MODELS OF MULTI–USER WRITE–EFFICIENT MEMORIES AND GENERAL DIAMETRIC THEOREMS

Rudolf Ahlswede and Ning Cai

Universität Bielefeld
Fakultät für Mathematik
Postfach 100131
33501 Bielefeld
Germany

Abstract

Write–efficient memories (WEM) were introduced by Ahlswede/Zhang as a model for storing and updating information on a rewritable medium.

We strengthen the capacity theorem by providing a full control of the rates of the spreads. Next we adress and settle the storage capacity region problem under the average costs constraint in the case of many users, who write on the memory in an arbitrary order, if neither the encoder nor the decoder knows the previous content of the memory. The combinatorial essence is a diametric theorem for several families.

Finally we present a storage capacity theorem for several persons using the memory in cyclic order.

1. Introduction

We continue the investigation of write–efficient memories (WEM), which were introduced in [AZ89] as a model for storing and updating information on a rewritable medium. Such a medium (memory) consists of n cells. Each such cell can carry some letter x from a finite alphabet $\mathcal{X}$. Thus in total a sequence $x^n \in \mathcal{X}^n$ can be stored. When a user wants to change the content x^n to a new content $y^n \in \mathcal{X}^n$ (updating) he must make changes in positions only, where the sequences are different. Very likely it is easier to make a few changes than to make many. Therefore we introduce a function $\varphi : \mathcal{X} \times \mathcal{X} \to [0, \infty)$, where $\varphi(x, y)$ measures the cost (time or energy) of a change from x to y. We assume that these costs add up, that is, the cost for changing sequence $x^n = (x_1, \dots, x_n) \in \mathcal{X}^n$ to sequence $y^n = (y_1, \dots, y_n) \in \mathcal{X}^n$ is given by

$$\varphi_n(x^n, y^n) := \sum_{t=1}^{n} \varphi(x_t, y_t). \tag{1.1}$$

Obviously, if there is no constraint on costs, one can have a set $\mathcal{M} = \{1, 2, \dots, M\}$ of messages with $M = |\mathcal{X}|^n$, represent each message $m \in \mathcal{M}$ by a sequence $u_m \in \mathcal{X}^n$ and the user can update m to m' by replacing u_m by $u_{m'}$. The problems start with cost constraints. We consider two kinds.

Criterion Max: In this model a user can never make an updating, whose cost exceeds a prescribed cost $D_{\max}$.

Criterion Ave: It is assumed that messages occur at random with equal probabilities and independently. The average cost is not allowed to exceed a prescibed cost D_{ave}.

To meet such criteria becomes increasingly difficult with an increasing number M of messages. Mathematically, we are led to diametric problems for the sequence space $\mathcal{X}^n$.

Question 1: $\mathcal{U} \subset \mathcal{X}^n$ has a diameter not exceeding $D_{\max}$, if

$$\varphi_n(x^n, y^n) \le D_{\max} \text{ for all } x^n, y^n \in \mathcal{U}. \tag{1.2}$$

What is the maximal cardinality of $\mathcal{U}$?

Results can be found in [ACZ92], [AK77], [K64]. Moreover, it has been shown in [Kl66] that for *binary* $\mathcal{X}$ and the Hamming distance d_H as cost function the Hamming sphere is optimal. The solution for *non–binary* $\mathcal{X}$ has been obtained only recently in [AKh]. Here the solution is a suitable product of a sphere and a cylinder set.

Question 2: $\mathcal{U} \subset \mathcal{X}^n$ has an average diameter not exceeding D_{ave}, if

$$\frac{1}{|\mathcal{U}|^2} \sum_{x^n \in \mathcal{U}} \sum_{y^n \in \mathcal{U}} \varphi_n(x^n, y^n) \le D_{\text{ave}}. \tag{1.3}$$

What is the maximal cardinality of $\mathcal{U}$?

For every φ, this has been answered in [AAl94] in an asymptotic sense (optimal rate for specified per letter average cost). It is stated as Corollary 1 to our more general Theorem 3 in Section 5.

We return to memories. Every $\mathcal{U}$ meeting (1.2) (resp. (1.3)) can be used as set of codewords representing the set of messages $\mathcal{M}$. Any labelling $\mathcal{U} = \{u_1, \dots, u_M\}$ serves our purposes. For given message m the encoder updates the memory to $u_m = (u_{m1}, \dots, u_{mn})$ when the present content is $u_{m'} = (u_{m'1}, \dots, u_{m'n})$. While updating the t–th cell there arises a cost $\varphi(u_{m't}, u_{mt})$; $1 \leq t \leq n$.

Now comes the key idea. If the encoder makes the representation for message m dependent on the content $u_{m'}$ of the memory, which he reads before he updates, then many more messages can be handled. Of course it must be guaranteed always that the decoder (a reader) can recover a message from the sequence in the memory!

Let us look at the

Example 1: $\mathcal{X} = \{0,1\}$, $n = 3$, $\varphi = d_H$, $D_{\max} = 1$.

Clearly, the maximal set in $\mathcal{X}^3$ with diameter 1 has only 2 elements and necessarily $M \leq 2$.

Now we use for every message m a set C_m of candidates for a representation. Let $\mathcal{M} = \{1,2,3,4\}$ and choose the disjoint sets

$$C_1 = \big\{(0,0,0),(1,1,1)\big\}, \; C_2 = \big\{(0,0,1),(1,1,0)\big\},$$

$$C_3 = \big\{(0,1,0),(1,0,1)\big\}, \; C_3 = \big\{(1,0,0),(0,1,1)\big\}.$$

One readily verifies that for every $m, m' \in \mathcal{M}$ and every $u \in C_m$ there is a $u' \in C_{m'}$ with $d_H(u,u') = 1$. Using the C_i's we have an updating device for 4 messages on the same 3 cells.

Generally speaking, in the $D_{\max}$–neighbourhood of every C_m there must be a member of every $C_{m'}$ to which the transition can be made. Intuitively, the C_m's must have members "everywhere" in $\mathcal{X}^n$ and therefore we call them "**spreads**".

In earlier work we have used the abbreviation (E_+, D_-) to indicate that the encoder can read the content of the memory before he chooses a close by representative for the new message. Here we just refer to Rule II whereas in the previous case, where the encoder does not have this possibility, we refer to Rule I. Now we give the formal definitions. When we shortly speak of a WEM, Rule II is tacitly assumed.

A collection of subsets (also called "spreads") $\mathcal{C} = \{C_m\}_{m=1}^M$ of $\mathcal{X}^n$ is an (n, M, D) WEM code, if $C_i \cap C_j = \varnothing$ for all $i \neq j$ and if

$$D_{\max} := \max_{1 \leq i,j \leq M} \max_{x^n \in C_i} \min_{y^n \in C_j} \sum_{t=1}^{n} \varphi(x_t, y_t) \leq D. \tag{1.4}$$

$D_{\max}$ is called the maximal updating cost with respect to the given cost function. The performance of a code $\mathcal{C}$ can also be measured by two parameters, namely, the maximal cost per letter $d_{\mathcal{C}} = n^{-1} D_{\max}$ and the rate of the size of the code $r_{\mathcal{C}} = n^{-1} \log M$. The rate achievable with a maximal per letter cost d is thus

$$R(d) = \sup_{\mathcal{C}: d_{\mathcal{C}} \leq d} r_{\mathcal{C}}. \tag{1.5}$$

This is the most basic quantity (the storage capacity) of a WEM $(\mathcal{X}^n, \varphi_n)_{n=1}^\infty$.

For a WEM code $\mathcal{C}$ the average updating cost D_{ave} can be defined as

$$D_{\text{ave}} = \frac{1}{M^2} \sum_{1 \le i,j \le M} \frac{1}{|C_i|} \sum_{x^n \in C_i} \min_{y^n \in C_j} \sum_{t=1}^{n} \varphi(x_t, y_t) \tag{1.6}$$

and the average cost per letter can be defined as

$$\overline{d}_{\mathcal{C}} = n^{-1}\ D_{\text{ave}}. \tag{1.7}$$

The rate achievable with an average per letter cost d is thus

$$\overline{R}(d) = \sup_{\mathcal{C}: \overline{d}_{\mathcal{C}} \le d} r_{\mathcal{C}}. \tag{1.8}$$

The achievable rates $R(d)$ and $\overline{R}(d)$ were characterized in [AZ89]. Actually, their values are the same for all $d \geq 0$.

However, as compared to Rule I the Rule II has the drawback that while trying to store message m' , when m is stored as $x^n \in C_m$, the encoder has to find a $y^n \in C_{m'}$ with $\varphi_n(x^n, y^n) \le dn$. This causes an extra effort, which is not present under Rule I. It can be kept smaller by working with small spreads. This leads to

(Problem 8 in [AZ94]) **Question 3:** What are the achiebable rates under the additional restriction that the sizes of spreads do not exceed $2^{n\rho}$?

The kernel of this question is a purely combinatorial problem of some independent interest.

(Problem 7 in [AZ94]) **Question 3a:** How small "is the smallest rich world"?

This question is formalized and answered in Theorem 1 in Section 3. The solution is a good demonstration of the use of informationtheoretical techniques in combinatorics. It yields the answer to question 3, which is stated as Theorem 2 in Section 4.

In the remainder of this paper we are concerned with a WEM with many users. Its study was initiated in [AZ94], where also suboptimal constructions can be found. Assume here that L users share a memory and each of them has his own messages, that is, the i–th user has his message set $\mathcal{M}_i$. Also, each of them may have his own cost function and his own cost constraint.

Per letter cost and rate are now replaced by per letter cost vectors and rate vectors. The interval between 0 and the optimal rate is now replaced by a rate region.

Under Criterion Ave this region is denoted by $\overline{\mathcal{R}}(\vec{d})$.

(Related to problem 9 in [AZ94]) **Question 4:** What is the region of achievable rates $\overline{\mathcal{R}}(\vec{d})$ for several users under Criterion Ave and Rule I?

It turns out that mathematically the same approach settles also cases, where cost functions and constraints depend not only on the active user, but also on the previous user. Our answer is Theorem 3 in Section 5 — a very general diametric theorem.

As an **example** catching this philosophy let us assume that in a problem session students write on a black board in the alphabet $\{0,1\}$. Whenever a student comes to the blackboard he can write down his news by changing certain zeroes and ones at the board. This is never done in practice, because the alphabets of natural languages are so big that there is hardly any advantage over erasing everything and writing all new again.

Finally, let us assume that professors use the board in their lectures according to a fixed schedule, for instance in cyclic order. This gives some advantage over an arbitrary moving order of the users.

Question 5: What is the region of achievable rates when all users follow a cyclic protocol and Rule II under Criterion Ave is used?

This question could be answered in Theorem 4 of Section 6, where we don't impose a constraint on the sizes of spreads.

The paper is organized as follows. To make itself contained, we present the necessary auxiliary concepts and results from information theory in Section 2. Questions 3a, 3, 4, and 5 are answered in Sections 3, 4, 5, and 6, respectively. Each of Sections 3 – 6 is divided into three or four subsections. First the problems are formulated and then the main results are stated. They are proved in the third parts.

In conclusion we mention that the paper [A71] was based on purely speculative ideas. Recently we learnt [O 1995] that they led to practical codes, which are used in mobile communication, and that a company makes now billions with it. The physicist Boltzmann once said that nothing is as practical as a good theory. We think that our models are natural. The results specify the theoretical optimal performances. We have no reasons to doubt that some day they become also practically relevant.

