JOURNAL OF COMBINATORICS,
INFORMATION & SYSTEM SCIENCES

Vol. 5, No. 3, 220–268 (1980)

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

RUDOLF AHLSWEDE

Fak. f. Mathematik, Universität Bielefeld, Universitätsstr. 1,
4800 Bielefeld, West Germany

1. INTRODUCTION

We continue to develop the approach to source coding taken in Part I, which consists in viewing coding problems abstractly as coloring and covering problems for hypergraphs. The coloring problems, which arise, are usually of a nature where it suffices to have good colorings in a high fraction of cases under consideration. This distinguishes them from coloring problems, which are usually studied in the Theory of Hypergraphs (see [47], [48]). Source Coding Theory is a rich source for new problems of this kind. We strongly believe that an emphasis on the connections between those two theories will result in a fruitful exchange of problems, methods, and ideas and will be to the benefit of both subjects.

The paper contains several results on colorings and coverings for hypergraphs, which go beyond those presented in Part I. They are all obtained again by elementary, often probabilistic arguments. Some of them are stated in Section 2 (Coloring Lemmas 6–9), others are given in other Sections in conjunction with a coding problem in which they arise or for which they are needed for the solution (Covering Lemmas 2, 3 in Section 6, Coloring Lemma 3A* in Section 7).

Let us now briefly indicate the new results on source coding.

(a) Our first result is a pseudo-new result, because it consists in correcting an error, which we found in Section 7 of Part I. The inequality (7.2) is wrong and this effects the validity of Coloring Lemmas 6, 7 and the proofs of Theorems 3, 4 in Sections 8, 9. We apologize to the reader of Part I for this mistake. Fortunately this error opens new doors (it corrects the theory): it turns out that orthogonal coloring of hypergraphs is a mathematically more serious and also interesting matter than we originally expected (see Section 2). In Section 3 we give a rigorous proof of Theorem 3. An example shows that the original Theorem 4 is

not true. However, the conclusion holds if we replace the condition $\min_{s \in \mathcal{S}} H(X(s), Y(s)) > 0$ by the somewhat stronger entropy positiveness condition

$$\min_{s \in \mathcal{S}} H(X(s) \mid Y(s))\, H(Y(s) \mid X(s)) > 0. \tag{1.1}$$

(Theorem 4')

(b) In Section 4 we consider again an AVS $\left(\{X(s^n) : s^n \in \mathcal{S}^n\}\right)_{n=1}^{\infty}$, where $X(s^n)$ has distribution $P(\cdot \mid s^n)$ as defined in Part I (1.1), of Section 1, § 1. In Theorem 1 we obtained the optimal rate of an AVS with side information about s^n at the decoder. We treat now a more general problem for this source, which is analog to the coding problem for correlated sources solved by Slepian and Wolf [15]. Notice that in their model of a source the joint distributions are known, whereas here only conditional distributions are specified. For encoding functions f_n(resp. g_n) defined on $\mathcal{X}^n$(resp. $\mathcal{S}^n$) a decoding function F_n shall be a mapping of the cartesian product of the ranges of f_n and g_n into $\mathcal{X}^n \times \mathcal{S}^n$. The error probability of the code (f_n, g_n, F_n) is defined by

$$e(f_n, g_n, F_n) = \max_{s^n \in \mathcal{S}^n} \text{Prob}\,\{F_n(f_n(X(s^n)), g_n(s^n)) \neq (X(s^n), s^n)\}. \tag{1.2}$$

In Theorem 5, Section 4, we characterize the region $\mathcal{R}_{\mathcal{XS}}$ of all achievable pairs of rates (R_1, R_2).

We explain now how this result relates to results obtained previously. The difference between the present problem and the partial side information problem described in Part I, Section 1, § 2 and solved in Theorem 2 is that now s^n has to be reproduced by the decoder.

Theorem 1 determines the optimal R_1, if $R_2 = \log |\mathcal{S}|$, that is the case in which the decoder knows the $\mathcal{S}$-outputs. However, this rate pair lies in general not on the boundary of $\mathcal{R}_{\mathcal{XS}}$.

Choosing $\mathcal{Y} = \mathcal{S}$ and $p(x, y \mid s) = p(x \mid s)\delta(y \mid s)$ we see that an AVS is a special case of an AVCS. However, Theorem 5 does not follow from Theorem 4', because the condition (1.1) does not hold.

(c) In Section 5 we present the general robustification technique, which we announced in Part I. This method in conjunction with the elimination technique of [13] makes it now possible to reduce coding problems for arbitrarily varying sources and channels to those for compound sources and channels (see [6]) *provided that randomized encoding is permitted for sources and channels or the average error concept is used for the channels.* That this can mean a restriction is explained in [13]. We think that an important aspect of the arbitrarily varying channel problem, which has puzzled several mathematicians for so long, is now understood. The method makes use of the fact that for stationary, memoryless sources

and channels the probabilities occurring are invariant under permutations of the time components. This property has not been exploited yet in Information Theory and has led us to the next result.

(d) It has been realized a long time ago by Ahlswede/Körner (see [39]), and likely also by others, that the *direct part* of our MAC coding theorem ([9]) can be derived from the Slepian/Wolf source coding theorem.

A new link between source- and channel-coding, which we found, is given by the mathematically rather simple Covering Lemma 2 in Section 6. It uses the idea of permutations mentioned in (c). Applying random permutations to a maximal error code one can obtain our old decomposition into maximal error codes ([14], presented in [17], [39], used in abstract graph theoretic set up in [44]), which is a (slightly stronger) version of the Slepian/Wolf Theorem. Thus we have the implications:

DMC Coding theorem $\Rightarrow$ DMCS Coding theorem
$\Rightarrow$ MAC Coding theorem (without converse).

Extensions of Covering Lemma 2 should be investigated. We think that this new link will help to greatly simplify multi-user communication theory and also lead to new results.†

(e) In Section 7 we present a candidate $\bar{R}_A(\theta)$ for the rate-distortion function $R_A^*(\theta)$ of an AVS in the case, where the decoder has complete knowledge of the states s^n. We can actually show that $\bar{R}_A(\theta) \geqslant R_A^*(\theta)$. However, since we have not yet proved the opposite inequality, we just state $\bar{R}_A(\theta) = R_A^*(\theta)$ as a conjecture. The validity of $\bar{R}_A(\theta) \leqslant R_A^*(\theta)$ depends on the validity of another conjecture concerning a natural generalisation of the result of Wyner/Ziv [20]. Our approach can best be explained by giving a new proof of the direct part in [20]. The AVS-problem is harder. We present here one essential tool (Coloring Lemma 3A*) because it is also useful otherwise for obtaining "direct" coding theorems. There are a few techniques to prove "converses" in certain cases, but there is no general method to prove "converses". To find such a method is probably the most important and challenging task in multi-user coding theory.

(f) In Sections 8, 9 we analyze how our techniques relate to the Graph Decomposition Theorem (GDT) of Lovász [36], which was recently used in source coding by Csiszár/Körner [38]. In Section 8 we show that the results of [38] can also be derived by our methods, which turn out to be even more general for the kind of balanced coloring problems studied in [38].

In Section 9 we explain that the GDT itself is essentially equivalent to our decomposition into maximal error codes (CDT). Another equiva-

†For recent program see "The best known codes are highly probable and can be produced by a few permutations" by R. Ahlswede and G. Dueck (to appear in the *IEEE Trans. on Inf. Theory*)

lent result is in terms of hypergraph coloring (HCP_1). Also a related hypergraph coloring problem (HCP_2) is formulated, its coding theoretic significance is explained and some open problems are stated. We also investigate the question of obtaining a converse to the GDT. It turns out that for this question the other formulations are more suitable. Our conclusion after the analysis is that the GDT is an elegant formulation of something which we already know in coding theory. Its proof is closely related to, but not identical with, Feinstein's [4] maximal coding method. We explain how one can use the type of proof for channel coding (Minimal Error Lemma).

(g) In Part I we promised results on the two helper side information problem, which originated in [41]. Suppose in (X, Y, Z) Y and Z are the helpers, then we solved the case where Y and Z are independent. This case is fairly easy and our attempt to get from here the general case by some kind of approximation has not been successful. We therefore decided not to write about the problem here. A very special, but very instructive case has been solved in [42]. This case shows that the decomposition phenomenon (see Part I, Section 6) does not occur for the helpers.

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

§ 1 Introductory Remarks

We recall that the pair $(\mathcal{V} \times \mathcal{W}, \mathcal{E})$ is a rectangular hypergraph, if $\mathcal{V}$, $\mathcal{W}$ are finite sets and $\mathcal{E}$ is a family of subsets of $\mathcal{V} \times \mathcal{W}$. $C = \bigcup_{E \in \mathcal{E}} E$ is the carrier of the hypergraph.

If φ (resp. ψ) is a coloring of $\mathcal{V}$ (resp. $\mathcal{W}$), then $\rho = (\varphi, \psi)$ is an orthogonal coloring of $\mathcal{V} \times \mathcal{W}$. The following two types of colorings are needed in coding an AVCS with and without side information at the decoder.

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.

A stronger notion of orthogonal coloring, denoted by ρ_λ^2, is defined by the requirement, that in every edge E, $E \in \mathcal{E}$, at least $(1 - \lambda)|E|$ colors occur, which *occur only once in C.*

More generally one can consider an orthogonal 2-hypergraph $(\mathcal{V} \times \mathcal{W}, \mathcal{E}, (\mathcal{E}_j)_{j=1}^J)$, that is a 2-hypergraph with rectangular vertex set. We denote again by $\rho_\lambda^2 = (\varphi_\lambda, \psi_\lambda)$ 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.

We study first hypergraphs with one edge E only. Using standard random coloring $(V^{|\mathcal{V}|}; W^{|\mathcal{W}|}) = (V_1, \ldots, V_{|\mathcal{V}|}; W_1, \ldots, W_{|\mathcal{W}|})$ with

L_1, L_2 colors we are interested in estimating the probability $P_{\lambda,E}$ of not obtaining a coloring ρ_λ.

In applications to coding we have sequences $(\mathcal{V}_n \times \mathcal{W}_n, \mathcal{E}_n)_{n \in \mathbb{N}} = (\mathcal{X}^n \times \mathcal{Y}^n, \mathcal{E}_n)_{n \in \mathbb{N}}$ of hypergraphs with $|\mathcal{E}_n|$ growing exponentially in n, and if for every $E \in \mathcal{E}_n P_{\lambda,E}$ is superexponentially small, then we know the existence of a coloring of type ρ_λ for $(\mathcal{X}^n \times \mathcal{Y}^n, \mathcal{E}_n)$.

Now notice that for an edge of the form

$$E = A \times B,\ A \subset \mathcal{X}^n,\ B \subset \mathcal{Y}^n \text{ with } |A| = 2$$

for instance

$P_{\lambda,E}$ is only exponentially small and the inequality (7.2) and the bound (7.5) in Part I are not valid. We shall see in the next Section that for the AVCS with side information at the decoder this "line type" edge occurs. However there the collection of all such edges has some additional structure, which makes a combined random-maximal coloring approach possible. For the AVCS without side information we need colorings of type ρ_λ^2 (coloring edges within the carrier) and it can be seen from Example 2 in Section 3 that "line type" edges become a crucial obstacle. Condition (1.1) in Theorem 4′ excludes exactly those edges. For them orthogonal random coloring performs poorly, because there is a strong dependency among the RV's $\{(V_i, W_j) : (i, j) \in E\}$.

It is important to notice that this is not the case for "diagonals" and that therefore "long diagonals" have a very small $P_{\lambda,E}$.

$F \subset \mathcal{V} \times \mathcal{W}$ is called a *diagonal*, if no two elements of F have the same first or second component. *We analyze now random coloring for certain types of edges by decomposing them into diagonals.* This leads us to Coloring Lemmas 6, 7, 8, 9.

Other approaches are conceivable and we propose the

PROBLEM 1. What can be said about the RV's $\{(V_i, W_j) : (i, j) \in E\}$ for a general set $E \subset \mathcal{V} \times \mathcal{E}$? Which probabilistic inequalities or laws can be established? In particular, given N, for which sets with $|E| = N$ does random coloring perform most poorly?

§ 2 TYPES OF EDGES AND PARTITIONING INTO DIAGONALS

We define now four types of edges which occur in the coding problems for AVCS's. Let ζ, ζ_1, ζ_2 be reals with $0 < \zeta_1, \zeta_2; \zeta_1 + \zeta_2 \leqslant \zeta$.

An edge $E \subset \mathcal{V} \times \mathcal{W} = \mathcal{X}^n \times \mathcal{Y}^n$ is said to be of *ζ-point type*, if for $D = |E|$

$$D \leqslant n^\zeta. \tag{2.1}$$

Define now $d_1 = \min\{|E_{|w}| : E_{|w} \neq \phi,\ w \in \mathcal{W}\}$, $d_2 = \min\{|E_{|v}| : E_{|v} \neq \phi,$

$v \in \mathcal{V}\}$, $D_1 = \max\{|E_{|w}| : w \in \mathcal{W}\}$, $D_2 = \max\{|E_{|v}| : v \in \mathcal{V}\}$, that is, the minimal and maximal sizes of cross-sections of E.

E is said to be of $(\zeta, \zeta_1, \zeta_2)$-*diagonal type*, if

$$D_1 \leqslant n^{\zeta_1},\ D_2 \leqslant n^{\zeta_2},\ D \geqslant n^{\zeta}, \tag{2.2}$$

and of $(\zeta, \zeta_1, \zeta_2)$-*rectangle type*, if

$$d_1 \geqslant n^{\zeta_1},\ d_2 \geqslant n^{\zeta_2},\ D \geqslant n^{\zeta}. \tag{2.3}$$

Finally, E is said to be of (ζ_1, ζ_2)-*column type*, if for $\mathcal{V}(E, v) = \{v' : \exists\, w$ with $(v, w), (v', w) \in E\}$

$$|\mathcal{V}(E, v)| \leqslant n^{\zeta_1} \quad \text{for all} \quad v \in \mathcal{V} \text{ and } d_2 \geqslant n^{\zeta_2}. \tag{2.4}$$

The (ζ_1, ζ_2)-*row type* is defined analogously. Those two types may be called *line type*.

Our first result concerns partitions of an arbitrary edge E into diagonals. With E we associate a graph $G(E)$ with vertex set E : the vertices (v, w) and (v', w') are connected, iff $v = v'$ or $w = w'$. deg (v, w) counts the number of vertices connected with (v, w).

PROPOSITION 1. *Let* $E \subset \mathcal{V} \times \mathcal{W}$ *satisfy* $\max_{(v, w) \in E} \deg(v, w) \leqslant T - 1$, *then there exists a partition* $\{F_1, \ldots, F_t\}$ of E into diagonals, such that

(a) $t \leqslant T$ and (b) $|F_i| \leqslant \dfrac{|E|}{2T}$, $1 \leqslant i \leqslant t$.

Proof. Clearly, by Coloring Lemma 2 one can color the vertices with T colors such that adjacent vertices have different colors. A set of vertices with the same color forms a diagonal and we have a partition of E into $t \leqslant T$ diagonals.

To show (b), let us choose among the partitions into T or fewer diagonals one, $\{F_1, \ldots, F_t\}$ say, with a minimal number of diagonals having a cardinality $< \dfrac{|E|}{2T}$. Suppose now that for instance $|F_1| = \alpha|E|T^{-1}$ for $0 < \alpha < \frac{1}{2}$. From $\sum_{i=1}^{t} |F_i| = |E|$ we conclude that for some $i \neq 1$, $|F_i| \geqslant |E|T^{-1}$. Let A_i be the set of vertices from F_i, which are connected with a vertex from F_1. The structure of $G(E)$ is such that $|A_i| \leqslant 2|F_1| = 2\alpha|E|T^{-1}$. Choose a subset $B_i \subset F_i \setminus A_i$ with $|B_i| = \overline{\lceil (1 - 2\alpha)|E|(2T)^{-1} \rceil}$ and define two new diagonals $F_1' = F_1 \cup B_i$, $F_i' = F_i \setminus B_i$. Then $|F_1'| \geqslant |E|(2T)^{-1}$ and $|F_1'| \geqslant \overline{\lceil |E|(2T)^{-1} \rceil} - \overline{\lceil |E|(2T)^{-1} \rceil} + \alpha|E|T^{-1} \geqslant |E|(2T)^{-1}$. This contradicts the definition of the partition $\{F_1, \ldots, F_t\}$ and (b) is proved. Q.E.D.

Our next result is for edges of rectangle or diagonal type.

PROPOSITION 2. $E \subset \mathcal{V} \times \mathcal{W} = \mathcal{X}^n \times \mathcal{Y}^n$ *can be partitioned into diagonals* $\{F_1, \ldots, F_t\}$ *such that*

(*a*) $t \leqslant 2|E|n^{-\min(\zeta_1, \zeta_2)}$, $|F_i| \geqslant 4^{-1}n^{\min(\zeta_1, \zeta_2)}$ *for* $1 \leqslant i \leqslant t$, *if* E *is of* $(\zeta, \zeta_1, \zeta_2)$*-rectangle type.*

(*b*) $t \leqslant 2n^{\max(\zeta_1, \zeta_2)}$, $|F_i| \geqslant 4^{-1}n^{\zeta-\max(\zeta_1, \zeta_2)}$ *for* $1 \leqslant i \leqslant t$, *if* E *is of* $(\zeta, \zeta_1, \zeta_2)$*-diagonal type.*

Proof. Apply Proposition 1 with:

(a) $T = D_1 + D_2$. Since $|E| \geqslant \max(D_2 n^{\zeta_1}, D_1 n^{\zeta_2})$, we have $D_1 \leqslant |E|n^{-\zeta_2}$, $D_2 \leqslant |E|n^{-\zeta_1}$ and therefore the bound on t. Also $|F_i| \geqslant |E|(2T)^{-1} \geqslant (2(n^{-\zeta_1} + n^{-\zeta_2}))^{-1} \geqslant 4^{-1}n^{\min(\zeta_1, \zeta_2)}$.

(b) $T = n^{\zeta_1} + n^{\zeta_2}$. Obviously, $t \leqslant T$ implies $t \leqslant 2n^{\max(\zeta_1, \zeta_2)}$ and $|F_i| \geqslant 2^{-1}n^{\zeta}(n^{\zeta_1} + n^{\zeta_2})^{-1} \geqslant 4^{-1}n^{\zeta-\max(\zeta_1, \zeta_2)}$. Q.E.D.

§ 3. Coloring Most Points Correctly in Their Neighbourhoods

If in addition to a hypergraph $(\mathcal{V}, \mathcal{E})$, $\mathcal{V} = \{1, \ldots, |\mathcal{V}|\}$, $\mathcal{E} = \{E_1, \ldots, E_d\}$ we are given $u_i \in E_i$, $1 \leqslant i \leqslant d$, $u_i < u_j$ for $i < j$, then we speak of a neighbourhood system or matching system (NS). Here E_i is the neighbourhood of u_i.

We are interested in colorings of the vertices $\mathcal{V}$, denoted by μ_λ, such that for at least $(1 - \lambda)d$ u_i's

$$\mu_\lambda(u_i) \neq \mu_\lambda(v) \text{ for all } v \in E_i,\ v \neq u_i. \tag{2.5}$$

Let $V^{|\mathcal{V}|} = V_1 \ldots V_{|\mathcal{V}|}$ be a standard random L-coloring of $\mathcal{V}$. For $1 \leqslant i \leqslant d$ define

$$g_i(V_1, \ldots, V_{u_i}) = \begin{cases} 1 & \text{if } V_{u_i} \neq V_j \text{ for all } j < u_i \text{ and } j \in E_i \\ 0 & \text{otherwise} \end{cases} \tag{2.6}$$

and

$$G_i(V_{u_i}, V_{u_i+1}, \ldots, V_{|\mathcal{V}|}) = \begin{cases} 1 & \text{if } V_{u_i} \neq V_j \text{ for all} \\ & j > u_i \text{ and } j \in E_i \\ 0 & \text{otherwise.} \end{cases} \tag{2.7}$$

Clearly, if $\sum_{i=1}^{d} g_i \geqslant (1 - \lambda)\, d$ and $\sum_{i=1}^{d} G_i \geqslant (1 - \lambda)d$, then we have a $\mu_{2\lambda}$ coloring.

Now observe that for $1 \leqslant i \leqslant d$

$$\begin{aligned} &\text{Prob}\,(g_i = 1 \mid g_{i-1} = \epsilon_{i-1}, \ldots, g_1 = \epsilon_1) \\ &\geqslant \frac{L - |E_i|}{L} \geqslant \frac{L - t}{L},\ t = \max_{1 \leqslant i \leqslant d} |E_i|, \end{aligned} \tag{2.8}$$

and by the usual arguments (see the derivation of (2.14) in Part I)

$$\text{Prob}\left(\sum_{i=1}^{d} g_i < (1 - \lambda)d\right) \leqslant \exp\left\{\left[h(\lambda) + \lambda \log \frac{t}{L}\right]d\right\},$$

if $t(L - t)^{-1}(1 - \lambda) \leqslant 1$ or, equivalently, if $L \geqslant t\lambda^{-1}$.

