and $\Omega_D = \{C_1, \ldots, C_{k_D}\}$ of $\mathcal{S}$, and at each time instant encoder 1 (resp. 2) knows in which A_i (resp. B_j) s_t is contained and so does the decoder with respect to his partition. This model covers all cases of the present paragraph and goes considerably beyond it.

§4 AVCS WITHOUT SIDE INFORMATION

f_n and g_n are defined as in §3. F_n is now a mapping from the cartesian product of the ranges of f_n and g_n into $\mathcal{X}^n \times \mathcal{Y}^n$. The error probability of the code (f_n, g_n, F_n) is defined by

$$\epsilon(f_n, g_n, F_n) = \max_{s^n \in \mathcal{S}^n} \text{Prob}\,\{F_n(f_n(X(s^n)), g_n(Y(s^n))) \neq (X(s^n), Y(s^n))\}. \tag{1.14}$$

The rate region, defined as usual, is denoted by $\mathcal{R}^{\mathcal{A}}$. $\bar{\mathcal{S}}$ is again the set of formal convex combinations of elements in $\mathcal{S}$,

$$P(x, y \mid \bar{s}) = \sum_j \alpha_j P(x, y \mid s_j) \quad \text{for } \bar{s} = \Sigma \alpha_j s_j$$

and $(X(\bar{s}), Y(\bar{s}))$ has distribution $P(\cdot, \cdot \mid \bar{s})$.

THEOREM 4. *If an AVCS $\mathcal{A}$ satisfies the entropy* **positiveness condition**

$$H(X(s), Y(s)) > 0 \quad \textit{for all } s \in \mathcal{S}, \tag{1.15}$$

then its rate region $\mathcal{R}^{\mathcal{A}}$ equals $\bar{\mathcal{R}}$:

$\bar{\mathcal{R}} = \{(R_1, R_2) : R_1, R_2$ *satisfy* (a), (b), (c)$\}$. *here*

(a) $$R_1 \geq \sup_{\bar{s} \in \bar{\mathcal{S}}} H(X(\bar{s}) \mid Y(\bar{s}))$$

(b) $$R_2 \geq \sup_{\bar{s} \in \bar{\mathcal{S}}} H(Y(\bar{s}) \mid X(\bar{s}))$$

(c) $$R_1 + R_2 \geq \sup_{\bar{s} \in \bar{\mathcal{S}}} H(X(\bar{s}), Y(\bar{s})).$$

The theorem says that

$$\mathcal{R}^{\mathcal{A}} = \bigcap_{\bar{s} \in \bar{S}} \mathcal{R}(\bar{s}).$$

This characterisation of the rate region is in general *not* valid without the positiveness condition.

EXAMPLE. (0-1-case) Let us consider the case

$$P(x, y \mid s) = 1 \text{ or } 0 \quad \text{for all } x \in \mathcal{X}, y \in \mathcal{Y}, s \in \mathcal{S}. \tag{1.16}$$

This can also be described by an $|\mathcal{X}| \times |\mathcal{Y}|$-matrix with 0's and 1's as entries. This matrix has a 1 in position (x, y) exactly when $P(x, y \mid s) = 1$ for some $s \in \mathcal{S}$.

Now let us look at the special case, in which the matrix is of the form $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$. It is convenient to use $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$. An easy calculation yields $\bar{\mathcal{R}} = \{(R_1, R_2) : R_1 \geqslant \log 2, R_2 \geqslant \log 2, R_1 + R_2 \geqslant \log 6\}$. For a code (f_n, g_n, F_n) with $e(f_n, g_n, F_n) \leqslant \lambda < 1$ in the 0-1-case by (1.14) necessarily $e(f_n, g_n, F_n) = 0$.

We show that then

$$\|f_n\| = 3^n, \|g_n\| = 3^n. \tag{1.15}$$

Suppose that (w.l.o.g. by symmetry) $\|f_n\| < 3^n$, then there exists a pair (x^n, x'^n) with $f(x^n) = f(x'^n)$. Endowing $\mathcal{X}^n$ with a vector space structure $GF(3)^n$ one readily verifies that there are vectors $\epsilon^n = (\epsilon_1, \ldots, \epsilon_n)$ and $\epsilon'^n = (\epsilon'_1, \ldots, \epsilon'_n)$ with $\epsilon_t, \epsilon'_t \in \{0, 1\}$, $1 \leqslant t \leqslant n$, such that

$$x'^n + \epsilon'^n = x^n + \epsilon^n.$$

The nth Kronecker product of the above matrix has a 1 exactly in the positions (x^n, y^n) with $y^n = x^n + \epsilon^{*n}$, $\epsilon^*_t \in \{0, 1\}$, $1 \leqslant t \leqslant n$, and there both positions $(x^n, x^n + \epsilon^n)$ and $(x'^n, x'^n + \epsilon'^n)$ have a 1. Since the second components are equal and the first components have the same color this contradicts $e(f_n, g_n, F_n) = 0$.

In the general 0-1-case the matrix is the vertex-vertex incidence matrix of a bipartite graph and we have the problem to determine the rate region of strict orthogonal colorings of the nth (Kronecker) product of bipartite graphs. Let us denote this region by $\mathcal{R}_0$. To give "single-letter" characterizations for this region is of comparable difficulty with the related 0-error capacity problem of Shannon [30], which comes from channel coding and is for *packing* rather than for coloring. For recent progress on this problem—in particular, the solution of the famous pentagon case—see [32].

As a *problem*, which deserves further study, we propose to decide whether for an arbitrary AVCS $\mathcal{A}$

$$\mathcal{R}^{\mathcal{A}} = \mathcal{R}^{\mathcal{A}_+} \cap \mathcal{R}^{\mathcal{A}_0},$$

if the AVCS $\mathcal{A}_0$ (resp. $\mathcal{A}_+$) is given by the set of PD's

$\{P(\cdot, \cdot \mid s) : H(X(s), Y(s)) = 0\}$ (resp. $\{P(\cdot, \cdot \mid s) : H(X(s), Y(s)) > 0\}$.

$\mathcal{R}^{\mathcal{A}_+}$ is known by Theorem 4.

In his doctoral thesis [26] J. H. Jahn showed that for an AVCS the rate region equals always $\overline{\mathcal{R}}$, if the encoders are allowed to use *randomized encoding*. His proof is based on a generalization of Cover's proof ([16]), which leads to *correlated random codes* (see [8], [13]) and the *elimination technique* of [13], which makes it possible to transform such codes into codes with independent randomisation at the encoders only. The wide applicability of this technique to arbitrarily varying multi-way channels and correlated sources was emphasized in [13]. Jahn's generalization of Cover's proof lacks symmetry in the error estimates and this accounts for the fact that the proof is very complicated. In Section 7 we give a proof which avoids these difficulties. In conclusion we remark that *all our theorems hold also for infinite sets* $\mathcal{S}$. This can be proved with the help of a lemma in [8], called Approximation Lemma in [13], in the same way as it was used in [13].

Section 2. Coloring Hypergraphs

The nature of this section is purely combinatorial. In the text some remarks about connections to coding theory are made. However, these connections will become fully clear only in the following sections.

Let $\mathcal{V} = \{1, 2, \ldots, I\}$ be a finite set and let $\mathcal{E} = \{E_j : 1 \leqslant j \leqslant J\}$ be a family of subsets of $\mathcal{V}$. "Hypergraph" is a fancy name for the pair $(\mathcal{V}, \mathcal{E})$. The elements of $\mathcal{V}$ are called vertices and the elements of $\mathcal{E}$ are called edges.

§1 Colorings which are good on all edges (universal colorings)

A *vertex coloring* of $(\mathcal{V}, \mathcal{E})$ with L colors is a map Φ: $\mathcal{V} \to \{1, 2, \ldots, L\}$. In [10] we proved by a simple counting argument the following

Coloring Lemma 1. *Let J, L and t be non-negative integers such that*

$$J \cdot L \leqslant t!. \tag{2.1}$$

For any hypergraph $(\mathcal{V}, \mathcal{E})$ with $|\mathcal{E}| = J$ and

$$|E_j| \leqslant L \ (1 \leqslant j \leqslant J) \tag{2.2}$$

there exists a vertex coloring Φ with L colors such that in every edge (universality) E_j $(1 \leqslant j \leqslant J)$ every color occurs at most t times, that is,

$$|\Phi^{-1}(l) \cap E_j| \leqslant t \text{ for all } l = 1, \ldots, L \text{ and all } j = 1, \ldots, J. \tag{2.3}$$

This result was used in [10] to prove a list code version of what is now called the Slepian-Wolf source coding theorem [15]. Since the hypothesis (2.1) holds for t's which are rather small as compared to J and L one

actually gets very small list sizes. That was enough for the purposes it served in [10]. Noticeable facts about this result are:

(1) No assumptions about the interdependencies of the E_j's are made.

(2) No assumption about I is made.

(3) The *universality* mentioned above is not guaranteed by the Slepian-Wolf Theorem, which says only that edges are colored "well in *average*". For the source coding problems solved in this paper universality is a key issue and our old coloring result encouraged us in solving them.

(4) Since for this result the number of colors used could be as small as $\max_{1 \leqslant j \leqslant J} |E_j|$, by *allowing L to be somewhat larger, better results can be obtained.*

We present now some results which seem to be of importance for multi-user source coding. The discussion is by no means exhaustive, that is, other coloring results are conceivable. However, the techniques used are adaptable to many situations, their power lies in their simplicity.

We denote by Φ_λ, $0 \leqslant \lambda < 1$, a coloring of $(\mathcal{V}, \mathcal{E})$ for which in every edge E_j, $1 \leqslant j \leqslant J$, at least $(1 - \lambda)|E_j|$ colors occur only once. Φ_0 is said to be strict.

Strict colorings usually require an enormous number of colors. Our first little result concerns strict colorings. It also plays an auxiliary role for the proof of Lemma 3C.

To a hypergraph $(\mathcal{V}, \mathcal{E})$ we can assign a graph $(\mathcal{V}, \mathcal{E}^*)$, where the vertex set is the same as before and 2 vertices are connected if they are both contained in an E_j for some j. A graph is a special hypergraph. A strict vertex coloring of $(\mathcal{V}, \mathcal{E}^*)$ is also a strict vertex coloring of $(\mathcal{V}, \mathcal{E})$, and vice versa. "deg (i)" denotes the number of vertices in the graph, which are connected with vertex i by an edge.

COLORING LEMMA 2. *Let $(\mathcal{V}, \mathcal{E}')$ be a graph with*

$$\max_{v \in \mathcal{V}} \deg(v) \leqslant D,$$

then $(\mathcal{V}, \mathcal{E}')$ can be strictly colored with L colors if

$$L \geqslant D + 1. \tag{2.4}$$

Proof. Color the vertices $1, \ldots, I$ iteratively in *any way* such that no two adjacent vertices get the same color. If the procedure stops before all vertices are colored, then necessarily one vertex i must have $\deg(i) \geqslant D + 1$, contradicting the hypothesis.

We present now 3 coloring lemmas of increasing complexity, the later ones imply the earlier ones. Therefore, we could just give the last one; however, for tutorial reasons and also in order to reflect the development of the ideas we prefer the prescribed setup. A coloring Φ is called an *L-coloring* if $\|\Phi\| \leqslant L$.

COLORING LEMMA 3A. *A hypergraph* $(\mathcal{V}, \mathcal{E})$ *has an* L*-coloring* $\Phi_{2\lambda}$, $0 < \lambda < \frac{1}{2}$, *if*

$$\max_{1 \leqslant j \leqslant J} |E_j|(L - |E_j|)^{-1}(1 - \lambda)\lambda^{-1} \leqslant 1 \tag{2.5}$$

and

$$\sum_{j=1}^{J} \exp\{|E_j|(h(\lambda) + \lambda \log(|E_j|L^{-1}))\} < 1. \tag{2.6}$$

Here, and elsewhere $h(\lambda) = -\lambda \log \lambda - (1 - \lambda) \log (1 - \lambda)$. *For* (2.6) *to hold necessarily* $L \geqslant \max_{1 \leqslant j \leqslant J} |E_j|$.

Proof. Let $X_1, \ldots, X_L$ be i.i.d. RV's with distribution Prob $(X_l = l) = \frac{1}{L}$ for $l = 1, \ldots, L$. Color the vertices at random such that vertex v_i gets color l if $X_i = l$.

We refer to this as the *standard random* L*-coloring of* $\mathcal{V}$ in the sequel. Define now for $i = 1, \ldots, I$; $j = 1, \ldots, J$ RV's

$$f_i^j(X_1, \ldots, X_i) = \begin{cases} 1 & \text{if } X_i \neq X_{i'} \text{ for all } i' \leqslant i, i' \in E_j \\ 0 & \text{otherwise} \end{cases}. \tag{2.7}$$

We can view the coloring procedure as an iterative coloring of vertices $v_1, \ldots, v_j$. Then f_i^j takes the value 1 if i gets a color which has not occurred until step i in E_j.

Clearly, if $\sum_{i \in E_j} f_i^j(X_1, \ldots, X_i) \geqslant (1 - \lambda)|E_j|$ then at most $\lambda|E_j|$ colors occur more than once in E_j, and therefore $(1 - 2\lambda)|E_j|$ vertices are colored correctly. We upperbound now

$$\text{Prob}\left\{\sum_{i \in E_j} f_i^j < (1 - \lambda)|E_j|\right\}.$$

It is clear from the definition (2.7) that this expression depends only on RV's X_i with $i \in E_j$. W.l.o.g. we can therefore consider the following problem:

$X_1, \ldots, X_t$, $t = |E_j|$, are distributed as before,

$$f_i(X_1, \ldots, X_i) = \begin{cases} 1 \text{ if } X_i \neq X_{i'} \text{ for all } i' < i \\ 0 \text{ otherwise} \end{cases}.$$

It suffices to show that under the hypothesis $t(L - t)^{-1}(1 - \lambda)\lambda^{-1} \leqslant 1$

$$\text{Prob}\left\{\sum_{i=1}^{t} f_i(X_1, \ldots, X_i) < (1 - \lambda)t\right\} \leqslant \exp\left\{t\left[h(\lambda) + \lambda \log \frac{t}{L}\right]\right\}, \tag{2.8}$$

because this inequality implies then

$$\text{Prob}\left\{\min_{1 \leqslant j \leqslant J} |E_j|^{-1} \sum_{i \in E_j} f_i^j < (1 - \lambda)\right\} \leqslant \sum_{j=1}^{J} \exp\left\{|E_j|\left[h(\lambda) + \lambda \log \frac{|E_j|}{L}\right]\right\}, \tag{2.9}$$

if (2.5) is satisfied, and hence (2.6) gives the desired result. In order to show (2.8) we use Bernstein's trick, which yields for $\alpha < 0$,

$$\text{Prob}\left\{\sum_{i=1}^{t} f_i - (1-\lambda)\,t < 0\right\} \leqslant \exp\{-\alpha(1-\lambda)t\}\mathbf{E}\prod_{i=1}^{t}\exp\{\alpha f_i\}. \tag{2.10}$$

Notice that the f_i's are not independent, however, since

$$\begin{aligned}&\text{Prob}\{f_t = \epsilon_t, \ldots, f_1 = \epsilon_1\}\\ &= \prod_{s=2}^{t}\text{Prob}\{f_s = \epsilon_s \mid f_{s-1} = \epsilon_{s-1}, \ldots, f_1 = \epsilon_1\}\,\text{Prob}\{f_1 = \epsilon_1\}\end{aligned} \tag{2.11}$$

and since

$$\text{Prob}\{f_s = 1 \mid f_{s-1} = \epsilon_{s-1}, \ldots, f_1 = \epsilon_1\} \geqslant \frac{L-s}{L} \geqslant \frac{L-t}{L}, \tag{2.12}$$

we can conclude

$$\text{Prob}\left\{\sum_{i=1}^{t} f_i < (1-\lambda)t\right\} \leqslant \exp\{-\alpha(1-\lambda)t\}\cdot\left(\frac{t}{L} + \frac{L-t}{L}\,e^{\alpha}\right)^{t} \tag{2.13}$$

Set $p = \dfrac{t}{L}$ and $q = \dfrac{L-t}{L}$.

