Coding for Write–Efficient Memory

R. Ahlswede and Z. Zhang

Abstract

We introduce write–efficient memories (WEM) as a new model for storing and updating information on a rewritable medium. There is a cost $\varphi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_\infty$ assigned to changes of letters. A collection of subsets $\mathcal{C} = \{C_i : 1 \leq i \leq M\}$ 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

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

$D_{\max}$ is called the maximal correction cost with respect to the given cost function. The performance of a code φ 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 $r_{\mathcal{C}} = n^{-1} \log M$. The rate achievable with a maximal per letter cost d is thus

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

This is the most basic quantity (the storage capacity) of a WEM $(\mathcal{X}^n, \varphi^n)_{n-1}^{\infty}$. We give a characterization of this and related quantities.

1 Introduction

The information–theoretic study of memory cells started with the work of Kusnetsov and Tsybakov (1974). The papers by Ahlswede (1986), Ahlswede and Han (1983), Costa (1983), El Gamal (1983), Gelfand and Pinsker (1980), Heegard and El Gamal (1983), and Tsybakov (1977) either continue this line of investigation or else contain results with strong connections to it. A new impetus came from a paper by Rivest and Shamir (1982), in which write–once memories (WOM) were introduced. They have been further analysed in Heegard (1983) and Wolf, Wyner, Ziv, and Körner (1984). In a WOM there are two states 0 and 1 for a cell (or position) in the memory (or medium); 1 is the state of a cell which has been used and 0 is the state of a cell which has not been used. Once a cell is in state 1 it will stay there forever. So among the four possible transitions of the states, $0 \to 1$, $0 \to 0$, $1 \to 1$, and $1 \to 0$, the first three are allowed, whereas the last one is forbidden. This means that the memory is somehow a permanent one.

Recently a new model of updating information stored on a rewritable medium has been introduced by Willems and Vinck (1986) and Borden (in press). There the memory is not permanent, all of the four transitions are permitted; however, for a rather ling time period of updating, it is allowed to write either 1's or 0's, but not the combinations of both of them. This assumption arose in laser technology, where the printing of the same letter can be done fast, but changing the directions of the magnets necessary for a variation of letters was rather slow. This storage medium has been named "write–unidirectional memory". The abbreviation "WUM" suggests itself. The papers of Cohen and Simonyi (in press) and Simonyi (1989) deal with or relate to WUMs.

Recently, we learnt from Frans Willems that technology has advanced to the extend that the changing of magnets is no longer a major obstacle. This stimulated us to consider still another memory (Hamming WEM) and then to set up a seemingly natural and quite general model for memories, which includes also WOMs and WUMs as special cases. We argued as follows. Since a rewriting (transition from 1 to 0 and from 0 to 1) always costs time and energy, whereas the other two transitions ($0 \to 0$ and $1 \to 1$) cost very little, it is reasonable to design codes of updating information stored on a rewritable medium which require as few changes as possible in the text written on the medium. In other words, we want to minimize the rate of transitions $0 \to 1$ and $1 \to 0$ in order to save time and energy. These ideas are made precise in the following definitions.

Let $\mathcal{C} = C_i : \{1 \leq i \leq M\}$ be a collection of subsets of $\{0,1\}^n$. All of the members of this system are disjoint, that is for $i \neq j$

$$C_i \cap C_j = \varnothing. \tag{1.1}$$

To use this system as a WEM (write–efficient memory) code, we use all the sequences in a subset C_i to represent the same message, say m_i. While m_i is stored on the medium, one of the sequences in C_i, say c_i, is written. If we want to update m_i to a new message m_j, choose one sequence in C_j, say c_j, which is nearest to c_i in Hamming distance, and correct only the positions where c_i and c_j are different. We need to write only $d_H(c_i, c_j)$ times. For a WEM code there are several parameters of interest: n is called the block length of the code, M is called the size of the code, and $D_{\max}$ is the maximal updating correction; that is,

$$D_{\max} = \max_{1 \le i,j \le 1} \max_{c_i \in C_i} \min_{c_j \in C_j} d_H(c_j, c_j). \tag{1.2}$$

If the parameters are as specified and $D_{\max} \le D$, then we speak of an (n, M, D) Hamming WEM code.

Now we go for more generality. Instead of using the alphabet $\{0,1\}$ and instead of counting the number of transitions $0 \to 1$ and $1 \to 0$ now we allow a general alphabet $\mathcal{X} = \{1, 2, \dots, \alpha\}$ and a function $\varphi = \mathcal{X} \times \mathcal{X} \to \mathbb{R}_\infty$ measuring the costs of the transitions from one letter to another one.