Since the same inequality holds, if g_i is replaced by G_i, we have proved

COLORING LEMMA 6. *For an NS* $(\mathcal{CV}, (E_i, u_i)_{1\leqslant i\leqslant d})$ *standard random L-coloring leads to a coloring* $\mu_{2\lambda}$, $0 < \lambda < \frac{1}{2}$, *with a probability greater than*

$$1 - 2\exp\left\{\left[h(\lambda) + \lambda \log \frac{t}{L}\right] d\right\}, \text{ if } L \geqslant \max_{1\leqslant i\leqslant d} |E_i|\lambda^{-1} = t\lambda^{-1}.$$

§ 4 ONE SIDED BALANCED COLORINGS OF RECTANGULAR HYPERGRAPHS

Let $E \subset \mathcal{CV} \times \mathcal{QW}$ be arbitrary, let L be any positive integer and let $D_2 = \max_{v\in\mathcal{CV}} |E_{|v}|$. For an L-coloring φ of $\mathcal{CV}$ we consider

$$b_\varphi(l) = \sum_{v\in\varphi^{-1}(l)} |E_{|v}| \text{ and } b_\varphi = \max_{1\leqslant l\leqslant L} b_\varphi(l).$$

COLORING LEMMA 7. *If* $V^{|\mathcal{CV}|}$ *denotes the standard random L-coloring on* $\mathcal{CV}$, *then for any* $\alpha > 0$

$$\text{Prob}\,(b_{V^{|\mathcal{CV}|}} > \alpha \max(|E|L^{-1}, D_2))$$
$$\leqslant L\cdot\exp\left\{-\frac{\alpha}{2}\max(|E|L^{-1}D_2^{-1}, 1) + |E|L^{-1}D_2^{-1}\right\}.$$

Proof. Define for $v \in \mathcal{CV}$, $1 \leqslant l \leqslant L$,

$$f_v^l = \begin{cases} 1 & \text{if } V_v = l \\ 0 & \text{otherwise.} \end{cases}$$

Then for $\gamma > 0$

$$\text{Prob}\,\Big(\sum_{v\in\mathcal{CV}} |E_{|v}| f_v^l > \alpha \max(|E|L^{-1}, D_2)\Big)$$
$$\leqslant \exp\{-\alpha\gamma \max(|E|L^{-1}, D_2)\} \prod_{v\in\mathcal{CV}} \mathbf{E}\exp\{\gamma|E_{|v}|f_v^l\}.$$

Now

$$\prod_v \mathbb{E}\exp\{\gamma|E_{|v}|f_v^l\} = \prod_v\left(\frac{1}{L}\exp(\gamma|E_{|v}| + \frac{L-1}{L}\right)$$
$$= \prod_v\left(1 + \frac{1}{L}\left(\gamma|E_{|v}| + \frac{\gamma^2}{2!}|E_{|v}|^2 + \dots\right)\right)$$
$$\leqslant \exp\left\{\sum_v \frac{1}{L}\left(\gamma|E_{|v}| + \frac{\gamma^2}{2!}|E_{|v}|^2 + \dots\right)\right\}$$
$$\leqslant \exp\left\{\sum_v \frac{1}{L}\gamma|E_{|v}|(1 + \gamma D_2 + \gamma^2 D_2^2 + \dots)\right\}. \tag{2.9}$$

For $\gamma = \frac{1}{2}D_2^{-1}$ this is equal to $\exp\{|E|L^{-2}D_2^{-1}\}$ and the probability in question is smaller than

$$\exp\{-\alpha\tfrac{1}{2}\max(|E|L^{-1}D_2^{-1}, 1) + |E|L^{-1}D_2^{-1}\}.$$

Since this holds for all $1 \leqslant l \leqslant L$, the result follows. Q.E.D.

Remark 1. In applications we choose $\alpha = 4 + 2e^{\epsilon n}$ and thus get the double exponential bound $L \exp\{-e^{\epsilon n}\}$.

§ 5 Orthogonal Coloring of a Long Diagonal Within an Edge

We consider the situation $F \subset E \subset \mathcal{V} \times \mathcal{W}$, where F is a diagonal, $|F| = d$.

We use the standard orthogonal random $L_1 \times L_2$-coloring $(V^{|\mathcal{V}|}; W^{|\mathcal{W}|})$:

$$L_1 \geqslant \gamma\lambda^{-1}D_1;\ L_2 \geqslant \gamma\lambda^{-1}\alpha \max(|E|L_1^{-1}, D_2). \tag{2.10}$$

Here $0 < \lambda < \frac{1}{4}$, $\alpha > 2$ and $\gamma > 1$. D_1, D_2 are the maximal sizes of cross-sections of E.

We estimate now

$$\text{Prob}\,((V^{|\mathcal{V}|}, W^{|\mathcal{W}|}) \text{ is not a } \rho^2_{4\lambda} \text{ coloring of } F \text{ within } E)$$

from above by a sum $p_1 + p_2 + p_3$ of three probabilities.

Step 1. Denote the elements of F by (a_i, b_i), $1 \leqslant i \leqslant d$, and consider the NS system $(\mathcal{V}, E_{|b_i}, a_i)_{1 \leqslant i \leqslant d})$.

Coloring Lemma 6 gives a bound on the probability p_1 that $V^{|\mathcal{W}|}$ is not a $\mu_{2\lambda}$ coloring of this NS system.

Step 2. Coloring Lemma 7 gives a bound on

$$p_2 = \text{Prob}\,(b_{V^{|\mathcal{V}|}} > \alpha \max(|E|L_1^{-1}, D_2)).$$

The property

$$b_{V^{|\mathcal{V}|}} \leqslant \alpha \max(|E|L_1^{-1}, D_2)$$

implies that for all l, $1 \leqslant l \leqslant L_1$,

$$H_l := \bigcup_{V_v = l} E_{|v} \quad \text{satisfies}$$

$$|H_l| \leqslant \alpha \max(|E|L_1^{-1}, D_2).$$

Define the NS system $(\mathcal{W}, (G_i, b_i)_{1 \leqslant i \leqslant d})$, where $G_i = H_l$ iff $b_i \in H_l$.

Step 3. Apply now Coloring Lemma 6 with $L = L_2$ to the NS system $(\mathcal{W}, (G_i, b_i)_{1 \leqslant i \leqslant d})$ in order to obtain a bound on p_3, the probability that $W^{|\mathcal{W}|}$ is not a $\mu_{2\lambda}$ coloring of $(\mathcal{W}, (G_i, b_i)_{1 \leqslant i \leqslant d})$.

L_1 and L_2 are chosen in (2.10) such that

$$p_1 + p_2 + p_3 \leqslant 2 \exp\left\{\left[h(\lambda) + \lambda \log \frac{D_1}{L_1}\right] d\right\}$$
$$+ L_1 \exp\left\{-\frac{\alpha}{2} \max(|E|L_1^{-1}D_2^{-1}, 1) + |E|L_1^{-1}D_2^{-1}\right\}$$
$$+ 2 \exp\left\{\left[h(\lambda) + \lambda \log \frac{D_2}{L_2}\right] d\right\},$$

where $D_2' = \alpha \max(|E|L_1^{-1}, D_2)$, and such that the right side expression is smaller than

$$4 \exp\left\{\left[h(\lambda) + \lambda \log \frac{\lambda}{\gamma}\right]\right\} + L_1 \exp\left\{-\left(\frac{\alpha}{2} - 1\right)\right\}.$$

We thus have proved

COLORING LEMMA 8. *Let $E \subset \mathcal{V} \times \mathcal{W}$ be an edge with D_1, D_2 as maximal sizes of cross-sections and let L_1, L_2 be integers with $L_1 \geq \gamma\lambda^{-1}D_1$,*

$$L_2 \geq \gamma\lambda^{-1}\alpha \max(|E|L_1^{-1}, D_2), \text{ where } 0 < \lambda < \tfrac{1}{4}, \alpha > 2 \text{ and } \gamma > 1.$$

Then the orthogonal random $L_1 \times L_2$-coloring $(V^{|\mathcal{V}|}, W^{|\mathcal{W}|})$ is a $\rho_{4\lambda}^2$ of a diagonal $F \subset E$ with a probability larger than

$$1 - 4 \exp\left\{\left[h(\lambda) + \lambda \log \frac{\lambda}{\gamma}\right]|F|\right\} - L_1 \exp\left\{-\left(\frac{\alpha}{2} - 1\right)\right\}.$$

As an immediate consequence of this Lemma we get

COLORING LEMMA 9. *Let $(\mathcal{V} \times \mathcal{W}, \mathcal{E}, (\mathcal{E}_j)_{j=1}^J)$ be a 2-hypergraph and let $D_1^* = \max_{1 \leq j \leq J} D_{1j}$, $D_2^* = \max_{1 \leq j \leq J} D_{2j}$, where D_{1j}, D_{2j} are the maximal sizes of cross-sections of $E_j \in \mathcal{E}$.*

For integers L_1, L_2 with $L_1 \geq \gamma\lambda^{-1}D_1^*$, $L_2 \geq \gamma\lambda^{-1}\alpha \max(\max_{1 \leq j \leq J} |E_j|L_1^{-1}, D_2^*)$ $(0 < \lambda < \frac{1}{4}, \alpha > 2, \gamma > 1)$ the orthogonal $L_1 \times L_2$-coloring $(V^{|\mathcal{V}|}, W^{|\mathcal{W}|})$ is a $\rho_{4\lambda}^2$ of the 2-hypergraph with a probability greater than

$$1 - N\left(4 \exp\left\{\left[h(\lambda) + \lambda \log \frac{\lambda}{\gamma}\right] d^*\right\} - L_1 \exp\left\{-\left(\frac{\alpha}{2} - 1\right)\right\}\right),$$

if every subedge E_j' can be partitioned into diagonals of length $\geq d^*$ and if $N \geq d^{*-1} \sum_{i,j} |E_j'|$.

3. THE AVCS WITH AND WITHOUT SIDE INFORMATION AT THE DECODER REVISITED

§ 1. STRUCTURE OF THE HYPERGRAPH $(\mathcal{X}^n, \mathcal{Y}^n, (\mathcal{G}_\delta(X^n, Y^n \mid s^n))_{s^n \in \mathcal{S}^n})$

The coding problem for the AVCS with side information at the decoder was formulated in Section 1, § 3 of Part I. Its rate region $\mathcal{R}_D^{\mathcal{A}}$ is characterized there in Theorem 3. For its proof we need a classification of the edges of $(\mathcal{V}, \mathcal{W}, \mathcal{E}) = (\mathcal{X}^n, \mathcal{Y}^n, (\mathcal{G}_\delta(X^n, Y^n \mid s^n))_{s^n \in \mathcal{S}^n})$ according to their $(\zeta, \zeta_1, \zeta_2)$-types introduced in Section 2. This requires estimates on the sizes of edges and their maximal and minimal cross-sections. We denote those by $D(s^n)$, $D_1(s^n)$, $D_2(s^n)$, $d_1(s^n)$, $d_2(s^n)$ (see Section 2 for the definitions) and frequently we write $E(s^n)$ instead of $\mathcal{G}_\delta(X^n, Y^n \mid s^n)$.

It follows from Lemma $G_1(c)$ in Section 3 of Part I that

$$D(s^n) = \exp\{H(X(s^n), Y(s^n)) + 0(\sqrt{n})\} \text{ for all } s^n \in \mathcal{S}^n. \qquad (3.1)$$

From Lemma *CS* in Section 3, Part I, and the fact that we actually have equality in (3.22) of Part I it follows that

$$|\mathcal{G}_0(X^n, Y^n \mid s^n)_{|x^n}| = \exp\{H(Y(s^n) \mid X(s^n)) + 0(\sqrt{n})\}$$
$$\text{for all } x^n \text{ with } \mathcal{G}_0(X^n, Y^n \mid s^n)_{|x^n} \neq \emptyset, \quad (3.2)$$

$$|\mathcal{G}_0(X^n, Y^n \mid s^n)_{|y^n}| = \exp\{H(X(s^n) \mid Y(s^n)) + 0(\sqrt{n})\}$$
$$\text{for all } y^n \text{ with } \mathcal{G}_0(X^n, Y^n \mid s^n)_{|y^n} \neq \emptyset. \quad (3.3)$$

(3.2) and (3.3) imply that

$$\begin{aligned} D_1(s^n) &= \exp\{H(X(s^n) \mid Y(s^n)) + 0(\sqrt{n})\}, \\ D_2(s^n) &= \exp\{H(Y(s^n) \mid X(s^n)) + 0(\sqrt{n})\}, \end{aligned} \quad (3.4)$$

and that

$$d_i(s^n) \exp\{0(\sqrt{n})\} \geqslant D_i(s^n) \geqslant d_i(s^n) \text{ for } i = 1, 2. \quad (3.5)$$

The inequality $\min\{H(X(s) \mid Y(s)) : s \in \mathcal{S}, H(X(s) \mid Y(s)) > 0\} > 0$ and (3.4) imply

$$|\{t : 1 \leqslant t \leqslant n, H(X(s_t) \mid Y(s_t)) = 0\}| \geqslant n - 0(\sqrt{n}) - 0(\log D_1(s^n)). \quad (3.6)$$

Similarly one shows

$$|\{t : 1 \leqslant t \leqslant n, H(Y(s_t) \mid X(s_t)) = 0\}| \geqslant n - 0(\sqrt{n}) - 0(\log D_2(s^n)), \quad (3.7)$$

$$|\{t : 1 \leqslant t \leqslant n, H(X(s_t), Y(s_t) = 0\}| \geqslant n - 0(\sqrt{n}) - 0(\log D(s^n)). \quad (3.8)$$

Now we start with the classification of edges. Choose $(\zeta^0, \zeta_1^0, \zeta_2^0) = (5\sqrt{n}\log n, \sqrt{n}\log n, \sqrt{n}\log n)$ and denote by $\mathcal{E}^0_{\text{point}}$ (resp. $\mathcal{E}^0_{\text{rect.}}$) the set of edges in $\mathcal{E}$ of $(\zeta^0, \zeta_1^0, \zeta_2^0)$-point (resp. rectangle) type. If now $E(s^n) \notin \mathcal{E}^0_{\text{point}} \cup \mathcal{E}^0_{\text{rect.}}$, then by (2.1) and (2.3)

$$|E(s^n)| \geqslant n^{5\sqrt{n}\log n} \quad \text{and} \quad d_i(s^n) < n^{\sqrt{n}\log n}$$
$$\text{for } i = 1 \text{ or for } i = 2 \text{ or for } i = 1, 2. \quad (3.8a)$$

If $d_i(s^n) < n^{\sqrt{n}\log n}$ for $i = 1, 2$, then, by (3.5), $E(s^n)$ is of $(5\sqrt{n}\log n, 2\sqrt{n}\log n, 2\sqrt{n}\log n)$-diagonal type for n large. Write $\mathcal{E}^1_{\text{diag.}}$ for the set of those edges.

If just $d_i(s^n) < n^{\sqrt{n}\log n}$ and therefore by (3.5) $D_1(s^n) < n^{2\sqrt{n}\log n}$ for n large, then (3.6) implies for the Hamming distance:

$$\text{dist}(x^n, x'^n) \leqslant c\sqrt{n}\log^2 n, \text{ if } x^n, x'^n \in E(s^n)_{|y^n} \quad (3.9)$$

for some y^n. c is a constant.

Therefore

$$|\mathcal{V}(E(s^n), x^n)| \leqslant n^{2c\sqrt{n}\log^2 n} \quad \text{for } n \geqslant |\mathcal{X}|, x^n \in \mathcal{X}^n \quad (3.10)$$

and we have a $(2c\sqrt{n}\log^2 n, \sqrt{n}\log n)$-column type, because also $d_2(s^n) \geqslant n^{\sqrt{n}\log n}$. Denoting the set of those edges by $\mathcal{E}^2_{\text{col}}$ and correspondingly

by $\mathcal{E}^2_{\text{row}}$ we can state

PROPOSITION 1.

$$\mathcal{E} = \mathcal{E}^0_{\text{point}} \cup \mathcal{E}^0_{\text{rect.}} \cup \mathcal{E}^0_{\text{diag.}} \cup \mathcal{E}^2_{\text{col.}} \cup \mathcal{E}^2_{\text{row}}.$$

Write $S^n = S^0_{\text{point}} \cup S^0_{\text{rect.}} \cup S^0_{\text{diag.}} \cup S^2_{\text{col.}} \cup S^2_{\text{row}}$ for the corresponding sets of indices s^n.

Notice that (3.9) holds for *all* $E(s^n) \in \mathcal{E}^2_{\text{col.}}$ and therefore the stronger statement follows:

$$|\cup\{\mathcal{V}(E(s^n), x^n) : s^n \in S^2_{\text{col.}}\}| \leqslant n^{2c\sqrt{n}\log^2 n}. \tag{3.11}$$

Also, since only $D_1(s^n) < n^{2\sqrt{n}\log n}$ was used to derive (3.9), we have for $\mathcal{E}^0_{\text{point}} \cup \mathcal{E}^2_{\text{col.}}$

$$|\cup\{\mathcal{V}(E(s^n), x^n) : s^n \in S^0_{\text{point}} \cup S^0_{\text{col.}}\}| \leqslant n^{5c\sqrt{n}\log^2 n} \tag{3.12}$$

for all $x^n \in \mathcal{X}^n$.

Correspondingly for $\mathcal{E}^0_{\text{point}} \cup \mathcal{E}^2_{\text{row}}$

$$|\cup\{\mathcal{W}(E(s^n), y^n) : s^n \in S^0_{\text{point}} \cup S^0_{\text{col.}}\}| \leqslant n^{5c\sqrt{n}\log^2 n} \tag{3.13}$$

for all $y^n \in \mathcal{Y}^n$.

§ 2. COLORING THE HYPERGRAPH

The graph $(\mathcal{X}^n, \mathcal{A})$, where $(x^n, x'^n) \in \mathcal{A}$ exactly if $x'^n \in \cup\{\mathcal{V}(E(s^n), x^n) : s^n \in S^0_{\text{point}} \cup S^0_{\text{col.}}\}$, has by (3.12) a maximal degree less than $n^{5c\sqrt{n}\log n}$. Coloring Lemma 2 implies therefore the existence of a strict coloring φ_0 for this graph with

$$\|\varphi_0\| \leqslant n^{5c\sqrt{n}\log^2 n} + 1. \tag{3.14}$$

The corresponding coloring on $\mathcal{Y}^n$ shall be ψ_0. By definition of $\mathcal{V}(E, x^n)$ (resp. $\mathcal{W}(E, y^n)$)

$$E_{|x^n} \cap E_{|x'^n} = \emptyset \quad \text{for} \quad x'^n \notin \mathcal{V}(E, x^n)$$

$$(\text{resp. } E_{|y^n} \cap E_{|y'^n} = \emptyset \quad \text{for} \quad y'^n \notin \mathcal{W}(E, y^n)).$$

Therefore $\rho_0 = (\varphi_0, \psi_0)$ is a strict orthogonal coloring of $(\mathcal{X}^n, \mathcal{Y}^n, \mathcal{E}^0_{\text{point}})$. Moreover,

$$\varphi_0(x^n) = \varphi_0(x'^n) \Rightarrow E(s^n)_{|x^n} \cap E(s^n)_{|x'^n} = \emptyset \quad \text{for all } s^n \in S^2_{\text{col.}} \tag{3.15}$$

ψ_0 plays the corresponding role for S^2_{row}.

We now use standard orthogonal random $L_1 \times L_2$-coloring on $\mathcal{X}^n \times \mathcal{Y}^n$. In order to achieve a coloring $\rho^1 = (\varphi^1, \psi^1)$, which is of type ρ_λ for $(\mathcal{X}^n, \mathcal{Y}^n, \mathcal{E}^1_{\text{diag.}} \cup \mathcal{E}^0_{\text{rect.}})$, and at the same such that φ^1 (resp. ψ^1) is of type Φ_λ(resp. ψ_λ) for

$$\mathcal{H}_{\text{row}} = (\mathcal{X}^n, \{E(s^n)_{|y^n} : y^n \in \mathcal{Y}^n, \quad s^n \in S^2_{\text{row}}\})$$

$$(\text{resp. } \mathcal{H}_{\text{col.}} = (\mathcal{Y}^n, \{E(s^n)_{|x^n} : x^n \in \mathcal{X}^n, \quad s^n \in S^2_{\text{col.}}\})).$$

(3.15) implies that then $((\varphi_0, \varphi^1), (\psi_0, \psi^1))$ is of type ρ_λ for $(\mathcal{X}^n, \mathcal{Y}^n, \mathcal{E})$.