The best choice for α is $\alpha = \log\left(\dfrac{p}{q}\cdot\dfrac{1-\lambda}{\lambda}\right)$, if $\dfrac{p}{q}\dfrac{1-\lambda}{\lambda} \leqslant 1$, which is true by hypothesis.

$$\begin{aligned}\text{Prob}\left\{\sum_{i=1}^{t} f_i < (1-\lambda)t\right\} &\leqslant \exp\{[h(\lambda) + \lambda\log p + (1-\lambda)\log q]t\}\\ &\leqslant \exp\{[h(\lambda) + \lambda\log p]t\} \leqslant \exp\left\{\left[h(\lambda) + \lambda\log\frac{t}{L}\right]t\right\}\end{aligned} \tag{2.14}$$

Before we consider more complex coloring problems, let us pause and demonstrate the significance of the result for source coding.

Remarks (1) Notice that in case $|E_j| = \exp\{an\}$ for all $j = 1, \ldots, J$, $L = \exp\{bn\}$, and $J = \exp\{cn\}$ for constants $a, b, c > 0$ with $b > a$, obviously (2.5) holds for $n \geqslant n_0(a, b, \lambda)$ and for a suitable $n_1(a, b, c, \lambda)$ $\exp\{cn\}\cdot\exp\{[h(\lambda) - n\lambda(b-a)]e^{an}\} < 1$ for $n \geqslant n_1$, and therefore also (2.6) holds. In applications to coding this result is useful if we have PD's P_j on each E_j which are very close to being uniform. The lack of complete uniformity is not crucial, because we can deal with it by choosing λ very small. This can best be seen in the following:

EXAMPLE 1 Let $(X_t, Y_t)_{t=1}^{\infty}$ be a discrete memoryless correlated source (DMCS). Write $X^n = X_1, \ldots, X_n$ and $Y^n = Y_1, \ldots, Y_n$.

As in Section 3, $\mathcal{T}_\delta(X^n)$ denotes the (X^n, δ)-typical sequences and $\mathcal{G}_\delta(Y^n \mid x^n)$ denotes the sequences generated by x^n. We know (see Section

3 or [6]) that

(a) $|\mathcal{T}_\delta(X^n)| = \exp\{H(X)n + 0(\sqrt{n})\}$.

(b) $|\mathcal{G}_\delta(Y^n \mid x^n)| = \exp\{H(Y \mid X)n + 0(\sqrt{n})\}$ for $x^n \in \mathcal{T}_\delta(X^n)$.

(c) $\text{Prob}\,(X^n \in \mathcal{T}_\delta(X^n)) = 1 - 0\left(\frac{1}{\delta^2}\right)$.

(d) $\text{Prob}\,(Y^n \in \mathcal{G}_\delta(Y^n \mid x^n) \mid X^n = x^n) = 1 - 0\left(\frac{1}{\delta^2}\right)$ for $x^n \in \mathcal{T}_\delta(X^n)$

and

(e) $\text{Prob}\,(Y^n = y^n \mid X^n = x^n) = \exp\{-H(Y \mid X)n + 0(\sqrt{n})\}$
for $x^n \in \mathcal{T}_\delta(X^n)$, $y^n \in \mathcal{G}_\delta(Y^n \mid x^n)$.

Choose as hypergraph $(\mathcal{V}, \mathcal{E}) = (\mathcal{Y}^n, (\mathcal{G}_\delta(Y^n \mid x^n)_{x^n \in \mathcal{T}_\delta(X^n)})$ and set $\lambda = \lambda(n) = \exp\{-3c\sqrt{n}\}$.

An easy calculation shows that with $L(n) = \exp\{H(Y \mid X)n + 7c\sqrt{n}\}$ (2.5) and (2.6) hold and the lemma implies the existence of a coloring $\Phi_{2\lambda}$ with $L(n)$ colors.

Let the $\mathcal{X}$-encoder report the elements of $\mathcal{T}_\delta(X^n)$ and the $\mathcal{Y}$-encoder the color of the element y^n observed.

If $\mathcal{G}_\delta(Y^n \mid x^n)_{\text{inc}}$ denotes the elements of edge $\mathcal{G}_\delta(Y \mid x^n)$ incorrectly colored, then

$$\begin{aligned}\text{Prob}\,(Y^n \in \mathcal{G}_\delta(Y^n \mid x^n)_{\text{inc}} \mid x^n) &\\ \leqslant 2\lambda \mid \mathcal{G}_\delta(Y^n \mid x^n) \mid &\exp\{-H(Y \mid X)n + c\sqrt{n}\}| \text{ (use } e) \\ \leqslant \exp\{-c\sqrt{n}\},&\end{aligned}$$

which is very small for n large.

This, (c) and (d) imply that the decoder can reproduce (X^n, Y^n) with arbitrarily small error probability. By (a), (b) and the choice of L the rates are less than $H(X) + \frac{c}{\sqrt{n}}$ resp. $H(Y \mid X) + \frac{7c}{\sqrt{n}}$.

Thus we have derived Slepian-Wolf's result ([15]). It would have been sufficient to show that for arbitrary small $\epsilon > 0$, $(H(X)+\epsilon, H(Y|X)+\epsilon)$ is achievable.

We gave the slightly improved form with $\frac{\text{const}}{\sqrt{n}}$ deviation to demonstrate that our approach could be used for sharper error (resp. rates) estimates, but this is a point of minor importance. Relevant is that we have actually proved more than the result of [15]: not only a large proportion (average), but strictly all edges are colored almost correctly (*Universality*). For colorings of average goodness Lemma 4 below, which is an abstract version of Cover's [16], already suffices. For results of that type the (AEP)-property (e) is not needed.

(2) The universality of the coloring makes the solution of Gallager's problem for AVS possible. However, in solving this problem an additional difficulty arises, because certain $|E_j|$ may be so small that (2.6) does not hold. This difficulty can be overcome with an additional coloring provided by Lemma 2. The coloring technique used there may be appropriately named maximal coloring. Since we emphasize general principles, we draw attention to the duality "(maximal coloring, random coloring) in source coding and (maximal coding, random coding) in channel coding".

We give now a refinement of Lemma 3A. There are two reasons for this:

I. In applying Lemma 3A we have to choose λ very small in order to cope with the lack of complete uniformity of the PD's on the E_j's. By coloring all subsets of the E_j's on which the PD's are uniform also essentially correctly we can keep λ constant and *base everything on counting.*

II. In Section 5 we treat AVS's with partial *side information* (in the sense of [17], [18]). It will be seen there that the solution of its coding problem requires "*proper coloring of subedges*".

The following concept turns out to be appropriate.

Suppose that in addition to a hypergraph $(\mathcal{V}, \mathcal{E})$ we are given with every edge E_j, $E_j \in \mathcal{E}$, a family $\mathcal{E}_j = \{E_j^1, \ldots, E_j^{M_j}\}$ of subsets of E_j, that is, $(E_j, \mathcal{E}_j)$ is again a hypergraph. Then we call $\mathcal{H}_2 = (\mathcal{V}, \mathcal{E}, (\mathcal{E}_j)_{j=1}^J)$ a second order or (shortly) a *2-hypergraph.* We denote by Φ_λ^2, $0 < \lambda < 1$, a vertex coloring of $\mathcal{H}_2$ for which in every subedge E_j^m $(m = 1, \ldots, M_j;$ $j = 1, \ldots, J)$ at least $(1 - \lambda)\,|E_j^m|$ colors occur, which *occur only once in* E_j. (2.15)

COLORING LEMMA 3B. *A 2-hypergraph* $\mathcal{H}_2 = (\mathcal{V}, \mathcal{E}, (\mathcal{E}_j)_{j=1}^J)$ *has an* L-coloring $\Phi_{2\lambda}^2$ *if* (2.5) *holds and*

$$\sum_{j=1}^{J} \sum_{m=1}^{M_j} \exp\{|E_j^m|(h(\lambda) + \lambda \log(|E_j|L^{-1}))\} < 1. \tag{2.16}$$

Proof. Use the standard random L-coloring $(X_1, \ldots, X_I)$ and define for $i = 1, \ldots, I$; $m = 1, \ldots, M_j$; $j = 1, \ldots, J$ RV's

$$f_i^{jm}(X_1, \ldots, X_I) = \begin{cases} 1 \text{ if } X_i \neq X_{i'} \text{ for all } i' \text{ in } (E_j^m \cap \{1, \ldots, i\}) \cup (E_j - E_j^m) \\ 0 \text{ otherwise} \end{cases} \tag{2.17}$$

We upperbound now $\text{Prob}\{\sum_{i \in E_j^m} f_i^{jm} < (1 - \lambda)|E_j^m|\}$.

Now for $i_1 < i_2 < \ldots < i_{M_j}$ with $\{i_s : 1 \leqslant s \leqslant M_j\} = E_j^m$ we have

$$\text{Prob}\{f_{i_s}^{jm} = 1 \mid f_{i_{s-1}}^{jm} = \epsilon_{s-1}, \ldots, f_{i_1}^{jm} = \epsilon_1\}$$

$$\geqslant \frac{L - s - (|E_j| - |E_j^m|)}{L} \geqslant \frac{L - |E_j|}{L}. \tag{2.18}$$

Now repeat the arguments which led to (2.14). The only difference which comes in now is that (2.13) is to be replaced by

$$\text{Prob}\left\{\sum_{i=1}^{t^*} f_i < (1-\lambda)t^*\right\} \leqslant \exp\{-\alpha(1-\lambda)t^*\}\left\{\frac{t}{L} + \frac{L-t}{L}e^{\alpha}\right\}^{t^*},$$

where $t^* = |E_j^m|$, and $t = |E_j|$, as before. Then (2.14) is to be replaced by $\text{Prob}\left\{\sum_{i=1}^{t^*} f_i < (1-\lambda)t^*\right\} \leqslant \exp\left\{t^*\left[h(\lambda) + \lambda \log \frac{t}{L}\right]\right\}$. The very same arguments which led to (2.9) yield now

$$\text{Prob}\left\{\min_{j=1,\ldots,J}\left(\min_{m=1,\ldots,M_j} \frac{1}{|E_j^m|} \sum_{i \in E_j^m} f_i^{jm}\right) < 1-\lambda\right\}$$

$$\leqslant \sum_{j=1}^{J}\sum_{m=1}^{M_j} \exp\{|E_j^m|[h(\lambda) + \lambda \log(|E_j|L^{-1})]\}$$

and hence the lemma.

Finally, we present now the most general Coloring Lemma of its kind, which includes both, Lemma 2 and Lemma 3B. Our motivation for aiming at this was again twofold:

I. It is desirable to base coding theory on as few principles as possible.

II. Gallager's problem can be solved with the composition of two colorings, one which is good for "*big edges*" (its existence is guaranteed by Lemma 3A or 3B) and one which is good for "*small edges*" (it can be constructed according to the proof of Lemma 2). This composition is no longer suitable for the limited side information problem of Section 5, because here both times many colors are needed. Its solution requires *one* coloring which is good for "*big*" *and for* "*small*" *subedges simultaneously*. Let $(\mathcal{V}, \mathcal{A}, (\mathcal{F}_E)_{E \in \mathcal{A}})$ and $(\mathcal{V}, \mathcal{B}, (\mathcal{F}_E)_{E \in \mathcal{B}})$ be two 2-hypergraphs with the same vertex set $\mathcal{V}$ and $\mathcal{A} \cap \mathcal{B} = \emptyset$. Define $\mathcal{H}_2 = (\mathcal{V}, \mathcal{A} \cup \mathcal{B}, (\mathcal{F}_E)_{E \in \mathcal{A} \cup \mathcal{B}})$. We are interested in colorings Φ_λ^2 of $\mathcal{H}_2$ which are strict on $(\mathcal{V}, \mathcal{A})$.

Those colorings automatically color all subedges out of $\bigcup_{E \in \mathcal{A}} \mathcal{F}_E$ strictly and we need not be concerned with them. Write $\mathcal{B}$ as $\mathcal{B} = \{E_1, \ldots, E_J\}$ and denote the subedges by E_j^m, $1 \leqslant m \leqslant M_j$, $1 \leqslant j \leqslant J$.

Let $(\mathcal{V}, \mathcal{A}^*)$ be the graph assigned to $(\mathcal{V}, \mathcal{A})$ as previously and let D denote the maximal degree of the vertices in this graph.