2. Concepts and facts from information theory

Write efficient memories (WEM) are purely combinatorial, that is, non–probabilistic models. However, in their mathematical analysis probability theory comes in twice, at first in existence proofs based on random selection of codewords representing messages and secondly in the *description of basic parameters* such as the updating capacity. Actually those descriptions are in terms of entropy or conditional entropy functionals of random variables (RV's) or (equivalently) their corresponding probability distributions (PD's). The situation is similar to the theory of error correcting codes, where for instance the asymptotic forms of Hamming's bound or Gilbert's bound can be expressed in terms of entropy. Even more instructive is Shannon's fundamental formula for the capacity of a noisy channel, which involves an optimisation over an *auxiliary class of PD's.* For readers familiar with rate–distortion theory let us emphasize that there the optimisation runs over an auxiliary class of *channels*, that is, conditional PD's. Now, in Theorem 2 in Section 4.2 (and the earlier Storage Capacity Theorem of [AZ89]) the optimisation runs over an auxiliary class of *bivariate distributions.* Other theorems of this paper use *multivariate distributions.*

Besides basic concepts from information theory, which can be found in standard text books (e.g. [CT91], [CsKö81], [G68], [W78]), we do need more advanced techniques from *multi–user* information theory ([CT91], [CsKö81], [W78]). We now explain our notation and known results used in the sequel. The logarithm "log" is always understood to be to the base 2. The letters P, Q stand for PD's and the letters $W, W', W_1, \dots$ for stochastic matrices (also called channels). We frequently use for an input distribution P and a channel W the conventions PW and $P \times W$ for the output distribution and the joint distribution, respectively. $X, Y, \dots$ denote RV's and their distributions, conditional distributions and joint distributions are written as P_X , $P_{Y|X}$, P_{XY} and so on.

$\mathcal{P}(\mathcal{X})$ is the set of PD's on a finite set $\mathcal{X}$ and

$$\mathcal{P}(n, \mathcal{X}) := \left\{ P \in \mathcal{P}(\mathcal{X}) : P(x) \in \left\{ 0, \frac{1}{n}, \frac{2}{n}, \dots, 1 \right\} \text{ for all } x \in \mathcal{X} \right\}.$$

We now introduce the entropy of a RV Z or its corresponding distribution $P_Z = P$. Viewing $(\mathcal{Z}, Z)$ as an experiment with chance outcome Z , the entropy can be viewed as measuring the uncertainty about the outcome before performing the experiment.

The entropy of a PD $P \in \mathcal{P}(\mathcal{Z})$ is

$$H(P) := -\sum_{z \in \mathcal{Z}} P(z) \log P(z), \tag{2.1}$$

the entropy of a RV Z is

$$H(Z) := H(P_Z) = \mathbb{E}\big(-\log P_Z(Z)\big), \tag{2.2}$$

where $\mathbb{E}$ is the expectation operator. Similarly for vectors of RV's $X^\ell = (X_1, \dots, X_\ell)$

$$H(X^\ell) = H(X_1, \dots, X_\ell) = H(P_{X^\ell}). \tag{2.3}$$

Elementary properties of entropy are

1.) $H : \mathcal{P}(\mathcal{Z}) \to \mathbb{R}_+$ is continuous.
2.) $0 \leq H(P) \leq \log |\mathcal{Z}|$ for $P \in \mathcal{P}(\mathcal{Z})$, where $H(P) = 0$ exactly if for some $z \in \mathcal{Z}$ $P(z) = 1$ and $H(P) = \log |\mathcal{Z}|$ exactly if $P(z) = |\mathcal{Z}|^{-1}$ for all $z \in \mathcal{Z}$.
3.) For two RV's Z_1 and Z_2

$$H(Z_1) \leq H(Z_1 Z_2) \leq H(Z_1) + H(Z_2),$$

where $H(Z_1) = H(Z_1 Z_2)$, exactly if Z_2 is a function of Z_1 , and $H(Z_1 Z_2) = H(Z_1) + H(Z_2)$, exactly if Z_1 and Z_2 are independent.
4.) H is a concave function.

Next we describe conditional entropy. For a $P \in \mathcal{P}(\mathcal{Z}_1)$ and a $\mathcal{Z}_1 \times \mathcal{Z}_2$–stochastic matrix W the conditional entropy is

$$\begin{aligned} H(W|P) : &= \sum_{z_1 \in \mathcal{Z}_1} P(z_1) H\big(W(\cdot|z_1)\big) \\ &= -\sum_{z_1 \in \mathcal{Z}_1} P(z_1) \sum_{z_2 \in \mathcal{Z}_2} W(z_2|z_1) \log W(z_2|z_1). \end{aligned} \tag{2.4}$$

In terms of RV's Z_1 and Z_2 the conditional entropy of Z_2 given Z_1, is

$$H(Z_2|Z_1) := H(P_{Z_2|Z_1}|P_{Z_1}) = \mathbb{E}\big[-\log P_{Z_2|Z_1}(Z_2|Z_1)\big]. \tag{2.5}$$

Furthermore, the conditional entropy of Z_2 given $Z_1 = z_1$ is

$$H(Z_2|Z_1 = z_1) := \mathbb{E}\big[-\log P_{Z_2|Z_1}(Z_2|Z_1)|Z_1 = z_1\big] = H\big(P_{Z_2|Z_1}(\cdot|z_1)\big), \tag{2.6}$$

and for RV's Z_1, Z_2, and Z_3

$$\begin{aligned} H(Z_3|Z_1, Z_2 = z_2) :&= \mathbb{E}\big[-\log P_{Z_3|Z_1Z_2}(Z_3|Z_1Z_2)|Z_2 = z_2\big] \\ &= \sum_{z_1} P_{Z_1|Z_2}(z_1|z_2)H\big(P_{Z_3|Z_1Z_2}(\cdot|z_1, z_2)\big) \end{aligned} \tag{2.7}$$

Then

$$H(Z_2|Z_1) = \sum_{z_1} P_{Z_1}(z_1)H(Z_2|Z_1 = z_1), \tag{2.8}$$

and

$$H(Z_3|Z_1, Z_2) = \sum_{z_2} P_{Z_2}(z_2)H(Z_3|Z_1, Z_2 = z_2). \tag{2.9}$$

Conditional entropy has the following properties:

5.) $H(W|P)$ is a continuous function of (P, W).

6.) $0 \le H(Z_2|Z_1) \le H(Z_2)$, (2.10)

where $H(Z_2|Z_1) = 0$ exactly if Z_2 is a function of Z_1, (2.11)

$H(Z_2|Z_1) = H(Z_2)$ exactly if Z_1 and Z_2 are independent. (2.12)

7.) $H(Z_3|Z_1, Z_2) \le H(Z_3|Z_2)$, (2.13)

where $H(Z_3|Z_1, Z_2) = H(Z_3|Z_2)$ exactly if (Z_1, Z_2, Z_3) forms a Markov's chain, that is, given Z_2 the RV's Z_1 and Z_3 are independent. (2.14)

8.) $H(Z^\ell) = \sum\limits_{t=1}^{\ell} H(Z_t|Z^{t-1})$, (2.15)

and

$H(Z^\ell|U) = \sum_{t=1}^{\ell} H(Z_t|Z^{t-1}, U)$. (2.16)

Here, when $t = 1$, $H(Z_t|Z^{t-1})$ and $H(Z_t|Z^{t-1}, U)$ are understood as $H(Z_1)$ and $H(Z_1|U)$, respectively.

9.) $H(W|P)$ is a concave function of W.

It immediately follows from properties 6.) – 8.) that

$$H(Z^\ell) \le \sum_{t=1}^{\ell} H(Z_t), H(Z^\ell|U) \le \sum_{t=1}^{\ell} H(Z_t|U). \tag{2.17}$$

Property 9.) actually follows from property 7.).

To see this, for P , W_1 , W_2 , and $\lambda \in (0,1)$, we define RV's X, Y, U with the joint distribution

$P_{XYU}(x,y,1) = P(x)W_1(y|x)\lambda$ and $P_{XYU}(x,y,2) = P(x)W_2(y|x)(1-\lambda)$,

and notice that by property 7.)

$\lambda H(W_1|P) + (1-\lambda)H(W_2|P) = H(Y|XU) \leq H(Y|X) = H\big(\lambda W_1 + (1-\lambda)W_2|P\big)$.

Moreover, by 2.), (2.6), and (2.8), we have that for all z_1 ,

$$\big|\big\{z_2 : P_{Z_2|Z_1}(z_2|z_1) > 0\big\}\big| \leq L \text{ implies } H(Z_2|Z_1) \leq \log L \tag{2.18}$$

and equality holds exactly if $P_{Z_2|Z_1}(z_2|z_1) = \frac{1}{L}$ for all z_1, z_2 with $P_{Z_2|Z_1}(z_2|z_1) > 0$.

Entropy and conditional entropy are the only information measures (or quantities) used in this paper. We don't need the mutual information.

We also need a special case of the well known Markov inequality. For $a_i \geq 0$; $i = 1, 2, \ldots, M$;

$$M^{-1}\sum_{i=1}^{M} a_i \leq A \text{ implies } |\{i : a_i \geq 2A\}| \leq \frac{1}{2}M. \tag{2.19}$$

Next we introduce concepts of "typicality" which make it possible to reduce the analysis of outcomes of sequences of RV's to counting.

For a finite set $\mathcal{Z}$ and $z^n \in \mathcal{Z}^n$ denote by P_{z^n} the empirical distribution, i.e. for all $z \in \mathcal{Z}$

$$P_{z^n}(z) := \frac{1}{n} \text{ (number of } z \text{ in } z^n \text{)}, \tag{2.20}$$

and call P_{z^n} type of z^n . Obviously for all z^n $P_{z^n} \in \mathcal{P}(n, \mathcal{Z})$. For $P \in \mathcal{P}(n, \mathcal{Z})$ the set $\mathcal{T}_P^n$ of all P–typical sequences in $\mathcal{Z}^n$ is given by $\mathcal{T}_P^n := \{z^n : P_{z^n} = P\}$.

Analogously we define the (joint) type $P_{y^n z^n}$ for pairs $(y^n, z^n) \in \mathcal{Y}^n \times \mathcal{Z}^n$ as the empirical distribution of (y^n, z^n) , (i.e. by counting the number of pairs (y, z) in the components of (y^n, z^n)) . Similarly, for $Q \in \mathcal{P}(n, \mathcal{Y} \times \mathcal{Z}), \mathcal{T}_Q^n := \big\{(y^n, z^n) : P_{y^n, z^n} = Q\big\}$.

We abbreviate $\mathcal{T}_X^n := \mathcal{T}_{P_X}^n$ and $\mathcal{T}_{XY}^n := \mathcal{T}_{P_{XY}}^n$, for RV's X and Y .

Let $Q \in \mathcal{P}(n, \mathcal{Y} \times \mathcal{Z})$ have a 1–dimensional marginal distribution P_{y^n} . We define a set of sequences Q–generated by y^n

$$G_Q(y^n) := \big\{z^n : (y^n, z^n) \in \mathcal{T}_Q^n\big\}. \tag{2.21}$$

We shall use the facts

$$|\mathcal{P}(n, \mathcal{Z})| \leq (n+1)^{|\mathcal{Z}|}, \tag{2.22}$$

and for $Q \in \mathcal{P}(n, \mathcal{Y} \times \mathcal{Z})$, $P \in \mathcal{P}(n, \mathcal{Y})$, $Q = P \times W$, and $y^n \in \mathcal{T}_P^n$

$$(n+1)^{-|\mathcal{Y}||\mathcal{Z}|} \exp_2\big\{nH(W|P)\big\} \leq |G_Q(y^n)| \leq \exp\big\{nH(W|P)\big\}. \tag{2.23}$$

Let now for $\delta \geq 0$

$$G_{Q,\delta}(y^n) = \bigcup_{Q':\|Q'-Q\|\leq n\delta} G_{Q'}(y^n), \tag{2.24}$$

where the operation " $\|\ \|$ " denotes the total variation and let for $P \in \mathcal{P}(n, \mathcal{Y})$ and $Q_1, Q_2 \in \mathcal{P}(n, \mathcal{Y} \times \mathcal{Z})$ with $Q_i(y,z) = P(y)W_i(z|y) \quad (i = 1,2)$ the distribution Q be defined by

$$Q(y,z,z') = P(y)W_1(z|y)W_2(z'|y) \text{ for } y \in \mathcal{Y}, z, z' \in \mathcal{Z},$$

then

$$|G_{Q,\delta}(y^n) \cap \big(G_{Q_1}(y^n) \times G_{Q_2}(y^n)\big)| \geq |G_{Q_1}(y^n)|\,|G_{Q_2}(y^n)|(1+o(1)) \tag{2.25}$$

(as $n \to \infty$).