STEP 1. For an edge $E(s^n) \in \mathcal{E}^1_{\text{diag.}}$ we know that $D_1(s^n)$, $D_2(s^n) \leqslant n^{\sqrt{n}\log n}$, $D(s^n) \geqslant n^{5\sqrt{n}\log n}$ and for $E(s^n) \in \mathcal{E}^0_{\text{rect.}}$ we know that

$$d_1(s^n), d_2(s^n) \geqslant n^{\sqrt{n}\log n}, \quad D(s^n) \geqslant n^{5\sqrt{n}\log \sqrt{n}}.$$

Proposition 2, Section 2 implies that in both cases $E(s^n)$ can be partitioned into $\mathcal{E}(s^n) = \{F_1(s^n), \ldots, F_{t(s^n)}(s^n)\}$ such that

$$|F_i(s^n)| \geqslant 4^{-1}n^{\sqrt{n}\log n}, \quad 1 \leqslant i \leqslant t(s^n), \quad s^n \in \mathcal{S}^0_{\text{rect.}} \cup \mathcal{S}^1_{\text{diag.}}. \tag{3.16}$$

Apply now Coloring Lemma 9 to the 2-hypergraph

$$\{\mathcal{X}^n, \mathcal{Y}^n, \{E(s^n) : s^n \in \mathcal{S}^0_{\text{rect.}} \cup \mathcal{S}^1_{\text{diag.}}\}, (\mathcal{E}(s^n))_{s^n \in \mathcal{S}^0_{\text{rect.}} \cup \mathcal{S}^1_{\text{diag.}}}\}$$

with

$$\lambda = \lambda_n = \exp\{-\tfrac{1}{4}\sqrt{n}\log n\}, \alpha = 2 + 2n^2, d^* = 4^{-1}n^{\sqrt{n}\log n},$$
$$\gamma = e^3, N = (|\mathcal{X}|\,|\mathcal{Y}|\,|\mathcal{S}|)^n. \tag{3.17}$$

Then for sufficiently large n

$$N(4\exp\{(-(1-\lambda)\log(1-\lambda) - \lambda\log\gamma)d^*\} + L_1\exp\left\{-\left(\frac{\alpha}{2} - 1\right)\right\}$$
$$\leqslant 4(|\mathcal{X}|\,|\mathcal{Y}|\,|\mathcal{S}|)^n\exp\{-\lambda_n d^*\} + |\mathcal{X}|^n\exp\{-n^2\}$$
$$\leqslant \exp\{-\tfrac{1}{8}\sqrt{n}\log^2 n\} < \tfrac{1}{4} \text{ for } n \text{ large enough.}$$

The conditions for L_1, L_2 are

$$L_1 \geqslant \exp\{3 + \sqrt{n}\log n + \max_s H(X(s) \mid Y(s))n + 0(\sqrt{n})\}$$

$$L_2 \geqslant (2 + 2n^2)\exp\{3 + \sqrt{n}\log n\}\max(L_1^{-1}\exp\{\max_s H(X(s),$$
$$Y(s))n + 0\sqrt{n})\}, \exp\{\max_s H(Y(s)\mid X(s))n + 0(\sqrt{n})\}).$$

They are fulfilled, if

$$L_1 \geqslant \exp\{\max_{s\in\mathcal{S}} H(X(s) \mid Y(s))n + 2\sqrt{n}\log n\}$$
$$L_2 \geqslant \exp\{\max_{s\in\mathcal{S}} H(Y(s) \mid X(s))n + 2\sqrt{n}\log n\} \tag{3.18}$$
$$L_1L_2 \geqslant \exp\{\max_{s\in\mathcal{S}} H(X(s), Y(s)) + 2\sqrt{n}\log n\}.$$

STEP 2. By Coloring Lemma 3A already the standard random L_1-coloring of $\mathcal{X}^n$ leads to a Φ_λ of $\mathcal{H}_{\text{row}}$ with probability $> 1 - \frac{1}{4}$ if

$$L_1 \geqslant \lambda^{-1}|E(s^n)_{|y^n}| \text{ for all } s^n \in \mathcal{S}^2_{\text{row}}, y^n \in \mathcal{Y}^n \tag{3.19}$$

and if

$$\sum_{E(s^n)|y^n \neq \emptyset} \exp\{|E(s^n)_{|y^n}|(h(\lambda) + \lambda\log(|E(s^n)|_{y^n}L_1^{-1})\} < \tfrac{1}{4}. \tag{3.20}$$

From (3.3) and $d_2(s^n) \geqslant \sqrt{n} \log n$ we know

$$\exp\{\max_s H(X(s) \mid Y(s))n + 0(\sqrt{n})\} \geqslant |E(s^n)_{|y^n}| \geqslant n^{\sqrt{n} \log n} \quad (3.21)$$

if $E(s^n)_{|y^n} \neq \emptyset$ and $s^n \in \mathcal{S}^2_{\text{row}}$.

For $\lambda = \lambda_n = \exp\{-\frac{1}{4}\sqrt{n} \log n\}$ and L_1 as in (3.18) therefore (3.19) holds. The left side expression in (3.20) is smaller than

$$(|\mathcal{Y}| \, |\mathcal{S}|)^n \exp\{n^{\sqrt{n} \log n}(h(\lambda_n) - \lambda_n \sqrt{n} \log n)\}.$$

Since $h(\lambda_n) \leqslant \frac{1}{2}\sqrt{n} \log n \exp(-\frac{1}{4}\sqrt{n} \log n\}$, this is smaller than $(|\mathcal{Y}| \, |\mathcal{S}|)^n \exp\{-\frac{1}{2}\sqrt{n} \log n \, n^{2^{-1}\sqrt{n} \log n}\} \leqslant \exp\{-e^{\sqrt{n}}\} < \frac{1}{4}$ for $n \geqslant |\mathcal{Y}| \, |\mathcal{S}|$.

For $\mathcal{H}_2$ the argument is symmetrically the same. We have proved that there exists an $L_1 \times L_2$-coloring $\rho^1 = (\varphi^1, \psi^1)$ which is of type ρ_{λ_n} for $(\mathcal{X}^n, \mathcal{Y}^n, \mathcal{E}^1_{\text{diag.}} \cup \mathcal{E}^0_{\text{rect.}})$ and such that φ^1(resp. ψ^1) is of type Φ_{λ_n} for $\mathcal{H}_1$(resp. $\mathcal{K}_2$). By the foregoing explanations $((\varphi_0, \varphi_1), (\psi_0, \psi_1))$ is a coloring of type ρ_{λ_n} and by (3.14)

$$\begin{aligned} \|(\varphi_0, \varphi^1)\| &\leqslant L_1 \cdot \exp\{5c\sqrt{n} \log^3 n\}, \\ \|(\psi_0, \psi^1)\| &\leqslant L_2 \cdot \exp\{5c\sqrt{n} \log^3 n\} \end{aligned} \quad (3.22)$$

with L_1, L_2 as in (3.18).

§ 3. The Error Probability of the Code

Define now the code (f_n, g_n, F_n) with $f_n = (\varphi_0, \varphi^1)$, $g_n = (\psi_0, \psi^1)$ and

$$F_n(l_1, l_2, s^n) = \mathcal{G}_\delta(X^n, Y^n \mid s^n) \cap (f_n^{-1}(l_1) \times g_n^{-1}(l_2)). \quad (3.23)$$

Lemma G_1 (a), (b) implies that the worst case error probability is bounded by

$$\lambda_n \exp\{0(\sqrt{n})\} + 0\left(\frac{1}{\delta^2}\right) \leqslant \exp\left\{-\frac{1}{2}\sqrt{n} \log n\right\} + 0\left(\frac{1}{\delta^2}\right)$$

for n large enough. From this and (3.22) the direct part of Theorem 3 follows by choosing δ sufficiently large.

The converse part is immediate, because for every s

$$R_1 \geqslant H(X(s) \mid Y(s)), \; R_2 \geqslant H(Y(s) \mid X(s)),$$

$$R_1 + R_2 \geqslant H(X(s), Y(s)) \text{ has to hold.} \qquad \text{Q.E.D.}$$

§ 4. Proof of Theorem 4'

After the foregoing analysis it is now easy to derive with the help of Coloring Lemma 9 the

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

$$H(X(s) \mid Y(s)) \cdot H(Y(s) \mid X(s)) > 0 \text{ for all } s \in \mathcal{S}, \quad (3.24)$$

then its rate region $\mathcal{R}^{\mathcal{A}}$ equals $\overline{\mathcal{R}}$.

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

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

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

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

(For the necessary definitions see (1.2), (1.3), (1.14) in Part I.)

Proof. The converse has been explained in Part I. We now consider again the hypergraph

$$(\mathcal{X}^n, {}^{Q}\mathcal{Y}^n, (G_\delta(X^n, Y^n \mid s^n)_{s^n \in S^n}) = ({}^{C}V, {}^{'}W, (E(s^n))_{s^n \in S^n})$$

with the carrier C:

$$C = \bigcup_{s^n \in S^n} E(s^n). \tag{3.25}$$

We know from Lemma *CS* and the Carrier Lemma in Section 3, Part I, that

$$|C| \leqslant \exp\{\max_{\bar{s} \in \bar{S}} H(X(\bar{s}), Y(\bar{s}))n + 0(\sqrt{n})\} \tag{3.26}$$

$$|C_{|y^n}| \leqslant \exp\{\max_{\bar{s} \in \bar{S}} H(X(\bar{s}) \mid Y(\bar{s}))n + 0(\sqrt{n})\}$$

$$|C_{|x^n}| \leqslant \exp\{\max_{\bar{s} \in \bar{S}} H(Y(\bar{s}) \mid X(\bar{s}))n + 0(\sqrt{n})\}$$

for all $(x^n, y^n) \in \mathcal{X}^n \times {}^{Q}\mathcal{Y}^n$.

Condition (3.24) and (3.2), (3.3), (3.5) imply that

$$d_i(s^n) \geqslant \exp\{cn\},\ c > 0, \text{ for } i = 1, 2,\ s^n \in \mathcal{S}^n. \tag{3.27}$$

This means that *all* edges are of $(2cn(\log n)^{-1}, cn(\log n)^{-1}, cn(\log n)^{-1})$-rectangle type. Proposition 2 in Section 2 implies that $E(s^n)$ can be partitioned into diagonals $\mathcal{E}(s^n) = \{F_1(s^n), \ldots, F_{t(s^n)}(s^n)\}$ such that

$$|F_j(s^n)| \geqslant 4^{-1} \exp\{cn\},\ 1 \leqslant j \leqslant t(s^n),\ s^n \in \mathcal{S}^n. \tag{3.28}$$

Apply now Coloring Lemma 9 to the 2-hypergraph $\{\mathcal{X}^n, {}^{Q}\mathcal{Y}^n, \{C\}, \mathcal{E}(C)\}$, where

$$\mathcal{E}(C) = \bigcup_{s^n \in S^n} \mathcal{E}(s^n), \tag{3.29}$$

with $\lambda = \lambda_n = \exp\{-\sqrt{n} \log n\}$, $\alpha = 2 + 2n^2$,

$$d^* = 4^{-1} \exp\{cn\},\ \gamma = e^3,\ N = (|\mathcal{X}|\,|{}^{Q}\mathcal{Y}|\,|\mathcal{S}|)^n.$$

Then $N(4 \exp\{(-(1-\lambda)\log(1-\lambda) - \lambda \log \gamma)d^*\}) + L_1 \exp\left\{-\left(\frac{\alpha}{2} - 1\right)\right\}$

$\leqslant \exp\{-\frac{1}{2}cn\} < 1$ for n large enough. The conditions on L_1, L_2 are satisfied if the inequalities in (3.18) hold also with $\mathcal{S}$ replaced by $\bar{\mathcal{S}}$.

The $L_1 \times L_2$-coloring obtained is of type $\rho_{4\lambda_n} = (\varphi_{4\lambda_n}, \psi_{4\lambda_n})$. Since *all* diagonals are colored properly within C also all edges $E(s^n)$ are colored properly within C.

Define the code (f_n, g_n, F_n) by $f_n = \rho_{4\lambda_n}$, $g_n = \psi_{4\lambda_n}$ and

$$F_n(l_1, l_2) = \begin{cases} (\varphi^{-1}_{4\lambda_n}(l_1), \psi^{-1}_{4\lambda_n}(l_2)) \text{ if this set has exactly} \\ \qquad\qquad\qquad\qquad\qquad\qquad 1 \text{ element in } C \\ \text{any decision otherwise.} \end{cases}$$

Since for all $s^n \in \mathcal{S}^n$, $\text{Prob}((X(s^n), Y(s^n)) \in C) = 1 - 0\left(\frac{1}{\delta^2}\right)$, it follows from Lemma G_1 that the worst case error probability is bounded by $4\lambda_n \exp\{0(\sqrt{n})\} + 0\left(\frac{1}{\delta^2}\right)$ and this can be made arbitrarily small by choosing δ and n large. Q.E.D.

§ 5. A Counterexample

We now show that condition (3.24) in Theorem 4′ cannot be replaced by

$$H(X(s), Y(s)) > 0 \quad \text{for all } s \in \mathcal{S}. \tag{3.30}$$

For this we modify the Example on page 82 of Part I. Recall that there $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$, $p(x, y \mid s) \in \{0, 1\}$ for all $(x, y, s) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{S}$. Furthermore, the AVCS is such that the matrix M, which has a 1 as entry exactly when $p(x, y \mid s) = 1$ for some $s \in \mathcal{S}$ and a 0 otherwise, is of the form

$$M = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}.$$

Replace now $\mathcal{Y}$ by $\mathcal{Y}^* = \mathcal{Y} \times \{1, 2\}$ and build M^* by replacing in M every 1 by $\begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$ and every 0 by $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$. The original row i is thus replaced by the two identical rows $(i, 1)$, $(i, 2)$. An elementary calculation shows that now

$$\bar{\mathcal{R}} = \{(R_1, R_2) : R_1 \geqslant \log 2, R_2 \geqslant \log 4, R_1 + R_2 \geqslant \log 12\}$$

and that $\min_s H(X(s), Y(s)) = \log 2 > 0$.

We show that $(R_1, R_2) = (\log 3, \log 6) \in \bar{\mathcal{R}}$ is not achievable. For $R_2 = \log 6$ the $\mathcal{Y}^*$-outputs are known exactly at the decoder.

Suppose now that (f_n, g_n, F_n) is a code for the AVCS without side in-

formation with $e(f_n, g_n, F_n) \leqslant \lambda < \frac{1}{2}$ and that $\|f_n\| < 3^n$. Then there exists a pair (x^n, x'^n) with $f_n(x^n) = f_n(x'^n)$. We now derive a contradiction as before. Endowing $\mathcal{X}^n$ with a vector space structure $GF(3)^n$ one sees that there are vectors $\epsilon^n = (\epsilon_1, \ldots, \epsilon_n)$ and $\epsilon'^n = (\epsilon'_1, \ldots, \epsilon'_n)$, $\epsilon_t, \epsilon'_t \in \{0, 1\}$ for $1 \leqslant t \leqslant n$, with

$$x'^n + \epsilon'^n = x^n + \epsilon^n. \tag{3.31}$$

The n-th Kronecker product of *matrix* M has a 1 exactly in the positions (x^n, y^n) with $y^n = x^n + \epsilon^{*n}$, $\epsilon^*_t \in \{0, 1\}$ for $1 \leqslant t \leqslant n$ and has therefore a 1 in both positions $(x^n, x^n + \epsilon^n)$, $(x'^n, x'^n + \epsilon'^n)$.

This implies that the nth Kronecker product of M^* has a $(\frac{1}{2})^n$ in all positions $(x^n, ((x^n + \epsilon^n), \alpha^n))$, $\alpha^n \in \{1, 2\}^n$, and all positions $(x'^n, ((x'^n + \epsilon'^n), \beta^n))$, $\beta^n \in \{1, 2\}^n$.

Since by (3.31)

$$\{((x^n + \epsilon^n), \alpha^n) : \alpha^n \in \{1, 2\}^n\} = \{((x'^n + \epsilon'^n), \beta^n) : \beta^n \in \{1, 2\}^n\}$$

and since $f_n(x^n) = f_n(x'^n)$, for a suitable s^n the error probability is $\geqslant \frac{1}{2}$.

Q.E.D.

§ 6. Some Remarks

1. Consider the AVCS with $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$, $\mathcal{S} = \{0, 1\}$ and

$$p(x, 0 \mid 0) = p(0, y \mid 1) = \tfrac{1}{2} \quad \text{for all } x, y \in \{1, 2\}.$$

Suppose the decoder knows the states s^n. By Theorem 3 or also directly one sees that the region equals $\{(R_1, R_2) : R_i \geqslant \log_2 2 \text{ for } i = 1, 2\}$. Now the channel is such that both encoders know the states s^n also. Suppose $s^n = 000\ldots0\ 111\ldots1$, then it suffices that the $\mathcal{X}$-encoder (resp. $\mathcal{Y}$-encoder) encodes his received sequence in the 0-block (resp. 1-block). If s^n has t 0's as components, for this the rate pair $(R_1(s^n), R_2(s^n)) = \left(\frac{t}{n}, \frac{n-t}{n}\right)$ suffices. Notice that $R_1(s^n) + R_2(s^n) = \log_2 2$ for all $s^n \in \mathcal{S}^n$.

Now imagine that the decoder wants to forward the outcome of the source to another person, who also knows the states, then here only a total rate $\log_2 2$ is needed. One may therefore distinguish between potential and actual rate.

2. The problem to find optimal orthogonal colorings of rectangular hypergraphs has some basic meaning for science. If one takes different projections of an object in an optical sense or "projections" of measurements, then one is interested in the question of reconstructing the original picture or measurement with a minimal amount of information about the "projections".

4. The Rate Region for Arbitrarily Varying Sources (AVS)

§ 1. The Rate Region

The coding problem for the AVS under consideration has been defined in (b) of the introduction. The region of achievable pairs of rates was denoted by $\mathcal{R}_{\mathcal{X}\mathcal{S}}$. We adopt the following notation: S is a RV with values in $\mathcal{S}$ and X is a RV with Prob $(X = x \mid S = s) = p(x \mid s)$, Prob $(X = x) \sum_{s \in \mathcal{S}} p(x \mid s)$ Prob $(S = s)$.

Theorem 5. *The rate region of the AVS can be characterized as follows:* $\mathcal{R}_{\mathcal{X}\mathcal{S}} = \{(R_1, R_2) : R_1, R_2$ *satisfy* (a), (b), (c)$\}$, *here*

(a) $R_1 \geqslant \max_S H(X \mid S)$ (b) $R_2 \geqslant \max_S H(S \mid X)$

(c) $R_1 + R_2 \geqslant \max_S H(X, S)$

and the maxima range over all RV's S with values in $\mathcal{S}$.

We know from Theorem 1† that for $R_2 = \log |\mathcal{S}|$, that is the case in which the decoder knows the $\mathcal{S}$-outputs, the optimal value for R_1 is $\max_s H(X(s)) = \max_S H(X \mid S)$. In the other extremal case, in which the decoder knows the $\mathcal{X}$-outputs, the optimal value for R_2 is $\max_S H(S \mid X)$.

We would like to thank I. Csiszár for pointing out to us that this case can be solved by the Graph Decomposition Theorem (GDT) of Lovász [36]. This is simply due to the fact that our decomposition into maximal error codes (CDT) ([14], included in [17], [39], for an abstract graph theoretic version, see [44]) follows from the GDT. In Section 9 it is shown that the GDT and the CDT are essentially equivalent.

It should be noticed that both rate pairs $(\max_S H(X \mid S), \log |\mathcal{S}|)$ and $(\max_S H(X), \max_S H(S \mid X))$ lie in general not on the boundary of $\mathcal{R}_{\mathcal{X}\mathcal{S}}$.

§ 2. A "mixed source-channel" coding problem

From the point of view of classical source (resp. channel) models with completely specified joint distributions (resp. conditional distributions)

---

†J. Wolfowitz kindly pointed out an error in the calculations of the original proof for Theorem 1 in Part I. This can be corrected by replacing (4.6)–(4.8) by (4.6′)–(4.8′):

$$S_\delta(n) = \{s^n : s^n \in \mathcal{S}^n \text{ with } |\mathcal{G}_\delta(X^n \mid s^n)| \leqslant \exp\{4c\sqrt{n}\}\} \tag{4.6′}$$

$$h(\lambda(n)) + \lambda(n) \log (|\mathcal{G}_\delta(X^n \mid s^n)| L(n)^{-1}) \leqslant -2c\sqrt{n} \exp\{-3c\sqrt{n}\} \quad \text{for } n \geqslant n_0 \tag{4.7′}$$

$$|\mathcal{S}^n - \mathcal{S}_\delta(n)| \exp\{-2c\sqrt{n} e^{-3c\sqrt{n}} e^{4c\sqrt{n}}\} < 1 \quad \text{for } n \geqslant n_0. \tag{4.8′}$$

In order to color $(\mathcal{X}^n, (\mathcal{G}_\delta(X^n \mid s^n))_{s^n \in \mathcal{S}_\delta(n)})$ strictly now $(n|\mathcal{X}|)^{0(\sqrt{n})}$ colors are needed.