We are now prepared to state

COLORING LEMMA 3C *Let* $\mathcal{H}_2 = (\mathcal{V}, \mathcal{A} \cup \mathcal{B}, (\mathcal{F}_E)_{E \in \mathcal{A} \cup \mathcal{B}})$ *be a 2-hypergraph with* $\mathcal{A}$, $\mathcal{B}$, E_j $(1 \leqslant j \leqslant J)$, E_j^m $(1 \leqslant m \leqslant M_j,\ 1 \leqslant j \leqslant J)$ *and* D *as just described.*

For $L \geqslant D + 1 + d$, $\mathcal{H}_2$ *has an* L*-coloring* $\Phi_{2\lambda}^2$, *which is strict on* $(\mathcal{V},$

$\mathcal{A}$), *if*

$$\max_{1 \leqslant j \leqslant J} |E_j|(d - |E_j|)^{-1}(1 - \lambda)\lambda^{-1} \leqslant 1 \tag{2.20}$$

and

$$\sum_{j=1}^{J} \sum_{m=1}^{M_j} 2 \exp \{|E_j^m|(h(\lambda) + \lambda \log (|E_j|d^{-1}))\} < 1. \tag{2.21}$$

Proof. The idea of the proof consists in a combination of the ideas for the proofs of Lemma 2 and Lemma 3B as follows. We color the vertices $v_1, v_2, \ldots$ iteratively as in the proof of Lemma 2 except that now we have, since $L \geqslant D + 1 + d$, at *each step at least d colors available one of which we choose at random according to the uniform distribution on any d available colors,* those with smallest values in $\{1, \ldots, L\}$ for instance. Thus we get a strict L-coloring of $(\mathcal{V}, \mathcal{A})$ as before. What do we get for $(\mathcal{V}, \mathcal{B}, (\mathcal{F}_E)_{E \in \mathcal{B}})$? This random coloring procedure can be described by a sequence of RV's $Y_1, \ldots, Y_I$.

Those RV's are not independent or identically distributed. We overcome this additional difficulty by substituting the functions f_j^m defined in (2.17) by the following two types of functions.

For $m = 1, \ldots, M_j$; $j = 1, \ldots, J$ and $i = 1, \ldots, I$ define RV's

$$g_i^j(X_1, \ldots, X_I) = \begin{cases} 1 \text{ if } Y_i \neq Y_{i'}, \text{ for all } i' < i \text{ with } i' \in E_j \\ 0 \text{ otherwise} \end{cases} \tag{2.22}$$

and

$$G_i^j(X_1, \ldots, X_I) = \begin{cases} -1 \text{ if } Y_i = Y_{i'}, \text{ for some } i' > i \text{ with } i' \in E_j \\ 0 \text{ otherwise} \end{cases} \tag{2.23}$$

Clearly, if

$$\sum_{i \in E_j^m} g_i^j + G_i^j \geqslant (1 - 2\lambda)|E_j^m| \tag{2.24}$$

then at least a fraction of $(1 - 2\lambda)$ vertices in E_j^m is colored correctly within E_j.

We can use

$$\begin{aligned} &\text{Prob} \{ \sum_{i \in E_j^m} g_i^j + G_i^j < (1 - 2\lambda)|E_j^m| \} \\ &\leqslant \text{Prob} \{ \sum_{i \in E_j^m} g_i^j < (1 - \lambda)|E_j^m| \} + \text{Prob} \{ \sum_{i \in E_j^m} G_i^j < -\lambda|E_j^m| \}. \end{aligned} \tag{2.25}$$

As in the previous proof one shows that

$$\text{Prob} \{ \sum_{i \in E_j^m} g_i^j < (1 - \lambda)|E_j^m| \} \leqslant \exp \left\{ \left[h(\lambda) + \lambda \log \frac{|E_j|}{d} \right] |E_j^m| \right\}. \tag{2.26}$$

By symmetry the same bound holds for the second term in (2.25).

For those "who are without belief", for $\alpha < 0$

$$\begin{aligned}\text{Prob}\{\Sigma G_i^j + \lambda|E_j^m| < 0\} &\leqslant \exp\{\alpha\lambda|E_j^m|\}\mathbf{E}\exp\{\alpha\Sigma G_i^j\} \leqslant \\ &\leqslant \exp\{\alpha\lambda|E_j^m|\}\left(\frac{|E_j|}{d}e^{-\alpha} + \frac{d-|E_j|}{d}\right)|E_j^m| \\ &= \exp\{-\alpha(1-\lambda)|E_j^m|\}\left(\frac{|E_j|}{d} + \frac{d-|E_j|}{d}e^{\alpha}\right)|E_j^m|,\end{aligned}$$

that is our old expression.

This and (2.26) imply the result. Q.E.D.

§2 Colorings which are good in average

We introduce a *weighted* hypergraph as a quadruple $(\mathcal{V}, \mathcal{E}, (Q_j)_{j=1}^J, Q)$ where $(\mathcal{V}, \mathcal{E})$: $\mathcal{V} = \{1, \ldots, I\}$, $\mathcal{E} = \{E_1, \ldots, E_J\}$ is a hypergraph, $Q: \mathcal{E} \to \mathbf{R}_+$, $Q_j: E_j \to \mathbf{R}_+$ such that for all i, $1 \leqslant i \leqslant I$, and j, $i \leqslant j \leqslant J$,

$$\sum_{i \in E_j} Q_j(i) \leqslant 1, \quad \sum_{i \in \mathcal{V}} Q(i) \leqslant 1. \tag{2.27}$$

For a coloring Φ of the vertices define for $i = 1, \ldots, I$; $j = 1, \ldots, J$

$$g_i^j = \begin{cases} 1 \text{ if } \Phi(i) = \Phi(i') \text{ for some } i' \in E_j - \{i\} \\ 0 \text{ otherwise} \end{cases}. \tag{2.28}$$

We say that Φ has an average goodness $\bar{\lambda}$ for the weighted hypergraph, if

$$\sum_{j=1}^{J} \sum_{i \in E_j} g_i^j Q_j(i) Q(j) \leqslant \bar{\lambda}. \tag{2.29}$$

Coloring Lemma 4. *The weighted hypergraph* $(\mathcal{V}, \mathcal{E}, (Q_j)_{j=1}^J, Q)$ *has an L-coloring of average goodness* $\bar{\lambda}$, $0 < \bar{\lambda} < 1$, *if*

$$\max_{1 \leqslant j \leqslant J} |E_j| L^{-1} \leqslant \bar{\lambda}. \tag{2.30}$$

Proof. Use standard random L-coloring. Then

$$\mathbf{E}g_i^j(X_1, \ldots, X_I) \leqslant |E_j|L^{-1}$$

and therefore

$$\mathbf{E}\left(\sum_{j=1}^{J} \sum_{i \in E_j} Q_j(i) Q(j) g_i^j(X_1, \ldots, X_I) \leqslant \max_{1 \leqslant j \leqslant J} |E_j| L^{-1}.\right.$$

Remark. This is an abstract version of Cover's argument [16]. Notice that not the AEP-property, but only the value of $\max_{1 \leqslant j \leqslant J} |E_j|$ is important. If we apply for instance the Lemma to Example 1 with the same choice of the hypergraph as there, then we don't need property (e). Adding to $\bar{\lambda}$ the small errors resulting from (c), (d), we immediately get the result of [15].

Before we become color blind we turn now to coverings and then we present the necessary results on typical sequences. As can be seen

already in the previous remarks, they are the *skeleton* for coding theory and make the application of our coloring techniques possible. The reader familiar with chapt. 3 of [6] just has to get used to our notation and can then proceed to Sections 4, 5. For the understanding of Sections 6, 8, 9 complete familiarity with Section 3 is necessary.

Further results on colorings are derived in Section 6, §2, and in Section 7.

§3 Coverings

We present a simple result about coverings, which is the "combinatorial cernel" of Shannon's rate distortion theorem for the DMC ([2]). For a hypergraph $(\mathcal{V}, \mathcal{E})$ denote by $\deg(v)$, $v \in \mathcal{V}$, the number of edges containing v.

$$\mathcal{C} \subset \mathcal{E} \text{ is a covering of } \mathcal{V} \text{ if } \mathcal{V} = \bigcup_{E \in \mathcal{C}} E.$$

Covering Lemma. *If for a hypergraph* $(\mathcal{V}, \mathcal{E})$ $\min_{v \in \mathcal{V}} \deg(v) \geqslant d$, *then there exists a covering* $\mathcal{C}$ *with* $|\mathcal{C}| \leqslant |\mathcal{E}| d^{-1} \log |\mathcal{V}| + 1$.

Proof. Choose edges $E^{(1)}, \ldots, E^{(k)}$ independently at random according to the uniform distribution on $\mathcal{E}$. The probability that v, $v \in \mathcal{V}$, is not covered in k drawings is less than

$$(1 - d|\mathcal{E}|^{-1})^k.$$

The probability that there exists a v which is not covered is less than

$$|\mathcal{V}|(1 - d|\mathcal{E}|^{-1})^k.$$

If this quantity is strictly smaller than 1, then there exists a covering with cardinality k and the result follows.

Remark. The most frequent application is to the hypergraph

$$(\mathcal{T}_{\delta_1}(Y^n), (\mathcal{G}_\delta(Y^n \mid x^n))_{x^n \in \mathcal{T}_\delta(X^n)})$$

(see Lemma G4).

From Section 3 we know

$$|\mathcal{V}| = \exp\{H(Y)n + 0(\sqrt{n})\} \qquad \text{(Lemma T1 (c))}$$
$$|d| = \exp\{H(X \mid Y)n + 0(\sqrt{n})\} \qquad \text{(Lemma G4 (c))}$$
$$|\mathcal{E}| = \exp\{H(X)n + 0(\sqrt{n})\} \qquad \text{(Lemma T1 (c)).}$$

Hence there is a covering $\mathcal{C} = (\mathcal{G}_{\delta_1}(Y^n \mid x_i^n))_{i=1}^c$ of $\mathcal{T}_\delta(Y^n)$ with $c = \exp\{I(X \wedge Y)n + 0(\sqrt{n})\}$ edges.

Section 3. Typical Sequences and Generated Sequences

For the ease of reference we give here the notions of typical sequences and generated sequences and those of their properties needed in the sequel.

They were intuitively described already by Shannon in [1] and also used by Feinstein in [4]. Wolfowitz was the first to rigorously formalize them ([5]) and emphasize in his work their importance for Information Theory (see [6]) which is simply due to the fact that those sequences carry essentially equal probabilities and thus everything can almost be reduced to counting. In [6] those notions are defined within $\sqrt{}$-deviations, here we also use typical sequences and generated sequences with exact compositions, and thus we can count exactly. With this skeleton in our back we can view many source coding problems just as coloring problems for hypergraphs. Lemmata, which can be found in [6] or can be proved with easy modifications of proofs given there, will be stated without proof. Thus only the Lemmata in paragraphs 3, 4, 5 require a proof.

§1. Typical sequences

Let $X^n = X_1, \ldots, X_n$ be a sequence of i.i.d. RV's with values in $\mathcal{X}$ and distribution p. For $x^n \in \mathcal{X}^n$ and $x \in \mathcal{X}$ denote by $n(x \mid x^n)$ the number of components in which x^n has x.

x^n is (X^n, δ)-typical (or (p, δ)-typical) if for $\delta \geqslant 0$

$$|np(x) - n(x \mid x^n)| \leqslant \delta\sqrt{np(x)(1 - p(x))} \text{ for all } x \in \mathcal{X}. \tag{3.1}$$

Denote the set of those sequences by $\mathcal{T}_\delta(X^n)$ or by $\mathcal{X}^n_\delta(p)$. Those two notations allow to focus on the RV's or the distribution. The set $\mathcal{T}_0(X^n)$ is of particular interest. Its elements are of exact type. Of course $\mathcal{T}_0(X^n)$ is non-empty only if for all x

$$p(x) = \frac{n_x}{n},\ n_x \text{ integral.} \tag{3.2}$$

We denote the set of those PD's by $\mathcal{P}_0(n, \mathcal{X})$. Clearly $|\mathcal{P}_0(n, \mathcal{X})| \leqslant (n+1)^{|\mathcal{X}|}$. Frequently, in order to save notation we make use of Landau's symbol 0:

For

$$f, g\colon \mathbf{R} \to \mathbf{R} \text{ (the reals) } f = 0(g) \text{ if } \overline{\lim_{r\to\infty}} \frac{|f(r)|}{|g(r)|} < \infty.$$

In all cases where we use those symbols, actually numerical bounds can easily be given. *Also all the $0(\sqrt{n})$ occurring below are less than a universal constant $c = c(|\mathcal{X}|, |\mathcal{Y}|, |S|)$ independent of PD's.*

LEMMA T1. *(a) For every n*

$$\text{Prob}\,(X^n \in \mathcal{T}_\delta(X^n)) = 1 - 0\left(\frac{1}{\delta^2}\right)$$

that is, the probability goes to 1 uniformly in n, if $\delta \to \infty$.

(b) $\text{Prob}\,(X^n = x^n) = \exp\{-H(X)n + 0(\sqrt{n})\}$ *for* $x^n \in \mathcal{T}_\delta(X^n)$.

(c) $|\mathcal{T}_\delta(X^n)| = \exp\{H(X)n + 0(\sqrt{n})\}$.

LEMMA T2. *If the distribution* $P_X \in \mathcal{P}_0(n, \mathcal{X})$, *then*

$$|\mathcal{T}_0(X^n)| = \exp\{H(X)n + 0(\log n)\}.$$

The elements of $\mathcal{T}_0(X^n)$ have equal probabilities of value

$$\exp\{-H(X)n + 0(\log n)\}. \quad \text{(Stirling!)}$$

LEMMA T3. *For* $p \in \mathcal{P}_0(n, \mathcal{X})$

$$|\mathcal{T}_\delta(X^n) \cap \mathcal{X}_0^n(p)| = \begin{cases} \text{or } 0 \\ \exp\{H(X)n + 0(\sqrt{n})\} \end{cases}$$

§2. GENERATED SEQUENCES

Let $\mathcal{S}$ be a finite set and suppose that for $s^n \in \mathcal{S}^n$ $X(s^n) = X_1(s_1), \ldots, X_n(s_n)$ is a sequence of independent RV's with distributions depending only on the s_t's. Denote by X^n as sequence of RV's with conditional distributions

$$\text{Prob}\,(X^n = x^n \mid s^n) = \text{Prob}\,(X(s^n) = x^n),\ s^n \in \mathcal{S}^n. \tag{3.3}$$

Abbreviate Prob $(X(s) = x)$ as $p(x \mid s)$ and Prob $(X(s^n) = x^n)$ as $p(x^n \mid s^n)$. For $x^n \in \mathcal{X}^n$, $s^n \in \mathcal{S}^n$, $x \in \mathcal{X}$, $s \in \mathcal{S}$ denote by $n(x, s \mid x^n, s^n)$ the number of positions in which x^n has x and s^n has s.