Moreover for all $\Psi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, $P_{XYU} \in \mathcal{P}(n, \mathcal{X} \times \mathcal{Y} \times \mathcal{U})$, $u^n \in \mathcal{T}_U^n$, and $(x^n, y^n) \in G_{P_{XYU}}(u^n)$,

$$\frac{1}{n}\sum_{t=1}^{n} \Psi(x_t, y_t) = \mathbb{E}\Psi(X,Y). \tag{2.26}$$

In fact, u^n has no direct relation with (2.26), that is, for all $(x^n, y^n) \in \mathcal{T}_{XY}^n$, (2.26) holds.

An essential ingredient of the Storage Capacity Theorem of [AZ89] for codes WEM is a combinatorial result. It is used in this paper for the proofs of the direct parts of Theorems 2 and 4, that is, whenever we deal with Rule II. We say that the hypergraph $(\Omega, \mathcal{E})$ carries M colors if there is a vertex coloring with M colors such that all these colors occur in every edge.

Coloring Lemma. *([AZ89])*

The hypergraph $(\Omega, \mathcal{E})$ *carries* M *colors if* $M \leq (\ell n|\mathcal{E}| \min_{E\in\mathcal{E}}|E|)^{-1} \min_{E\in\mathcal{E}} |E|$.

As mentioned above, in information theory a quantity often is characterized as an extremal value of an information quantity over a region of PD's on a finite set and a region is characterized in a similar way by a group of inequalities for information quantities. By the continuity and differentiability properties of information quantities, they are, in principle, computable by standard analytical methods.

The following result of Ahlswede and Körner plays a very important role in reducing an incomputable quantity (region) to a computable one in multi–user information theory. It is used in our converse proofs of Theorems 1, 2, and 3.

Support Lemma. *(Lemma 3 of [AKö75])*

Let $f_j (j = 1, \ldots, k) : \mathcal{P}(\mathcal{Z}) \to \mathbb{R}$ *be continuous functions. Then to any PD* μ *on the Borel* σ*–algebra of* $\mathcal{P}(\mathcal{Z})$ *there exist* k *elements* P_i *of* $\mathcal{P}(\mathcal{Z})$ *and non–negative numbers* $\alpha_1, \ldots, \alpha_k$ *with* $\sum_{i=1}^{k} \alpha_i = 1$ *such that for every* $j = 1, \ldots, k$

$$\int_{\mathcal{P}(\mathcal{Z})} f_j(P)\mu(dP) = \sum_{i=1}^{k} \alpha_i f_j(P_i). \tag{2.27}$$

Proof: The map $f = (f_1, \ldots, f_k) : \mathcal{P}(\mathcal{Z}) \to \mathbb{R}^k$ is continuous and since $\mathcal{P}(\mathcal{Z})$ is compact and connected so is the image $J = f\big(\mathcal{P}(\mathcal{Z})\big)$.

Clearly, the point $\left(\int_{\mathcal{P}(\mathcal{Z})} f_1(P)\mu(dP), \ldots, \int_{\mathcal{P}(\mathcal{Z})} f_k\mu(dP)\right)$ belongs to the convex closure of J , and thus by the Eggleston–Carathéodory theorem (cf. [E58], Theorem 18) there are k points in J , say, $f(P_1), \ldots, f(P_k)$, satisfying (2.27).

Remarks:

1.) Originally, in [AKö75], Carathéodory's theorem was used, which does not require connectedness and gives the weaker conclusion that $k+1$ instead of k points are needed.

2.) Notice that in the proof above only compactness and connectedness of $\mathcal{P}(\mathcal{Z})$ was used. Therefore $\mathcal{P}(\mathcal{Z})$ can be replaced by any set A with these topological properties. In particular, for finite sets $\mathcal{X}_1, \ldots, \mathcal{X}_L$ the set of product distributions $\mathcal{P}(\mathcal{X}_1) \times \mathcal{P}(\mathcal{X}_2) \times \cdots \times \mathcal{P}(\mathcal{X}_L)$ could serve as A .

3. A smallest rich world

3.1 On the sizes of the spreads in a WEM code.

Let $\{C_m\}_{m=1}^M$ be a WEM code with length n under the Criterion Max. We assume that each $x^n \in \Omega := \bigcup\limits_{m=1}^{M} C_m$ may appear on the memory, because otherwise we can simply delete it from the spread to which it belongs. Let $x^n \in C_m$, then for all $m' \neq m$ there must be a $y^n \in C_{m'}$ with $\varphi_n(x^n, y^n) \leq D_{\max}$ so that one can update m to m' under the Criterion Max and the constraint $D_{\max}$. Thus for all $x^n \in \Omega$

$$|\{y^n : y^n \in \Omega,\ \frac{1}{n}\varphi_n(x^n, y^n) \leq d\}| \geq M, \tag{3.1}$$

where we write $d := \frac{1}{n}D_{\max}$.

We say that $\Omega \subset \mathcal{X}^n$ satisfying (3.1) is a rich world, because each member of the world has "enough neighbours" (in the sense that their φ_n–distance is not too large). To keep the world rich, the size of the "world" cannot be too small.

On the other hand, since the spreads are pairwise disjoint, $|\Omega| = \sum\limits_{m=1}^{M} |C_m|$ and therefore $M^{-1}|\Omega| \leq \max\limits_m |C_m|$. That is, the restriction on the sizes of the spreads requires that the world cannot be too large. For this reason, to find the smallest size of "a rich world" is a first step towards answering Question 3 in the Introduction.

Out of mathematical interest we formulate the "rich world"–problem slightly more general. On the other hand we assume that $\varphi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is symmetric, that is $\varphi(x, y) = \varphi(y, x)$ for all $x, y \in \mathcal{X}$. Define

$$\alpha = \min_{x,y\in\mathcal{X}} \varphi(x, y) \quad\text{and}\quad \beta = \max_{x,y\in\mathcal{X}} \varphi(x, y). \tag{3.2}$$

Now for any closed interval $\mathcal{L} \subset [\alpha, \beta]$, any positive integer n , and any $S \subset \mathcal{X}^n$ we define

$$B(x^n, \mathcal{L}, S) := \left\{ y^n \in S : \frac{1}{n} \varphi_n(x^n, y^n) \in \mathcal{L} \right\}. \tag{3.3}$$

(In the case $\alpha = 0$ and $\mathcal{L} = [0, \beta]$ it is the intersection of S with a ball of center x^n and φ–radius β . This is appropriate for WEM, if $\varphi(x, y) = \begin{cases} 0, & \text{if } x = y \\ > 0, & \text{if } x \neq y. \end{cases}$)

We call S $(n, \mathcal{L}, \rho)$–good for any positive number ρ , if

$$|B(x^n, \mathcal{L}, S)| \geq 2^{n\rho} \text{ for all } x^n \in S. \tag{3.4}$$

This says that every point in S has $2^{n\rho}$ points of S in its neighbourhood. In this sense S is a "rich world". Denote by $N(n, \mathcal{L}, \rho)$ the *smallest* cardinality of $(n, \mathcal{L}, \rho)$–good sets. This definition catches the goal to make the "world small".

Since for an $(n_1, \mathcal{L}, \rho)$–good S_1 and an $(n_2, \mathcal{L}, \rho)$–good S_2 the cartesian product $S_1 \times S_2$ is $(n_1 + n_2, \mathcal{L}, \rho)$–good , we have $N(n_1 + n_2, \mathcal{L}\rho) \leq N(n_1, \mathcal{L}, \rho) \cdot N(n_2, \mathcal{L}, \rho)$ and therefore $\lim\limits_{n \to \infty} \frac{1}{n} \log N(n, \mathcal{L}, \rho)$ exists. We denote the limit by $\sigma(\mathcal{L}, \rho)$.

3.2 The main result of this Section.

The characterisation of $\sigma(\mathcal{L}, \rho)$ requires a few concepts.

Let (U, X, Y) be a triple of RV's with values in $\mathcal{U} \times \mathcal{X} \times \mathcal{X}$ for a finite set $\mathcal{U}$. We say that (X, Y) is matched through U , if

$$H(X|U) = H(Y|U) \text{ and } H(Y|X, U) = H(X|Y, U). \tag{3.5}$$

We set

$$\begin{aligned} \mathcal{Q}(\mathcal{L}, \rho) = \big\{ (X, U) : \quad & \text{for some } Y \text{ the pair } (X, Y) \\ & \text{is matched through } U, \mathbb{E}\varphi(X, Y) \in \mathcal{L}, \text{ and } H(Y|X, U) \geq \rho \big\}. \end{aligned} \tag{3.6}$$

Theorem 1. *For symmetric* $\varphi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

$$\sigma(\mathcal{L}, \rho) = \min_{(X,U) \in \mathcal{Q}(\mathcal{L}, \rho)} H(X|U). \tag{3.7}$$

Actually, we can bound the cardinality of $\mathcal{U}$ *by* $2|\mathcal{X}| + 2$.

Furthermore, we can limit the distributions P_{XY} *to those with equal marginals.*

Remarks:

3.) The structure of the characterisation for $\sigma(\mathcal{L}, \rho)$ is typical in multi–user information theory. Here U is an auxiliary RV with values in an arbitrary finite set $\mathcal{U}$. By bounding its cardinality the formula describes an optimisation of a continuous function over a compact set in an euclidean space. By considering ε–nets arbitrarily good approximation is principally possible. Of course, the complexity of this task decreases with $|\mathcal{U}|$.

4.) The direct part of the proof (the proof of the lower bound) is constructive and based on typical and generated sequences (see Section 2).

5.) Symmetry of φ is essential, because otherwise there may not exist $(n, \mathcal{L}, \rho)$–good sets. A simple and extremal example is related to the Write–Once–Memory (WOM) model of [RSh82]. Choose $\mathcal{X} = \{0, 1\}$, $\varphi(1, 0) = 1$, $\varphi(x, y) = 0$ for $(x, y) \neq (1, 0)$, and $\mathcal{L} = \{0\}$. Since for all $S \subset \{0, 1\}^n$ the element $x^n \in S$ with a maximal number of 1's has no neighbour y^n with $\frac{1}{n}\varphi_n(x^n, y^n) \in \mathcal{L}$, S cannot be $(n, \mathcal{L}, \rho)$–good for any n and ρ.