Thus $\|f_n\| \leqslant \exp\{H^* n + 0(\log n \cdot \sqrt{n})\}$.

the AVS is a mixture of a source and a channel coding problem. Even though we try to avoid channel coding in source coding (see Introduction to Part I) for problems of the present type this may not be possible in a simple way. Our proof of Theorem 5 uses the CDT in conjunction with colorings of 2-hypergraphs. The CDT was originally used to prove the Slepian/Wolf source coding theorem ([15]). For that theorem an average goodness is enough and Cover's nice proof ([16]), compare also Coloring Lemma 4 in Part I, Section 2, § 2 for an abstract version) is much simpler. However, for the present problem the CDT finds a natural use. We include here a new proof of it, because it is so simple and contradicts a common belief. It is often believed that random coding is linked with the average error concept and that maximal coding is linked with the maximal error concept. In [43] we showed how to do random coding with maximal errors and here we show how to do maximal coding with average errors.

§ 3. An abstract maximal code method for "average errors"

Let $(\mathcal{X}, \mathcal{Y}, \mathcal{F})$ be a bipartite graph and let

$$G(x) = \{y \in \mathcal{Y} : (x, y) \in \mathcal{F}\} \neq \emptyset.$$

An abstract (M, σ)-code is a family $\{u_1, \ldots, u_M\} \subset \mathcal{X}$ with

$$\frac{1}{M} \sum_{l=1}^{M} \sum_{j \neq l} |G(u_l) \cap G(u_j)| \, |G(u_l)|^{-1} < \sigma. \tag{4.1}$$

In [43] $\sum_{j \neq l} |G(u_l) \cap G(u_j)| \, |G(u_l)|^{-1}$ was used as measure of performance for code word u_l.

Maximal Code Lemma. *For every $\sigma > 0$ there exists an (M, σ)-code for the bipartite graph $(\mathcal{X}, \mathcal{Y}, \mathcal{F})$ with*

$$M \geqslant \sigma |\mathcal{X}| \left(\frac{1}{|\mathcal{X}|} \sum_{x, x' \in \mathcal{X}} |G(x) \cap G(x')| \, |G(x)|^{-1} \right)^{-1}. \tag{4.2}$$

Proof. If N is the maximal integer for which an $(N-1, \sigma)$-code exists, then for all vectors $(x_1, \ldots, x_N) \in \mathcal{X}^N$:

$$\sum_{l=1}^{N} \sum_{j \neq l} |G(x_l) \cap G(x_j)| \, |G(x_l)|^{-1} \geqslant \sigma N$$

and therefore

$$\sum_{(x_1, \ldots, x_N) \in \mathcal{X}^N} \sum_{l=1}^{N} \sum_{j \neq l} |G(x_l) \cap G(x_j)| \, |G(x_l)|^{-1} \geqslant \sigma N |\mathcal{X}|^N$$

Since for every $x, x' \in \mathcal{X}$ the term $|G(x) \cap G(x')| \, |G(x)|^{-1}$ occurs $N(N-1)|\mathcal{X}|^{N-2}$ times in the sum, we conclude

$$N(N-1)|\mathcal{X}|^{N-2} \sum_{x, x' \in \mathcal{X}} |G(x) \cap G(x')| \, |G(x)|^{-1} \geqslant \sigma N |\mathcal{X}|^N, \tag{4.3}$$

which implies (4.2). Q.E.D.

In order to analyze the quantity

$$f(x, x') = \frac{1}{|\mathcal{X}|} \sum_{x, x' \in \mathcal{X}} |G(x) \cap G(x')| \, |G(x)|^{-1}$$

we define the vertex-vertex degrees

$$\deg(y) = |\{x : (x, y) \in \mathcal{F}\}|, \quad \deg(x) = |G(x)|,$$

and

$$\mathrm{Deg}(x) = \sum_{y \in G(x)} \deg(y). \tag{4·4}$$

Observe that

$$\mathrm{Deg}(x) = \sum_{x' \in \mathcal{X}} |G(x) \cap G(x')|, \tag{4.5}$$

and that for

$$g = \min_{x \in \mathcal{X}} |G(x)| \quad \text{and} \quad D = \max_{x \in \mathcal{X}} \mathrm{Deg}(x)$$

$$\frac{1}{|\mathcal{X}|} \sum_{x, x' \in \mathcal{X}} |G(x) \cap G(x')| \, |G(x)|^{-1} \leqslant Dg^{-1}. \tag{4.6}$$

Thus we have the

COROLLARY. *For every $\sigma > 0$ there exists an (M, σ)-code for $(\mathcal{X}, \mathcal{Y}, \mathcal{F})$ with*

$$M \geqslant \sigma \, |\mathcal{X}| \, gD^{-1}. \tag{4.7}$$

Remark 1. For comparison let us repeat Feinstein's [4] argument. Fix a positive $\lambda < \frac{1}{2}$ and consider a system $\{(u_i, A_i) : 1 \leqslant i \leqslant M\}$ with

(a) $u_i \in \mathcal{X}$, $A_i \subset \mathcal{Y}$ for $1 \leqslant i \leqslant M$; $A_i \cap A_j = \emptyset$ for $i \neq j$.

(b) $A_i \subset G(u_i)$, $1 \leqslant i \leqslant M$.

(c) $|A_i| \geqslant (1 - \lambda) \, |G(u_i)|$, $1 \leqslant i \leqslant M$.

(d) M is maximal for a system with properties (a), (b), (c).

Set $A = \bigcup_{i=1}^{M} A_i$. Then for all $x \in \mathcal{X}$, $|G(x) \cap A| \geqslant \lambda \, |G(x)|$ and therefore $\sum_{x \in \mathcal{X}} |G(x) \cap A| = \sum_{y \in A} |G(y)| \geqslant \lambda \sum_{x \in \mathcal{X}} |G(x)|$. This implies

$$|A| \geqslant (\max_{y \in \mathcal{Y}} |G(y)|)^{-1} \lambda \sum_{x \in \mathcal{X}} |G(x)|$$

and from (b) we obtain $|A| \leqslant M \max_{x \in \mathcal{X}} |G(x)|$.

Therefore

$$M \geqslant (\max_{(x, y) \in \mathcal{X} \times \mathcal{Y}} |G(x)| \, |G(y)|)^{-1} \cdot \lambda \sum_{x \in \mathcal{X}} |G(x)|.$$

Remark 2. (Application to the *channel graph*). For a DMC with input (resp. output) variable X (resp. Y) the triple $(\mathcal{T}_\delta(X^n), \mathcal{T}_\delta(Y^n), \bigcup_{x^n \in \mathcal{T}_\delta(X^n)} x^n \times \mathcal{G}_\delta(Y^n \mid x^n)$ forms a bipartite graph. Using the decoding rule

$B_l = G(u_l) - \bigcup_{j \neq 1} G(u_j)$ we derive from (4.1) $\frac{1}{M} \sum_{i=1}^{M} |B_i|\,|G(u_i)|^{-1} \geqslant 1 - \sigma$.

This and the properties of the generated sets $\mathcal{G}_\delta(Y^n \mid x^n)$ as stated in Lemma G1, Part I, guarantee that $\{(u_i, B_i) : 1 \leqslant i \leqslant M\}$ is a code for the DMC with an average error probability $\bar{\lambda} \leqslant \sigma + 0\left(\frac{1}{\delta^2}\right)$.

Moreover, in this case

$$D = \exp\{(H(Y \mid X) + H(X \mid Y))n + 0(\sqrt{n})\}$$
$$g = \exp\{H(Y \mid X)n + 0(\sqrt{n})\}$$
$$|\mathcal{X}| = \exp\{H(X)n + 0(\sqrt{n})\}$$

and therefore by the Corollary, $M \geqslant \sigma \exp\{I(X \wedge Y)n + 0(\sqrt{n})\}$. This is the coding theorem.

§ 4. Decomposition of $\mathcal{X}$ into Codes (CDT)

For $(\mathcal{X}, \mathcal{Y}, \mathcal{F})$ and $A \subset \mathcal{X}$ define $D_A = \max_{x \in A} \mathrm{Deg}\,(x)$ and $g_A = \min_{x \in A} |G(x)|$. Of course

$$D_A \leqslant D \quad \text{and} \quad g_A \geqslant g. \tag{4.8}$$

Denote the restriction of $(\mathcal{X}, \mathcal{Y}, \mathcal{F})$ to A by $(A, \mathcal{Y}, \mathcal{F}_A)$. By the Corollary and (4.8) there exists an (M_A, σ)-code for this bipartite graph with

$$M_A \geqslant \sigma |A|\, g_A\, D_A^{-1} \geqslant \sigma\, |A|\, g\, D^{-1}. \tag{4.9}$$

Now we describe the iterative construction of codes. Abbreviate $\sigma g\, D^{-1}$ by α and set $\mathcal{X}_1 = \mathcal{X}$.

By the Corollary there exsits an (M_1, σ)-code $\mathcal{U}_1$ with $\alpha\,|\mathcal{X}_1| \geqslant M_1 \geqslant \alpha\,|\mathcal{X}| - 1$. (Throw unnecessary code words out). Then $|\mathcal{X}_2| = |\mathcal{X}_1 - \mathcal{U}_1| \geqslant (1 - \alpha)\,|\mathcal{X}|$ and there is an (M_2, σ)-code $\mathcal{U}_2$:

$$\alpha(1 - \alpha)\,|\mathcal{X}| \geqslant M_2 \geqslant \alpha(1 - \alpha)\,|\mathcal{X}| - 1.$$

We show inductively that the construction can be repeated such that for all $t \geqslant 1$

$$\alpha\,|\mathcal{X}|\,(1 - \alpha)^{t-1} \geqslant M_t \geqslant \alpha\,|\mathcal{X}|\,(1 - \alpha)^{t-1} - 1. \tag{4.10}$$

Now $|\mathcal{X} - \bigcup_{s=1}^{t} \mathcal{U}_s| \geqslant |\mathcal{X}| - \sum_{s=0}^{t-1} \alpha\,|\mathcal{X}|\,(1 - \alpha)^s = |\mathcal{X}|\left(1 - \sum_{s=0}^{t-1} \alpha(1 - \alpha)^s\right)$ $= |\mathcal{X}|\,(1 - \alpha)^t$ and therefore we can find an (M_{t+1}, σ)-code $\mathcal{U}_{t+1}$ with $\alpha\,|\mathcal{X}|\,(1 - \alpha)^t \geqslant M_{t+1} \geqslant \alpha\,|\mathcal{X}|\,(1 - \alpha)^t - 1$.

If now for a T,

$$\alpha\,|\mathcal{X}| \sum_{t=1}^{T} (1 - \alpha)^{t-1} - T \geqslant |\mathcal{X}| - T - 1, \tag{4.11}$$

$$\sum_{t=1}^{T} M_t \geqslant |\mathcal{X}| - T - 1.$$

In this case the remaining elements can be trivially partitioned into $T+1$ codes and $\mathcal{X}$ is then partitioned into at most $2T+1$ codes. Equivalent with (4.11) are $1-(1-\alpha)^T \geqslant 1-|\mathcal{X}|^{-1}$ or $T\log(1-\alpha)^{-1} \geqslant \log|\mathcal{X}|$. Sufficient for this is

$$T \geqslant \alpha^{-1}\log|\mathcal{X}| = \sigma^{-1}Dg^{-1}\log|\mathcal{X}|.$$

From every (M, σ)-code one can extract a subcode of length $M' \geqslant \left(1-\frac{1}{\gamma}\right)M$ and worst case performance $\gamma\sigma$, $\gamma > 1$. Choosing in the iteration $\alpha = \left(1-\frac{1}{\gamma}\right)\sigma g D^{-1}$ this leads to a decomposition into

$$T' \geqslant \gamma(\gamma-1)^{-1}\sigma^{-1}Dg^{-1}$$

codes of worst case performance $\gamma\sigma$, $\gamma > 1$. Thus we have proved the *Code Decomposition Theorem (CDT).*

For the bipartite graph $(\mathcal{X}, \mathcal{Y}, \mathcal{F})$, $G(x) \neq \emptyset$ *for* $x \in \mathcal{X}$, *there exist partitions of* $\mathcal{X}$ *into*

(a) $2\sigma^{-1}Dg^{-1}\log|\mathcal{X}|+1$ *codes with (average) performance smaller than* σ.

(b) $4\sigma^{-1}Dg^{-1}\log|\mathcal{X}|+1$ *codes with (maximal) performance smaller than* 2σ.

PROBLEM. Can one achieve those results also with codes of essentially equal length? One may call this a balanced decomposition. In Section 6 we show how to get this result for the channel graph, which has a nice product space structure, by a simple argument using randomly chosen permutations from Π_n, the symmetric group on $\{1, \ldots, n\}$.

The random method of [43] gives this result for more general classes of graphs. The Hamming distance used in [43] is to be replaced by abstract graphic parameters.

§ 5. PROOF OF THEOREM 5

The *converse* follows immediately from the converse for the standard correlated source [15] and the fact that for any code (f_n, g_n, F_n) with error probability λ definition (1.2) implies that

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

for all RV's S^n with values in $\mathcal{S}^n$ and in particular for all $S^n = (S_1, \ldots, S_n)$, where the S_i's are independent and identically distributed.

For the direct part we make use of the CDT in conjunction with Coloring Lemma 3C. In the proof of Theorem 2 we used the Covering Lemma of Section 2, Part I, in order to get a suitable partition of $\mathcal{S}^n$, and then we defined a 2-hypergraph based on this partition, which could be properly colored with the help of Coloring Lemma 3C. Now we get a

suitable partition of $\mathcal{S}^n$ from the CDT and then we define a 2-hypergraph based on this partition and color it. Again there are the two slightly different proofs: one using the almost uniformity of PD's and the other based on counting alone. We choose the first one.

STEP 1 *Partitioning of* $\mathcal{S}^n$

Fix a $p \in \mathcal{P}_0(n, \mathcal{S})$ and a RV S_p with distribution $p \cdot X_p$ is a RV with Prob $(X_p = x) = \sum_{s \in \mathcal{S}} p(s)p(x \mid s)$. From the CDT and Remark 2 we know that for every λ, $0 < \lambda < 1$, there exists a partition $\mathcal{C}(p, L_p) = \{C_p^1, \ldots, C_p^{L_p}\}$ of $\mathcal{T}_0(S_p^n)$ with

$$L_p \leqslant \exp\{H(S_p \mid X_p)n + 0(\sqrt{n})\}, \tag{4.12}$$

$$|C_p^i| \leqslant \exp\{I(X_p \wedge S_p)n + 0(\sqrt{n})\}, \ 1 \leqslant i \leqslant L_p, \tag{4.13}$$

and such that C_p^i is the set of code words of a code for the DMC with transmission matrix $(p(x \mid s))_{x \in \mathcal{X}; s \in \mathcal{S}}$. This code has maximal error probability less than λ. The decoding sets can be chosen as follows:

$$\mathcal{G}(s^n) = \mathcal{G}_\delta(X^n \mid s^n) - \cup\{\mathcal{G}_\delta(X^n \mid s'^n) : s'^n \in C_p^i, s'^n \neq s^n\} \text{ for all } s^n \in C_p^i. \tag{4.14}$$

For any integer $l_p = \exp\{r_p n\}$ with $0 \leqslant r_p \leqslant I(X_p \wedge S_p)$ we can find a refinement of $\mathcal{C}(p, L_p)$:

$$\mathcal{C}(p, L_p, l_p) = \{C_p^{ij} : 1 \leqslant i \leqslant L_p, \ 1 \leqslant j \leqslant l_p\} \tag{4.15}$$

with the properties

$$C_p^{ij} \subset C_p^i \text{ for } 1 \leqslant j \leqslant l_p, \ 1 \leqslant i \leqslant L_p \tag{4.16}$$

and

$$|C_p^{ij}| \leqslant \exp\{(I(X_p \wedge S_p) - r_p)n + 0(\sqrt{n})\} \text{ for } 1 \leqslant j \leqslant l_p, \ 1 \leqslant i \leqslant L_p. \tag{4.17}$$

Clearly,

$$\mathcal{C} = \{C_p^{ij} : 1 \leqslant i \leqslant L_p, \ 1 \leqslant j \leqslant l_p, \ p \in \mathcal{P}_0(n, \mathcal{S})\} \tag{4.18}$$

is a partition of $\mathcal{S}^n$.

STEP 2. *Definition of a 2-hypergraph* $\mathcal{H}_2$

Choose $\mathcal{H}_2 = (\mathcal{V}, \mathcal{A} \cup \mathcal{B}, (\mathcal{E}_E)_{E \in \mathcal{A} \cup \mathcal{B}})$, $\mathcal{A} \cap \mathcal{B} = \emptyset$, as follows:

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

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

$$\mathcal{P}_a(n, \mathcal{S}) = \{p \in \mathcal{P}_0(n, \mathcal{S}) \text{ with } H(X_p \mid S_p)n \leqslant 2c\sqrt{n}\}. \tag{4.19}$$

Define now

$$E_p^{ij} = \cup\{\mathcal{G}(s^n) : s^n \in C_p^{ij}\}, \tag{4.20}$$

$$\mathcal{A} = \{E_p^{ij} : p \in \mathcal{P}_a(n, \mathcal{S}), 1 \leqslant i \leqslant L_p, 1 \leqslant j \leqslant l_p\}, \tag{4.21}$$

$$\mathcal{B} = \{E_p^{ij} : p \in \mathcal{P}_b(n, \mathcal{S}), 1 \leqslant i \leqslant L_p, 1 \leqslant j \leqslant l_p\}. \tag{4.22}$$

Finally define for $E_p^{ij} \in \mathcal{A} \cup \mathcal{B}$ the set of subedges

$$\mathcal{C}_{E_p^{ij}} = \{\mathcal{Q}(s^n) : s^n \in C_p^{ij}\}. \tag{4.23}$$

STEP 3 *Choice of the r_p's*

Suppose that (R_1, R_2) is given and that

$$R_1 \geqslant \max_{\mathcal{S}} H(X \mid \mathcal{S}),\ R_2 \geqslant \max_{\mathcal{S}} H(\mathcal{S} \mid X),\ R_1 + R_2 \geqslant \max_{\mathcal{S}} H(X, \mathcal{S}).$$

Define for $p \in \mathcal{P}_0(n, \mathcal{S})$

$$r_p = \min\,(I(S_p \wedge X_p),\ R_2 - H(S_p \mid X_p)), \tag{4.24}$$

$$R_{2p} = H(S_p \mid X_p) + r_p \leqslant R_2, \tag{4.25}$$

and

$$R_{1p} = H(X_p \mid S_p) + (I(S_p \wedge X_p) - r_p). \tag{4.26}$$

Now also $R_{1p} \leqslant R_1$ holds. In case $r_p = I(S_p \wedge X_p)$ this is obvious, because $R_1 \geqslant H(X_p \mid S_p)$. In case $r_p = R_2 - H(S_p \mid X_p)$ we have $R_1 \geqslant H(X_p, S_p) - R_2 = H(X_p, S_p) - r_p - H(S_p \mid X_p) = H(X_p \mid S_p) + (I(S_p \wedge X_p) - r_p) = R_{1p}$. Since the $\mathcal{S}$-encoder can encode the type p at a negligible rate, it suffices to show that for *fixed* p the rates R_{1p}, R_{2p} as defined in (4.24), (4.25) are achievable for $0 \leqslant r_p \leqslant I(S_p \wedge X_p)$.

STEP 4. *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 $(\mathcal{V}, \mathcal{A}^*)$, the graph assigned to the hypergraph $(\mathcal{V}, \mathcal{A})$ (see page 84 in Part I), and a suitable *lower* bound on the cardinality of subedges in $(\mathcal{V}, \mathcal{B}, (\mathcal{C}_E)_{E \in \mathcal{B}})$. It follows from definition (4.20), (4.17) and Lemma G1 (c), page 94 of Part I, that

$$|E_p^{ij}| \leqslant \exp\{(H(X_p \mid S_p) + I(X_p \wedge S_p) - r_p)n + 0(\sqrt{n})\} \quad \text{for } 1 \leqslant i \leqslant L_p;\ 1 \leqslant j \leqslant l_p. \tag{4.27}$$

We now derive an upper bound on D. Recall that an AVS is defined by a *set* $\{p(\cdot \mid s) : s \in \mathcal{S}\}$, which means that

$$p(\cdot \mid s) \neq p(\cdot \mid s') \quad \text{for } s \neq s'. \tag{4.28}$$

Now for $p \in \mathcal{P}_a(n, \mathcal{S})$ an $s^n \in \mathcal{T}_0(S_p^n)$ has fewer than $H_*^{-1}2c\sqrt{n}$ components t with $H(X(s^t)) > 0$, if $H_* = \min_{s \in \mathcal{S}} \{H(X(s)) : H(X(s)) > 0\}$. Here we used $\sum_{i=1}^{n} H(X(s_i)) = nH(X_p \mid S_p)$. Keeping (4.28) in mind we see that any $x^n \in \mathcal{X}^n$ can be contained in at most $T = \binom{n}{H_*^{-1}2c\sqrt{n}} |\mathcal{S}|^{H_*^{-1}2c\sqrt{n}}$ sets $\mathcal{Q}(s^n)$, $s^n \in C_p^i$, $1 \leqslant i \leqslant L_p$, $p \in \mathcal{P}_a(n, \mathcal{S})$.