A sequence $x^n \in \mathcal{X}$ is $(X^n \mid s^n, \delta)$-generated (or $(X(s^n), \delta)$-typical) if

$$|n(x, s \mid x^n, s^n) - n(s \mid s^n)p(x \mid s)| \leqslant \delta\sqrt{n(s \mid s^n)p(x \mid s)(1 - p(x))} \quad \text{for all } x \in \mathcal{X},\ s \in \mathcal{S}. \tag{3.4}$$

Denote the set of those sequences by

$$\mathcal{G}_\delta(X^n \mid s^n) \text{ or } \mathcal{T}_\delta(X(s^n)) \text{ or } \mathcal{X}_\delta^n(p(\cdot \mid s^n)).$$

LEMMA G1. (*a*) *For every n*

$$\text{Prob}\,(X^n \in \mathcal{G}_\delta(X^n \mid s^n) \mid s^n) = \text{Prob}\,(X(s^n) \in \mathcal{T}_\delta(X(s^n)) = 1 - 0\left(\frac{1}{\delta^2}\right)$$

(*b*) $\text{Prob}\,(X(s^n) = x^n) = \exp\{-H(X(s^n)) + 0(\sqrt{n})\}$

(*c*) $|\mathcal{G}_\delta(X^n \mid s^n)| = \exp\{H(X(s^n)) + 0(\sqrt{n})\}$.

Clearly,

$$H(X(s^n)) = \sum_{t=1}^n H(X_t(s_t)) = \sum_{t=1}^n H(X \mid s_t). \tag{3.5}$$

If S has distribution $(P_S(s))_{s \in \mathcal{S}} = \left(\frac{n(s \mid s^n)}{n}\right)_{s \in \mathcal{S}}$, then by the linearity of conditional entropies

$$\sum_{t=1}^n H(X \mid s_t) = nH(X \mid S). \tag{3.6}$$

Also, if $s^n \in \mathcal{T}_\delta(S^n)$, $S^n = S_1, \ldots, S_n$ i.i.d. RV's, then

$$\sum_{t=1}^n H(X \mid s_t) = nH(X \mid S) + 0(\sqrt{n}). \tag{3.7}$$

LEMMA G2. *If for every* $s \in \mathcal{S}$ $p(\cdot \mid s) \in \mathcal{P}_0(n, \mathcal{X})$, *then*

$$|\mathcal{G}_0(X^n \mid s^n)| = |\mathcal{T}_0(X(s^n))| = \exp\{H(X(s^n)) + 0(\log n)\}.$$

The elements of $\mathcal{G}_0(X^n \mid s^n)$ have *equal* probabilities of value

$$\exp\{-H(X(s^n)) + 0(\log n)\}.$$

LEMMA G3 *If for every* $s \in \mathcal{S}$ $p(\cdot \mid s) \in \mathcal{P}_0(n(s \mid s^n), \mathcal{X})$, *then*

$$|\mathcal{G}_\delta(X^n \mid s^n) \cap \mathcal{X}_0^n(p(\cdot \mid s^n))| = \begin{cases} 0 \\ \text{or} \\ \exp\{H(X(s^n)) + 0(\sqrt{n})\}. \end{cases}$$

LEMMA G4 *Let* $(X_t, S_t)_{t=1}^\infty$ *be a DMCS. For every* $\delta \geqslant 0$ *there is a* $\delta_1 = \delta_1(|\mathcal{X}|, |\mathcal{Y}|, \delta)$ *such that for all* n

(a) $\mathcal{G}_\delta(X^n \mid s^n) \subset \mathcal{T}_{\delta_1}(X^n)$ for $s^n \in \mathcal{T}_\delta(S^n)$,

(b) $\bigcup_{s^n \in \mathcal{T}_\delta(S^n)} \mathcal{G}_{\delta_1}(X^n \mid s^n) \supset \mathcal{T}_\delta(X^n)$.

(c) *If* $x^n \in \mathcal{T}_{\delta_1}(X^n)$, *then* x^n *is contained in*

$\exp\{H(S \mid X)n + 0(\sqrt{n})\}$ *sets* $\mathcal{G}_\delta(X^n \mid s^n)$ *with* $s^n \in \mathcal{T}_\delta(S^n)$.

LEMMA G5 *Let* $(X_t, S_t)_{t=1}^\infty$ *be a DMCS. For every* $\delta \geqslant 0$ *there is a* $\delta_2(|\mathcal{X}|, |\mathcal{S}|, \delta)$ *such that for all* n:

$$x^n \in \mathcal{G}_\delta(X^n \mid s^n) \text{ for } s^n \in \mathcal{T}_\delta(S^n) \text{ implies } (x^n, s^n) \in \mathcal{T}_{\delta_2}(X^n, S^n).$$

§3 GENERATED SEQUENCES IN A MARKOV CHAIN $X \to S \to U$

The results of this paragraph are needed in Section 5, and only there.

Let X, S, U be RV's forming a Markov chain $X \to S \to U$. We consider triples $(X^n, S^n, U^n) = (X_1, \ldots, X_n, S_1, \ldots, S_n, U_1, \ldots, U_n)$, where the $(X_t, S_t, U_t)_{t=1}^n$ are i.i.d. with (X_t, S_t, U_t) having the same distribution as (X, S, U).

Using the abbreviations

$$\begin{aligned} w(x \mid s) &= \text{Prob}\,(X = x \mid S = s), \\ v(s \mid u) &= \text{Prob}\,(S = s \mid U = u), \\ p(x \mid u) &= \text{Prob}\,(X = x \mid U = u), (x, s, u) \in \mathcal{X} \times \mathcal{S} \times \mathcal{U}, \end{aligned} \tag{3.8}$$

we can write

$$p(x \mid u) = \sum_s w(x \mid s) v(s \mid u) \tag{3.9}$$

and by the Markov property

$$\text{Prob}\,(X = x \mid S = s, U = u) = w(x \mid s). \tag{3.10}$$

LEMMA M1 *With the assumptions of this paragraph* $x^n \in \mathcal{G}_\delta(X^n \mid s^n, u^n)$, $s^n \in \mathcal{G}_\delta(S^n \mid u^n)$ *imply* $x^n \in \mathcal{G}_{\delta^*}(X^n \mid u^n)$ *for a suitable constant* $\delta^* = \delta^*(\delta)$.

Proof. One just has to use the definitions of the sets involved and

make the necessary substitutions. Since to our knowledge this has nowhere appeared in print, we carry it out.

Our assumptions say that for all $(x, s, u) \in \mathcal{X} \times \mathcal{S} \times \mathcal{U}$

$$|n(xsu \mid x^n s^n u^n) - n(su \mid s^n u^n) w(x \mid s)| \leqslant \delta[n(su \mid s^n u^n) w(x \mid s)(1 - w(x \mid s))]^{1/2}, \tag{3.11}$$

$$|n(su \mid s^n u^n) - n(u \mid u^n) v(s \mid u)| \leqslant \delta[n(u \mid u^n) v(s \mid u)(1 - v(s \mid u))]^{1/2}. \tag{3.12}$$

We have to show that for $(x, u) \in \mathcal{X} \times \mathcal{U}$

$$|n(xu \mid x^n u^n) - n(u \mid u^n) p(x \mid u)| \leqslant \delta^*[n(u \mid u^n) p(x \mid u)(1 - p(x \mid u))]^{1/2}. \tag{3.13}$$

CASE 1. $0 < p(x \mid u) < 1$.

It follows from (3.11) and (3.12) that

$$|n(xsu \mid x^n s^n u^n - n(u \mid u^n) w(x \mid s) v(s \mid u)| \leqslant \delta[n(su \mid s^n u^n) w(x \mid s)(1 - w(x \mid s))]^{1/2} + \delta[n(u \mid u^n) v(s \mid u)(1 - v(s \mid u))]^{1/2}$$

and therefore by (3.9) that

$$\begin{aligned} |n(xu \mid x^n u^n) - n(u \mid u^n) p(x \mid u)| &\leqslant \sum_s |n(xsu \mid x^n s^n u^n) - n(u \mid u^n) w(x \mid s) v(s \mid u)| \\ &\leqslant \delta \sum_s [n(su \mid s^n u^n) w(x \mid s)(1 - w(x \mid s))]^{1/2} \\ &\quad + [n(u \mid u^n) v(s \mid u)(1 - v(s \mid u))]^{1/2}. \end{aligned}$$

Since $p(x \mid u)(1 - p(x \mid u)) > 0$, there is a constant $\delta^*(x, u)$ such that the last quantity is upper bounded by

$$\delta^*(x, u)[n(u \mid u^n) p(x \mid u)(1 - p(x \mid u))]^{1/2}.$$

For $\delta^* = \max \{\delta^*(x, u) : (x, u) \text{ with } 0 < p(x \mid u) < 1\}$ (3.13) holds in this case.

CASE. 2 $p(x \mid u)(1 - p(x \mid u)) = 0$.

SUBCASE 1. $p(x \mid u) = 0$.

Now (3.9) implies for every s $w(x \mid s)\, v(s \mid u) = 0$. If $w(x \mid s) = 0$, then (3.11) implies for those s

$$n(xsu \mid x^n s^n u^n) = 0, \tag{3.14}$$

and if $v(s \mid u) = 0$, then (3.12) gives for those s

$$0 = n(su \mid s^n u^n) = \sum_x n(xsu \mid x^n s^n u^n) \tag{3.15}$$

and hence again (3.14). Therefore, $n(xu \mid x^n u^n) = \sum_s n(xsu \mid x^n s^n u^n) = 0$ and (3.13) holds.

SUBCASE 2. $p(x \mid u) = 1$.

Here for every s, either $v(s \mid u) = 0$ and we have by (3.15)

$$n(su \mid s^n u^n) = n(xsu \mid x^n s^n u^n), \tag{3.16}$$

or $w(x \mid s) = 1$, and then by (3.11) we get again (3.16). This implies that

$$n(u \mid u^n) = \sum_s n(su \mid s^n u^n) = \sum_s n(xsu \mid x^n s^n u^n) = n(xu \mid x^n u^n)$$

and therefore again (3.13).

LEMMA M2. *If* $u^n \in \mathcal{T}_\delta(U^n)$ *and* $s^n \in \mathcal{G}_\delta(S^n \mid u^n)$, *then*

(a) $|\mathcal{G}_\delta(X^n \mid s^n u^n)| = \exp\{H(X \mid S)n + 0(\sqrt{n})\}$,

(b) $\text{Prob}\,(X(s^n) \in \mathcal{G}_\delta(X^n \mid s^n, u^n)) = 1 - 0\left(\frac{1}{\delta^2}\right)$.

Proof. By Lemma G5 $(s^n, u^n) \in \mathcal{T}_{\delta_2}(S^n, X^n)$. This, (3.7) and Lemma G1 (a) imply $|\mathcal{G}_\delta(X^n \mid s^n u^n)| = \exp\{H(X \mid SU)n + 0(\sqrt{n})\}$. Since by the Markov property $H(X \mid SU) = H(X \mid S)$, (a) follows. Lemma G1 (a) implies that $\text{Prob}\,(X(s^n, u^n) \in \mathcal{G}_\delta(X^n \mid s^n, u^n)) = 1 - 0\left(\frac{1}{\delta^2}\right)$. By the Markov property for all $x^n \in \mathcal{X}^n$

$$\text{Prob}\,(X(s^n, u^n) = x^n) = \text{Prob}\,(X(s^n) = x^n)$$

and hence (b).

LEMMA M3. *If for every* $(s, u) \in \mathcal{S} \times \mathcal{U}$ $q(\cdot \mid s, u) \in \mathcal{P}_0(n(su \mid s^n u^n), \mathcal{X})$, *then*

$$|\mathcal{G}_\delta(X^n \mid s^n, u^n) \cap \mathcal{X}_0(q(\cdot \mid s^n, u^n))| = \begin{cases} 0 \\ \text{or} \\ \exp\{H(X(s^n) + 0(\sqrt{n})\} \end{cases}.$$

Proof. By the Markov property $H(X(s^n, u^n)) = H(X(s^n))$. With this observation the result follows from Lemma G3.

§4 CROSS-SECTIONS

The results of this paragraph are needed in Sections 6, 8, 9. Let $(\{P(\cdot, \mid s^n) : s^n \in \mathcal{S}^n\})_{n=1}^\infty$ be an AVCS as defined in §3 of Section 1. X^n, Y^n are RV's with conditional distributions

$$\text{Prob}\,(X^n = x^n,\ Y^n = y^n \mid s^n) = p(x^n, y^n \mid s^n),\ (x^n, y^n, s^n) \in \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{S}^n.$$

Again we use RV's $X(s^n)$, $Y(s^n)$ with distribution

$$\text{Prob}\,(X(s^n) = x^n,\ Y(s^n) = y^n) = p(x^n, y^n \mid s^n).$$

The *cross-sections* $\mathcal{G}_\delta(X^n, Y^n \mid s^n)_{\mid x^n}$, $x^n \in \mathcal{X}^n$, are defined by

$$\mathcal{G}_\delta(X^n, Y^n \mid s_n)_{\mid x^n} = \{y^n : (y^n, x^n) \in \mathcal{G}_\delta(X^n, Y^n \mid s^n)\}. \tag{3.17}$$

LEMMA CS. *For all* $x^n \in \mathcal{X}^n$

$$|\mathcal{G}_\delta(X^n, Y^n \mid s^n)_{\mid x^n}| \leqslant \exp\{H(Y(s^n) \mid X(s^n)) + 0(\sqrt{n})\}.$$

Proof. $(x^n, y^n) \in \mathcal{G}_\delta(X^n, Y^n \mid s^n)$ means by definition that for all $(x, y, s) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{S}$

$$|n(xys \mid x^n y^n s^n) - n(s \mid s^n) p(x, y \mid s)| \leqslant \delta\,[n(s \mid s^n) p(x, y \mid s)(1 - p(x, y \mid s))]^{1/2}. \tag{3.18}$$

This implies that for all $(x, s) \in \mathcal{X} \times S$

$$|n(xs \mid x^n s^n) - n(s \mid s^n)p(x \mid s)| \leqslant \delta \sum_y [n(s \mid s^n)p(x, y \mid s)(1 - p(x, y \mid s))]^{1/2}$$

$$\leqslant \delta_1[n(s \mid s^n)p(x \mid s)]^{1/2}, \ \delta_1 = \delta_1(\delta) \text{ suitable.} \tag{3.19}$$

Thus x^n is not necessarily $(X^n \mid s^n, \delta)$-generated in the sense of definition (3.4), instead we have a slightly different notion of generation here. This alternation is necessary, because the case $p(x \mid s) = 1$ causes a technical difficulty. However, denoting the set of sequences x^n generated in the sense of (3.19) by $\mathcal{G}^*(X^n \mid s^n)$ one readily shows with Stirling's formula for instance that

$$|\mathcal{G}^*(X^n \mid s^n)| \leqslant \exp\{H(X(s^n)) + 0(\sqrt{n})\}. \tag{3.20}$$

Now, since $p(y \mid xs)p(x \mid s) = p(y, x \mid s)$, (3.19) implies

$$|n(xys \mid x^n y^n s^n) - n(xs \mid x^n s^n)p(y \mid xs)|$$

$$\leqslant |n(xys \mid x^n y^n s^n) - n(s \mid s^n)p(x, y \mid s)| + \delta_1[n(s \mid s^n)p(x \mid s)]^{1/2}p(y \mid xs)$$

which by (3.18) is smaller than

$$\delta[n(s \mid s^n)p(x, y \mid s)]^{1/2} + \delta_1[n(s \mid s^n)p(x \mid s)]^{1/2}p(y \mid xs).$$

Since $p(x, y \mid s) = 1$ implies $p(y \mid xs) = 1$ there is a constant δ_2 such that

$$|n(xys \mid x^n y^n s^n) - n(xs \mid x^n s^n)p(y \mid xs)| \leqslant \delta_2[n(s \mid s^n)p(y \mid xs)]^{1/2}. \tag{3.21}$$

This, (3.19) and Stirling's formula yield

$$|\mathcal{G}_\delta(X^n, Y^n \mid s^n)_{|x^n|}| \leqslant \exp\{H(Y(s^n) \mid X(s^n)) + 0(\sqrt{n})\} \tag{3.22}$$

Q.E.D.