3.3 The Proof of Theorem 1.

Converse part:

Let S be $(n, \mathcal{L}, \rho)$–good. We introduce the set

$$B(n, \mathcal{L}, S) := \bigcup_{x^n \in S} \{x^n\} \times \big\{B(x^n, \mathcal{L}, S)\big\} \tag{3.8}$$

and the RV's $\hat{X}^n = (\hat{X}_1, \dots, \hat{X}_n)$, $\hat{Y}^n = (\hat{Y}_1, \dots, \hat{Y}_n)$ with the joint distribution

$$\Pr(\hat{X}^n = x^n, \hat{Y}^n = y^n) := \begin{cases} \frac{1}{|B(n,\mathcal{L},S)|} & \text{for } (x^n, y^n) \in B(n, \mathcal{L}, S) \\ 0 & \text{otherwise.} \end{cases} \tag{3.9}$$

By (3.3), (3.8), and the symmetry of φ

$$(x^n, y^n) \in B(n, \mathcal{L}, S) \text{ exactly if } (y^n, x^n) \in B(n, \mathcal{L}, S). \tag{3.10}$$

Therefore,

$$\Pr(\hat{X}^n = x^n) = \Pr(\hat{Y}^n = x^n) = \begin{cases} \frac{|B(x^n,\mathcal{L},S)|}{|B(n,\mathcal{L},S)|} & \text{for } x^n \in S \\ 0 & \text{otherwise.} \end{cases} \tag{3.11}$$

By property 2.) in Section 2, (2.3), (2.13), and (2.15), we have,

$$\begin{aligned} \log|S| \geq H(\hat{X}^n) &= \sum_{t=1}^{n} H(\hat{X}_t|\hat{X}^{t-1}) \\ &\geq \sum_{t=1}^{n} H(\hat{X}_t|\hat{X}^{t-1}, \hat{Y}^{t-1}) = \sum_{t=1}^{n} H(\hat{X}_t|V^{t-1}), \text{ if } V^{t-1} := (\hat{X}^{t-1}, \hat{Y}^{t-1}). \end{aligned} \tag{3.12}$$

Moreover, by (3.9) and (3.10), $P_{\hat{X}^n,\hat{Y}^n}$ is symmetric, namely,

$$P_{\hat{X}^n,\hat{Y}^n}(x^n,y^n) = P_{\hat{X}^n,\hat{Y}^n}(y^n,x^n). \tag{3.13}$$

The desired matching properties of the auxiliary variables in the set $\mathcal{Q}(\mathcal{L},\rho)$ shall be shown now to be a consequence of symmetry properties of the distribution $P_{\hat{X}^n,\hat{Y}^n}$. As a space saving notation we set $v = (x^{t-1}, y^{t-1})$ and $\overline{v} = (y^{t-1}, x^{t-1})$ (here t is understood by context).

Since $\Pr(\hat{X}^{t-1} = x^{t-1}, \hat{Y}^{t-1} = y^{t-1}) = \Pr(\hat{X}^{t-1} = y^{t-1}, \hat{Y}^{t-1} = x^{t-1})$, we have

$$\Pr(V^{t-1} = v) = Pr(V^{t-1} = \overline{v}). \tag{3.14}$$

Since $\Pr(\hat{X}_t = x_t, \hat{Y}_t = y_t | \hat{X}^{t-1} = x^{t-1}, \hat{Y}^{t-1} = y^{t-1}) = \Pr(\hat{X}_t = y_t, \hat{Y}_t = x_t | \hat{X}^{t-1} = y^{t-1}, \hat{Y}^{t-1} = x^{t-1})$, we also have

$$\Pr(\hat{X}_t = x_t, \hat{Y}_t = y_t | V^{t-1} = v) = \Pr(\hat{X}_t = y_t, \hat{Y}_t = x_t | V^{t-1} = \overline{v}). \tag{3.15}$$

Therefore

$$\Pr(\hat{X}_t = x_t | V^{t-1} = v) = \sum_{y_t} \Pr(\hat{X}_t = x_t, \hat{Y}_t = y_t | V^{t-1} = v)$$

$$= \sum_{y_t} Pr(\hat{X}_t = y_t, \hat{Y}_t = x_t | V^{t-1} = \overline{v}) = \Pr(\hat{Y}_t = x_t | V^{t-1} = \overline{v}) \tag{3.16}$$

and by combination of (3.15) and (3.16)

$$\Pr(\hat{X}_t = y_t | \hat{Y}_t = x_t, V^{t-1} = \overline{v}) = \Pr(\hat{Y}_t = y_t | \hat{X}_t = x_t, V^{t-1} = v). \tag{3.17}$$

(It is understood that for $t = 1$ the distributions are unconditional.)

We remember as a rule: Exchanging $\hat{X}_t$ and $\hat{Y}_t$ is permitted, if simultaneously we exchange v and $\overline{v}$.

Now comes the harvest.

$$H(\hat{Y}_t | V^{t-1}) = \sum_v \Pr(V^{t-1} = \overline{v}) H(\hat{Y}_t | V^{t-1} = \overline{v}) \quad \text{(by (2.8))}$$

$$= \sum_v \Pr(V^{t-1} = v) H(\hat{Y}_t | V^{t-1} = \overline{v}) \quad \text{(by (3.14))}$$

$$= \sum_v \Pr(V^{t-1} = v) H(\hat{X}_t | V^{t-1} = v) \quad \text{(by (3.16))} = H(\hat{X}_t | V^{t-1}) \quad \text{(by (2.8))}.$$

and thus

$$H(\hat{Y}_t | V^{t-1}) = H(\hat{X}_t | V^{t-1}). \tag{3.18}$$

Similarly,

$$
\begin{aligned}
H(\hat{Y}_t|\hat{X}_t, V^{t-1}) &= \sum_v \Pr(V^{t-1}=v)H(\hat{Y}_t|\hat{X}_t, V^{t-1}=v) \quad (\text{by } (2.9)) \\
&= \sum_v \Pr(V^{t-1}=\overline{v})H(\hat{Y}_t|\hat{X}_t, V^{t-1}=v) \quad (\text{by } (3.14)) \\
&= \sum_v \Pr(V^{t-1}=\overline{v})H(\hat{X}_t|\hat{Y}_t, V^{t-1}=\overline{v}) \quad (\text{by } (3.17)) \\
&= H(\hat{X}_t|\hat{Y}_t, V^{t-1}) \quad (\text{by } (2.9)) \text{ and thus}
\end{aligned}
$$

$$
H(\hat{Y}_t|\hat{X}_t, V^{t-1}) = H(\hat{X}_t|\hat{Y}_t, V^{t-1}). \tag{3.19}
$$

Now we use a standard technique in multi–user information theory (see [AKö75]). Let T be a RV uniformly distributed over $\{1,2,\dots,n\}$ and independent of $(\hat{X}^n, \hat{Y}^n)$. We choose

$$
(X, Y, \tilde{U}) = \big(\hat{X}_T, \hat{Y}_T, (T, V^{T-1})\big) \tag{3.20}
$$

and notice that

$$
\begin{aligned}
&H(X|\tilde{U}) = H(\hat{X}_T|T, V^{T-1}) \\
&= \sum_{t=1}^n \Pr(T=t)H(\hat{X}_T|T=t, V^{T-1}) \quad (\text{by } (2.9)) \\
&= \frac{1}{n}\sum_{t=1}^n H(\hat{X}_t|V^{t-1}) \ (\text{by } \Pr(\hat{X}_t=x|T=t, V^{t-1}=v^{t-1}) = \Pr(\hat{X}_t=x|V^{t-1}=v^{t-1})) \\
&= \frac{1}{n}\sum_{t=1}^n H(\hat{Y}_t|V^{t-1}) \quad (\text{by } (3.18)) \\
&= H(Y|\tilde{U}).
\end{aligned} \tag{3.21}
$$

Similarly, by replacing the roles of V^{t-1} in (3.21) with $(\hat{Y}_t, V^{t-1})$,

$$
\begin{aligned}
H(X|Y,\tilde{U}) &= \frac{1}{n}\sum_{t=1}^n H(\hat{X}_t|\hat{Y}_t, V^{t-1}) \\
&= \frac{1}{n}\sum_{t=1}^n H(\hat{Y}_t|\hat{X}_t, V^{t-1}) \quad (\text{by } (3.19)) \\
&= H(Y|X,\tilde{U}).
\end{aligned} \tag{3.22}
$$

We have seen that (X,Y) is matched through $\tilde{U}$ and that by (3.12) and (3.20)

$$
\log|S| \geq n\ H(X|\tilde{U}). \tag{3.23}
$$

To complete the proof of the converse we have to show that $(X,\tilde{U}) \in \mathcal{Q}(\mathcal{L},\rho)$. By (3.9)

$$\mathbb{E}\ \varphi(X,Y) = \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\ \varphi(\hat{X}_t,\hat{Y}_t) \in \mathcal{L} \tag{3.24}$$

as consequence of the definition of S (based even on worst case constraint!). Now by the definition of an $(n,\mathcal{L},\rho)$–good set we have $n\rho \leq \log|B(x^n,\mathcal{L},S)|$ for $x^n \in S$ and therefore

$$n\rho \leq \sum_{x^n} \Pr(\hat{X}^n = x^n)\log|B(x^n,\mathcal{L},S)| \ = \sum_{x^n} Pr(\hat{X}^n = x^n) H(\hat{Y}^n|\hat{X}^n = x^n),$$

because by (3.9) and (3.11) $\Pr(\hat{Y}^n = y^n|\hat{X}^n = x^n) = \frac{1}{|B(x^n,\mathcal{L},S)|}$ for $y^n \in B(x^n,\mathcal{L},S)$, which with (2.18) implies that for all x^n $H(\hat{Y}^n|\hat{X}^n = x^n) = \log|B(x^n,\mathcal{L},S)|$. Now

$$\begin{aligned} n\rho \leq H(\hat{Y}^n|\hat{X}^n) &= \sum_{t=1}^{n} H(\hat{Y}_t|\hat{X}^n,\hat{Y}^{t-1}) \quad \text{(by (2.16))} \\ &\leq \sum_{t=1}^{n} H(\hat{Y}_t|\hat{X}_t,\hat{X}^{t-1},\hat{Y}^{t-1}) \quad \text{(by (2.13))} \\ &= \sum_{t=1}^{n} H(\hat{Y}_t|\hat{X}_t,V^{t-1}) = n\ H(Y|X\tilde{U}), \end{aligned}$$

as was to be shown.

The application of the Support Lemma to bound the cardinality of the range of U is as originally in [AKö75]. It will be done in subsection 3.4.

Direct part:

Since the empirical distributions are dense in $\mathcal{P}(\mathcal{X}\times\mathcal{X}\times\mathcal{U})$, we can consider distributions $P_{\overline{X},\overline{Y},\overline{U}} \in \mathcal{P}(n,\mathcal{X}\times\mathcal{X}\times\mathcal{U})$ with $\mathbb{E}_\varphi(\overline{X},\overline{Y}) \in \mathcal{L}$ and

$$P_{\overline{X},\overline{Y},\overline{U}} \sim P_{XYU},\ (X,U) \in \mathcal{Q}(\mathcal{L},\rho).$$

Here and throughout this paper we write $A \sim B$, if $A = B(1+o(1))$ and analogously $A \lesssim B$ means that $A \leq B(1+o(1))$.

We fix (any) $u^n \in \mathcal{T}^n_{\overline{U}}$, define the generated sets (see (2.21)) $G_{P_{\overline{X},\overline{U}}}(u^n)$ and $G_{P_{\overline{Y},\overline{U}}}(u^n)$ and choose

$$S = G_{P_{\overline{X},\overline{U}}}(u^n) \cup G_{P_{\overline{Y},\overline{U}}}(u^n). \tag{3.25}$$

By (2.23) we have, since $H(X|U) = H(Y|U)$ and $H(\overline{X}|\overline{U}) \sim H(\overline{Y}|\overline{U})$,

$$\frac{1}{n}\log|S| \lesssim H(X|U). \tag{3.26}$$

Furthermore, for $x^n \in G_{P_{\overline{X},\overline{U}}}(u^n)$, since by (2.21), (2.26), and (3.3), for all $y^n \in G_{P_{\overline{X},\overline{Y},\overline{U}}}(x^n, y^n)$,

$$\varphi_n(x^n, y^n) = \sum_{t=1}^{n} \varphi(x_t, y_t) = n\mathbb{E}\ \varphi(\overline{X}, \overline{Y}) \in \mathcal{L}\ ,$$

$$\log|B(x^n, \mathcal{L}, S)| \geq \log|G_{P_{\overline{X}\,\overline{Y}\,\overline{U}}}(x^n, u^n)| \sim n\ H(Y|XU) \quad \text{(by (2.23))} \ \geq n\rho$$

and, symmetrically, for $y^n \in G_{P_{\overline{X}\,\overline{U}}}(u^n)$,

$$\log|B(y^n, \mathcal{L}, S)| \geq \log|G_{P_{\overline{X}\,\overline{Y}\,\overline{U}}}(y^n, u^n)| \sim n\ H(X|Y, U) = n\ H(Y|XU) \geq n\rho.$$

3.4 Bounding the range of the auxiliary random variable by application of the Support Lemma.

The Support Lemma [AKö75] (see Section 2) has been widely used in multi–user information theory. With it one can get also a simpler proof of the result in [AAl94]. It is also used in Section 5 of the present paper. However it is not familiar to most scientists working in other areas. Therefore we feel that it is necessary to explain its application in the converse proof of Theorem 1 in a separate subsection. Readers, who are familiar with it or not interested in it, may skip this subsection.