This implies that x^n is contained in at most T edges in $\mathcal{A}$. Using

(4.27) we get

$$D \leqslant \exp\{(H(X_p \mid S_p) + I(X_p \wedge S_p) - r_p)n + 0(c\sqrt{n}\log n)\}. \quad (4.29)$$

Finally, it follows from the definition of $\mathcal{G}(s^n)$, $s^n \in C_p^i$, Lemma G1 on page 94, Part I, and the definition of $\mathcal{P}_b(n, \mathcal{S})$ that for $c = \log^2 n$ and n sufficiently large

$$|\mathcal{G}(s^n)| \geqslant \exp\{\sqrt{n}\log^2 n\} \text{ for } s^n \in C_p^i,\ 1 \leqslant i \leqslant L_p,\ p \in \mathcal{P}_b(n, \mathcal{S}). \quad (4.30)$$

STEP 5. *Application of Coloring Lemma 3C.*

Choose

$$\lambda_n = \exp\{-\sqrt{n}\log n\},\ L(n) = \exp\{R_1 n + \sqrt{n}\log^4 n\},$$
$$d(n) = L(n) - D - 1,$$

and apply Lemma 3C (pages 89, 90 of Part I). (2.20) there is equivalent with

$$|E_p^{ij}| \leqslant \lambda_n d(n) \quad \text{for all } i, j, p. \quad (4.31)$$

Since $D \leqslant \exp\{R_{1p}n + 0(\sqrt{n}\log^3 n)\}$ and since $R_1 \geqslant R_{1p}$, we have for n large

$$d(n) \geqslant \exp\{R_1 n + \tfrac{1}{2}\sqrt{n}\log^4 n\}. \quad (4.32)$$

This, the choice of λ_n and (4.27) imply (4.31). It remains for (2.21) to be verified. Using (4.30), (4.27) and (4.32) we see that for n large the left hand expression in (2.21) is smaller than

$$(n+1)^{|\mathcal{S}|}\,|\mathcal{S}|^{2n} \exp\{e^{\sqrt{n}\log^2 n}(h(\lambda_n) - \lambda_n \tfrac{1}{4}\sqrt{n}\log^4 n)\} < 1.$$

By Lemma 3C there exists an $L(n)$-Coloring $\Phi^2_{2\lambda_n}$ of $\mathcal{H}_2$, which is strict on $(\mathcal{CV}, \mathcal{A})$.

STEP 6. *The code and its error probability.* Define the encoding functions $f_n(x^n) = \Phi^2_{2\lambda_n}(x^n)$ for $x^n \in \mathcal{X}^n$, $g_n(s^n) = (i, j, p)$ if $s^n \in C_p^{ij}$, and the decoding function F_n by

$$F_n(l, (i, j, p)) = \begin{cases} (x^n, s^n) \text{ if (a) } E_p^{ij} \cap f_n^{-1}(l) = \{x^n\} \\ \qquad \text{and (b) } x^n \in \mathcal{G}(s^n) \text{ for } s^n \in C_p^{ij} \\ \text{any decision otherwise.} \end{cases}$$

Thus clearly,

$$\|f_n\| \leqslant \exp\{R_1 n + \sqrt{n}\log^4 n\},$$

$$\|g_n\| \leqslant (n+1)^{|\mathcal{S}|} \exp\Big(\max_{p \in \mathcal{P}_0(n, \mathcal{S})} R_{2p} \cdot n + 0(\sqrt{n})\Big\} \leqslant \exp\{R_2 n + 0(\sqrt{n})\}.$$

The decoding error probability is readily estimated:

By construction of the C_p^i's Prob $(X(s^n) \in \mathcal{G}(s^n)) \geqslant 1 - \lambda$ for *all* $s^n \in \mathcal{S}^n$. Our coloring is such that every subedge $\mathcal{G}(s^n)$ is properly colored within its edge in at least $(1 - 2\lambda_n)\,|\mathcal{G}(s^n)|$ of its elements. Denoting this set by

$\mathcal{K}(s^n) \subset \mathcal{G}(s^n)$ we conclude from Lemma G1 that

$$\text{Prob}\,(X(s^n) \in \mathcal{K}(s^n)) \geqslant 1 - \lambda - 2\lambda_n \exp\{0(\sqrt{n})\} \geqslant 1 - 2\lambda \quad \text{for } n \text{ large.}$$

Q.E.D.

Actually by bothering about several constant factors of $\sqrt{n}$-terms one can also achieve $\|f_n\| \leqslant \exp\{R_1 n + 0(\sqrt{n})\}$.

Remark. The assumption $p(\cdot \mid s) \neq p(\cdot \mid s')$ for $s \neq s'$ in the definition of an AVS can be dropped. The proof of Theorem 1 does not use this condition. We used it in the present proof in order to get the bound on D and also in the proof of Theorem 2. There one can replace every s by its equivalence class without further modifications. In the present proof we can make the code constructions for the equivalence classes and encode the elements within an equivalence class separately. The formula for the rate region remains valid.

5. A General Robustification Technique

§ 1. Introductory Remarks

In this Section we assume that the reader is familiar with [13]. There we introduced a general method, called *elimination technique*, to obtain from a correlated code for arbitrarily varying channels (AVC) with error probability $\lambda_n' \leqslant e^{-\epsilon n}$, $\epsilon > 0$, an ordinary code of essentially the same rate and average error probability $\lambda_n = o(1)$. (Actually, by a more careful analysis one can achieve $\lambda_n \leqslant e^{-n^\delta}$, $\delta < 1$). Since for correlated codes the capacity was known ([8], called Random Code Theorem in [13]), we thus obtained the ordinary capacity. More precisely, we obtained this result, if the ordinary capacity is known to be positive.

Jahn ([26], [49]) has applied this technique to several multi-user source and channel coding problems. The approach is always as follows:

(a) one establishes a Random Code Theorem for correlated codes,

(b) one eliminates the correlation.

Here we are concerned with step (a). The method described below makes it possible to derive from a Capacity Theorem for compound channels ([7], [6]) or sources a Capacity Theorem for arbitrarily varying channels or sources, which are the more robust models. We describe the method right away abstractly and recommend to look at Example 4 below for the motivation to our concepts.

§ 2. Th Robustification Technique

Let Π_n be the symmetric group (the group of permutations) acting on $\{1, 2, \ldots, n\}$. We consider functions $f : \Pi_n \times \mathcal{S}^n \to [0, 1]$, where $\mathcal{S}$ is a finite set. For $s^n = (s_1, \ldots, s_n) \in \mathcal{S}^n$ and $\pi \in \Pi_n$ we denote $(s_{\pi(1)}, \ldots, s_{\pi(n)})$ by $\pi(s^n)$. $\pi'\pi$ is the permutation obtained by multiplying π from the left

with π'. We say *f is invariant*, if

$$f(\pi'\pi, \pi'(s^n)) = f(\pi, s^n) \text{ for all } \pi, \pi' \in \Pi_n \text{ and } s^n \in \mathcal{S}^n. \tag{5.1}$$

The following sets were defined in (3.1), (3.2) of Part I:

$\mathcal{P}_0(n, \mathcal{S})$ is the set of PD's on $\mathcal{S}$ with $p(s) = \frac{n_s}{n}$, n_s integral, for $s \in \mathcal{S}$. Clearly,

$$|\mathcal{P}_0(n, \mathcal{S})| \leqslant (n+1)^{|\mathcal{S}|}. \tag{5.2}$$

$\mathcal{S}_0^n$ are the $(p, 0)$-typical sequences in $\mathcal{S}^n$. We need three elementary propositions.

PROPOSITION 1. *Let $p \in \mathcal{P}_0(n, \mathcal{S})$, $s^n \in \mathcal{S}^n$ and $A \subset \mathcal{S}_0^n(p)$ be fixed, $|A| \geqslant \alpha|\mathcal{S}_0^n(p)|$, then*

$$|\{\pi \in \Pi_n : \pi^{-1}(s^n) \in A\}| = |A| \prod_{s \in \mathcal{S}} (p_s n)! \geqslant \alpha|\Pi_n| = \alpha n!$$

Proof. Notice that for $s'^n \neq s''^n$, $\{\pi : \pi(s'^n) = s^n\} \cap \{\pi : \pi(s''^n) = s^n\} = \emptyset$ and that

$$\{\pi : \pi(s'^n) = s^n\} = \prod_{s \in \mathcal{S}} (p_s n)! \text{ for all } s'^n, s^n \in \mathcal{S}_0^n(p).$$

This proves the first equality and the last one is obvious. The inequality follows now from $|\mathcal{S}_0^n(p)| = n! \left(\prod_{s \in \mathcal{S}} (p_s n)!\right)^{-1}$ and the assumption on A.

Q.E.D.

For the PD $P = \prod_1^n{}^* p$ on $\mathcal{S}^n$ one can show with Stirling's formula that $P(\mathcal{S}_0^n(p)) \geqslant \text{const. } n^{-1/2}$. We use here a weaker inequality, because it can be verified by a short calculation.

PROPOSITION 2.

$$P(\mathcal{S}_0^n(p)) \geqslant (n+1)^{-|\mathcal{S}|} \text{ for all } p \in \mathcal{P}_0(n, \mathcal{S}).$$

Proof. Show that $P(\mathcal{S}_0^n(p)) \geqslant P(\mathcal{S}_0^n(q))$ for all $q \in \mathcal{P}_0(n, \mathcal{S})$ and use (5.2). This argument and the details can be found in [39], Chapt. 1.

PROPOSITION 3. *If for $B \subset \mathcal{S}^n$ $P(B) \geqslant 1 - \beta(n+1)^{-|\mathcal{S}|}$, $0 < \beta < 1$, then*

$$P(B \cap \mathcal{S}_0^n(p)) \geqslant (1-\beta)P(\mathcal{S}_0^n(p)).$$

Proof.

$$P(B \cap \mathcal{S}_0^n(p)) \geqslant P(\mathcal{S}_0^n(p)) - \beta(n+1)^{-|\mathcal{S}|} \geqslant (1-\beta)P(\mathcal{S}_0^n(p))$$

(by Proposition 2). Q.E.D.

We now have all tools to prove.

THEOREM 6 (Robustification technique). *If an invariant $f : \Pi_n \times \mathcal{S}^n \to [0, 1]$ satisfies*

$$\sum_{s^n \in \mathcal{S}^n} f(\pi^*, s^n)P(s^n) > 1 - \gamma, \ 0 < \gamma < 1, \text{ for a } \textit{fixed } \pi^* \in \Pi_n \tag{5.3}$$

and *all*

$$P = \prod_1^n{}^* p,\ p \in \mathcal{P}_0(n, \mathcal{S}),$$

then

$$\frac{1}{n!} \sum_{\pi \in \Pi_n} f(\pi, s^n) > 1 - 3\gamma^{1/2}(n+1)^{|\mathcal{S}|} \text{ for all } s^n \in \mathcal{S}^n. \tag{5.4}$$

Proof. For $0 < \delta < 1$ define $B = \{s^{*n} : f(\pi^*, s^{*n}) > (1-\delta)(1-\gamma)\}$. From (5.3) we obtain

$$(1 - P(B))(1-\delta)(1-\gamma) + P(B) > \delta(1-\gamma)$$

or

$$P(B)(1 - (1-\delta)(1-\gamma)) > \delta(1-\gamma).$$

Therefore $P(B) \geqslant \dfrac{\delta(1-\gamma)}{\delta+\gamma}$ and with the choice $\delta = \gamma^{1/2}$ this yields

$$P(B) \geqslant \frac{\gamma^{1/2}(1-\gamma)}{\gamma^{1/2}(1+\gamma^{1/2})} = 1 - \gamma^{1/2} = 1 - \delta.$$

With the choice $\beta = \gamma^{1/2}(n+1)^{|\mathcal{S}|}$ Proposition 3 implies $P(B \cap \mathcal{S}_0^n(p)) \geqslant (1-\beta)P(\mathcal{S}_0^n(p))$, and since the elements in $\mathcal{S}_0^n(p)$ have equal probabilities also

$$|B \cap \mathcal{S}_0^n(p)| \geqslant (1-\beta)\,|\mathcal{S}_0^n(p)|.$$

Now apply Proposition 1 with $A = B \cap \mathcal{S}_0^n(p)$ and $\alpha = 1 - \beta$ in order to estimate $\dfrac{1}{n!} \sum_{\pi \in \Pi_n} f(\pi, s^n)$ from below.

Clearly, for any $s^n \in \mathcal{S}_0(p)$

$$\begin{aligned}\frac{1}{n!} \sum_{\pi \in \Pi_n} f(\pi, s^n) &= \frac{1}{n!} \sum_{\pi \in \Pi_n} f(\pi\pi^*, s^n) \ (\Pi_n \text{ is a group}) \\ &= \frac{1}{n!} \sum_{\pi \in \Pi_n} f(\pi^*, \pi^{-1}(s^n)) \text{ (by invariance)} \\ &\geqslant \frac{1}{n!} \sum_{\pi :\, \pi^{-1}(s^n) \in A} f(\pi^*, \pi^{-1}(s^n)).\end{aligned}$$

From the definitions of B and A, $f(\pi^*, \pi^{-1}(s^n)) > (1-\delta)(1-\gamma)$ for $\pi^{-1}(s^n) \in A$, and by Proposition 1 $|\{\pi : \pi^{-1}(s^n) \in A\}| \geqslant (1-\beta)n!$ Therefore $\dfrac{1}{n!} \sum_{\pi \in \Pi_n} f(\pi, s^n) \geqslant (1-\beta)(1-\gamma)(1-\delta) \geqslant 1 - \beta - \gamma - \delta \geqslant 1 - \gamma^{1/2}((n+1)^{|\mathcal{S}|} + \gamma^{1/2} + 1)$ and (5.4) follows. Q.E.D.

§ 2. Application of the Method

We now show for a typical example how the method works.

Let $\mathcal{K} = \{w(\cdot|\cdot|s) : s \in \mathcal{S}\}$, $|\mathcal{S}| < \infty$, be the transmission matrices of

an AVC. Define the convex hull

$$\bar{\mathcal{K}} = \{\sum_s w(\cdot|\cdot|s)q(s) : q \text{ PD on } \mathcal{S}\}$$

and consider a compound channel with $\bar{\mathcal{K}}$ as class of matrices. Let $\{u_i, D_i) : 1 \leqslant i \leqslant N\}$ be an n-length block code with average error probability γ for this channel:

$$\frac{1}{N}\sum_{i=1}^{N} w(D_i \mid u_i) > 1 - \gamma \quad \text{for all } w \in \bar{\mathcal{K}}. \tag{5.5}$$

This implies

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{s^n \in \mathcal{S}^n} w(D_i|u_i|s^n)P(s^n) > 1 - \gamma \quad \text{for all } P = \prod_1^{n} {}^{*} p,\ p \text{ PD on } \mathcal{S}. \tag{5.6}$$

Define now

$$f(\pi, s^n) = \frac{1}{N}\sum_{i=1}^{N} w(\pi(D_i)\,|\pi(u_i)|\,s^n) \text{ where } \pi(D_i) = \bigcup_{y^n \in D_i} \pi(y^n), \tag{5.7}$$

and let π^* be the identity in Π_n.

The function f is *invariant*, because the channel is memoryless, and (5.6) implies (5.3). Therefore Theorem 6 implies

$$\frac{1}{n!}\sum_{\pi \in \Pi_n}\frac{1}{N}\sum_{i=1}^{n} w(\pi(D_i)\ |\pi(u_i)|s^n) > 1 - 3\gamma^{1/2}(n+1)^{|\mathcal{S}|} = 1 - \lambda, \text{ say,}$$

for *all* $s^n \in \mathcal{S}^n$. (5.8)

The system $\{(\pi(D_i), \pi(u_i)) : 1 \leqslant i \leqslant N;\ \pi \in \Pi_n, \mu\}$, where μ is the uniform distribution on Π_n, is a correlated code for the AVC with an average error probability less than λ.

For $\gamma \leqslant e^{-\epsilon n}$, $\epsilon > 0$, the elimination technique [13] can be applied. The invariance of f is the crucial property for the present method to work. All discrete memoryless multi-user sources and channels have this property. If $\{(u_i, v_j, D_{ij}) : 1 \leqslant i \leqslant N_1;\ 1 \leqslant j \leqslant N_2\}$ is a code for the MAC for instance, simply define

$$f(\pi, s^n) = \frac{1}{N_1 N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} w(\pi(D_{ij})\,|\pi(u_i), \pi(v_j)|\,s^n) \quad \text{for all } s^n \in \mathcal{S}^n.$$

In source coding π has to be applied to all encoding and decoding functions. There the elimination technique leads to randomized encoding functions and a deterministic decoding function (see [26]).

This approach does *not* give our Theorems 1–5, which are for deterministic encoding and for a more robust model (see [13], Section 8, and [6], [39]). The main purpose of the present Section is to make a general phenomenon in Information Theory understood.

In conclusion we propose some

PROBLEMS.

(1) Find techniques to eliminate randomisation in the encoding.

(2) Given $M \leqslant n!$, what is the minimal number $t(M, n)$ such that for *every* $A \subset \Pi_n$, $|A| \geqslant M$, there exist t permutations $\rho_1, \ldots, , \rho_t$ with $\{\rho_i\pi : 1 \leqslant i \leqslant t, \pi \in A\} = \Pi_n$.

(3) Given $M \leqslant |\mathcal{S}_0^n(p)|$, what is the minimal number $k(M, n, p, \epsilon)$ such that for *every* $B \subset \mathcal{S}_0^n(p)$, $|B| \geqslant M$, there exist permutations $\rho_1, \ldots \rho_k$ with $|\{\rho_i(s^n) : 1 \leqslant i \leqslant k, s^n \in B\}| \geqslant (1 - \epsilon)\,|\mathcal{S}_0^n(p)|$ $(0 \leqslant \epsilon \leqslant 1)$.

6. A NEW LINK BETWEEN CHANNEL AND SOURCE CODING

§ 1. THE LINK

As in Section 5 we consider the set of sequences $\mathcal{S}_0^n(p)$, $p \in \mathcal{P}_0(n, \mathcal{S})$, and the symmetric group Π_n.

COVERING LEMMA 2. *For any* $A \subset \mathcal{S}_0^n(p)$ *there exist permutations* $\tau_1, \ldots, \tau_k \in \Pi_n$ *with*

$$\bigcup_{i=1}^{k} \tau_i(A) = \mathcal{S}_0^n(p),$$

if

$$k > |A|^{-1}\,|\mathcal{S}_0^n(p)| \log |\mathcal{S}_0^n(p)|. \tag{6.1}$$

Proof. The hypergraph $(\mathcal{S}_0^n(p), (\pi(A))_{\pi \in \Pi_n})$ has the property that deg (s^n), the number of edges containing s^n, equals

$$|A|\, n!\, |\mathcal{S}_0^n(p)|^{-1} \quad \text{for all } s^n \in \mathcal{S}_0^n(p).$$

The result follows therefore from the Covering Lemma in Part I, Section 2, § 3. Q.E.D.

It has been realized a long time ago by Ahlswede/Körner (see [39]), and likely also by others, that the direct part of our MAC coding theorem ([9]) can bederived from the Slepian/Wolf source coding theorem.

The new link between source and channel coding is given by Covering Lemma 2. We explain in the next paragraph, that with its help one can easily *derive from the coding theorem for the DMC the CDT* (*see Section* 4), which is a stronger version of the Slepian/Wolf theorem. Thus we have the implications:

DMC coding theorem $\Rightarrow$ *DMCS coding theorem* $\Rightarrow$ *MAC coding theorem* (without converse)

§ 2. THE DECOMPOSITION INTO CODES

Consider a DMC with input alphabet $\mathcal{S}$, and let $\{(u_i, D_i) : 1 \leqslant i \leqslant M\}$ be a code such that $\mathcal{U} = \{u_1, \ldots, u_m\} \subset \mathcal{S}_0^n(p)$, $p \in \mathcal{P}_0(n, \mathcal{S})$. We know from Covering Lemma 2 that for

$$k > |\mathcal{U}|^{-1} |\mathcal{S}_0^n(p)| \log |\mathcal{S}_0^n(p)| \tag{6.2}$$

permutations $\tau_1, \ldots, \tau_k$ exist with $\bigcup_{l=1}^{k} \tau_l(\mathcal{U}) = \mathcal{S}_0^n(p)$.