§5 Carriers

In Sections 6 and 9 we make use of the

Carrier Lemma. *For an AVCS*

(*a*) $$\bigcup_{s^n \in \mathcal{S}^n} \mathcal{G}_\delta(X^n, Y^n \mid s^n) \subset \bigcup_{\bar{s} : \bar{s} = \Sigma p_i s_i,\ p \in \mathcal{P}_0(n, S)} \mathcal{T}_{\delta_1}(X^n(\bar{s}), Y^n(\bar{s}))$$

for a suitable constant $\delta_1 = \delta_1(\delta)$.

(*b*) $$\left|\bigcup_{s^n \in \mathcal{S}^n} \mathcal{G}_\delta(X^n, Y^n \mid s^n)\right| \leqslant \exp\{\max_{\bar{s} \in \bar{S}} H(X(\bar{s}), Y(\bar{s}))n + 0(\sqrt{n})\}.$$

Proof. Since $|\mathcal{P}_0(n, \mathcal{S})| \leqslant (n+1)^{|S|}$ (b) is a consequence of (a) and Lemma T1 (c).

To show (a) is again a routine matter. If $(x^n, y^n) \in \mathcal{G}_\delta(X^n, Y^n \mid s^n)$ then (3.18) holds and hence for $\bar{s} = \sum_{t=1}^n \frac{1}{n} s_t$ and all x, y

$$|n(x, y \mid x^n, y^n) - np(x, y \mid \bar{s})| \leqslant \delta \sum_s [n(s \mid s^n)p(x, y \mid s)(1 - p(x, y \mid s))]^{1/2}.$$

Now realize that $p(x, y \mid \bar{s}) = 0$ implies $np(x, y \mid \bar{s}) = \sum_s n(s \mid s^n)p(x, y \mid s) = 0$ and hence that the right side in (3.23) equals 0.

Also, if $p(x, y \mid \bar{s}) = 1$, then for every s either $p(x, y \mid s) = 1$ or $n(s \mid s^n) = 0$ and again the right side in (3.23) vanishes. Therefore there exists a δ_1 such that for all x, y

$$|n(x, y \mid x^n, y^n) - np(x, y \mid \bar{s})| \leqslant \delta_1[np(x, y \mid \bar{s})(1 - p(x, y \mid \bar{s}))]^{1/2}, \quad (3.24)$$

which means that $(x^n, y^n) \in \mathcal{T}_{\delta_1}(X^n(\bar{s}), Y^n(\bar{s}))$ and the proof is complete.

Section 4. Proof of Theorem 1

That $R_D \geqslant H^* = \max_{s \in S} H(X(s))$ follows from the source coding theorem for the standard DMS, the heart of the matter is to show the opposite inequality. For tutorial reasons we first make the supposition

$$H_* = \min_{s \in S} H(X(s)) > 0, \quad (4.1)$$

which we then remove.

Consider the hypergraph $\mathcal{H} = (\mathcal{X}^n, (\mathcal{G}_\delta(X^n(s^n))_{s^n \in \mathcal{S}^n}$.

We know from Lemma G1 and (3.5) that for all $s^n \in \mathcal{S}^n$

$$\exp\{H_* n + 0(\sqrt{n})\} \leqslant |\mathcal{G}_\delta(X^n \mid s^n)| \leqslant \exp\{H^* n + 0(\sqrt{n})\}. \quad (4.2)$$

Choose $\lambda = \lambda(n) = \exp\{-3c\sqrt{n}\}$ and $L(n) = \exp\{H^* n + 7c\sqrt{n}\}$, then, as in example 1, (2.5) and (2.6) hold and Lemma 3A guarantees the existence of a coloring $\Phi_{2\lambda}$ with $L(n)$ colors. Therefore for *every* $s^n \in \mathcal{S}^n$ the set $\mathcal{G}_\delta(X^n \mid s^n)_{\text{inc}}$ of incorrectly colored elements in $\mathcal{G}_\delta(X^n \mid s^n)$ satisfies

$$|\mathcal{G}_\delta(X^n \mid s^n)_{\text{inc}}| \leqslant 2\lambda |\mathcal{G}_\delta(X^n \mid s^n)|. \quad (4.3)$$

This and (b), (c) of Lemma G1 imply

$$\text{Prob}(X^n \in \mathcal{G}_\delta(X^n \mid s^n)_{\text{inc}} \mid s^n) \leqslant \exp\{-3c\sqrt{n}\} \exp\{H(X(s^n)) + c\sqrt{n}\} \exp\{-H(X(s^n)) + c\sqrt{n}\} \leqslant \exp\{-c\sqrt{n}\}. \quad (4.4)$$

This, and (a) of Lemma G1 imply that the decoder, knowing s^n, can reproduce $X(s^n)$ for every s^n with an arbitrarily small error probability for large n, if he uses the decoding function

$$F_n(l, s^n) = \begin{cases} \mathcal{G}_\delta(X^n \mid s^n) \cap f_n^{-1}(l) \text{ if this set contains exactly one element} \\ \text{any decision otherwise} \end{cases} \quad (4.5)$$

Here, $f_n = \Phi_{2\lambda}$.

Now we remove the assumption $H_* > 0$, that is, we allow conditional PD's to be concentrated on a single letter. An information theorist may feel that this is just a small technical point. However, we know from §4 in Section 1 that for AVCS this point is crucial. Since edges with very small cardinality—even with cardinality 1—now occur, (2.6) is no longer satisfied. This is not only due to the bounds, but lies in the nature of random coloring: the probability of coloring a small edge essentially correctly is not large enough to cope simultaneously with all the edges. Of course in problems for which colorings, which are good in average, can

be applied, this difficulty does not arise. The situation is similar to the fact that in channel coding for small rates random coding performs poorly.

We overcome the present difficulty with the help of Coloring Lemma 2.

Define

$$\mathcal{S}_\delta(n) = \{s^n : s^n \in \mathcal{S}^n \text{ with } |\mathcal{G}_\delta(X^n \mid s^n)| \leqslant \exp\{\log^2 n\}\}. \tag{4.6}$$

Then the hypergraph

$$(\mathcal{X}^n, (\mathcal{G}_\delta(X^n \mid s^n))_{s^n \in \mathcal{S}^n - \mathcal{S}_\delta(n)})$$

can be colored, with the same choice of $\lambda(n)$ and $L(n)$ as before, appropriately, because for $s^n \in \mathcal{S}^n - \mathcal{S}_\delta(n)$

$$h(\lambda(n)) + \lambda(n) \log (|\mathcal{G}_\delta(X^n \mid s^n)| L(n)^{-1}) \leqslant -1, \tag{4.7}$$

as before, and clearly

$$|\mathcal{S}^n - \mathcal{S}_\delta(n)| \exp\{-e\log^3 n\} < 1 \text{ for } n \geqslant n_0, \text{ suitable.} \tag{4.8}$$

We color now $(\mathcal{X}^n, (\mathcal{G}_\delta(X^n \mid s^n))_{s^n \in \mathcal{S}_\delta(n)})$ strictly by applying Coloring Lemma 2. What is the maximal vertex degree D of the associated graph?

Define $H_+ = \min\{H(X(s)) : s \in \mathcal{S}, H(X(s)) > 0\}$.

If $k(s^n) = |\{s_t : 1 \leqslant t \leqslant n, H(X(s_t)) > 0\}|$, then by definition (3.4)

$$|\mathcal{G}_\delta(X^n \mid s^n)| \geqslant \exp\{H_+ k(s^n) - c\sqrt{k(s^n)}\}$$

and for $s^n \in \mathcal{S}_\delta(n)$

$$\exp\{\log^2 n\} \geqslant \exp\{H_+ k(s^n) - c\sqrt{k(s^n)}\}. \tag{4.9}$$

This implies

$$k(s^n) \leqslant 0(\log^2 n) \text{ for } s^n \in \mathcal{S}_\delta(n).$$

Therefore two connected vertices x^n and x'^n in the graph necessarily have Hamming distance less than $0(\log^2 n)$. This implies that

$$\deg(x^n) \leqslant \binom{n}{0(\log^2 n)} |\mathcal{X}|^{0(\log^2 n)} \leqslant (n|\mathcal{X}|)^{0(\log^2 n)}. \tag{4.10}$$

The graph can therefore be colored strictly with $(n|\mathcal{X}|)^{0(\log^2 n)}$ colors. Denote such a coloring by ψ_0 and define the encoding function by

$$f_n(x^n) = (\Phi_{2\lambda}(x^n), \psi_0(x^n)), \; x^n \in \mathcal{X}^n. \tag{4.11}$$

Finally, define the decoding function by

$$F_n((l_1, l_2), s^n) = \begin{cases} \mathcal{G}_\delta(X^n \mid s^n) \cap \Phi_{2\lambda}^{-1}(l_1) \text{ if } s^n \in \mathcal{S}^n - \mathcal{S}_\delta(n) \\ \text{and this set contains exactly 1 element} \\ \mathcal{G}_\delta(x^n \mid s^n) \cap \psi_0^{-1}(l_2) \text{ if } s^n \in \mathcal{S}_\delta(n) \\ \text{any decision otherwise.} \end{cases} \tag{4.12}$$

Since $\|f_n\| \leqslant \|\Phi_{2\lambda}\| \, \|\psi_0\| \leqslant \exp\{H^* n + 7c\sqrt{n} + 0(\log^3 n)\}$, the proof is complete.

Remark. An alternate proof based on Coloring Lemma 3B can be given. The argument is carried out in detail in the proof of the more general Theorem 3 in Section 8. Here we outline the proof to make the reader familiar with the idea, which we also use in Sections 5, 8 and 9. Let us consider the case $H_* > 0$. Define the 2-hypergraph

$$\mathcal{H}_2 = (\mathcal{V}, \mathcal{E}, (\mathcal{E}_{s^n})_{s^n \in \mathcal{S}^n}) \text{ with } \mathcal{V} = \mathcal{X}^n,\ \mathcal{E} = (\mathcal{G}_\delta(X^n \mid s^n))_{s^n \in \mathcal{S}^n},$$

and
$$\mathcal{E}_{s^n} = \{\mathcal{G}_\delta(X^n \mid s^n) \cap \mathcal{X}_0^n(p(\cdot \mid s^n)) : p(\cdot \mid s) \in \mathcal{P}_0(n(s \mid s^n), \mathcal{X})\}. \quad (4.13)$$

Choose $L(n)$ as before and keep λ constant. Lemma 3B can be applied, because by Lemma G3 the cardinality of the non-empty subedges is not smaller than $\exp\{H_* n - c\sqrt{n}\}$,

$$h(\lambda) + \lambda \log (|\mathcal{G}_\delta(X^n \mid s^n)|L^{-1}) \leqslant h(\lambda) - \lambda 6c\sqrt{n} \leqslant -1 \ (n \geqslant n_0),$$

and because $\sum_{s^n} |\mathcal{E}_{s^n}| \leqslant |\mathcal{S}|^n (n+1)^{|\mathcal{X}||\mathcal{S}|}$. Since the subedges partition the edges and since Prob $(X^n = x^n \mid s^n)$ is the *same* for all x^n in a subedge of $\mathcal{G}_\delta(X^n \mid s^n)$, we conclude that

$$\text{Prob}\,(X^n \in \mathcal{G}_\delta(X^n \mid s^n)_{\text{inc}} \mid s^n) \leqslant 2\lambda. \quad (4.14)$$

Define f_n and F_n as before, then (4.14) and Lemma G1 imply that the decoding error probability is smaller than $0\left(\frac{1}{\delta^2}\right) + 2\lambda$ uniformly for all s^n. The case $H_* = 0$ goes essentially as in the previous proof. Here small *subedges* make difficulties for random coloring, but if if they are small, then by Lemma G3 their edge is of the same magnitude. The proof can be completed as before.

Section 5. Proof of Theorem 2

§ 1 The converse: $\mathcal{R}_{\text{DP}} \subset \mathcal{R}^{**}$

This is a simple consequence of the results of [17]. For let (f_n, g_n, F_n) be a code with error probability λ, the uniformity in the error concept (1.8) implies that also for any RV S^n with values in $\mathcal{S}^n$

$$\text{Prob}\,\{F_n(f_n(X^n(S^n)), g_n(S^n)) \neq X^n(S^n)\} \leqslant e(f_n, g_n, F_n) \leqslant \lambda. \quad (5.1)$$

Let us use this fact in the special case $S^n = ((S_p)_1, \ldots, (S_p)_n)$, where the $(S_p)_t$'s are independent with distribution p.

Now we apply the converse to the source coding problem with side information of [17] and obtain:

There exists a RV U_p, $\|U_p\| \leqslant |\mathcal{S}| + 2$, $U_p \to S_p \to X_p$, such that

$$\log \|f_n\| \geqslant n(H(X_p \mid U_p) - h(\lambda) - \lambda \log |\mathcal{X}|) \quad (5.2)$$

and
$$\log \|g_n\| \geqslant nI(S_p \wedge U_p). \quad (5.3)$$

Since these inequalities hold for all p, $p \in \mathcal{P}(\mathcal{S})$, the inclusion $\mathcal{R}_{\text{DP}} \subset \mathcal{R}^{**}$ follows.

(Obviously, if one applies the strong converse for the above source coding problem of [19], one gets also the strong converse for the present problem.)

§2 THE DIRECT PART: $\mathcal{R}_{DP} \supset \mathcal{R}^{**}$

This more difficult result is derived with the help of our coloring techniques (Lemma 3C) in conjunction with the Covering Lemma of Section 2. After we have those tools, we just have to set up the appropriate 2-hypergraph. For this we need the definitions and results of Section 3.