Recall that for any $(n, \mathcal{L}, \rho)$ –good set S we are given a triple $(X, Y, \tilde{U})$ of RV's with values in $\mathcal{X} \times \mathcal{Y} \times \tilde{\mathcal{U}}$, satisfying (3.21) – (3.24). Notice that there $\tilde{\mathcal{U}}$ is a finite but not (uniformly) bounded set and its size may increase with n . Thus $\min H(X|\tilde{U})$ is not computable. Our task here is to bound it and thus reduce it to a computable quantity.

Since by (3.20) $\Pr(X = x) = \frac{1}{n} \sum_{t=1}^{n} \Pr(\hat{X}_t = x)$ and $\Pr(Y = y) = \frac{1}{n} \sum_{t=1}^{n} \Pr(\hat{Y} = y)$, (3.11) implies

$$P_X = P_Y. \tag{3.27}$$

This is the additional requirement on the marginals in Theorem 1. Notice here P_X and P_Y depend on $\tilde{U}$ and so we have to show that (3.27) keeps unchanged when we replace $\tilde{U}$ by a suitable new random variable U with range bounded by $2|\mathcal{X}| + 2$.

Now we apply the Support Lemma in Section 2 to the set of PD's $\mathcal{P}(\mathcal{X} \times \mathcal{X})$, where the measure μ is given by

$$\mu(P) := \begin{cases} \Pr(\tilde{U} = u), & \text{if } \ P = P_{XY|\tilde{U}}(\cdot|u) \\ 0, & \text{otherwise.} \end{cases} \tag{3.28}$$

The continuous functions f_j in the lemma are defined as follows.

For all $P \in \mathcal{P}(\mathcal{X} \times \mathcal{X})$,

$$f_1(P) := \sum_{(x,y)\in\mathcal{X}\times\mathcal{X}} P(x,y)\varphi(x,y), \tag{3.29}$$

$$f_2(P) := H(P_1), \text{ where } P_1 \text{ is the marginal of } P \text{ for the first component}, \tag{3.30}$$

$$f_3(P) := H(P_2), \text{ where } P_2 \text{ is the marginal of } P \text{ for the second component}, \tag{3.31}$$

and

$$f_4(P) := H(P). \tag{3.32}$$

Moreover, assume $\mathcal{X} = \{0, 1, \ldots, |\mathcal{X}| - 1\}$ and P_1 , P_2 are as in (3.30), (3.31). Then

$$f_{x+4}(P) := P_1(x) \quad \text{for} \quad x \in \mathcal{X} \setminus \{0\}, \tag{3.33}$$

and

$$f_{x+|\mathcal{X}|+3}(P) := P_2(x) \quad \text{for} \quad x \in \mathcal{X} \setminus \{0\}. \tag{3.34}$$

Applying the Support Lemma to f_j , $1 \le j \le 2|\mathcal{X}|+2$, we are guaranteed the existence of non–negative α_i and $P^{(i)} \in \mathcal{P}(\mathcal{X} \times \mathcal{X})$, $1 \le i \le 2|\mathcal{X}| + 2$, with $\sum\limits_{i=1}^{2|\mathcal{X}|+2} \alpha_i = 1$ and such that

$$\begin{aligned}
\mathbb{E}\varphi(X,Y) &= \mathbb{E}\big(\mathbb{E}(\varphi(X,Y)|\tilde{U})\big) = \sum_u \Pr(\tilde{U}=u)\left(\sum_{x,y} P_{XY|\tilde{U}}(x,y|u)\varphi(x,y)\right)\\
&= \int f_1(P)\mu(dP) \quad (\text{by (3.28) and (3.29)})\\
&= \sum_{i=1}^{2|\mathcal{X}|+2} \alpha_i f_1(P^{(i)}) = \sum_{i=1}^{2|\mathcal{X}|+2} \alpha_i \left(\sum_{x,y} P^{(i)}(x,y)\varphi(x,y)\right) \quad (\text{by (3.29)}),
\end{aligned} \tag{3.35}$$

$$\begin{aligned}
H(X|\tilde{U}) &= \sum_u \Pr(\tilde{U}=u) H\big(P_{X|\tilde{U}}(\cdot|u)\big) \quad (\text{by (2.8)})\\
&= \int f_2(P)\mu(dP) \quad (\text{by (3.28) and (3.30)})\\
&= \sum_{i=1}^{2|\mathcal{X}|+2} \alpha_i f_2(P^{(i)}) = \sum_{i=1}^{2|\mathcal{X}|+2} \alpha_i H(P_1^{(i)}) \quad (\text{by (3.30)}),
\end{aligned} \tag{3.36}$$

$$
\begin{aligned}
H(Y|\tilde{U}) &= \sum_u \Pr(\tilde{U}=u) H\big(P_{Y|\tilde{U}}(\cdot|u)\big) \\
&= \int f_3(P)\mu(dP) \quad \text{(by (3.28) and (3.31))} \\
&= \sum_{i=1}^{2|\mathcal{X}|+2} \alpha_i f_3(P^{(i)}) = \sum_{i=1}^{2|\mathcal{X}|+2} \alpha_i H(P_2^{(i)}) \quad \text{(by (3.31))} \qquad (3.37)
\end{aligned}
$$

$$
\begin{aligned}
H(XY|\tilde{U}) &= \sum_u \Pr(\tilde{U}=u) H\big(P_{XY|\tilde{U}}(\cdot|u)\big) \\
&= \int f_4(P)\mu(dP) \quad \text{(by (3.28) and (3.32))} \\
&= \sum_{i=1}^{2|\mathcal{X}|+2} \alpha_i f_4(P^{(i)}) = \sum_{i=1}^{2|\mathcal{X}|+2} \alpha_i H(P^{(i)}) \quad \text{(by (3.32))}, \qquad (3.38)
\end{aligned}
$$

$$
\begin{aligned}
P_X(x) &= \sum_u \left(\sum_{y\in\mathcal{X}} P_{XY|\tilde{U}}(x,y|u) \right) \Pr(\tilde{U}=u) = \int f_{x+4}(P)\mu(dP) \text{ (by (3.28) and (3.33))} \\
&= \sum_{i=1}^{2|\mathcal{X}|+2} \alpha_i f_{x+4}(P^{(i)}) = \sum_{i=1}^{2|\mathcal{X}|+2} \alpha_i \left(\sum_{y\in\mathcal{X}} P^{(i)}(x,y) \right) \text{ (by (3.33)), for } x \in \mathcal{X} \setminus \{0\}, \qquad (3.39)
\end{aligned}
$$

and

$$
\begin{aligned}
P_Y(y) &= \sum_u \left(\sum_{x\in\mathcal{X}} P_{XY|\tilde{U}}(x,y|u) \right) \Pr(\tilde{U}=u) = \int f_{y+|\mathcal{X}|+3}(P)\mu(dP) \text{ (by (3.28) and (3.34))} \\
&= \sum_{i=1}^{2|\mathcal{X}|+2} \alpha_i f_{y+|\mathcal{X}|+3}(P^{(i)}) = \sum_{i=1}^{2|\mathcal{X}|+2} \alpha_i \left(\sum_{x\in\mathcal{X}} P^{(i)}(x,y) \right) \text{ (by (3.34)), for } y \in \mathcal{X} \setminus \{0\}. \qquad (3.40)
\end{aligned}
$$

Now we let $\mathcal{U} = \{1, 2, \dots, 2|\mathcal{X}|+2\}$ and (X, Y, U) be the RV's with distribution $P_{XYU}(x,y,u) = \alpha_u P^{(u)}(x,y)$.

Then (3.35) shows that $\mathbb{E}\varphi(XY)$ is unchanged, (3.39) and (3.40) guarantee that $P_X(x)$, $x \neq 0$, and $P_Y(y)$, $y \neq 0$, (and therefore P_X and P_Y) are unchanged, if we replace $\tilde{U}$ by U . (3.36) – (3.38) say that $H(X|\tilde{U}) = H(X|U)$, $H(Y|\tilde{U}) = H(Y|U)$ and $H(X,Y|\tilde{U}) = H(X,Y|U)$. Thus, if we replace $(X,Y,\tilde{U})$ by (X,Y,U) , (3.21) – (3.24) and (3.27) still hold. This completes our proof.

Finally we emphasize that only the bound $2|\mathcal{X}|+2$ was obtained, because we insist on (3.27). When we give it up, only the functions f_i , $i = 1, \dots, 4$, are needed. Consequently, we then get the much better bound 4 for $|\mathcal{U}|$.

4. WEM WITH BOUNDED SIZES OF SPREADS

4.1 The definition of the achievable region for WEM with bounded sizes of spreads.

In the Introduction we have mentioned the model of an n–length WEM code with the Criterion Max (without restriction to sizes of the spreads).

Next, we turn to a new model, in which we require that the size of the spreads C_m, $m = 1, \ldots, M$ is bounded, i.e., for all m,

$$|C_m| \leq K. \tag{4.1}$$

We call spreads $\{C_m\}_{m=1}^M$, which satisfy (1.4) and (4.1), an (n, M, D, K) WEM–code.

The achievable region $\mathcal{R}(d)$ for these codes is a set of pairs non–negative reals (R, κ) such that for all $\varepsilon > 0$, when n is large enough (depending on ε), there is an (n, M, D, K) WEM–code with $\frac{1}{n}\log M \geq R - \varepsilon$, $\frac{1}{n}\log K \leq \kappa + \varepsilon$, and $D = nd$.

4.2 A storage capacity theorem under the constraint of spreads.

To describe our result, we define $\mathcal{R}^*(d)$ as set of pairs $(R, \kappa) \in \mathbb{R}^+ \times \mathbb{R}^+$ for which a triple of RV's (X, Y, U) with values in $\mathcal{X} \times \mathcal{X} \times \mathcal{U}$ exists such that

$$(X, Y) \text{ is matched through } U \tag{4.2}$$

$$\mathbb{E}\varphi(X, Y) \leq d \tag{4.3}$$

$$R \leq H(Y|XU) \tag{4.4}$$

$$R + \kappa \geq H(X|U). \tag{4.5}$$

Theorem 2. *(Storage capacity under spreads constraint)*

For the WEM with symmetric sum–type cost function we have for any $d > 0$

$$\mathcal{R}(d) = \mathcal{R}^*(d). \tag{4.6}$$

The auxiliary RV U *in the description of* $\mathcal{R}^*(d)$ *needs to take at most* $2|\mathcal{X}| + 2$ *values, if we insist upon the condition* $P_X = P_Y$. *Otherwise, 4 values for* U *suffice.*

To prove Theorem 2, we need Theorem 1 and the Coloring Lemma in Section 2.

Remark 6:

Theorem 2 characterizes the achievable region for the Criterion Max, but from the following proof one can see that the achievable region will not change, if one changes the Criterion Max to Criterion Ave. We argue as follows. Since a WEM code satisfying the Criterion Max must satisfy the Criterion Ave (with the same parameters), the achievable region for the Criterion Max is contained in the achievable region for the Criterion Ave. On the other hand, to prove the converse part, we just have to observe that (3.24) uses only the average number of neighbours in $\mathcal{L}$ and therefore Theorem 1 has an "analogue for averages" and that this is the only place where the criterions matter in the proof of Theorem 2.