From the covering $\{\tau_l(\mathcal{U}) : 1 \leqslant l \leqslant k\}$ we can pass to a partition of $\mathcal{S}_0^n(p)$ by choosing for instance

$$A_l = \tau_l(\mathcal{U}) - \bigcup_{j<l} \tau_j(\mathcal{U}),\ 1 \leqslant l \leqslant k.$$

The codes

$$\{(\tau_l(u_i), \tau_l(D_i)) : 1 \leqslant i \leqslant M,\ \tau_l(u_i) \in A_l\},\ 1 \leqslant l \leqslant k$$

form the desired decomposition into codes.

If $|\mathcal{U}| = \exp\{I(S \wedge X)n + 0(\sqrt{n})\}$, S has distribution p, then we see from (6.2) that $k = \exp\{H(S \mid X)n + 0(\sqrt{n})\}$ codes suffice for the decomposition. Since $p \in \mathcal{P}_0(n, S)$ was arbitrary, we immediately get the CDT for the DMC and therefore also the Slepian/Wolf Theorem

It can't be emphasized enough that we now have a quite general tool to pass from Multi-user channel coding theorems to Multi-user source coding theorems.

§ 3. Balanced Coverings and Partitions

The iterative construction in the proof of the CDT in Section 4 is such that codes become shorter and shorter. However, most of $\mathcal{X}$ is covered by long codes. We show here that for the channel graph one can achieve that all codes are long. This result follows from a general statement about coverings of hypergraphs (Covering Lemma 3) by random partitioning.

We call a covering $\mathcal{C} = \{E_1, \ldots, E_k\}$ of a hypergraph $(\mathcal{V}, \mathcal{E})$, $\mathcal{C} \subset \mathcal{E}$, *c-balanced* if

$$|\{E \in \mathcal{C} : v \in E\}| \leqslant c \quad \text{for all } v \in \mathcal{V}. \tag{6.3}$$

Covering Lemma 3. *A hypergraph* $(\mathcal{V}, \mathcal{E})$ with

$$D = \max_{v \in \mathcal{V}} \deg(v) \geqslant \min_{v \in \mathcal{V}} \deg(v) = d > 0$$

has a c-balanced covering with k edges, if

(a) $k \geqslant |\mathcal{E}|\, d^{-1}(\log |\mathcal{V}| + 1)$

(b) $c \leqslant k \leqslant c|\mathcal{E}|\, D^{-1}$

(c) $\exp\{(h(\lambda) + \lambda \log D|\mathcal{E}|^{-1})k + \log |\mathcal{V}|\} < \frac{1}{2}$ for $\lambda = ck^{-1}$

(*In case* (a) *holds and* $c > k$ *the result is trivially true*).

Proof. Choose edges $E^{(1)}, \ldots, E^{(k)}$ independently at random according to the uniform distribution on $\mathcal{E}$. Condition (a) guarantees that with probability $> \frac{1}{2}$ this leads to a covering of $\mathcal{V}$. (See proof of the Covering Lemma on page 2 in Part I).

It suffices now to show that Prob $(\{E^{(1)}, \ldots, E^{(k)}\}$ is not c-balanced)

$< \frac{1}{2}$. Define $g_v^i = \begin{cases} 1 \text{ if } v \notin E^{(i)} \\ 0 \text{ if } v \in E^{(i)} \end{cases}$ and observe that the probability for v to be covered by more than c edges is given by

$$\text{Prob}\left(\sum_{i=1}^{k} g_v^i < k - c\right) = \text{Prob}\left(\sum_{i=1}^{k} g_v^i < k\left(1 - \frac{c}{k}\right)\right).$$

With $\lambda = ck^{-1}$ and the choice $p = D|\mathcal{E}|^{-1}$ condition (b) insures $p \leqslant \lambda$. By the very same arguments as used on page 86 in Part I we can now upper bound this probability by $\exp\{(h(\lambda) + \lambda \log D|\mathcal{E}|^{-1})k\}$. Since there are $|\mathcal{V}|$ vertices, (c) and (a) guarantee that with positive probability we obtain a c-balanced covering. Q.E.D.

COROLLARY. *The hypergraph* $(\mathcal{S}_0^n(p), (\pi(A))_{\pi \in \Pi_n}, A \subset \mathcal{S}_0^n(p), |A| \geqslant 1$ *has for* $c = 8 \log_2 |\mathcal{S}_0^n(p)| + 1)$ *and* $k = \lfloor |\mathcal{S}_0^n(p)|\, |A|^{-1} c \rfloor$ *a c-balanced covering with k edges.*

Proof. Recall that $D = d = |A|\, |\mathcal{E}|\, |\mathcal{V}|^{-1}$.

First observe that k satisfies (a) and that therefore the result follows in case $c > k$.

In case $c \leqslant k$ (b) holds by the choice of k. It remains for (c) to be verified. We have to show that

$$h(ck^{-1})k + c \log D|\mathcal{E}|^{-1} + \log |\mathcal{V}| < -1$$

or that

$$h(ck^{-1})k - c \log k + c \log (\log |\mathcal{V}| + 1) + \log |\mathcal{V}| < -1. \quad (6.4)$$

Since $h(ck^{-1})k \leqslant c \log k - c \log c + 2c$, $c \log c > c \log (\log |\mathcal{V}| + 1) + \log |\mathcal{V}| + 2c + 1$ is sufficient for (6.4) to hold. Since $\mathcal{V} = \mathcal{S}_0^n(p)$, this inequality holds for

$$c = 8(\log_2 |\mathcal{S}_0^n(p)| + 1). \qquad \text{Q.E.D.}$$

Next we show how to obtain from a c-balanced covering with edges of equal sizes a b-balanced partition:

$\mathcal{B} = \{B_1, \ldots, B_k\}$ is b-balanced, if

$$|B_i|\, |B_j|^{-1} \leqslant b \quad \text{for all } i, j. \quad (6.5)$$

Let $\mathcal{C} = \{E_1, \ldots, E_k\}$ be a c-balanced covering of $(\mathcal{V}, \mathcal{E})$ and let $\deg_{\mathcal{C}}(v) = |\{E \subset \mathcal{C} : v \in E\}|$. Define independent RV's $X_1, \ldots, X_{|\mathcal{V}|} = X^{|\mathcal{V}|}$ by

$$\text{Prob}\,(X_v = i) = (\deg_{\mathcal{C}}(v))^{-1} \quad \text{if } v \in E_i. \quad (6.6)$$

With every set $E_j \in \mathcal{C}$ associate a random set $E_j(X^{|\mathcal{V}|}) = \{v \in E_j : X_v = j\}$. $\{E_j(X^{|\mathcal{V}|}) : 1 \leqslant j \leqslant k\}$ is a partition of $\mathcal{V}$ with $E_j(X^{|\mathcal{V}|}) \subset E_j$, $1 \leqslant j \leqslant k$. If we define $f_v^j = \begin{cases} 1 \text{ if } X_v = j \\ 0 \text{ if } X_v \neq j \end{cases}$, then

$$\text{Prob}\,(|E_j(X^{|\mathcal{V}|})| < (1 - \lambda)|E_j|) = \text{Prob}\,(\sum_{v \in E_j} f_v^j < (1 - \lambda)|E_j|).$$

By the usual arguments (page 86, Part I) we see that for $(1-\lambda) \leqslant c^{-1}$ this probability is smaller than

$$\exp\{(h(\lambda) + \lambda \log(1 - c^{-1}))|E_j|\}.$$

Therefore, if

$$\sum_{j=1}^{k} \exp\{(h(\lambda) + \lambda \log(1 - c^{-1})|E_j|\} < 1, \tag{6.7}$$

then there exists a partition $\{B_1, \ldots, B_k\}$ of $\mathcal{V}$ with $B_j \subset E_j$ and $|B_j| \geqslant (1-\lambda)|E_j|$ for $1 \leqslant j \leqslant k$. In particular, for $1 \leqslant i, j \leqslant k$

$$|B_i|\,|B_j|^{-1} \leqslant (1-\lambda)^{-1}, \text{ if } |E_j| = |A| \text{ for } 1 \leqslant j \leqslant k. \tag{6.8}$$

Evaluation of (6.7) and the Corollary lead to

THEOREM 7. *The hypergraph* $(\mathcal{S}_0^n(p), (\pi(A))_{\pi \in \Pi_n})$, $A \subset \mathcal{S}_0^n(p)$, $|A| \geqslant 1$ *has for* $c = 16(\log_2|\mathcal{S}_0^n(p)| + 1)$ *and* $k = \lfloor |\mathcal{S}_0^n(p)|\,|A|^{-1}c \rfloor$

(a) *a* c*-balanced covering* $C = \{\tau_i(A) : \tau_i \in \Pi_n, 1 \leqslant i \leqslant k\}$, and

(b) *a* c^2*-balanced partition* $\{B_1, \ldots, B_{k'}\}$ with $B_i \subset \tau_i(A)$ for $1 \leqslant i \leqslant k' \leqslant k$.

Proof. (a) restates the Corollary and we now show (b).

Case 1. $1 \leqslant |A| < 2^3c$.

From the covering $C = \{E_1, \ldots, E_k\}$ pass to the partition

$$\{B_i' = E_i - \bigcup_{j<i} E_j : 1 \leqslant i \leqslant k\}.$$

List the *non-empty* sets among the B_i''s as $B_1, \ldots, B_{k'}$, $k' \leqslant k$. Thus $|B_i|\,|B_j|^{-1} \leqslant |A| < 2^3c < c^2$.

Case 2. $|A| \geqslant 2^3c$.

Choose $1 - \lambda = c^{-2} \leqslant c^{-1}$ and verify (6.7) for $|E_j| = |A|$, $k = |\mathcal{S}_0^n(p)| \times |A|^{-1}c$. It suffices to show that $(h(c^{-2}) + (1-c^{-2})\log(1-c^{-1}))|A| + \log c + \log|\mathcal{S}_0^n(p)| - \log|A| < 0$.

For this calculate $h(c^{-2}) + (1-c^{-2})\log(1-c^{-1}) = 2c^{-2}\log c - (1-c^{-2}) \times \log(1+c^{-1}) \leqslant 2c^{-2}\log c - (1-c^{-2})2^{-1}c^{-1}$.

For $c \geqslant 2^5$ this quantity is smaller than -2^{-3}, and for $|A| \geqslant 2^3c$:

$$-2^{-3}|A| + \log c + \log|\mathcal{S}_0^n(p)| - \log|A| < 0. \qquad \text{Q.E.D.}$$

7. ON THE RATE-DISTORTION FUNCTION FOR THE AVS WITH SIDE-INFORMATION AT THE DECODER

§ 1. THE RESULT OF WYNER/ZIV

We describe the result of [20], because we need all the definitions anyhow. Let $(X_t, Y_t)_{t=1}^{\infty}$ be a DMCS with alphabets $\mathcal{X}$ and $\mathcal{Y}$, $\Theta : \mathcal{X} \times \hat{\mathcal{X}}$

$\to [0, \infty)$ is the distortion function. A code (n, M, Δ) is defined by a pair of mappings (f, F), where $f: \mathcal{X}^n \to \{1, \ldots, M\}$ is an encoding function, $F: \mathcal{Y}^n \times \{1, \ldots, M\} \to \hat{\mathcal{X}}^n$ is a decoding function, and

$$\mathbf{E} \frac{1}{n} \sum_{t=1}^{n} \Theta(X_t, \hat{X}_t) \leqslant \Delta, \text{ where } \hat{X}^n = (\hat{X}_1, \ldots, \hat{X}_n) = F(Y^n, f(X^n)). \quad (7.1)$$

It is clear from these definitions that they are for the case where the decoder has complete information about the $\mathcal{Y}$-outputs, and based on this and a coded version of $\mathcal{X}$-outputs tries to reproduce the $\mathcal{X}$-outputs within a fidelity Δ.

A pair (R, θ) is said to be *achievable*, if for any $\gamma > 0$ there exists for all sufficiently large n a code (n, M, Δ) with

$$\log M \leqslant (R + \gamma)n, \ \Delta \leqslant \theta + \gamma. \quad (7.2)$$

It follows from these definitions that $\mathcal{R}$, the set of all achievable pairs (R, θ), is closed and convex, and that $\mathcal{R}$ is known, if

$$R^*(\theta) = \min \{R : (R, \theta) \in \mathcal{R}\} \quad (7.3)$$

is known.

One is interested in characterisations of the rate-distortion function R^*, which are such that the function can in principal be numerically evaluated. The characterisation found in [20] is as follows:

Let U be an auxiliary RV with values in $\mathcal{U}$, $|\mathcal{U}| = |\mathcal{X}| + 1$. Denote by $\mathcal{M}(\theta)$ the set of those U's: $U \to X \to Y$, which satisfy for some function $g: \mathcal{Y} \times \mathcal{U} \to \hat{\mathcal{X}}$

$$\mathbf{E}\Theta(X, \hat{X}) \leqslant \theta \text{ with } \hat{X} = g(Y, U). \quad (7.4)$$

Finally, define for $\theta > 0$ the function

$$\bar{R}(\theta) = \min_{U \in \mathcal{M}(\theta)} I(U \wedge X \mid Y). \quad (7.5)$$

$\bar{R}(\theta)$ is non-increasing for $\theta \in (0, \infty)$ and therefore $\bar{R}(0)$ can be defined as $\lim_{\theta \to 0} \bar{R}(\theta)$.

THEOREM WZ [20]. *For* $\theta \geqslant 0$: $R^*(\theta) = \bar{R}(\theta)$.

The result can be looked at as follows: a covering of the $\mathcal{X}$-source is described with the help of a test channel (auxiliary RV U) exactly as in Shannon's classical [2]. (In Part I this was also used in the proof of Theorem 2). *Now the $\mathcal{Y}$-source plays two roles*: it is (obviously) used in the decoding, but it also can be used to reduce the rate $I(U \wedge X)$ of the covering to $I(U \wedge X \mid Y)$. For this reduction the Markovity is needed.

§ 2. The Rate-Distortion Function for the AVS with Side-Information at the Decoder

Recall the definition of an AVS $\left\{X(s^n): s^n \in \mathcal{S}^n\right\}_{n=1}^{\infty}$ given in (1.1) of Part I. The alphabets are now $\mathcal{X}$ (as before) and $\mathcal{S}$ instead of $\mathcal{Y}$. The difference to the previous source model is that now no distribution for the $\mathcal{S}$-outputs is known. With Θ as before, a code (n, M, Δ) is now a pair of mappings (f, F), where $f: \mathcal{X}^n \to \{1, \ldots, M\}$ is an encoding function (as before), $F: \mathcal{S}^n \times \{1, \ldots, M\} \to \hat{\mathcal{X}}^n$ is a decoding function, and

$$\mathbf{E}\frac{1}{n}\sum_{t=1}^{n}\Theta(X(s_t), \hat{X}_t) \leqslant \Delta \text{ for } \textit{all } s^n \in \mathcal{S}^n, \tag{7.6}$$

where

$$\hat{X}^n = (\hat{X}_1, \ldots, \hat{X}_n) = F(s^n, f(X(s^n))).$$

The set $\mathcal{R}_\Delta$ of achievable pairs (R, θ) is again closed and convex, and is characterized through its boundary

$$R_\Delta^*(\theta) = \min\{R: (\mathcal{R}, \theta) \in \mathcal{R}_\Delta\}. \tag{7.7}$$

For the formulation of our results we adopt the following notation:

$\mathcal{P}(\mathcal{S})$ is the set of all PD's on S, S_p is a RV with values in $\mathcal{S}$ and distribution p. X_{q_p} takes values in $\mathcal{X}$ and has the distribution q_p, where

$$q_p(X) = \sum_{s \in \mathcal{S}} p(x \mid s)p(s). \tag{7.8}$$

Denote by $\mathcal{M}_p(\theta)$ the set of those RV's $U: U \to X_{q_p} \to S_p$, which satisfy for some function $g_p: \mathcal{S} \times \mathcal{U} \to \hat{\mathcal{X}}$

$$\mathrm{E}\Theta(X_{q_p}, \hat{X}) \leqslant \theta, \text{ where } \hat{X} = g(S_p, U). \tag{7.9}$$

Set $\mathcal{Q}(\mathcal{X}) = \{q: q = q_p \text{ for some } p \in \mathcal{P}(\mathcal{S})\}$, $\mathcal{P}_q(\mathcal{S}) = \{p: q_p = q\}$, and define for $q \in \mathcal{Q}(\mathcal{X})$

$$\mathcal{N}_q(\theta) = \bigcap_{p \in \mathcal{P}_q(\mathcal{S})} \mathcal{M}_q(\theta). \tag{7.10}$$

Finally, define for $\theta > 0$ the functions

$$\bar{R}_q(\theta) = \min_{U \in \mathcal{N}_q(\theta)} \max_{p \in \mathcal{P}_q(\mathcal{S})} I(U \wedge X_{q_p} \mid S_p), \quad q \in \mathcal{Q}(\mathcal{X}). \tag{7.11}$$

$$\bar{R}_\Delta(\theta) = \max_{q \in \mathcal{Q}(\mathcal{X})} R_q(\theta), \quad \bar{R}_\Delta(0) = \lim_{\theta \to 0} \bar{R}_\Delta(\theta). \tag{7.12}$$

Conjecture 1. For the rate-distortion function $R_\Delta^*(\theta)$ of the AVS with side-information at the decoder

$$R_\Delta^*(\theta) = \bar{R}_\Delta(\theta), \quad \theta \geqslant 0. \tag{7.13}$$

Conjecture 2. For the conjecture $R_\Delta^*(\theta) \geqslant \bar{R}_\Delta(\theta)$, $\theta \geqslant 0$, to be true it

is *sufficient* to prove the following *Conjecture*: Let $\left(X_t, Y_{1t}, \ldots, Y_{ct}\right)_{t=1}^{\infty}$ be a discrete memoryless correlated source with $c+1$ components. Assume that there are c decoders, that the j-th decoder observes $\left(Y_{jt}\right)_{t=1}^{n}$, and that the value $f(X^n)$ of the encoding function is known to all decoders. Based on this they want to reproduce X^n with a certain fidelity θ. The distortion function $\Theta: \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty]$ shall be the same for all decoders. The rate-distortion function $R_c^*(\theta)$ is defined in the canonical way. Denote by $\mathcal{M}_c(\theta)$ the set of those U's: $U \to X \to (Y_1, \ldots, Y_c)$, which satisfy for some functions $g_j: \mathcal{Y}_j \times \mathcal{U} \to \hat{\mathcal{X}}$, $1 \leqslant j \leqslant c$, $\mathrm{E}\Theta(X, \hat{X}_j) \leqslant \theta$ with $\hat{X}_j = g_j(Y_j, U)$, $1 \leqslant j \leqslant c$. We conjecture that $R_c^*(\theta)$ equals

$$\bar{R}_c(\theta) = \min_{U \in \mathcal{M}_c(\theta)} \max_{1 \leqslant j \leqslant c} I(U \wedge X \mid Y_j).$$

The inequality $R_c^*(\theta) \leqslant \bar{R}_c(\theta)$ can easily be proved with the approach of § 3. It is perhaps surprising that no converse proof seems to exist until now for this natural generalisation of the case solved by Wyner/Ziv.

§ 3. A New Proof of $R^*(\theta) \geqslant \bar{R}(\theta)$, $\theta \geqslant 0$

The argument below is based on Covering Lemma 3 (Section 6) and Coloring Lemma 4 (Section 2, Part I). *There is no need for using codes for this type of problems.* Let $U \to X \to Y$ and $g: \mathcal{Y} \times \mathcal{U} \to \hat{\mathcal{X}}$ be such that

$$\mathrm{E}\Theta(X, \hat{X}) \leqslant \theta \quad \text{for} \quad \hat{X} = g(Y, U).$$

Since $\max_{x, \hat{x}} \Theta(x, \hat{x}) < \infty$ it suffices to show that there are functions $f_n: \mathcal{X}^n \to \{1, \ldots, M\}$, $F_n: \mathcal{Y}^n \times \{1, \ldots, M\} \to \hat{\mathcal{X}}^n$ such that

$$\sum_{(x^n, y^n) \in B_n} \frac{1}{n} \sum_{t=1}^{n} \Theta(x_t, \hat{x}_t) \leqslant \theta, \tag{7.14}$$

where $\mathrm{Prob}\,((X^n, Y^n) \notin B_n) = o(1)$, $\hat{x}^n = (\hat{x}_1, \ldots, \hat{x}_n) = F(y^n, f(x^n))$,

and
$$M \leqslant \exp\{I(U \wedge X \mid Y)n + 0(\sqrt{n})\}. \tag{7.15}$$

By the Markov property we have

$$I(U \wedge X \mid Y) = I(U \wedge X) - I(U \wedge Y) \geqslant 0 \tag{7.16}$$

and hence also $H(U \mid Y) \geqslant H(U \mid X)$.