Step 1. Partitioning of $\mathcal{S}^n$.

Fix a $p, p \in \mathcal{P}_0(n, \mathcal{S})$, and a RV S_p with distribution p. Let U_p be such that $U_p \to S_p \to X_p$. By the Covering Lemma there exist $(u_p^i)_{i=1}^{N_p} \subset \mathcal{T}_\delta(U_p^n)$ with

$$N_p = \exp\{I(U_p \wedge S_p)n + 0(\sqrt{n})\} \tag{5.2}$$

such that

$$(\mathcal{G}_\delta(S_p^n \mid u_p^i))_{i=1}^{N_p} \text{ covers } \mathcal{T}_0(S_p^n).$$

From this covering we pass to a *partition*

$$(\mathcal{D}_p^i)_{i=1}^{N_p} \text{ of } \mathcal{T}_0(S_p^n)$$

$$\text{with } \mathcal{D}_p^i \subset \mathcal{G}_\delta(S_p^n \mid u_p^i) \text{ for } i = 1, \ldots, N_p. \tag{5.3}$$

Clearly,

$$\{\mathcal{D}_p^i\colon 1 \leqslant i \leqslant N_p, p \in \mathcal{P}_0(n, \mathcal{S})\} \text{ is a partition of } \mathcal{S}^n.$$

Step 2. Definition of the 2-hypergraph $\mathcal{H}_2$.

Choose

$$\mathcal{H} = (\mathcal{V}, \mathcal{A} \cup \mathcal{B}, (\mathcal{F}_E)_{E \in \mathcal{A} \cup \mathcal{B}}), \mathcal{A} \cap \mathcal{B} = \emptyset, \text{ as follows:}$$

(1) $\mathcal{V} = \mathcal{X}^n$.

(2) Partition $\mathcal{P}_0(n, \mathcal{X})$ into 2 sets, $\mathcal{P}_a(n, \mathcal{S})$ and $\mathcal{P}_b(n, \mathcal{S})$, where

$$\mathcal{P}_a(n, \mathcal{S}) = \{p : p \in \mathcal{P}_0(n, \mathcal{S}) \text{ with } H(X_p \mid S_p)n \leqslant 2c\sqrt{n}\}. \tag{5.4}$$

Define now for $E_p^i = \bigcup_{s^n \in \mathcal{D}_p^i} \mathcal{G}_\delta(x_p^n \mid s^n, u_p^i)$

$$\mathcal{A} = (E_p^i)_{\substack{i=1,\ldots,N_p \\ p \in \mathcal{P}_a(n, \mathcal{S})}} \tag{5.5}$$

$$\mathcal{B} = (E_p^i)_{\substack{i=1,\ldots,N_p \\ p \in \mathcal{P}_b(n, \mathcal{S})}} \tag{5.6}$$

(3) Finally choose for $1 \leqslant i \leqslant N_p$, $p \in \mathcal{P}_0(n, \mathcal{S})$, $s^n \in \mathcal{D}_p^i$ and

$q(\cdot \mid s, u) \in \mathcal{P}_0(n(s, u \mid s^n, u_p^i))$ for all $(s, u) \in \mathcal{S} \times \mathcal{U}$

$$ {}_qF_p^i(s^n) = \mathcal{G}_\delta(X_p^n \mid s^n, u_p^i) \cap \mathcal{X}_0(q(\cdot \mid s^n, u_p^i)) \tag{5.7}$$

as subedges of edge E_p^i.

Step 3 The parameters of the 2-hypergraph.

In order to apply Lemma 3C we need suitable *upper* bounds on the size of the edges and on the maximal vertex degree D of $(\mathcal{V}, \mathcal{A}^*)$, and a suitable *lower* bound on the cardinality of the subedges in $(\mathcal{V}, \mathcal{B}, (\mathcal{F}_E)_{E \in \mathcal{B}})$.

It follows from Lemmas M1, G1(c) and (3.7) in Section 3 that for $p \in \mathcal{P}_0(n, \mathcal{S})$, $1 \leqslant i \leqslant N_p$,

$$ |E_p^i| \leqslant \exp\{H(X_p \mid U_p)n + 0(\sqrt{n})\}. \tag{5.8}$$

We now show that

$$ D \leqslant \exp\{\max_{p \in \mathcal{P}_a(n, \mathcal{S})} H(X_p \mid U_p)n + c_1\sqrt{n}\log n\}. \tag{5.9}$$

Recall that $H_* = \min_{s \in \mathcal{S}} \{H(X(s)) : H(X(s)) > 0\}$. Now (5.4) and (3.7) imply that for $p \in \mathcal{P}_a(n, \mathcal{S})$ an s^n, $s^n \in \mathcal{T}_0(S_p^n)$, has fewer than $H_*^{-1}2c\sqrt{n}$ components t with $H(X(s_t)) > 0$. Therefore an $x^n \in \mathcal{X}^n$ can be contained in at most $T = \binom{n}{H_*^{-1}2c\sqrt{n}}|\mathcal{S}|^{H_*^{-1}2c\sqrt{n}}$ sets $\mathcal{G}_\delta(X_p^n \mid s^n, u_p^i)$, $s^n \in \mathcal{D}_p^i$, $1 \leqslant i \leqslant N_p$, $p \in \mathcal{P}_a(n, \mathcal{S})$.

This implies that x^n is contained in at most T edges in $\mathcal{A}$, and (5.9) follows from this fact and (5.8), if constant c_1 is properly chosen.

Finally, it follows from (5.4), Lemma M3 and (3.7) in Section 3 that the non-empty subedges satisfy for $p \in \mathcal{P}_b(n, \mathcal{S})$

$$ |{}_qF_p^i(s^n)| \geqslant \exp\{c\sqrt{n}\}. \tag{5.10}$$

Step 4 Application of Lemma 3C.

For any λ, $0 < \lambda < \frac{1}{2}$, choose

$$ L(n) = \exp\{n \max_{p \in \mathcal{P}_0(n, \mathcal{S})} H(X_p \mid U_p) + 2c_1\sqrt{n}\log n\}. \tag{5.11}$$

Certainly (2.20) holds, and (2.21) holds, because the left side of it is smaller than

$$ \sum_{p \in \mathcal{P}_b(n, \mathcal{S})} \sum_{i=1}^{N_p} \sum_{s^n \in \mathcal{D}_p^i} \sum_{q(\cdot \mid u, s) \in \mathcal{P}_0(n(s, u \mid s^n, u_p^i), \mathcal{X})} 2 \cdot \exp\{r\} \tag{5.12}$$

with

$$ \begin{aligned} r &= |{}_qF_p^i(s^n)|(h(\lambda) + \lambda \log(2|\mathcal{G}_\delta(X_p^n \mid s^n, u_p^i)|L(n)^{-1}) \\ &< e^{c\sqrt{n}}(h(\lambda) - \lambda c_1\sqrt{n}\log n) < -e^{c\sqrt{n}} \end{aligned}$$

for n sufficiently large.

The expression in (5.11) is smaller than

$$(n+1)^{|S|}|\mathcal{S}|^n(n+1)^{|\mathcal{X}||S|(|S|+2)}2\exp\{-e^{c\sqrt{n}}\},$$

which of course is *much* smaller than 1 for n large.

(Actually, by a more careful calculation and a slightly different definition of $\mathcal{P}_a(n, S)$ one can see that the term $2c_1\cdot\sqrt{n}\log n$ can be replaced by $c_3\sqrt{n}$).

Step 5 The code and its error probability.

Set $f_n(x^n) = \Phi_{2\lambda}(x^n)$, $g_n(s^n) = (i, p)$ if $s^n \in \mathcal{D}_p^i$.

Thus clearly,

$$\|f_n\| \leqslant \exp\{\max_p H(X_p \mid U_p)n + 0(\sqrt{n}\log n)\},$$

$$\|g_n\| \leqslant \exp\{\max_p I(U_p \wedge S_p)n + 0(\sqrt{n})\}.$$

Define the decoding function F_n by

$$F_n(l, (i, p)) = \begin{cases} E_p^i \cap f_n^{-1}(l) \text{ if this intersection contains exactly 1 element} \\ \text{any decision otherwise} \end{cases}. \tag{5.13}$$

The decoding error probability is readily calculated. Given any s^n, $s^n \in \mathcal{S}^n$, then there is exactly one $\mathcal{D}_p^i$ containing s^n.

From Lemma M1(b) we know that

$$\text{Prob}\,(X(s^n) \in \mathcal{G}_\delta(X_p \mid s^n, u_p^i)) = 1 - 0\left(\frac{1}{\delta^2}\right). \tag{5.14}$$

Since $\mathcal{G}_\delta(X_p \mid s^n, u_p^i) = \bigcup_q {}_qF_p^i(s^n)$, and since every ${}_qF_p^i(s^n)$ is colored properly in at least $(1-\lambda)|{}_qF_p^i(s^n)|$ elements the probabilities of wrong decoding are less than $0\left(\frac{1}{\delta^2}\right) + \lambda$.

Q.E.D.

Section 6. The Decomposition Phenomenon and Edge Coloring of Bipartite Graphs via Vertex Coloring

§ 1 The decomposition phenomenon

We recall the reasoning which led us to find the capacity region of the multiple-access channel (MAC) in [9]. Since the results of [3] are incomplete (see [34]) this is the first coding theorem in multi-user communication (see also [29]). Suppose we have 2 senders and 1 receiver, and the senders send independent messages. If the two senders were at the same terminal and could cooperate, then we would have an ordinary DMC with input alphabet $\mathcal{X}\times\mathcal{Y}$ and output alphabet $\mathcal{Z}$, say. The capacity of this channel is $C = \max_{(X,Y)} I(XY \wedge Z)$.

If we require now that only "rectangular codes"

$$\{(w_{ij}, D_{ij}) : 1 \leqslant i \leqslant M_1,\ 1 \leqslant j \leqslant M_2\}$$

with $$w_{ij} = (u_i, v_j),\ u_i \in \mathcal{X}^n,\ v_j \in \mathcal{Y}^n$$

are permitted (which is the case in the original problem !), then the "capacity" does not exceed $\max_{X, Y \text{ ind.}} I(XY \wedge Z)$.

The question then was whether $I(XY \wedge Z)$, X,Y independent, can be achieved by $R_1 + R_2$ with encoders at different terminals and if so, how $I(XY \wedge Z)$ *decomposes* into rate pairs.

A guide for the answer was the simple identity

$$I(XY \wedge Z) = I(X \wedge Z) + I(Y \wedge Z \mid X). \tag{6.1}$$

The analogous question for two correlated sources was asked subsequently in [15]. Here the source can be encoded by two separate encoders such that the total rate needed is the same as if both encoders were at one terminal. The *decomposition* of the total rate into the rate pairs can here be understood from an even simpler identity

$$H(X,Y) = H(X) + H(Y \mid X). \tag{6.2}$$

The identities are closely related and so are the coding theorems for the two problems.

After Shannon had initiated multi-user communication theory with his [3] many years no progress was made. The decomposition phenomenon made certain problems tractable and accounts for recent activity in the field. (cf. [29]).

Let us pursue the similarity between the two coding problems for the MAC and the DMCS a little further by looking at their proofs. In [9] the following approach was chosen. Select code words $U_1, \ldots, U_{M_1}$; $V_1, \ldots, V_{M_2}$ independently at random, the U_i's (resp. V_j's) being identically distributed. This is a standard random coding approach, used already in [3]. The question is: "How do we choose the decoding sets?"

Maximum likelihood decoding (MLD) is symmetric in X and Y but the formula (6.1) isn't.

The answer given in [9] was to decode the U_i's against the "*average* of the V_j's" and then use the knowledge of U_i to decode, conditionally on this, the V_j's with MLD. In a strict sense this decoding rule is suboptimal, but it turns out to be still good enough. The capacity region is then found by exchanging the roles of X and Y and time-sharing.

This approach was extended in [9] also to the case of 3 senders. The proof then becomes awfully complicated in the converse part, because of those time-sharing arguments. Subsequently Liao [25] announced those

results, with the same description of the capacity region, also for an arbitrary number of senders. Since our 3-case was so complicated, we are curious about his proof, which until now has not appeared in print.

Because this complication seemed unnatural and also because we were unable to do the 2-sender-2-receiver problem of [11] ("compound MAC") with this approach, which is due to lack of symmetry in (6.1), we gave in [11] an alternate proof for the standard MAC, which is symmetric in X and Y and also simpler and easier to generalize (see [27]).

The idea: use MLD, the best decoding rule. Originally we had doubts whether the error probability calculations could be done.

But they are even simpler, once one frees the mind from thinking in terms of equality (6.1) and passes on to inequalities instead. Then for X, Y independent

$$\begin{aligned} R_1 &\leqslant I(X \wedge Z \mid Y) \\ R_2 &\leqslant I(Y \wedge Z \mid X) \\ R_1 + R_2 &\leqslant I(XY \wedge Z) \end{aligned} \tag{6.3}$$

is achievable.

Denoting this set by $\mathcal{R}(X, Y)$, the capacity region equals

$$\operatorname{conv}\Big(\bigcup_{X, Y \text{ ind.}} \mathcal{R}(X, Y)\Big).$$

That it is really necessary to take the convex hull was recently shown in [28].

Once one has these two characterisations of the capacity region it is easy to show directly their equivalence.

We know of 4 proofs for the DMCS-coding theorem ([15], [14] included in [17], [16], [22]), not counting the new ones of this paper.

They all have one part in common: one source, say X, is first coded completely with rate $H(X)$ and the rest is coded conditionally on X with rate $H(Y \mid X)$.

Our proof consists of an iteration of Feinstein's maximal code construction ([4]) until one source, say X, is up to a set of small probability partitioned into systems of code words. This approach has been frequently used and extended by Csiszár, Körner, and Marton (cf. [21], [24]); and also by Sgarro [25]. A drawback to that proof is that it does source coding via channel coding. The other proofs use a random selection argument. In all 4 cases the regions are obtained by time-sharing. With some abuse of language let us call those proofs "asymmetric". The corresponding proof for the MAC coding theorem is our original proof of [9]. Where is the source coding parallel of our "symmetric" second proof of [11]? We have mentioned that the "symmetric" proof for the MAC-coding theorem is better suited for extensions to more complex

situations. The same should be true in source coding, and as a matter of fact the AVCS-problems (§3 and §4 of Section 1) seem non-tractable with "asymmetric" proofs of this kind.

§2 A "SYMMETRIC" AND ABSTRACT VERSION OF COVER'S PROOF ([16])

Let $\mathcal{V}$ and $\mathcal{W}$ be finite sets and let C be a subset of $\mathcal{V} \times \mathcal{W}$. $(\mathcal{V}, \mathcal{W}, C)$ is a bipartite graph.