4.3 The proof of Theorem 2.

Converse part:

We show first that

$$\mathcal{R}(d) \subset \mathcal{R}^*(d). \tag{4.7}$$

For an (n, M, D, K) code $\mathcal{C} = \{C_m\}_{m=1}^M$ set $S = \bigcup_{m=1}^{M} C_m$. Then $N = |S| \leq M \cdot K$ and as we explained in Subsection 3.1 S is $(n, \mathcal{L}_d, R)$–good with $\mathcal{L}_d = [0, d]$, $d = \frac{1}{n} D$, and $R = \frac{1}{n} \log M$.

By Theorem 1

$$\frac{1}{n} \log N \geq \min_{(X,U) \in \mathcal{Q}(\mathcal{L}_d, R)} H(X|U) \tag{4.8}$$

and therefore for $\kappa = \frac{1}{n} \log K$

$$R + \kappa = \frac{1}{n} \log M + \frac{1}{n} \log K \geq \min_{(X,U) \in \mathcal{Q}(\mathcal{L}_d, R)} H(X|U).$$

Let (X, Y, U) assume this minimum. Then, by (3.6), (4.2) – (4.4) hold.

Direct part:

By Theorem 1 we can construct for R and (X, Y, U) in (4.2) – (4.5) an $(n, \mathcal{L}_d, R)$–good subset S in $\mathcal{X}^n$ with

$$\frac{1}{n} \log |S| \sim H(X|U). \tag{4.9}$$

Consider now the hypergraph $\big(S, \big(B(x^n, \mathcal{L}_d, S)\big)_{x^n \in S}\big)$ (see (3.3)) and apply the Coloring Lemma to

$$M = \left\lceil (\ell n |S| \min_{x^n \in S} |B(x^n, \mathcal{L}_d, S)|)^{-1} \min_{x^n \in S} |B(x^n, \mathcal{L}_d, S)| \right\rceil \geq (\ell n |S| \min_{x^n \in S} |B(x^n, \mathcal{L}_d, S)|)^{-1} 2^{nR} .$$

This results in an (n, M, d) WEM code $\{C_m\}_{m=1}^M$ with $\bigcup_{m=1}^{M} C_m = S$ and

$$\frac{1}{n} \log M \sim R, \tag{4.10}$$

because $\frac{1}{n} \log(\ell n |S| \min_{x^n} |B(x^n \mathcal{L}_d, S)|) \to 0$ as $n \to \infty$.

The average spread size is

$$\frac{1}{M} \sum_{i=1}^{M} |C_i| = \frac{1}{M} |S|. \tag{4.11}$$

By (2.19), at least half of the spreads, say $C_m, 1 \leq m \leq \lfloor \frac{M}{2} \rfloor$, have a cardinality of at most $2 \cdot \frac{1}{M} |S|$, or in rate by (4.5), (4.9), and (4.10),

$$\frac{1}{n} \log \frac{2}{M} |S| \sim H(X|U) - R \leq \kappa.$$

5.1 A generalization of average diametric problems.

Here we show how a model of multi–user WEM leads to general diametrical problem in the average, which generalizes that of [AAl94].

Let us consider Question 4 in the Introduction. L users share a rewritable memory with n cells, and each of them has his own message set. So user i has message set $\mathcal{M}_i$ with size M_i . They injectively map their message sets to subsets of $\mathcal{X}^n$, S_i , say, $1 \leq i \leq L$, according to Rule I.

Since each $i \in \{1, \ldots, L\}$ and each $m \in \mathcal{M}_i$ may be updated to any m' in any $\mathcal{M}_j$ $j \in \{1, 2, \ldots, L\}$ (i can be equal to j), for all $i, j \in \{1, 2, \ldots, L\}$, and all $x^n \in S_i$ and $y^n \in S_j$, x^n may be rewritten to y^n . For the transition from user i to user j there is a cost function

$$\varphi_{i,j} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+.$$

Thus we have a cost matrix

$$\Phi := (\varphi_{i,j})_{1 \leq i,j \leq L}. \tag{5.1}$$

When

$$\varphi_{i,j}(x, y) = \varphi_{j,i}(y, x) \text{ for all } i, j, x, y, \tag{5.2}$$

we say the cost functions are symmetric.

The cost for updating $x^n \in S_i$ to $y^n \in S_j$ is of sum–type, i.e.

$$\varphi_{i,j}(x^n, y^n) := \sum_{t=1}^{n} \varphi_{i,j}(x_t, y_t). \tag{5.3}$$

Assume that the constraint on the average cost for updating the messages of user i to those of user j is $n\delta_{i,j}$ (here $\delta_{i,j}$ is a positive constant), that is,

$$\frac{1}{M_i}\frac{1}{M_j} \sum_{x^n \in S_j} \sum_{y^n \in S_i} \varphi_{i,j}(x^n, y^n) \leq n\delta_{i,j} \tag{5.4}$$

where

$$|S_i| = M_i \text{ for } 1 \leq i \leq L. \tag{5.5}$$

We write $\Delta = (\delta_{ij})_{1 \leq i,j \leq L}$, $M = (M_1, \ldots, M_L)$ and we call $(S_1, \ldots, S_L)$ satisfying (5.4) and (5.5) an (n, M, Φ, Δ)–system . When $L = 1$, S_1 is a set with average diameter $n\delta_1$ (the case of [AAl94]).

Furthermore, we call $R = (R_1, \ldots, R_L)$, $R_i \geq 0$ for $1 \leq i \leq L$ (Φ, Δ) achievable, if for all $\varepsilon > 0$ and for $n > n_\varepsilon$ (suitable) there are (n, M, Φ, Δ)–systems with

$$\frac{1}{n} \log M_i \geq R_i - \varepsilon \text{ for } i = 1, \ldots, L. \tag{5.6}$$

Finally $\mathcal{R}(\Phi, \Delta)$ is the region of (Φ, Δ) achievable vectors and our goal is to characterize it.

5.2 The (Φ,Δ)–achievable region.

Define

$$\beta_{ij}=\max_{x',y'\in\mathcal{X}}\varphi_{i,j}(x',y')\quad\text{for}\quad 1\le i,j\le L, \tag{5.7}$$

$$\|\Delta\|=\big|\big\{(i,j):\delta_{i,j}\neq\beta_{i,j},\quad\text{where}\quad 1\le i,j\le L\big\}\big|, \tag{5.8}$$

$$\theta=\|\Delta\|+L,\quad\text{and}\quad\Theta=\{1,2,\ldots,\theta\}. \tag{5.9}$$

Moreover, define $\mathcal{R}^*(\Phi,\Delta)$ as set of vectors $R=(R_1,\ldots,R_L)$ for which a PD Q on $\Theta=\{1,2,\ldots,\theta\}$ and conditional probability distributions $W_j(\cdot|\xi)$ on $\mathcal{X}$ can be found for $\xi\in\Theta$, $j=1,\ldots,L$ such that

$$R_j\le H(W_j|Q)\quad\text{for}\quad 1\le j\le L \tag{5.10}$$

and

$$\sum_{\xi\in\Theta}Q(\xi)\sum_{x,y\in\mathcal{X}}\varphi_{i,j}(x,y)W_i(x|\xi)W_j(y|\xi)\le\delta_{i,j}\quad\text{for}\quad 1\le i,j\le L. \tag{5.11}$$

More elegantly (5.11) can be written in the form

$$\sum_{\xi\in\Theta}Q(\xi)\sum_{x,y\in\mathcal{X}}W^\tau(x|\xi)\Phi(x,y)W(y|\xi)\le\Delta, \tag{5.12}$$

where $W(\cdot|\xi)=\big(W_1(\cdot|\xi),\ldots,W_L(\cdot|\xi)\big)$, W^τ is the transpose of W , and $\Phi(x,y)=\big(\varphi_{i,j}(x,y)\big)_{1\le i,j\le L}$.

In some cases, people only are interested in the total rates

$$R_\sigma(\Phi,\Delta)=\max_{R\in\mathcal{R}(\Phi,\Delta)}\sum_{i=1}^L R_i, \tag{5.13}$$

rather that the regions. Here we have together characterisations in terms of

$$R^*_\sigma(\Phi,\Delta)=\max_{R\in\mathcal{R}^{*\prime}(\Phi,\Delta)}\sum_{i=1}^L R_i, \tag{5.14}$$

where $\mathcal{R}^{*\prime}(\Phi,\Delta)$ is the region obtained by replacing Θ by $\Theta'=\{1,2,\ldots,\|\Delta\|+1\}$ in the definition of $\mathcal{R}^*(\Phi,\Delta)$. We also have a tighter description in symmetric situations.

Theorem 3.

i) $\mathcal{R}(\Phi,\Delta)=\mathcal{R}^*(\Phi,\Delta)$, (5.15)

ii) *If the cost functions and the matrix Δ are both symmetric, one may replace the Θ used in* i) *by a smaller set of a size* $\big|\big\{\{i,j\}:\delta_{i,j}\neq\beta_{i,j}\big\}\big|+L$.

iii) $R_\sigma(\Phi,\Delta)=R^*_\sigma(\Phi,\Delta)$. (5.16)

Theorem 3 has the following two special cases as consequences.

Corollary 1. *([AAl94])*

For a cost function φ *on a finite set* $\mathcal{X}$ *let* $A_n \subset \mathcal{X}^n$ *have size not smaller than* a_n *, minimal average cost* $\overline{\varphi}_n$ *and let* $\lim\limits_{n\to\infty} \frac{1}{n}\log a_n$ *exist, then*

$$\lim_{n\to\infty} \frac{1}{n}\log \overline{\varphi}_n = \min\left[\lambda \sum_{x,y} \varphi(x,y)P(x)P(y) + \overline{\lambda}\sum_{x,y}\varphi(x,y)P'(x)P'(y)\right],$$

where the minimum is over all $\lambda \in [0,1]$ *and* P, P' *satisfying* $\lambda H(P) + \overline{\lambda} H(P') \geq \lim\limits_{n\to\infty}\frac{1}{n}\log a_n$ *.*

Corollary 2.

Suppose that φ *is a symmetric cost function and* $A_n, B_n \subset \mathcal{X}^n$ *have average cross–cost not exceeding* nd *, then*

(a) *Whenever* $R_A = \lim\limits_{n\to\infty}\frac{1}{n}\log|A_n|$ *and* $R_B = \lim\limits_{n\to\infty}\frac{1}{n}\log|B_n|$ *exist, then* (R_A, R_B) *satisfies for some* $P_i, Q_i, \lambda_i \quad (i = 1,2,3)$ *,* $\sum\limits_{i=1}^{3}\lambda_i = 1$

$$R_A \leq \sum_{i=1}^{3}\lambda_i H(P_i),\ R_B \leq \sum_{i=1}^{3}\lambda_i H(Q_i), \quad \textit{and} \quad \sum_{i=1}^{3}\lambda_i \sum_{x,y}\varphi(x,y)P_i(x)Q_i(y) \leq d$$

and this bound is best possible.

(b) *When* (A_n, B_n) *maximizes* $|A_n||B_n|$ *, then*

$$\lim_{n\to\infty}\frac{1}{n}\log|A_n||B_n| = \max\bigl[\lambda\bigl(H(P)+H(Q)\bigr) + \overline{\lambda}\bigl(H(P')+H(Q')\bigr)\bigr],$$

where the maximum is over all $\lambda \in [0,1]$ *,* P, P' *,* Q, Q' *with*

$$\lambda\sum_{x,y}\varphi(x,y)P(x)Q(y) + \overline{\lambda}\sum_{x,y}\varphi(x,y)P'(x)Q'(y) \leq d.$$

5.3 The proof of Theorem 3.

Since ii) and iii) can be proved in the same way as i) (the only difference being that in former cases one needs to count less equations, in the application of the Support Lemma), we only prove i).