Consider now the two hypergraphs

$$\mathcal{H}_1 = (\mathcal{V}_1, \mathcal{E}_1) = (\mathcal{T}_\delta(X^n), (\mathcal{G}_\delta(X^n \mid u^n))_{u^n \in \mathcal{T}_\delta(U^n)})$$

and

$$\mathcal{H}_2 = (\mathcal{V}_2, \mathcal{E}_2) = (\mathcal{T}_\delta(Y^n), (\mathcal{G}_\delta^*(Y^n \mid u^n))_{u^n \in \mathcal{T}_\delta(U^n)})$$

with

$$\mathcal{G}_\delta^*(Y^n \mid u^n) = \bigcup_{x^n \in \mathcal{G}_\delta(X^n \mid u^n)} \mathcal{G}_\delta(Y^n \mid x^n, u^n).$$

Since both edge sets are indexed by the same set $\mathcal{T}_\delta(U^n)$ we can choose M elements of $\mathcal{T}_\delta(U^n)$ independently according to the uniform distribution and thus have a random selection of edges, as described in the proof of Covering Lemma 3, in both hypergraphs. If M is properly chosen we get with positive probability coverings $\mathcal{C}_1, \mathcal{C}_2$ with properties stated in that Lemma. Its application is a now routine matter. By the properties of typical sequences and generated sequences stated in Section 3 of Part I:

$$|\mathcal{E}_1| = \exp\{H(U)n + 0(\sqrt{n})\} = |\mathcal{E}_2|$$
$$|\mathcal{V}_1| = \exp\{H(X)n + 0(\sqrt{n})\}, \quad |\mathcal{V}_2| = \exp\{H(Y)n + 0(\sqrt{n})\}$$
$$D_1 = \exp\{H(U \mid X)n + 0(\sqrt{n})\}, \; D_2 = \exp\{H(U \mid Y)n + 0(\sqrt{n})\}$$
$$d_1 = \exp\{H(U \mid X)n + 0(\sqrt{n})\}, \; d_2 = \exp\{H(U \mid Y)n + 0(\sqrt{n})\}.$$

With the choice

$$M \geqslant |\mathcal{E}_2|\, d_2^{-1}(\log |\mathcal{V}_2| + 1)$$

and

$$M = k = |\mathcal{E}_1|\, d_1^{-1}(\log |\mathcal{V}_1| + 1) \exp\{(0\sqrt{n})\} = \exp\{I(U \wedge X)n + 0(\sqrt{n})\}$$

(a) holds for *both* hypergraphs.

With the choice

$$c_2 = MD_2\, |\mathcal{E}_2|^{-1} = \exp\{I(U \wedge X \mid Y)n + 0(\sqrt{n})\}$$

(b) holds for $\mathcal{H}_2$.

Since

$$\lambda \log D_2\, |\mathcal{E}_2|^{-1} = c_2 k^{-1} \log D_2\, |\mathcal{E}_2|^{-1}$$
$$= -(I(U \wedge Y)n + 0(\sqrt{n})) \exp\{I(U \wedge Y)n + 0(\sqrt{n})\}$$

(c) is also satisfied for $\mathcal{H}_2$, if $I(U \wedge Y) > 0$. In case $I(U \wedge Y) = 0$ by (7.16) $I(U \wedge X \mid Y) = I(U \wedge X)$; and there is no need for using $\mathcal{H}_2$ at all (this is the Shannon case). Covering Lemma 3 implies therefore the existence of coverings $\mathcal{C}_1 = \{\mathcal{G}_\delta(X^n \mid u_i) : 1 \leqslant i \leqslant M\}$, $\mathcal{C}_2 = \{\mathcal{G}_\delta^*(Y^n \mid u_i) : 1 \leqslant i \leqslant M\}$, where $\mathcal{C}_2$ is c_2-balanced. From $\mathcal{C}_1$ we pass to a *partition* $\{A_i : 1 \leqslant i \leqslant M\}$, $A_i \subset \mathcal{G}_\delta(X^n \mid u_i)$, in any way. We can define for instance

$$A_i = \mathcal{G}_\delta(X^n \mid u_i) - \bigcup_{j<i} \mathcal{G}_\delta(X^n \mid u_j), \quad 1 \leqslant i \leqslant M. \tag{7.17}$$

If now $X^n = x^n \in A_i$ is the output of the $\mathcal{X}$-source, then by Lemma $M_2(b)$, page 97 of Part I,

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

For $y^n \in \mathcal{G}_\delta(Y^n \mid x^n, u_i) \cap \mathcal{T}_\delta(Y^n)$ define

$$\mathcal{U}(y^n) = \{u_i : 1 \leqslant i \leqslant M,\; y^n \in \mathcal{G}_\delta^*(Y^n \mid u_i)\}. \tag{7.18}$$

We know that $|\mathcal{U}(y^n)| \leqslant c_2$ *and also that* $u_i \in \mathcal{U}(y^n)$ by definition of $\mathcal{G}_\delta^*(y^n \mid u^n)$.

With a probability greater than $1 - 0\left(\frac{1}{\delta^2}\right)$ we have therefore the following situation: the decoder having observed $y^n \in \mathcal{T}_\delta(Y^n)$ knows that x^n lies in one of the sets A_i, $u_i \in \mathcal{U}(y^n)$. *If he would also know* the set A_i, $x^n \in A_i$, then (u_i, x^n, y^n) would lie in $\mathcal{T}_\delta(U^n X^n Y^n)$. In this case for $\hat{x}_t = g(y_t, u_t)$

$$\frac{1}{n}\sum_{t=1}^{n} \Theta(x_t, \hat{x}_t) \leqslant \frac{1}{n}\sum_{t=1}^{n} \Theta(x, g(y, u))p(u, x, y) + n^{-1}0(\sqrt{n})$$

$$\leqslant \theta + \gamma \text{ for any } \gamma > 0 \text{ and } n \text{ large.} \tag{7.19}$$

Now we apply Coloring Lemma 4 in order to provide this knowledge with high probability. Choose the weighted hypergraph $(\mathcal{V}, \mathcal{E}, (Q_j)_{j=1}^J, Q)$, where

$$\mathcal{V} = \{u_1, \ldots, u_N\}, \mathcal{E} = \{\mathcal{U}(y^n) : y^n \in \mathcal{T}_\delta(Y^n)\},$$

$$Q_{y^n}(u_i) = \text{Prob}\,(X^n \in A_i \mid Y^n = y^n) \text{ for } y^n \in \mathcal{T}_\delta(y^n)$$

and

$$u_i \in \mathcal{U}(y^n),\ Q = (\text{Prob}\,(Y^n = y^n))_{y^n \in \mathcal{T}_\delta(Y^n)}.$$

Then for $L \geqslant c_2\bar{\lambda}^{-1}$ there is an L-coloring Φ of this hypergraph, which is erroneous with probability less than $\bar{\lambda}$.

The encoder uses the encoding function $f(x^n) = \Phi(u_i)$ for $x^n \in A_i$, the decoder having received $y^n \in \mathcal{T}_\delta(Y^n)$ knows now u_i with probability greater than $1 - \bar{\lambda} - 0\left(\frac{1}{\delta^2}\right)$ and defines

$$\hat{x}_t = g((u_i)_t, y_t) \text{ for } t = 1, \ldots, n. \tag{7.20}$$

Q.E.D.

§ 4. Universal colorings of internally-weighted hypergraphs

We give a generalisation of Coloring Lemma 3A to internally-weighted hypergraphs $\left(\mathcal{V}, \mathcal{E}, \left(Q_j\right)_{j=1}^J\right)$: $\mathcal{V} = \{1, \ldots, I\}$, $\mathcal{E} = \{E_1, \ldots, E_J\}$ family of subsets of $\mathcal{V}$, $Q_j : E_j \to \mathbf{R}_+$, $1 \leqslant i \leqslant J$. For a coloring Φ of $\mathcal{V}$ define for $i = 1, \ldots, I$; $j = 1, \ldots, J$

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

We say that Φ has *goodness* λ^* for the internally-weighted hypergraph, if

$$\sum_{i \in E_j} g_i^j Q_j(i) \leqslant \lambda^* Q_j(E_j) \text{ for all } j = 1, \ldots, J. \tag{7.21}$$

(Compare also Coloring Lemma 4 on page 91, Part I).

Φ^{λ^*} denotes a coloring Φ of goodness λ^*. If $Q_j(i) = |E_j|^{-1}$ for all $1 \leqslant i \leqslant I$, $1 \leqslant j \leqslant J$, then we are in the situation considered in Coloring Lemma 3A. Clearly, in this case a coloring Φ^{λ^*} is the same as a coloring Φ_λ for $\lambda = \lambda^*$.

COLORING LEMMA 3A*. *Assume that the internally-weighted hypergraph* $\left(\mathcal{V}, \mathcal{E}, \left(Q_j\right)_{j=1}^{J}\right)$ *satisfies the uniformity condition*

$$Q_j(i) \leqslant bQ_j(E_j) \text{ for all } i \in E_j \text{ and } j = 1, \ldots, J. \tag{7.22}$$

Then it has for $L \geqslant \max_{1 \leqslant j \leqslant J} |E_j|$ *an L-coloring* $\Phi^{2\lambda^*}$, $0 < \lambda^* < \frac{1}{2}$, *if for some* $\alpha < 0$

$$\sum_{j=1}^{J} \exp\left\{\alpha(\lambda^* - |E_j|L^{-1})Q_j(E_j) + \frac{\alpha^2}{4} bQ_j(E_j)^2\right\} < \tfrac{1}{2}. \tag{7.23}$$

Proof. We use standard random L-coloring of $\mathcal{V}$ and define for an edge $E \in \mathcal{E}$

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

and

$$F_i(X_i, \ldots, X_J) = \begin{cases} 1 & \text{if } X_i \neq X'_i \text{ for all } i' > i,\ i' \in E \\ 0 & \text{otherwise.} \end{cases}$$

If Q is the weight on E, then

$$\sum_{i \in E} Q(i)f_i \geqslant (1 - \lambda^*)Q(E) \text{ and } \sum_{i \in E} Q(i)F_i \geqslant (1 - \lambda^*)Q(E)$$

implies that the weight of the correctly colored vertices in E is greater than $(1 - 2\lambda^*)Q(E)$. Clearly, for $\alpha < 0$

$$\text{Prob}\left(\sum_{i \in E} Q(i)f_i < (1 - \lambda^*)Q(E)\right) \leqslant \exp\{-\alpha(1 - \lambda^*)Q(E)\} \times \prod_{i \in E} \mathrm{E} \exp\{\alpha Q(i)f_i\}$$

and

$$\begin{aligned}
\prod_{i \in E} \mathrm{E} \exp\{\alpha Q(i)f_i\} &\leqslant \prod_{i \in E} \left(\frac{|E|}{L} + \frac{L - |E|}{L} e^{\alpha Q(i)}\right) \\
&\leqslant \prod_{i \in E} \left(\frac{|E|}{L} + \frac{L - |E|}{L}\left(1 + 2Q(i) + \frac{(\alpha Q(i))^2}{2}\right)\right) \\
&= \prod_{i \in E} \left(1 + \frac{L - |E|}{L}\left(\alpha Q(i) + \frac{(\alpha Q(i))^2}{2}\right)\right) \\
&= \exp \sum_{i \in E} \log\left(1 + \frac{L - |E|}{L}\left(\alpha Q(i) + \frac{(\alpha Q(i))^2}{2}\right)\right) \\
&\leqslant \exp\left\{\alpha Q(E) \frac{L - |E|}{L} + \sum_{i \in E} \frac{(\alpha Q(i))^2}{2}\right\}.
\end{aligned}$$

Since $\sum_{i\in E} Q(i)^2 \leqslant bQ(E)^2$, and since the same estimate holds for $\sum_{i\in E} Q(i)F_i$, summation over all edges gives (7.23). Q.E.D.

8. Balanced Colorings and Graph Decomposition

We need here an easy consequence of Coloring Lemma 7, which we state for the ease of reference as

Coloring Lemma 10 (Balanced colorings of hypergraph). *Let $(\mathcal{V}, \mathcal{E})$ be a hypergraph and let L be an arbitrary positive integer. If for $\alpha > 0$*

$$L \sum_{i=1}^{M} \exp\left\{-\frac{\alpha}{2} \max\left(|E_j|L^{-1}, 1\right) + |E_j|L^{-1}\right\} < 1, \tag{8.1}$$

then there exists an L-coloring Φ of $\mathcal{V}$ with

$$|E_j \cap \Phi^{-1}(l)| \leqslant \alpha \max\left(|E_j|L^{-1}, 1\right) \text{ for } 1 \leqslant j \leqslant M,\ 1 \leqslant l \leqslant L.$$

Proof. Apply Coloring Lemma 7 with $D_2 = 1$, $E = E_j$ and sum over j. Q.E.D.

Remark 1. Using in (2.9) and (8.1) the exp-function to the basis 2 and $\gamma = 1$ one easily verifies that in (8.1) $\frac{\alpha}{2}$ can be replaced by α. Moreover, then (8.1) can be replaced by

$$\text{LM } 2^{1-\alpha} < 1 \text{ for } \alpha > 1. \tag{8.3}$$

Koševlev [37] and Gallager [22] have derived a "random coding" error exponent for the Slepian/Wolf network in case the decoder is informed about the outputs of one of the sources.

Using Lovász's Graph Decomposition Theorem [36] (GDT), Csiszár/Körner [38] recently improved on those results by establishing what might be considered as the counterpart of the expurgated bound for source coding. We now explain that one could use Coloring Lemma 10 instead of the GDT, which we now state. For a detailed analysis see Section 9.

Graph Decomposition Theorem (GDT) *Let $G = (\mathcal{V}, \mathcal{E}, r)$ be a graph with a weight function r on the edges:*

$$0 \leqslant r(a, b) = r(b, a),\ r(a, a) = 0 \text{ for all } a, b \in \mathcal{V}.$$

If (a) $\sum_{b\in\mathcal{V}} r(a, b) < t$ *for all* $a \in \mathcal{V}$, *and*

(b) $\sum_{1\leqslant l\leqslant L} t_l \geqslant t,\ t_l \geqslant 0,$

then there exists an L-coloring of $\mathcal{V}$ such that

$$\sum_{b\in\Phi^{-1}(l)} r(a, b) < t_l \text{ for all } a \in \Phi^{-1}(l),\ 1 \leqslant l \leqslant L. \tag{8.4}$$

In [36] r is assumed to be integral. Using rational approximation the present form of [38] is readily seen to be equivalent.

We now formulate the key result for the derivation of the error exponents in [38], which follows from the GDT.

Let w denote an $\mathcal{X} \times \mathcal{X}$-stochastic matrix. Recall the definitions of $\mathcal{P}_0(n, \mathcal{X})$, $\mathcal{X}_0^n(p)$ and $\mathcal{X}_0^n(w(\cdot \mid x^n))$ given in Section 3, § 1, 2 of Part I. (In the notation of [38] $\mathcal{X}_0^n(p) = \mathcal{T}_p$, $\mathcal{X}_0^n(w(\cdot \mid x^n)) = \mathcal{T}_w(x^n)$).

KEY LEMMA 1 (Lemma 2 of [38]). *Given any finite set $\mathcal{X}$ and a $p \in \mathcal{P}_0(n, \mathcal{X})$. For any positive integer L there exists an L-coloring Φ of $\mathcal{X}_0^n(p)$ such that*

$$|\mathcal{X}_0^n(w(\cdot \mid x^n)) \cap \Phi^{-1}(l)| \leqslant L^{-1} |\mathcal{X}_0^n(w(\cdot \mid x^n))| \, (n+1)^{|\mathcal{X}|^2} \tag{8.5}$$

for every $x^n \in \Phi^{-1}(l)$, every $1 \leqslant l \leqslant L$, and for every $\mathcal{X} \times \mathcal{X}$-stochastic matrix unequal to the identity matrix.

As a consequence of Coloring Lemma 10 *we obtain*

KEY LEMMA 2. *Given any finite sets $\mathcal{X}$, $\mathcal{Y}$. For any positive integer $L \leqslant |\mathcal{X}|^n$ there exists an L-coloring Φ of $\mathcal{X}^n$ such that*

$$|\mathcal{X}_0^n(w(\cdot \mid y^n)) \cap \Phi^{-1}(l)| \leqslant \max\left(|\mathcal{X}_0^n(w(\cdot \mid y^n))| \, L^{-1}, 1\right) 2|\mathcal{X}| \, |\mathcal{Y}| n \tag{8.6}$$

for every $y^n \in \mathcal{Y}^n$, every $1 \leqslant l \leqslant L$, and every $\mathcal{Y} \times \mathcal{X}$-stochastic matrix.

Proof. Consider the hypergraph $(\mathcal{V}, \mathcal{E}) = (\mathcal{X}^n, \{\mathcal{X}_0^n(w(\cdot \mid y^n)) : y^n \in \mathcal{Y}^n, w \ \mathcal{Y} \times \mathcal{X}\text{-stochastic}\})$.

By definition of $\mathcal{X}_0^n(w(\cdot \mid y^n))$ clearly $M = |\mathcal{E}| \leqslant (n+1)^{|\mathcal{X}| \, |\mathcal{Y}|} |\mathcal{Y}|^n$. For $\alpha = 2 \, |\mathcal{X}| \, |\mathcal{Y}| \, n$ (8.3) holds and Coloring Lemma 10 applies. Q.E.D.

Remark 2 (Comparison of the Key Lemmas). The assumptions in Key Lemma 1:

(a) w not the identity matrix

(b) $x^n \in \Phi^{-1}(l)$ in (8.5)

are without information theoretic significance. They are made to make the GDT applicable. If one allows (artificially) L to exceed $|\mathcal{X}|^n$, they are needed to insure $x^n \notin \mathcal{X}_0^n(w(\cdot \mid x^n))$ and

$$|\mathcal{X}_0^n(w(\cdot \mid x^n)) \cap \Phi^{-1}(l)| = 0 \quad \text{for } x^n \in \Phi^{-1}(l).$$

In our Lemma *all edges* are colored in a balanced way for *all colors* and not just every edge with respect to the color of its "center" x^n.

Key Lemma 2 makes a statement for $\mathcal{Y} \times \mathcal{X}$-stochastic matrices (and not just for $\mathcal{X} \times \mathcal{X}$-stochastic matrices) and is therefore a more appropriate tool for deriving error estimates for more complex source coding problems. There seems to be no immediate way to obtain it from the GDT.

In case $\mathcal{X} = \mathcal{Y}$, $|\mathcal{X}| \geqslant 2$ we have $2|\mathcal{X}|^2 n \leqslant (n+1)^{|\mathcal{X}|^2}$ and (8.6) gives

a somewhat sharper bound than (8.5), if $|\mathcal{X}_0^n(w(\cdot \mid x^n))|\ L^{-1} \geqslant 1$, that is, for relatively few colors. In case $|\mathcal{X}_0^n(w(\cdot \mid x^n))|\ L^{-1} < 1$ (8.5) can be sharper than (8.6), but never by more than $2|\mathcal{X}|\ |\mathcal{Y}|n$.

Remark 3. (Concerning [38]). Inspection of the proofs of [38] shows that a factor or summand growing polynomially in n can be added in (8.5) without effecting the error exponents.

Key Lemma 2 can therefore replace key Lemma 1 for the purposes of [38].

Section 9. An Analysis of Coding Methods

The Graph Decomposition Theorem (GDT) of Lovász [36], which we stated in the preceding Section, has recently been used by Csiszár/Körner [38] to derive channel and source coding theorems. Since one aim of our paper is to present and understand basic methods in the subject, we now analyze the proof of the GDT and then its structure and its impact for coding theory.

§ 1. A Look at the Proof of the GDT

We repeat first the argument, which can be found in [38].

Clearly, every vertex L-coloring Φ of $G = (\mathcal{V}, \mathcal{E}, r)$ is equivalent to a partition

$$\mathcal{A} = \{A_1, \ldots, A_L\} \text{ of } \mathcal{V} : A_i = \{v : \Phi(v) = i\}.$$

Consider now any partition $\mathcal{A}$ for which the functional

$$\sum_i t_i |A_i| - \tfrac{1}{2} \sum_{a \in A_i,\, b \in A_i} r(a, b) \text{ is } \textit{maximal}.$$

If we now exchange for any $a \in A_i$ and any j the set A_i with $A_i - \{a\}$ and the set A_j with $A_j \cup \{a\}$, then the functional's value changes by a necessarily non-positive quantity

$$-t_i + t_j + \sum_{b \in A_i} r(a, b) - \sum_{b \in A_j} r(a, b) \leqslant 0.$$

Summation over j gives

$$\begin{aligned} 0 &\geqslant Lt_i + \sum_j t_j + \sum_j \sum_{b \in A_i} r(a, b) - \sum_j \sum_{b \in A_j} r(a, b) \\ &> -Lt_i + t + L \sum_{b \in A_i} r(a, b) - t, \end{aligned}$$

and hence

$$\sum_{b \in A_i} r(a, b) < t_i \quad \text{for all } a \in A_i. \qquad \text{Q.E.D.}$$

Remark 1. As a minor observation we mention that the assumptions on r : $r(a, a) = 0$, $r(a, b) = r(b, a)$ for all $a, b \in \mathcal{V}$, which are made to describe the multi-graph situation in [36], can be *dropped*, if one formu-

lates the Theorem as follows: Assume that

(a) $\sum_{b \in \mathcal{V}} r(a, b) < t$ *and* $\sum_{a \in \mathcal{V}} r(a, b) < t$ for all $a, b \in \mathcal{V}$,

(b) $\sum_{l=1}^{L} t_l \geqslant t$, $t_l \geqslant 0$ (as before),

then there exists a partition $\mathcal{A} = \{A_1, \ldots, A_L\}$, such that $\frac{1}{2} \sum_{b \in A_l} (r(a, b) + r(b, a)) < t_1$ for all $a \in A_l$, $1 \leqslant l \leqslant L$.