If P is a PD on $\mathcal{V} \times \mathcal{W}$ concentrated on C, that is, $P(v, w) \neq 0$ implies $(v, w) \in C$, then we call $(\mathcal{V}, \mathcal{W}, C, P)$ a *stochastic bipartite graph.* Let φ (resp. ψ) be a coloring of $\mathcal{V}$ (resp. $\mathcal{W}$). Then $\rho = (\varphi, \psi)$ is an *orthogonal* coloring of $\mathcal{V} \times \mathcal{W}$ and colors in particular all edges in C. Clearly, if for $(v, w) \in C$

$$\rho(v, w) \neq \rho(v', w') \text{ for all } (v', w') \in C - \{(v, w)\}, \tag{6.4}$$

then knowing $\rho(v, w)$ we can identify (v, w). Conversely, if (6.4) does not hold, then (v, w) cannot be identified or decoded correctly. Suppose (v, w) occurs with probability $P(v, w)$, then the *average* error probability $\mathbf{e}(\rho)$ is given by

$$\mathbf{e}(\rho) = P(\{(v, w) : |\rho^{-1}(\rho(v, w)) \cap C| > 1\}). \tag{6.5}$$

We call $\rho = (\varphi, \psi)$ an $L_1 \times L_2$-coloring if

$$\|\varphi\| \leqslant L_1 \text{ and } \|\psi\| \leqslant L_2.$$

Finally, $C_{|w}$ (resp. $C_{|v}$) denotes the *cross-section*

$$\{v : (v, w) \in C\} \text{ (resp. } \{w : (v, w) \in C\}).$$

Given L_1 and L_2 let us now color $\mathcal{V}$ at random in the standard way with $X^{|\mathcal{V}|} = X_1, \ldots, X_{|\mathcal{V}|}$ (i.i.d. with Prob $(X_i = l_1) = \dfrac{1}{L_1}$, $1 \leqslant l_1 \leqslant L_1$) and independently of this also $\mathcal{W}$ with $Y^{|\mathcal{W}|} = Y_1, \ldots, Y_{|\mathcal{W}|}$

$$\left(\text{i.i.d. with Prob } (Y_j = l_2) = \frac{1}{L_2},\ 1 \leqslant l_2 \leqslant L_2\right).$$

The expected average error probability $\mathbf{E}\mathbf{e}(X^{|\mathcal{V}|}, Y^{|\mathcal{W}|})$ can be upper-bounded as follows:

$$\begin{aligned}
&\text{Prob } ((X_v, Y_w) = (X_{v'}, Y_{w'}) \text{ for some } (v', w') \in C - \{(v, w)\}) \\
&\leqslant \text{Prob } ((X_v, Y_w) = (X_v, Y_{w'}) \text{ for some } w' \in C_{|v} - \{w\}) \\
&+ \text{Prob } ((X_v, Y_w) = (X_{v'}, Y_w) \text{ for some } v' \in C_{|w} - \{v\}) \\
&+ \text{Prob } ((X_v, Y_w) = (X_{v'}, Y_{w'}) \text{ for some } (v', w') \in C,\ v' \neq v,\ w' \neq w) \\
&\leqslant \frac{|C_{|v}|}{L_2} + \frac{|C_{|w}|}{L_1} + \frac{|C|}{L_1 L_2}.
\end{aligned} \tag{6.6}$$

Therefore,

$$\mathbf{E}\mathbf{e}(X^{|\mathcal{V}|}, Y^{|\mathcal{W}|}) \leqslant \sum_{(v, w)} p(v, w)\left[\frac{C_{|v}}{L_2} + \frac{C_{|w}}{L_1} + \frac{|C|}{L_1 L_2}\right].$$

With $N_2 = \max_{v \in \mathcal{V}} |C_{|v}|$, $N_1 = \max_{w \in \mathcal{W}} |C_{|w}|$, and $N = |C|$, we can write

$$\mathbf{E}e(X^{|\mathcal{V}|}, Y^{|\mathcal{W}|}) \leqslant \frac{N_2}{L_2} + \frac{N_1}{L_2} + \frac{N}{L_1 \cdot L_2}.$$

Thus we have proved.

COLORING LEMMA 5 *The standard orthogonal $L_1 \times L_2$-coloring of the bipartite stochastic graph $(\mathcal{V}, \mathcal{W}, C, P)$ has an expected average error probability less than*

$$\frac{N_2}{L_2} + \frac{N_1}{L_1} + \frac{N}{L_1 \cdot L_2},$$

where $N = |C|$ and N_1, N_2 are the respective maximal cardinalities of cross sections of C.

Notice that only the parameters of the *carrier* C are important and no AEP-property is used.

Also, if *P is an additive set function with $P(C) \leqslant 1$*, the above result holds.

§ 3 APPLICATION TO AVCS WITHOUT SIDE INFORMATION

We now derive Jahn's result, which is for randomized encoding ([26]). He uses the standard random $L_1 \times L_2$-coloring; however, he computes error probabilities in a non-symmetric way and this is the main reason for the fact that his proof is so complicated. As stochastic bipartite graph choose $\mathcal{V} = \mathcal{X}^n$, $\mathcal{W} = \mathcal{Y}^n$,

$$C = \bigcup_{s^n \in \mathcal{S}^n} \mathcal{G}_\delta(X^n, Y^n \mid s^n) \tag{6.7}$$

$$P_{s^n}(x^n, y^n) = \begin{cases} \text{Prob}\,(X^n = x^n, Y^n = y^n \mid s^n) \\ \text{if} \quad (x^n, y^n) \in C \\ 0 \quad \text{otherwise} \end{cases} \tag{6.8}$$

Clearly, $$P_{s^n}(C) \leqslant 1.$$

By Lemma CS and the Carrier Lemma of Section 3

$$N_1 \leqslant \exp\{\max_{\bar{s}} H(X(\bar{s}) \mid Y(\bar{s}))n + 0(\sqrt{n})\}$$

$$N_2 \leqslant \exp\{\max_{\bar{s}} H(Y(\bar{s}) \mid X(\bar{s}))n + 0(\sqrt{n})\}$$

$$N \leqslant \exp\{\max_{\bar{s}} H(X(\bar{s}), Y(\bar{s}))n + 0(\sqrt{n})\}.$$

Let $\mathbf{e}_{s^n}(V^{|\mathcal{V}|}, W^{|\mathcal{W}|})$ be the expected error probability of the standard $L_1 \times L_2$-coloring with $L_1 \geqslant N_1 \cdot n$, $L_2 \geqslant N_2 \cdot n$, $L_1 \cdot L_2 \geqslant N \cdot n$.

Then

$$e_{s^n}(V^{|\mathcal{V}|}, W^{|\mathcal{W}|} \leqslant \frac{3}{n} \text{ for } \textit{all } s^n \in \mathcal{S}^n. \tag{6.9}$$

Since by Lemma G_1 and definition (6.7) C carries for all s^n the PD $P(\cdot, \cdot \mid s^n)$ up to $1 - 0\left(\frac{1}{\delta^2}\right)$, the total *expected* error probability satisfies

$$\mathbf{E}\lambda_{s^n} \leqslant \frac{3}{n} + 0\left(\frac{1}{\delta^2}\right) \text{ for } \textit{every } s^n.$$

Now make use of the *elimination technique* of [13] to obtain a code with independent randomisation at the encoders only. Since this has been done in [26] in the canonical way, we save our space for the proof of the mathematically deeper Theorem 4.

Section 7. Orthogonal Coloring of Rectangular Hypergraphs $(\mathcal{V} \times \mathcal{W}, \mathcal{E})$

Again $\mathcal{V}$ and $\mathcal{W}$ are finite sets. $\mathcal{E}$ is a *family of subsets* of $\mathcal{V} \times \mathcal{W}$. We call $(\mathcal{V} \times \mathcal{W}, \mathcal{E})$ a rectangular hypergraph, because the vertex set is a general rectangle. If $|\mathcal{E}| = 1$ we get the case discussed in the previous section. We denote $\underset{E \in \mathcal{E}}{U} E$ by C and call it the *carrier* of the hypergraph.

We denote by $\rho_\lambda = (\varphi_\lambda, \psi_\lambda)$ an orthogonal coloring of $(\mathcal{V} \times \mathcal{W})$ for which in every edge E, $E \in \mathcal{E}$, at least $(1 - \lambda)|E|$ colors occur, which *occur only once in* C.

Use random coloring $V_1, \ldots, V_{|\mathcal{V}|}$; $W_1, \ldots, W_{|\mathcal{W}|}$ as in the previous Section. Order the elements of $\mathcal{V} \times \mathcal{W}$ lexicographically. We proceed similar as in the proof of Lemma 3B.

Write $\mathcal{E} = \{E_1, \ldots, E_J\}$ and define for $j = 1, \ldots, J$ and $(v, w) \in \mathcal{V} \times \mathcal{W}$ RV's

$$f^j_{vw}(V^{|\mathcal{V}|}; W^{|\mathcal{W}|}) = \begin{cases} 1 & \text{if } (V_v, W_w) \neq (V_{v'}, W_{w'}) \text{ for all} \\ & (v', w') \in E_j - \{v, w\} \text{ with } (v', w') < (v, w) \\ & \text{and for all } (v', w') \in C - E_j \\ 0 & \text{otherwise} \end{cases} \tag{7.1}$$

We upperbound now

$$\text{Prob}\Big\{\sum_{(v, w) \in E_j} f^j_{v, w} < (1 - \lambda)|E_j|\Big\}.$$

Use the notation

$$E_j = \{(v, w)_1, \ldots, (v, w)_{M_j}\},$$
$$\text{where } (v, w)_1 < \ldots < (v, w)_{M_j},\ (v, w)_s = (v_s, w_s).$$

$$\text{Prob}\,\{f^j_{(v, w)_s} = 0 \mid f^j_{(v, w)_{s-1}} = \epsilon_s - 1, \ldots, f^j_{(v, w)_1} = \epsilon_1\}$$
$$\leqslant \frac{|C|}{L_1 L_2} + \frac{|C_{|v_s}|}{L_2} + \frac{|C_{|w_s}|}{L_1}. \tag{7.2}$$

By the arguments which led to (2.9) or (2.19), we obtain now

$$\text{Prob}\left\{\min_{j=1,\ldots,J}\frac{1}{|E_j|}\sum_{(v,w)\in E_j} f^J_{v,w} < 1-\lambda\right\} \leqslant \sum_{j=1}^{J}\exp\left\{|E_j|\left[h(\lambda)+\lambda\log\frac{\beta}{L_1L_2}\right]\right\} \tag{7.3}$$

with $\beta = |C| + L_1 \max_v |C_{|v}| + L_2 \max_w |C_{|w}|$.

Thus we have proved

COLORING LEMMA 6. *The rectangular hypergraph* $(\mathcal{V}\times\mathcal{W}, \mathcal{E})$ *with carrier* C *has an orthogonal* $L_1\times L_2$*-coloring* $\rho_{2\lambda}$ *if*

$$\frac{\beta}{L_1L_2-\beta}\,\frac{1-\lambda}{\lambda}\leqslant 1 \tag{7.4}$$

and

$$\sum_{j=1}^{J}\exp\left\{|E_j|\left[h(\lambda)+\lambda\log\frac{\beta}{L_1L_2}\right]\right\} < 1. \tag{7.5}$$

We shall also need a slight generalisation of this result to *orthogonal* 2-hypergraphs, that is a 2-hypergraph with rectangular vertex set. Denote such a 2-hypergraph by $(\mathcal{V}\times\mathcal{W}, \mathcal{E}, (\mathcal{E}_j)_{j=1}^J)$. We denote by $\omega_\lambda = (\varphi_\lambda, \psi_\lambda)$ an orthogonal coloring of $(\mathcal{V}\times\mathcal{W})$ for which in every subedge E_j^m $(m = 1,\ldots, M_j;\ j = 1,\ldots, J)$ at least $(1-\lambda)|E_j^m|$ colors occur, which occur only once in E_j.

Define $\beta_j = \beta_j(L_1, L_2)$ by

$$\beta_j = (|E_j| + L_1\max_v |E_{j|v}| + L_2\max_w |E_{j|w}|). \tag{7.6}$$

Then by the same arguments which led to Lemma 6 we obtain now

COLORING LEMMA 7 *The rectangular 2-hypergraph* $(\mathcal{V}\times\mathcal{W}, \mathcal{E}, (\mathcal{E})_{j=1}^J)$ *has an orthogonal* $L_1\times L_2$*-coloring* $\omega_{2\lambda}$ *if*

$$\max_j \frac{\beta_j}{L_1L_2-\beta_j}\,\frac{1-\lambda}{\lambda}\leqslant 1 \tag{7.7}$$

and

$$\sum_{j=1}^{J}\sum_{m=1}^{M_j}\exp\left\{|E_j^m|\left[h(\lambda)+\lambda\log\frac{\beta_j}{L_1L_2}\right]\right\} < 1. \tag{7.8}$$

This is of course also a generalisation of Coloring Lemma 3B, the analogue of Coloring Lemma 3A is contained in it as a special case.

SECTION 8 PROOF OF THEOREM 3

This theorem is a generalisation of Theorem 1. Again there are also the two slightly different proofs: one using the almost uniformity of PD's and the other based on counting alone.

We now choose the second one. Define the 2-hypergraph as follows:

(a) $\mathcal{V} = \mathcal{X}^n$, $\mathcal{W} = \mathcal{Y}^n$

(b) $\mathcal{G}_\delta(X^n, Y^n \mid s^n) \cap (\mathcal{X} \times \mathcal{Y})_0^n(q(.,. \mid s^n))$,

where for all $s \in \mathcal{S}$ $q(\ldots \mid s) \in \mathcal{P}_0(n(s \mid s^n), \mathcal{X} \times \mathcal{Y})$, are the subedges of $\mathcal{G}_\delta(X^n, Y^n \mid s^n)$.

(Let us keep in mind that by our notation $q(.,. \mid s_t) = q(.,. \mid s_{t'})$ if $s_t = s_{t'}$.)