Direct part: $\mathcal{R}(\Phi,\Delta) \supset \mathcal{R}^*(\Phi,\Delta)$

Fix $R \in \mathcal{R}^*(\Phi,\Delta)$ and $\varepsilon > 0$. By continuity, for sufficiently large n , there exist $Q \in \mathcal{P}(n,\Theta)$ and $Q \times W_j \in \mathcal{P}(n, \Theta\times\mathcal{X})$, with

$$R_j \leq H(W_j|Q) + \frac{\varepsilon}{2} \quad \text{for} \quad 1 \leq j \leq L, \tag{5.17}$$

and

$$\sum_{\xi\in\Theta} Q(\xi) \sum_{x,y\in\mathcal{X}} \varphi_{i,j}(x,y) W_i(x|\xi) W_j(y|\xi) \leq \delta_{i,j} - \frac{\varepsilon}{2}. \tag{5.18}$$

By (2.25), (2.26), and (5.18), for sufficiently large n, any $u^n \in \mathcal{T}_Q^n$, and $M := \big(|G_{Q\times W_1}(u^n)|, \ldots, |G_{Q\times W_L}(u^n)|\big)$, $\big(G_{Q\times W_1}(u^n), \ldots, G_{Q\times W_L}(u^n)\big)$ is an (n, M, Φ, Δ)–system . By (2.23) and (5.17),

$$\frac{1}{n} \log |G_{Q\times W_j}(u^n)| \geq R_j - \varepsilon \quad \text{for} \quad 1 \leq j \leq L.$$

Converse part: $\mathcal{R}(\Phi, \Delta) \subseteq \mathcal{R}^*(\Phi, \Delta)$

For an (n, M, Φ, Δ)–system $\{S_i : 1 \leq i \leq L\}$ let $(\hat{X}_1^n, \ldots, \hat{X}_L^n, \hat{X}_1'^n, \ldots, \hat{X}_L'^n)$ be independent RV's with distributions

$$\Pr(\hat{X}_\ell^n = x^n) = P(\hat{X}_L'^n = x^n) = \begin{cases} \frac{1}{M_\ell} & \text{for} \quad x^n \in S_\ell \\ 0 & \text{otherwise.} \end{cases} \tag{5.19}$$

Then by (2.1), (2.3), and (2.17),

$$\log M_\ell = H(\hat{X}_\ell^n) \leq \sum_{t=1}^{n} H(\hat{X}_{\ell,t}^n), \tag{5.20}$$

where $\hat{X}_\ell^n = (\hat{X}_{\ell,1}, \ldots, \hat{X}_{\ell,n})$ for $\ell = 1, \ldots, L$, and by the definition of the (n, M, Φ, Δ)–system, for $1 \leq \ell, \ell' \leq L$

$$\mathbb{E}\ \varphi_{\ell,\ell'}(\hat{X}_\ell^n, \hat{X}_{\ell'}'^n) = \sum_{t=1}^{n} \mathbb{E}\ \varphi_{\ell,\ell'}(\hat{X}_{\ell,t}, \hat{X}_{\ell',t}') = \frac{1}{M_\ell}\frac{1}{M_{\ell'}} \sum_{x^n\in S_\ell} \sum_{y^n\in S_{\ell'}} \varphi_{\ell\ell'}(x^n, y^n) \leq n\delta_{\ell\ell'} \tag{5.21}$$

We define now μ on $\mathcal{P}(\mathcal{X}) \times \cdots \times \mathcal{P}(\mathcal{X})$ by

$$\mu(P_1, \ldots, P_L) = \begin{cases} \frac{1}{n}, & \text{if} \quad (P_1, \ldots, P_L) = (P_{\hat{X}_{1,t}}, \ldots, P_{\hat{X}_{L,t}}), 1 \leq t \leq n, \\ 0 & \text{otherwise.} \end{cases}$$

and apply the Support Lemma as indicated in Remark 2 for the sets of functions of $(P_1, \ldots, P_L)$,

$$\left\{ \sum_{x,y} \varphi_{\ell,\ell'}(x,y) P_\ell(x) P_{\ell'}(y) : 1 \leq \ell, \ell' \leq L \text{ and } \delta_{\ell,\ell'} \neq \beta_{\ell,\ell'} \right\} \text{ and } \big\{H(P_\ell) : 1 \leq \ell \leq L\big\}.$$

Notice that $P_{\hat{X}_{\ell,t}} = P_{\hat{X}_{\ell,t}'}$.

Then there are $\lambda_\xi (1 \leq \xi \leq \theta)$ with $\lambda_\xi \geq 0$, $\sum_{\xi \in \Theta} \lambda_\xi = 1$, and

$P_\ell^\xi \in \mathcal{P}(\mathcal{X}) \times \cdots \times \mathcal{P}(\mathcal{X}) \quad (1 \leq \ell \leq L; \xi \in \Theta)$ such that

$$\sum_{t=1}^{n} \frac{1}{n} \mathbb{E}\, \varphi_{\ell,\ell'}(\hat{X}_{\ell,t}, \hat{X}'_{\ell',t}) = \int \mathbb{E}\, \varphi_{\ell,\ell'} d\mu(dP_1, \ldots, dP_L) = \sum_{\xi \in \Theta} \lambda_\xi \left[\sum_{x^n, y^n} \varphi_{\ell,\ell'}(x,y) P_\ell^\xi(x) P_{\ell'}^\xi(y) \right] \tag{5.22}$$

for $1 \leq \ell, \ell' \leq L$ with $\delta_{\ell,\ell'} \neq \beta_{\ell,\ell'}$ and such that

$$\sum_{t=1}^{n} \frac{1}{n} H(\hat{X}_{\ell,t}) = \int H(P_\ell) \mu(dP_1, \ldots, dP_L) = \sum_{\xi \in \Theta} \lambda_\xi H(P_\ell^\xi) \text{ for } 1 \leq \ell \leq L. \tag{5.23}$$

Set now $Q(\xi) = \lambda_\xi$ for $\xi \in \Theta$ and $W_\ell(\cdot|\xi) = P_\ell^\xi(\cdot)$ for $\xi \in \Theta$ and $1 \leq \ell \leq L$. W_ℓ is a conditional distribution for all ℓ and (5.20), (5.23) imply

$$\frac{1}{n} \log M_\ell \leq H(W_\ell | Q) \tag{5.24}$$

and (5.21), (5.22) imply (5.11).

Remark 7:

A much harder problem is to characterize the achievable vectors $M = (M_1, \ldots, M_L)$ exactly. Already in the seemingly simple case, where $L = 1$, $\mathcal{X} = \{0, 1\}$, and φ_{11} is the Hamming distance, it is unsolved (e.g. [AlS92]).

6. Several users follow a cyclic protocol

6.1 The cyclic WEM code.

In this last section we answer Question 5 of the Introduction. Assume that the Criterion Ave and Rule II are used. J users, labelled by $0, 1, \ldots, J-1$, share a memory. They are well organized. Each time one user uses the memory and user $j \oplus 1$ follows user j , where the addition is modulo J . There is a vector $\varphi = (\varphi_0, \ldots, \varphi_{J-1})$ of sum–type cost functions and a positive cost constraint vector $d = (d_0, \ldots, d_{J-1})$. For convenience of notation here the cost function φ_j and constraint d_j are used for user $j \oplus 1$ to update the messages of user j . Denote by $M^{(j)}$, the number of messages of user j . Then an (n, M, φ, d) cyclic WEM code for this system is a family of spreads $\{C_i^{(j)} : 1 \leq i \leq M^{(j)}, 0 \leq j \leq J-1\}$ such that

$$C_i^{(j)} \cap C_{i'}^{(j)} = \varnothing \text{ if } , i \neq i',$$

and for all j , all $x^n \in \Omega_j := \bigcup_i C_i^{(j)}$ there is for every i' a $y^n \in C_{i'}^{(j \oplus 1)}$ with

$$\frac{1}{n} \varphi_j(x^n, y^n) \leq d_j \ (j = 0, 1, \ldots, J-1). \tag{6.1}$$

Thus for given J , the achievable region $\mathcal{R}(d)$ of cyclic WEM code is the set of vectors $(R^{(1)},\dots,R^{(J-1)})$ with non–negative components, such that for all $\varepsilon>0$ and $n>n_\varepsilon$ (suitable) there is an (n,M,φ,d) cyclic WEM code with $\frac{1}{n}\log M^{(j)}\geq R^{(j)}-\varepsilon$ for $j=0,1,\dots,J-1$.

6.2 The characterization of $\mathcal{R}(d)$.

We introduce now two regions, whose significance becomes apparent soon.

$\mathcal{R}'(d)=\big\{\big(R^{(0)},\dots,R^{(J-1)}\big):R^{(j)}\leq H\big(W^{(j\ominus 1)}\mid P^{(j\ominus 1)}\big)\big\}$ for stochastic matrices $W^{(j)}$ and PD's $P^{(j)}$ with $\sum\limits_{x,y}\varphi_j(x,y)P^{(j)}(x)W^{(j)}(y|x)\leq d_j$ and

$$P^{(j)}W^{(j)}=P^{(j\oplus 1)}. \tag{6.2}$$

The other region $\mathcal{R}''(d)$ is defined in terms of families like $\{\mathcal{P}^{(j)}:0\leq j\leq J-1\}$ of closed sets $\mathcal{P}^{(j)}$ of PD's on $\mathcal{X}$ and the related quantity

$$H^*(\mathcal{P}^{(j)})=\min_{P^{(j)}\in\mathcal{P}^{(j)}}\max_{W^{(j)}}H(W^{(j)}|P^{(j)}), \tag{6.3}$$

where the maximum is taken over matrices $W^{(j)}$ satisfying

$$\sum_{x,y}\varphi_j(x,y)P^{(j)}(x)W^{(j)}(y|x)\leq d_j \tag{6.4}$$

and

$$P^{(j)}W^{(j)}\in\mathcal{P}^{(j\oplus 1)}. \tag{6.5}$$

Theorem 4. *For cyclic WEM*

$$\mathcal{R}(d)=\mathcal{R}'(d)=\mathcal{R}''(d).$$

We base the proof on three lemmas.

Actually, they are direct generalizations of Theorems 1, 2 and Lemma 1 of [AZ89].

Lemma 1. $\mathcal{R}'(d)=\mathcal{R}''(d)$.

Proof: The relation $\mathcal{R}'(d)\subset\mathcal{R}''(d)$ is obvious since $\mathcal{R}'(d)$ can be obtained by choosing $\mathcal{P}^{(j)}$ in the definition of $\mathcal{R}''(d)$ as singletons. We start with a fixed family $\{\mathcal{P}^{(j)}:1\leq j\leq J-1\}$. Choosing any $P_1^{(0)}\in\mathcal{P}^{(0)}$ we look for a $W_1^{(0)}$, which achieves $\max_{W^{(0)}}H(W^{(0)}\mid P^{(0)})$ in (6.3). Thus we obtain $P_1^{(1)}=P_1^{(0)}W_1^{(0)}\in\mathcal{P}^{(1)}$ and we continue in the same way. This results in a matrix $W_1^{(1)}$ and $P_1^{(2)}\in\mathcal{P}^{(2)}$ and finally in a path

$$\begin{aligned}P_1^{(0)}\to W_1^{(0)}\to P_1^{(1)}\to W_1^{(1)}\to\dots\to P_1^{(J-1)}\to W_1^{(J-1)}&\\ \to P_2^{(0)}\to W_2^{(0)}\to P_2^{(1)}\to\dots&\end{aligned} \tag{6.6}$$

It produces $2J$ sequences $(P_i^{(j)})_{i=1}^\infty$ and $(W_i^{(j)})_{i=1}^\infty$ $(j = 0, 1, \dots, J-1)$ with the properties

$$\sum_{x,y} \varphi_j(x,y) P_i^{(j)}(x) W_i^{(j)}(y|x) \le d_j \quad \text{for} \quad i = 1, 2, \dots \tag{6.7}$$

and

$$P_i^{(j)} W_i^{(j)} = \begin{cases} P_i^{(j\oplus 1)} & \text{for} \quad i \ne J-1 \\ P_{i+1}^{(0)} & \text{for} \quad i = J-1. \end{cases} \tag{6.8}$$

For large k define

$$P^{(j)} = \frac{1}{k} \sum_{j=1}^{k} P_i^{(j)} \ (j = 0, 1, \dots, J-1), \tag{6.9}$$

$$\hat{P}^{(0)} = \sum_{i=2}^{k+1} P_i^{(0)}, \tag{6.10}$$

$$Q^{(j)} = \frac{1}{k} \sum_{i=1}^{k} P_i^{(j)} \times W_i^{(j)}. \tag{6.11}$$

Thus by our construction $Q^{(j)}$ has marginal distributions $P^{(j)}$ (for the first RV and all j) and $P^{(j\oplus 1)}$ (for the second RV and $j < J-1$) or $\hat{P}^{(0)}$ (for the second RV and $j = J-1$).