In graphic language this means that we have directed edges and permit loops. (For the proof use the same functional and the same exchange argument as before!). The role of the functional can best be understood in the special case $t_l = t_0$ for $1 \leqslant l \leqslant L$ (which is the case used in [38]), because then one has to *minimize* the functional $\sum_i \sum_{a \in A_i} \sum_{b \in A_i} r(a, b)$.

This means that one minimizes the total number of "inner connections." The proof described is elegant and of a combinatorially simple nature, because one uses "local optimisation". It reminded us right away of Feinstein's maximal code construction [4], and we shall see soon that this is not just a vague analogy.

We now investigate what this type of argument means for channel coding.

§ 2. A Minimal Error Code Construction

In Feinstein's construction, which was explained in Section 4, the *maximal error probability* λ, say, is *fixed in advance*, and *then the code length N is maximized.* In our version of this construction the *average error probability* is fixed in advance (see Section 4).

The argument above is such that the "global" parameter L is fixed in advance and the individual performances $\sum_{b \in A_l} r(a, b)$, $1 \leqslant l \leqslant L$, are minimized. The natural question now is "Can one give the analogue argument for channel coding, that is, fix N and minimize the individual error probabilities?" We now show how this can be done. As in Section 4 an abstract channel is a bipartite graph $(\mathcal{X}, \mathcal{Y}, \mathcal{F})$. $G(x) = \{y \in \mathcal{Y} : (x, y) \in \mathcal{F}\} \neq \emptyset$. An abstract $(M, (\sigma_i)_{1 \leqslant i \leqslant M})$-code is a family $\{u_1, \ldots, u_M\} \subset \mathcal{X}$ with

$$\sum_{j \neq i} |G(u_i) \cap G(u_j)| \, |G(u_i)|^{-1} = \sigma_i, \quad 1 \leqslant i \leqslant M. \tag{9.1}$$

Set $r(x, x') = |G(x) \cap G(x')| \, |G(x')|^{-1}$.

(Recall Remark 1!)

In analogy to Lovász's proof of the GDT we now fix any M and mini-

mize the functional

$$\sum_{l=1}^{M}\sum_{k=1}^{M} r(x_l, x_k) \quad \text{over all } \textit{families } \{x_1, \ldots, x_M\} \subset \mathcal{X}.$$

Assume that for *all* $x \in \mathcal{X}$:

$$\sum_{x' \in \mathcal{X}} r(x, x') + \sum_{x' \in \mathcal{X}} r(x', x) < t. \tag{9.2}$$

Let now $\{u_1, \ldots, u_M\}$ minimize the functional. Then, by exchanging u_i *by any* $u_i' \in \mathcal{X}$:

$$\begin{aligned} &- \sum_{j \neq i} r(u_i, u_j) - \sum_{j \neq i} r(u_j, u_i) - r(u_i, u_i) \\ &+ \sum_{j \neq i} r(u_i', u_j) + \sum_{j \neq i} r(u_j, u_i') + r(u_i', u_i') \geqslant 0. \end{aligned} \tag{9.3}$$

By definition of r we have

$$r(u_i, u_i) = r(u_i', u_i') = 1. \tag{9.4}$$

(Notice that this property is not needed in the case discussed in Remark 1.)

It follows from (9.3) and (9.4) that

$$\sum_{u_i' \in \mathcal{X}} \sum_{j \neq i} (r(u_i', u_j) + r(u_j, u_i')) \geqslant |\mathcal{X}| \sum_{j \neq i} (r(u_i, u_j) + r(u_j, u_i)),$$

and therefore by (9.2)

$$t(M-1) > |\mathcal{X}| \sum_{j \neq i} (r(u_i, u_j) + r(u_j, u_i)) > |\mathcal{X}| \sum_{j \neq i} r(u_i, u_j). \tag{9.5}$$

This and (9.1) imply

$$\sigma_i < |\mathcal{X}|^{-1} t(M-1), \quad 1 \leqslant i \leqslant M. \tag{9.6}$$

Thus we have arrived at the

MINIMAL ERROR LEMMA. *Given an abstract channel* $(\mathcal{X}, \mathcal{Y}, \mathcal{F})$, $G(x) \neq \phi$ *for* $x \in \mathcal{X}$, *with*

$$\sum_{x' \in \mathcal{X}} |G(x) \cap G(x')|\, |G(x)|^{-1} + \sum_{x' \in \mathcal{X}} |G(x) \cap G(x')|\, |G(x')|^{-1} < t$$

for all $x \in \mathcal{X}$. (9.7)

Then for every $M \geqslant 1$ *there exists an* $(M, (\sigma_i)_{1 \leqslant i \leqslant M})$*-code with*

$$\sigma_i < |\mathcal{X}|^{-1} tM, \; 1 \leqslant i \leqslant M. \tag{9.8}$$

Remark 3 This result shows that we are on the right track. The Lovász type argument differs from Feinstein's argument only in so far as different parameters are optimized. One might also look for a proof of the GDT (in case $t_i = t_0$) by minimizing L subject to the constraint

$$\sum_{b \in A_l} r(a, b) < t_0 \text{ for all } a \in A_l, \; 1 \leqslant l \leqslant L.$$

Remark 4. With $\sigma = \frac{1}{M}\sum_{i=1}^{M} \sigma_i$ this result implies

$$M > \sigma |\mathcal{X}| t^{-1}, \tag{9.9}$$

and from (9.7) we conclude that

$$\frac{1}{|\mathcal{X}|}\sum_{x,x'\in\mathcal{X}}|G(x)\cap G(x')|\,|G(x)|^{-1}<t.$$

This means that (4.2) in the Maximal Code Lemma is a little bit sharper than (9.9). This is due to the fact that in the present approach we could have used instead of σ_i also

$$\sum_{j\neq i}|G(u_i)\cap G(u_j)|\,|G(u_i)|^{-1}+\sum_{j\neq i}|G(u_j)\cap G(u_i)|\,|G(u_j)|^{-1}$$

as performance criterion. (9.5) shows that we had to give something away. Since for the channel graph $|G(x)|\,|G(x')|^{-1}=\exp\{0(\sqrt{n})\}$ for all x, x' the difference noticed is not essential (see Section 4, Remark 2). Furthermore, in *this case it is sufficient to use* $\sum_{j\neq i}|G(u_i)\cap G(u_j)|$ *as performance criterion.* We can here use $r(x,x')=|G(x)\cap G(x')|=r(x',x)$ and the previous argument gives then (9.8) under the assumption

$$\sum_{x'\in\mathcal{X}}|G(x)\cap G(x')|<t \text{ for all } x'\in\mathcal{X}. \tag{9.7'}$$

We shall see that this criterion is the one most suited to understand the structure of the GDT.

§ 3. Graph Decomposition Implies Code Decomposition

Let $(X,\mathcal{Y},\mathcal{E})$, $G(x)\neq\emptyset$ for all $x\in\mathcal{X}$, be a bipartite graph. Recall that $\mathrm{Deg}\,(x)=\sum_{y\in G(x)}\deg(y)$, that $\mathrm{Deg}\,(x)=\sum_{x'\in\mathcal{X}}|G(x)\cap G(x')|$, and that $D=\max_{x\in\mathcal{X}}\mathrm{Deg}\,(x)$.

Clearly, a partition of $\mathcal{X}$ is equivalently described by a coloring. By Remark 4 it is thus clear that the CDT can be formulated as follows:

*CDT** For every $L\geqslant 2(\log|\mathcal{X}|)D\sigma^{*-1}$ there exists an L-coloring Φ of $\mathcal{X}$ with

$$\sum_{x'\in\Phi^{-1}(l),\,x'\neq x}|G(x)\cap G(x')|<\sigma^* \text{ for all } x\in\Phi^{-1}(l),\ 1\leqslant l\leqslant L. \tag{9.10}$$

If we define

$$r(x,x')=\begin{cases}|G(x)\cap G(x')| & \text{for } x\neq x'\\ 0 & \text{for } x=x'\end{cases}$$

then

$$\sum_{x'\in\mathcal{X}}r(x,x')=\sum_{x'\in\mathcal{X}}|G(x)\cap G(x')|-|G(x)|$$

$$=\mathrm{Deg}\,(x)-\deg(x)\leqslant\max_x\mathrm{Deg}\,(x)-\min_x\deg(x)<D.$$

Choosing $t_l=\sigma^*$ for $1\leqslant l\leqslant L$, the GDT says that for $L\geqslant D\sigma^{*-1}$

there is an L-coloring with

$$\sum_{x' \in \Phi^{-1}(l)} r(x, x') < \sigma^* \text{ for all } x \in \Phi^{-1}(l),\ 1 \leqslant l \leqslant L. \tag{9.11}$$

This is exactly inequality (9.10).

Remark 5. Our factor $2 \log |\mathcal{X}|$ is the prize for the iterative construction. For the channel graph $\log |\mathcal{J}_\delta(X^n)| = 0(n)$ does not effect exponential error bounds. Instead of the GDT the CDT could have also been used in [38].

Remark 6 Our Theorem 7, Section 6, says that for the channel graph one can also achieve that $|\Phi^{-1}(l)|\ |\Phi^{-1}(l')|^{-1}$ is small for all l, l'. This sharper result does *not* follow from the GDT.

§ 4. The Code Decomposition Problem as a Hypergraph Coloring Problem

$T : (\mathcal{X}, \mathcal{Y}, \mathcal{F}) \to (\mathcal{X}, T(\mathcal{F}))$ with

$$T(\mathcal{F}) = \{G(y) : y \in G(x),\ x \in \mathcal{X}\} \tag{9.12}$$

transforms bipartite graphs into hypergraphs and

$$T' : (\mathcal{X}, \mathcal{E}) \to (\mathcal{X}, \mathcal{E}, T'(\mathcal{E}))$$

with

$$T'(\mathcal{E}) = \{(x, E) : x \in \mathcal{X}, E \in \mathcal{E} \text{ with } x \in E\} \tag{9.13}$$

transforms hypergraphs into bipartite graphs.

Clearly, $TT' : (\mathcal{X}, \mathcal{Y}, \mathcal{F}) \to (\mathcal{X}, \mathcal{Y}, \mathcal{F})$, $T'T : (\mathcal{X}, \mathcal{E}) \to (\mathcal{X}, \mathcal{E})$ and therefore $T' = T^{-1}$.

We now interpret the code decomposition problem for bipartite graphs as a hypergraph problem.

Let $\Phi : \mathcal{X} \to \{1, \ldots, L\}$ be a coloring (or code decomposition) of $(\mathcal{X}, \mathcal{Y}, \mathcal{F})$ with performances

$$g_1(x, \Phi) = \sum_{x' \in \Phi^{-1}(l) - \{x\}} |G(x) \cap G(x')| \text{ for } x \in \Phi^{-1}(l),\ 1 \leqslant l \leqslant L. \tag{9.14}$$

One readily verifies that

$$g_1(x, \Phi) = \sum_{E : x \in E,\ E \in T(\mathcal{F})} |\{x' : x' \in E - \{x\},\ \Phi(x') = \Phi(x)\}|. \tag{9.15}$$

Therefore $g_1(x, \Phi)$ means the following: count for every edge E, $x \in E$, the number of elements in $E - \{x\}$ with the same color as x and sum these numbers obtained for these edges. (In list code terminology this means that the numbers of code words occurring in all lists containing x are added. The lists are exactly the edges of the hypergraph.)

HCP_1 suggests the following concept: for any hypergraph $(\mathcal{X}, \mathcal{E})$ a coloring Φ is of type x_λ if *every* $x \in \mathcal{X}$ is colored correctly *within* at least a fraction $(1 - \lambda)$ of the edges containing x.

Another measure of performance would be the number of edges E containing x with the property:

$$\Phi(x') = \Phi(x) \quad \text{for } \textit{at least one } x' \in E - \{x\}.$$

Denoting this number by $g_2(x, \Phi)$ we can write this formally as

$$g_2(x, \Phi) = |\{E \in T(\mathcal{F}) : \exists\, x' \in E - \{x\} \text{ with } \Phi(x') = \Phi(x)\}|. \quad (9.16)$$

Notice that

$$g_1(x, \Phi) \geqslant g_2(x, \Phi) \quad \text{for all } x \in \mathcal{X} \text{ and all colorings } \Phi. \quad (9.17)$$

Therefore an upper bound on g_1 gives also an upper bound on g_2. For the bipartite graph $(\mathcal{X}, \mathcal{E}, \sigma(\mathcal{E}))$, $g_2(x, \Phi)$ can be written as

$$g_2(x, \Phi) = |G(x) \cap \cup\{G(x') : \Phi(x') = \Phi(x),\ x' \neq x\}|. \quad (9.18)$$

This corresponds to the maximal error concept for ordinary (*non list*) codes.

Summarizing our findings we can state: The code decomposition problem (CDP) is equivalent to the hypergraph coloring problem (HCP_1) with $g_1(x, \Phi)$ as performance measure. We have raised another hypergraph coloring problem (HCP_2) with $g_2(x, \Phi)$ as performance measure, which has also a coding interpretation.

§ 5. Comparison of the Graph Decomposition Problem (GDP) and the Hypergraph Coloring Problem (HCP_1)

A graph with multiple edges is a hypergraph $(\mathcal{X}, \mathcal{M})$ with edges of *size* 2. In this case clearly

$$g_1(x, \Phi) = g_2(x, \Phi) \quad \text{for all } x \in \mathcal{X} \text{ and all } \Phi, \quad (9.19)$$

and therefore here HCP_1 and HCP_2 are identical. Since $r(x, x')$ counts the multiplicity of edges (x, x'), we have for every $x \in \mathcal{X}$ the identity

$$\sum_{x' : \Phi(x') = \Phi(x)} r(x, x') = |\{E \in \mathcal{M} : E = (x, x'), \Phi(x') = \Phi(x)\}| = g_2(x, \Phi) = g_1(x, \Phi). \quad (9.20)$$

We have observed in § 4 that the CDP and the HCP_1 are equivalent. In § 3 we saw that the CDP is a special case of the GDP and now we have explained that the GDP is a special case of the HCP_1 (and also the HCP_2). Therefore we have the

Equivalence Lemma. *The three problems HCP_1, GDP, and CDP are all equivalent.*

Remark 7. It is instructive to see the equivalence between the GDP and the HCP_1 directly. Given a hypergraph $(\mathcal{X}, \mathcal{E})$ define

$$r(x, x') = \begin{cases} |\{E \in \mathcal{E} : x, x' \in E\}| & \text{if } x \neq x' \\ 0 & \text{if } x = x'. \end{cases} \quad (9.21)$$

Then

$$g_1(x, \Phi) = \sum_{E \in \mathcal{E}} |\{x' : x' \neq x;\ x, x' \in E;\ \Phi(x') = \Phi(x)\}|$$

$$= \sum_{x' \neq x} |\{E \in \mathcal{E} : x, x' \in E;\ \Phi(x') = \Phi(x)\}|$$

$$= \sum_{x' \neq x\,:\,\Phi(x')=\Phi(x)} |\{E \in \mathcal{E} : x, x' \in E\}| = \sum_{x'\,:\,\Phi(x')=\Phi(x)} r(x,x').$$

Conversely, given $(\mathcal{X}, r)$ choose for instance the graph with multiple edges as hypergraph $(\mathcal{X}, \mathcal{M})$, $\mathcal{M}$ = family of all edges $\{x, x'\} \subset \mathcal{X}$ taken $r(x, x')$ times. The weighted graph $(\mathcal{X}, r^*)$ associated with it according to (9.21) is again $(\mathcal{X}, r)$. *However*, instead of $(\mathcal{X}, \mathcal{M})$ we could use *any* decomposition of $(\mathcal{X}, r)$ into *cliques*. This fact is important, if one wants to prove a *converse* to the GDT or the CDT. There is this *freedom of choosing a suitable hypergraph representation* (or, equivalently, bipartite graph representation) in order to obtain upper bounds on L. $(\mathcal{X}, \mathcal{M})$ itself is usually not suited for this.

Since in the CDT the sets $A_l = \Phi^{-1}(l)$ are codes one can obtain a lower bound on L from an upper bound for the code length.

Suppose that $\mathcal{U} = \{u_1, \ldots, u_N\} \subset \mathcal{X}$ satisfies $\sum_{u_j \in \mathcal{U}\,:\,j \neq i} |G(u_i) \cap G(u_j)| < \sigma^*$ for all $1 \leqslant i \leqslant N$.

Then

$$\sum_{i=1}^{N} (\mathrm{Deg}\,(u_i) - \sigma^*) \leqslant \sum_{x} |G(x)|$$

and hence

$$N \leqslant |\mathcal{X}|\left(\frac{1}{|\mathcal{X}|}\sum_{x} |G(x)|\right)(\min_{x \in \mathcal{X}} \mathrm{Deg}\,(x) - \sigma^*)^{-1}.$$

Therefore

$$L \geqslant |\mathcal{X}|N^{-1} \geqslant (\min_{x \in \mathcal{X}} \mathrm{Deg}\,(x) - \sigma^*)\left(\frac{1}{|\mathcal{X}|}\sum_{x} |G(x)|\right)^{-1}. \qquad (9.22)$$

$\mathrm{Deg}\,(x)$ is determined by r, but $\frac{1}{|\mathcal{X}|}\sum_{x} |G(x)|$ depends on the representation as CDP. The bound is best if this quantity is minimal.

Remark 8. The CDT (and also the GDT) is based on a "two code words" performance criterion. We therefore know how far those results can carry. In deriving bounds for error exponents for instance one therefore has to use the "union bound". It is therefore not to be expected that improvements of the expurgated bound can be obtained this way. For progress in this direction one has to get better results on HCP_2, which is generally not equivalent to HCP_1. We therefore propose the

PROBLEM. Find better results on HCP_2, that is for colorings of type x_λ, in general hypergraphs and especially for the channel hypergraph.

Final Remark. In extremal hypergraph theory ([47], [48]) usually strict colorings, packings etc. are studied. Coding Theory can be viewed as extremal hypergraph theory for colorings, packings etc. where the

"strict" is replaced by "almost strict". We expect that an interchange of concepts, problems and methods available in those two areas will lead to progress in both of them.

REFERENCES

[1]-[35] as in R. Ahlswede (1979), "Coloring hypergraphs: a new approach to multi-user source coding—I", *Journ. of Combinatorics, Information & System Sciences*, **4** (1), 76–115.

[36] L. Lovász (1966), "On decomposition of graphs", *Studia Sci. Math. Hung.*, **1**, 237–238.

[37] V. N. Koševlev (1977), "On a problem of separate coding of two dependent sources" (in Russian), *Probl. Per. Inf.*, **13**, 26–32.

[38] I. Csiszár and J. Körner (June 1979), "Graph decomposition: a new key to coding theorems" presented at the IEEE Internat. Symposium on Inf. Theory, Grignano, Italy.

[39] ———, *Information Theory. Coding Theorems for Discrete Systems*, Forthcoming book.

[40] M. Salehi (1978), "Cardinality bounds on auxiliary variables in multiple-user theory via the method of Ahlswede and Körner", Stanford Technical Report.

[41] R. Ahlswede and J. Körner (1974), "On common information and related characteristics of correlated information sources", presented at the 7th Prague Conf. on Inf. Th., Stat. Dec. Fct's and Rand. Proc. Preprint.

[42] J. Körner and K. Marton (1979), "How to encode the modulo 2 sum of two binary sources", *IEEE Trans. Inf.*, **IT-25**, 219–221.

[43] R. Ahlswede, "A method of coding and an application to arbitrarily varying channels", to appear in *JCISS*.

[44] ——— (Oct. 1977), "Remarks on cryptography", presented at the IEEE Int. Symp. on Inf. Theory, Ithaca.

[45] T. Berger (Jan. 1971), "The source coding game", *IEEE Trans. Inf. Theory*, **IT-17** (1).

[46] ——— (1971), *Rate Distortion Theory*, Prentice-Hall Inc., Englewood Cliffs, N.J.

[47] C. Berge (1973), *Graphs and Hypergraphs*, North-Holland, American Elsevier.

[48] L. Lovász (1979), *Combinatorial Problems and Exercises*, North-Holland.

[49] J. H. Jahn, "Coding of arbitrarily varying multi-user channels", Technical Report No. 37, Dept. of Statistics, Stanford University, to appear in *IEEE Trans. on Inf. Theory*.

[Received : November, 1979]