By Lemma G1(c) and Lemma CS of Section 3 this 2-hypergraph has the parameters

$$|\mathcal{G}_\delta(X^n, Y^n \mid s^n)| \leqslant \exp\{H(X(s^n), Y(s^n)) + c\sqrt{n}\} \text{ for all } s^n \in \mathcal{S}^n \quad (8.1)$$

$$|\mathcal{G}_\delta(X^n, Y^n \mid s^n)_{|x^n}| \leqslant \exp\{H(Y(s^n) \mid X(s^n)) + c\sqrt{n}\} \text{ for all } s^n \in \mathcal{S}^n, x^n \in \mathcal{X}^n. \quad (8.2)$$

$$|\mathcal{G}_\delta(X^n, Y^n \mid s^n)_{|y^n}| \leqslant \exp\{H(X(s^n) \mid Y(s^n)) + c\sqrt{n}\} \text{ for all } s^n \in \mathcal{S}^n, x^n \in \mathcal{X}^n. \quad (8.3)$$

Given a *constant* λ, $0 < \lambda < \frac{1}{2}$ choose $L_1(n)$, $L_2(n)$ such that

$$L_1(n) \geqslant \exp\{\max_{s \in \mathcal{S}} H(X(s) \mid Y(s))n + 2c\sqrt{n}\}, \quad (8.4)$$

$$L_2(n) \geqslant \exp\{\max_{s \in \mathcal{S}} H(Y(s) \mid X(s))n + 2c\sqrt{n}\},$$

$$L_1(n) \cdot L_2(n) \geqslant \exp\{\max_{s \in \mathcal{S}} H(X(s), Y(s))n + 2c\sqrt{n}\}.$$

Let $K = K(n)$ be a number less than $\frac{n}{2}$ to be fixed later.

Define $\mathcal{S}_K(n) = \{s^n : H(X(s_t), Y(s_t)) > 0 \text{ for at most } K \text{ indices } t\}$.

Then for $s^n \in \mathcal{S}^n$ we derive directly from definition (3.4) that

$$|\mathcal{G}_\delta(X^n, Y^n \mid s^n)| \leqslant (|\mathcal{X}|\,|\mathcal{Y}|)^K \quad (8.5)$$

and for $s^n \in \mathcal{S}^n - \mathcal{S}_K(n)$ by Lemma G3

$$|\mathcal{G}_\delta(X^n, Y^n \mid s^n) \cap (+ \times \mathcal{Y})_0^n(q(.,. \mid s^n))| \geqslant \exp\{H_* K - e\sqrt{K}\}, \quad (8.6)$$

if the subedge is non-empty and

$$H_* = \min\{H(X(s), Y(s)) : s \in \mathcal{S}, H(X(s), Y(s)) > 0\}.$$

Inequalities (8.1)–(8.4) and (8.6) make Lemma 7 applicable for the 2-hypergraph obtained by restriction to the edges with $s^n \in \mathcal{S}^n - \mathcal{S}_K(n)$.

It says that there exists an $L_1(n) \times L_2(n)$-coloring $\rho_{2\lambda} = (\varphi_{2\lambda}, \psi_{2\lambda})$ with $\|\varphi_{2\lambda}\| \leqslant L_1(n)$, $\|\psi_{2\lambda}\| \leqslant L_2(n)$ if K satisfies

$$\exp\{e^{H_* K - c\sqrt{K}}(\lambda c\sqrt{n} - h(\lambda))\} > |\mathcal{S}|^n (n+1)^{|\mathcal{X}|\,|\mathcal{Y}|\,|\mathcal{S}|}. \quad (8.7)$$

Clearly (8.7) holds for $K = n^{1/2}(|\mathcal{X}|\,|\mathcal{Y}|\log n)^{-1}$.

Now we find a second orthogonal coloring of $\mathcal{X}^n \times \mathcal{Y}^n$ which is strict for the small edges ($s^n \in \mathcal{S}_K(n)$).

For this define the projections of the edges on $\mathcal{X}^n$ (resp. $\mathcal{Y}^n$) by

$$\text{Proj.}_{\mathcal{X}^n}(\mathcal{G}_\delta(X^n, Y^n \mid s^n)) = \{x^n : \text{exists } y^n \text{ with } (x^n, y^n) \in \mathcal{G}_\delta(X^n, Y^n \mid s^n)\}$$
$$(\text{resp. Proj.}_{\mathcal{Y}^n}(\mathcal{G}_\delta(X^n, Y^n \mid s^n)). \quad (8.8)$$

Those projections give rise to hypergraphs $(\mathcal{X}^n, \mathcal{A})$ and $(\mathcal{Y}^n, \mathcal{B})$, say. Their edges are by (8.5) again smaller than $(|\mathcal{X}|\,|\mathcal{Y}|)^K$ and the maximal degrees of the corresponding graphs $(\mathcal{X}^n, \mathcal{A}^*)$, $(\mathcal{Y}^n, \mathcal{B}^*)$ are by definition of $S_K(n)$ smaller than

$$\binom{n}{K}(\max(|\mathcal{X}|, |\mathcal{Y}|))^K \leqslant e^{\sqrt{n}} \quad (8.9)$$

and have, by Coloring Lemma 2, colorings φ_0 resp. ψ_0 with $\|\varphi_0\| \leqslant e^{2\sqrt{n}}$, $\|\psi_0\| \leqslant e^{2\sqrt{n}}$.

Now, $\rho_0 = (\varphi_0, \psi_0)$ colors all the small edges correctly, and $\tau = ((\varphi_0, \varphi_1), (\psi_0, \psi_1))$ is an orthogonal coloring with all the desired properties. The total worst case error probability is bounded by $\lambda + 0\left(\frac{1}{\delta^2}\right)$. Q.E.D.

Section 9. Proof of Theorem 4

§1 The converse

The proof of the converse is trivial. Suppose we have a code (f, g, F) with Prob $(F(f(X(s^n)), g(Y(s^n))) = (X(s^n), Y(s^n))) = 1 - \lambda(s^n) \geqslant 1 - \lambda$ for all $s^n \in \mathcal{S}^n$.

Put a PD $q^n = qx \ldots xq$ on $\mathcal{S}^n$ and define

$$\bar{s} = \sum_s q(s)s, \; s^n = (\bar{s}, \ldots, \bar{s}).$$

Then Prob $(F(f(X(\bar{s}^n)), g(Y(\bar{s}^n))) = (X(\bar{s}^n), Y(\bar{s}^n))) \geqslant 1 - \lambda$.

We have now a code for the standard DMCS of Slepian-Wolf. Therefore by Fano's Lemma

$$\log |f| \geqslant nH(X(\bar{s}) \mid Y(\bar{s})) - 1 - \lambda n \log |\mathcal{X}|$$
$$\log |g| \geqslant nH(Y(\bar{s}) \mid X(\bar{s})) - 1 - \lambda n \log |\mathcal{Y}|$$
$$\log |f| + \log |g| \geqslant nH(X(\bar{s}), Y(\bar{s})) - 1 - \lambda n \log |\mathcal{X}|\,|\mathcal{Y}|.$$

Since those inequalities hold for all $\bar{s}$, the converse part of Theorem 4 is proved.

§2 The direct part of Theorem 4

In the last section Coloring Lemma 7 was used. Here we need only Coloring Lemma 6.

Define an orthogonal hypergraph with $\mathcal{V} = \mathcal{X}^n$, $\mathcal{W} = \mathcal{Y}^n$, and as *edges* choose all the *subedges* of the 2-hypergraph of Section 8 (see (b) there). The *carrier* C is simply defined as the union over all the edges.

Choose $L_1(n)$ and $L_2(n)$ such that the inequalities in (8.4) hold also if the maximisations there are over $\bar{S}$. We know from Lemma CS and the Carrier Lemma in Section 3 that

$$\begin{aligned}\beta \leqslant & \exp\{\max_{\bar{s}\in\bar{S}} H(X(\bar{s}), Y(\bar{s}))n + c\sqrt{n}\} \\ & + L_1(n)\exp\{\max_{\bar{s}\in\bar{S}} H(X(\bar{s}) \mid Y(\bar{s}))n + c\sqrt{n}\} \\ & + L_2(n)\exp\{\max_{\bar{s}\in\bar{S}} H(Y(\bar{s}) \mid X(\bar{s}))n + c\sqrt{n}\}. \end{aligned} \tag{9.1}$$

Therefore,

$$\frac{\beta}{L_1L_2} \leqslant \exp\{-c\sqrt{n}\} \tag{9.2}$$

and (7.4) holds.

Since here all non-empty edges have cardinality bigger than $\exp\{H_*n - c\sqrt{n}\}$, and since $|J| \leqslant |\mathcal{S}|^n(n+1)^{|\mathcal{X}|\,|\mathcal{Y}|\,|\mathcal{S}|}$, (7.5) holds also. Now use as encoding functions the colorings $\varphi_{2\lambda}$ and $\psi_{2\lambda}$, which exist by Lemma 6. Define as decoding function

$$F_n(l_1, l_2) = \begin{cases}(\varphi_{2\lambda}^{-1}(l_1), \psi_{2\lambda}^{-1}(l_2)) \text{ if this set has exactly 1 element in } C \\ \text{any decision otherwise.}\end{cases} \tag{9.3}$$

Since for all $s^n \in \mathcal{S}^n$

$$\text{Prob}\,((X(s^n), Y(s^n)) \in C) = 1 - 0\left(\frac{1}{\delta^2}\right),$$

the code $(\varphi_{2\lambda}, \psi_{2\lambda}, F_n)$ has an error probability less than $2\lambda + 0\left(\frac{1}{\delta^2}\right)$.

Q.E.D.

References

[1] C. E. Shannon (1948), "The mathematical theory of communication", *Bell System Techn. J.*, 27, 379–423, 623–656.

[2] ——— (1959), "Coding theorems for a discrete source with a fidelity criterion". *IRE Nat. Conv. Rec.*, Pt. 4, 142–163.

[3] ——— (1961), "Two-way communication channels", *Proc. Fourth Berkely Sympos. on Math. Stat. and Prob.*, Vol. 1, 611–694, University of California Press, Berkeley.

[4] A. Feinstein (1954), "A new basic theorem of information theory", *IRE Trans. Inform. Theory*, Vol. PGIT-4, 2–22.

[5] J. Wolfowitz (1957), "The coding of messages subject of change errors", *Illinois J. Math.*, 1, 591–606.

[6] ——— (1964), *Coding Theorems of Information Theory*, 2nd edn., Springer-Verlag, Berlin-Heidelberg-New York.

[7] D. Blackwell, L. Breiman and A. J. Thomasian (1959), "The capacity of a class of channels", *Ann. Math. Stat.*, 30 (4), 1229–1241.

[8] ———— (1960), "The capacities of certain channel classes under random coding" *Ann. Math. Stat.*, **31**, 558–561.

[9] R. Ahlswede (1973), "Multi-way communication channels", Second International Symposium on Inform. Theory, Thakadsor (1971), Publishing House of the Hungarian Acad. of Sc., 23–52.

[10] ———— (1973), "Channel capacities for list codes", *Journal of Applied Probability*, **10** (4), 824–836.

[11] ———— (1974), "The capacity region of a channel with two senders and two receivers", *Ann. of Probability*, **2** (5), 805–814.

[12] ———— (1973), "Channels with arbitrarily varying channel probability functions in the presence of noiseless feedback", *Z. Wahrscheinlichkeitstheorie u. verw. Geb.*, **25**, 239–252.

[13] ———— (1978), "Elimination of correlation in random codes for arbitrarily varying channels", *Z. Wahrschinlichkeitsthearie u. verw. Geb.*, **44**, 159–175.

[14] ———— "Universal coding for correlated sources", Lecture at the 7th Hawaii Int. Conf. System Sciences, Jan. 1974.

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

[16] T. M. Cover (1975), "A proof of the data compression theorem of Slepian and Wolf for ergodic sources", *IEEE Trans. Inf. Th.*, Vol. IT-21, 226–228.

[17] R. Ahlswede and J. Körner (1975), "Source coding with side information and a converse for degraded broadcast channels". *IEEE Trans. Inf. Th.*, Vol. IT-21, 629–637.

[18] A. D. Wyner (1975), "On source coding with side information at the decoder", *IEEE Trans. Inf. Th.*, Vol. IT-21, 294–300.

[19] R. Ahlswede, P. Gàcs and J. Körner (1976), "Bounds on conditional probabilities with applications in multi-user communication." *Z. Wahrscheinlichkeitstheorie u. verw. Geb.*, **34**, 157–177.

[20] A. D. Wyner and J. Ziv (1976), "The rate-distortion function for source coding with side information at the decoder", *IEEE Trans. Inf. Th.*, Vol. IT-22, No. 1, 1–10.

[21] J. Körner and K. Marton (1977), "Images of a set via two channels and their role in multi-user communication", *IEEE Trans. Inf. Th.*, Vol. IT-23, No. 6, 751–761.

[22] R. G. Gallager, "Source coding with side information and universal coding", presented at the IEEE Int. Symposium on Inf. Th., Renneby (Sweden), 1976.

[23] A. Sgarro (1977), "Source coding with side information at several decoders", *IEEE Trans. Inf. Th.*, Vol. IT-23, 179–182.

[24] I. Csiszàr and J. Körner (1976), "Source networks with unauthorized users", *Journal of Combinatorics, Information & System Sciences*, **1** (1), 25–40.

[25] H. Liao (1972), "A coding theorem for multiple access communications", presented at International Symposium on Information Theory, Asilomar, California.

[26] J. H. Jahn (1978), "Kodierung beliebig variierender korrelierter Quellen", Dissertation, Bielefeld.

[27] M. Ulrey (1975), "A coding theorem for a channel with *s* senders and *r* receivers", *Information and Control*, **29**, 185–203.

[28] M. Bierbaum and M. Wallmeier, "A note on the capacity region of the multiple-access channel", to appear in *IEEE Trans. Inf. Th.*, V ol. IT-25 (1979).

[29] E. C. Van der Meulen (1977), "A survey of multi-way channels in Information Theory: 1961–1976", *IEEE Trans. Inf. Th.*, Vol. IT–23, 1–37.

[30] C. E. Shannon (1956), "The zero-error capacity of a noisy channel", *Comp. Inf. Theory, IRE Trans.*, 3, 3–15.

[31] J. Kiefer and J. Wolfowitz (1969), "The capacity of a channel with arbitrarily varying channel probability functions", *Information and Control*, **14**, 451–473.

[32] L. Lovasz, "On the Shannon capacity of a graph". *IEEE Trans. Inf. Th.*, Vol. IT–25, 1–7.

[33] J. Wolfowitz (1960), "Simultaneous channels", *Arch. Rational Mech. Anal.*, 4 (4), 371–38 .

[34] G. Dueck (1979), "The capacity region of the two-way channel can exceed the inner bound", *Inform. Control*, 40, — .

[35] R. Ahlswede, and J. Wolfowitz (1970), "The capacity of a channel with arbitrarily varying channel probability functions and binary output alphabet", *Z. Wahrscheinlichkeitstheorie u. verw. Geb.*, **15** (3), 186–194.

[Received : Sept., 1978]