Let now

$$W^{(j)} = \frac{Q^{(j)}}{P^{(j)}} \ (j = 0, \dots, J-1), \tag{6.12}$$

then it follows from (6.6) – (6.12) that

$$\begin{aligned} \sum_{x,y} \varphi_j(x,y) P^{(j)}(x) W^{(j)}(y|x) &= \sum_{x,y} \varphi_j(x,y) Q^{(j)}(x,y) \\ &= \frac{1}{k} \sum_{i=1}^{k} \sum_{x,y} \varphi_j(x,y) P_i^{(j)}(x) W_i^{(j)}(y|x) \le d_j \end{aligned} \tag{6.13}$$

and

$$P^{(j)} W^{(j)} = \begin{cases} P^{(j\oplus 1)}, & \text{if} \quad j \ne J-1 \\ \hat{P}^{(0)}, & \text{if} \quad j = J-1. \end{cases} \tag{6.14}$$

Let $(X_i^{(j)}, X_i^{(j\oplus 1)})$ have distribution $P_i^{(j)} \times W_i^{(j)}$ for $j \ne J-1$, let K have uniform distribution on $\{1, 2, \dots, k\}$, and let $(\tilde{X}^{(j)}, \tilde{X}^{(j\oplus 1)})$ have distribution $Q^{(j)}$, defined in (6.11). Then

$$\begin{aligned} H(W^{(j)}|P^{(j)}) &= H(\tilde{X}^{(j\oplus 1)}|\tilde{X}^{(j)}) \ge H(X_K^{(j\oplus 1)}|X_K^{(j)}, K) \ (\text{by } (2.13)) \\ &= \sum_{i=1}^{k} \frac{1}{k} H(W_i^{(j)}|P_i^{(j)}) \ge H^*(\mathcal{P}^{(j)}) - \varepsilon. \end{aligned}$$

However,

$$\|P^{(0)} - \hat{P}^{(0)}\| := \sum_x |P^{(0)}(x) - \hat{P}(x)| = O\left(\frac{1}{k}\right). \tag{6.15}$$

These facts and continuity of conditional entropies complete our proof.

Lemma 2. $\mathcal{R}'(d) \subset \mathcal{R}(d)$.

Proof: Let $\{W^{(j)} : 0 \le j \le J-1\}$ and $\{P^{(j)} : 0 \le j \le J-1\}$ satisfy the constraints in (6.2). Define $\Omega_j = \mathcal{T}^n_{P^{(j)}}$ and color it at random with uniform distribution with $M^{(j)}$ colors to get, as usual, $\{C_i^{(j)}\}_{i=1}^{M^{(j)}}$. Do this independently for $j = 0, 1, \dots, J-1$. Denote by $E(x^n, i)$ the event that there is no $y^n \in C_i^{(j\oplus 1)}$ with $\frac{1}{n}\varphi_j(x^n, y^n) \le d_j$ for $x^n \in \Omega_j$ and $i = 1, 2, \dots, M^{(j)}$.
Then, since by (2.26) and (6.4), for all $y^n \in G_{P^{(j)} \times W^{(j)}}(x^n)$, $\frac{1}{n}\varphi_j(x^n, y^n) \le d_j$,

$$\Pr\big(E(x^n, i)\big) \le \left(1 - \frac{1}{M^{(j\oplus 1)}}\right)^{|G_{P^{(j)} \times W^{(j)}}(x^n)|} \le \exp_e\{-2^{nH(W^{(j)}|P^{(j)}) - \log M^{(j\oplus 1)}}\} (\text{by } (2.23)). \tag{6.16}$$

Therefore

$$\Pr\left(\bigcup_j \bigcup_{x^n \in \Omega_j} \bigcup_{1 \le i \le M^{(j\oplus 1)}} E(x^n, i)\right) \lesssim \sum_{j=0}^{J-1} |\ \mathcal{T}^n_{P^{(j)}}\ |\ M^{(j\oplus 1)} \exp_e\{-2^{nH(W^{(j)}|P^{(j)}) - \log M^{(j\oplus 1)}}\} < 1,$$

if we choose $\log M^{(j\oplus 1)} < n\big(H(W^{(j)} \mid P^{(j)}) - \varepsilon\big)$, and sufficiently large n , and this probabilistic argument implies the existence of $\{C_i^{(j)} : 1 \le i \le M^{(j)}\ ,\ 1 \le j \le L\}$ such that $\frac{1}{n}\log M^{(j)} \sim H(W^{(j\ominus 1)}|P^{(j\ominus 1)})$ and for all j, i, i' , and $x^n \in C_j^{(j)}$ there is a $y^n \in C_i^{(j\oplus 1)}$ with $\frac{1}{n}\varphi_j(x^n, y^n) < d_j$.

Lemma 3. $\mathcal{R}(d) \subset \mathcal{R}''(d)$.

Proof: For a cyclic WEM code $\{C_i^{(j)} : 1 \le i \le M^{(j)}\ ,\ 1 \le j \le L\}$, set $\Omega_j = \bigcup_i C_i^{(j)}$ and let $\mathcal{P}^{(j)}$ be a minimal set of PD's with

$$\Omega_j \subset \bigcup_{P^{(j)} \in \mathcal{P}^{(j)}} \mathcal{T}^n_{P^{(j)}}\ (j = 0, \dots, J-1).$$

Let $P^{(j)} \in \mathcal{P}^{(j)}$ achieve the minimum in (6.3) and $x^n \in \mathcal{T}^n_{P^{(j)}}$. By the definition of the cyclic WEM code

$$M^{(j\oplus 1)} \le \left| \left\{ y^n \in \Omega_{j\oplus 1} : \frac{1}{n}\varphi_j(x^n, y^n) \le d_j \right\} \right| = |B(x^n, j\oplus 1, d_j)|, \tag{6.17}$$

where $B(x^n, j\oplus 1, d_j) = \left\{y^n \in \Omega_{j\oplus 1} : \frac{1}{n}\varphi_j(x^n, y^n) \le d_j\right\}$.

However, by (2.22) there are at most $(n+1)^{|\mathcal{X}|^2}$ $Q^{(j)}$'s with $G_{Q^{(j)}}(x^n) \neq \varnothing$. Thus there is a $Q^{(j)}$ such that

$$|B(x^n, j\oplus 1, d_j) \cap G_{Q^{(j)}}(x^n)| \geq (n+1)^{-|\mathcal{X}|^2} |B(x^n, j\oplus 1, d_j)|. \tag{6.18}$$

Let

$$W^{(j)} = \frac{Q^{(j)}}{P^{(j)}}. \tag{6.19}$$

Notice that (6.18) implies $B(x^n, j\oplus 1, d_j) \cap G_{Q^{(j)}}(x^n) \neq \varnothing$. Thus (6.4) follows from (2.26) and the definition of $B(x^n, j\oplus 1, d_j)$. Further, (6.5) holds since there is a $y^n \in \Omega_{j\oplus 1} \cap G_{Q^{(j)}}(x^n)$ and therefore by (2.21) and (6.19) $P_{y^n} = P^{(j)} W^{(j)}$.

Finally by (2.23), (6.3), and (6.17) – (6.19), we get

$$\frac{1}{n} \log M^{(j\oplus 1)} \leq \frac{|\mathcal{X}|^2}{n} \log(n+1) + \frac{1}{n} \log |G_{Q^{(j)}}(x^n)| \lesssim H(W^{(j)}|P^{(j)}) \leq H^*(\mathcal{P}^{(j)}).$$

References

[A71] Ahlswede, R. (1971), Multi–way communication, in “Proceesing of 2nd International Symposium on Information Theory”, Thakadsor, Armenian SSR, Akademiai Kiado, Budapest, pp. 23–52.

[AAl94] Ahlswede, R. and Althöfer, I. (1994), The asymptotic behaviour of diameters in the average, *J. Combin. Theory*, Ser. B 61 (2), 167–177.

[AC94] Ahlswede, R. and Cai, N. (1994), General edge–isoperimetric inequalities, to appear in European J. Combin.

[ACZ92] Ahlswede, R., Cai, N., and Zhang, Z. (1992), Diametric theorems in sequence spaces, *Combinatorica* 12 (1), 1–17.

[AK77] Ahlswede, R. and Katona, G. (1977), Contributions to the geometry of Hamming spaces, *Discrete Math.* 17, 1–22.

[AKh] Ahlswede, R. and Khachatrian, L., The diametric theroem in Hamming space — optimal anticodes, submitted to Advances in Mathematics.

[AKö75] Ahlswede, R. and Körner, J. (1975), Source coding with side information and a converse for degraded broadcast channels, *IEEE Trans. Inform. Theory* 21, 629–637.

[AZ89] Ahlswede, R. and Zhang, Z. (1989), Coding for write–efficient memory, *Inform. and Comput.* 83 (1), 80–97.

[AZ94] Ahlswede, R. and Zhang, Z. (1994), On multi–user write–efficient memories, *IEEE Trans. Inform. Theory* 40 (3), 674–686.

[AlS92] Althöfer, I. and Silke, T. (1992), An average distance inequality for large subsets of the cube, *J. Combin. Theory* Ser. B 56, 296–301.

[CT91] Cover, T.M. and Thomas, L.H. (1991), "Elements of Information Theory", Wiley, New York.

[CsKö81] Csiszár, I. and Körner, J. (1981), "Information Theory, Coding Theorems for Discrete Memoryless Systems", Academic Press, New York.

[E58] Eggleston, H.G. (1958), "Convexity", *Cambridge University Press*, Cambridge.

[G68] Gallager, R.G. (1968), "Information Theory and Reliable Communication", Wiley, New York.

[K64] Katona, G. (1964), Intersection theorems for systems of finite sets, *Acta Math. Hungar.* 15, 329–337.

[Kl66] Kleitman, D.J. (1966), On a combinatorial conjecture of Erdös, *J. Combin. Theory* 1, 209–214.

[O95] Orlitzky, A. (1995), Oral communication.

[RSh82] Rivest, R.L. and Shamir, A. (1982), How to use a write–once memory, Inform. and Control 55, 1–19.

[S78] Salehi, M. (1978), Cardinality bounds on auxiliary variables in multiple–user theory via the method of Ahlswede and Körner, *Stanford Technical Report*.

[Si86] Simonyi, G. (1986), On write–unidirectional memory, "Report Interne Enst 86 D 007".

[W78] Wolfowitz, J. (1978), "Coding Theorems of Information Theory", the 3rd edition, Springer Verlag, Heidelberg.

[WiV86] Willems, F.M.J. and Vinck, A.J. (1986), Repeated recordings for an optical disk, in "*Proceedings, 7th Symposium on Information Theory in the Benelux*", Delft University Press, pp. 49–53.

[WoWyZK84] Wolf, T.K., Wyner, A.D., Ziv, J., and Körner, J. (1984), Coding for write once memory, AT & T Lab. Tech. J. 63 (6), 1084–1112.