We call a collection of subsets of $\mathcal{X}^n$, $\mathcal{C} = \{C_i : 1 \le i \le M\}$ an (n, M, D) WEM code, if $C_i \cap C_j = \varnothing$ for all i, j with $i \ne j$ and if

$$D_{\max} \triangleq \max_{1 \le i,j \le M} \max_{x^n \in C_i} \min_{y^n \in C_j} \varphi^n(x^n, y^n) \le D, \tag{1.3}$$

where $\varphi^n(x, y^n) \triangleq \sum_{t=1}^{n} \varphi(x_t, y_t)$.

$D_{\max}$ is called the *maximal* correction 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 $r_{\mathcal{C}} = n^{-1} \log M$.

The rate achievable with a maximal per letter cost d is thus

$$R(d) = \sup_{\mathcal{C} : d_{\mathcal{C}} \le d} r_{\mathcal{C}}. \tag{1.4}$$

This is the most basic quantity (the storage capacity) of a WEM $(\mathcal{X}^n, \varphi^n)_{n=1}^{\infty}$. When we apply a WEM code $\mathcal{C} = \{C_i | 1 \le i \le M\}$ repeatedly in a process of rewriting on a rewritable medium, it seems natural to consider not only the maximal correction cost, but also the cost arising in "average". A reasonable concept of average cost can be introduced in the following model: It is assumed that the sequence of messages $(M_t)_{t=1}^{\infty}$, which will be written on the medium, are the values of a sequence of independent random variables, which are uniformly distributed over the set $\{1, \dots, M\}$. The rewriting process is defined by a mapping $\Psi : \Omega \times \{1, 2, \dots, M\} \to \Omega$, where

$$\Omega = \bigcup_{i=1}^{M} C_i \tag{1.5}$$

and

$$\Psi(x^n, i) \in C_i \quad \text{for} \quad x^n \in \Omega. \tag{1.6}$$

The understanding is as follows. Suppose that before the t–th rewriting a word x^n is written on the medium and that in the t–th rewriting we have to change the message to i, then we choose $\Psi(x^n, i)$ to represent i.

For our purposes any Ψ as defined above is suitable. One could choose, for instance, the "greedy" mapping

$$\gamma : \Omega \times \{1, \dots, M\} \to \Omega$$

with $\gamma(x^n, i) \in C_i$ and $\varphi\big(x^n, \gamma(x^n, i)\big) = \min_{z^n \in C_i} \varphi(x^n, z^n)$. The sequence of words $(U_t)_{t=1}^{\infty} = \big((X^n)_t\big)_{t=1}^{\infty}$ written on the medium during the rewriting process forms a Markov chain. Its transition probability matrix $\big(\mu(y^n|x^n)\big)_{x^n \in \Omega, y^n \in \Omega}$ is given by

$$\mu(y^n|x^n) = \begin{cases} 1/M & \text{if } \exists i \text{ with } y^n = \Psi(x^n, i) \\ 0 & \text{otherwise.} \end{cases} \tag{1.7}$$

The state space of this Markov chain can be partitioned into disjoint classes:

$$\Omega = \bigcup_{j=0}^{s} \Omega_j, \tag{1.8}$$

where in the terminology of Chung (1967), Ω_0 is the set of inessential states and the $\Omega_j (j = 1, \dots, s)$ are the sets of essential states. In the long run, Ω_0 makes no contribution to the average correcting cost and can therefore be deleted from Ω.

Let Π_j be the stationary distribution for the essential class Ω_j; that is,

$$\sum_{x^n \in \Omega_j} \Pi_j(x^n) = 1 \tag{1.9}$$

and

$$\sum_{x^n \in \Omega_j} \Pi_j(x^n)\mu(y^n|x^n) = \begin{cases} \Pi_j(y^n) & \text{for } y^n \in \Omega_j \\ 0 & \text{for } y^n \notin \Omega_j. \end{cases} \tag{1.10}$$

In order to make the Markov chain stationary the initial distribution must be of the form

$$\Pi(x^n) = \sum_{j=1}^{s} \lambda_j \Pi_j(x^n), \ \sum_{j=1}^{s} \lambda_j = 1, \ \lambda_j \geq 0. \tag{1.11}$$

If we start with any initial distribution p and define the average correcting cost by

$$\overline{D} = \lim_{m \to \infty} \frac{1}{m} \sum_{t=1}^{m} \mathbb{E}\varphi(U_t, U_{t+1}), \tag{1.12}$$

then from the theory of Markov chanins we can conclude that

$$\begin{aligned} \overline{D} &= \sum_{j=1}^{s} p(\Omega_j) \sum_{x^n \in \Omega_j} \sum_{y^n \in \Omega_j} \Pi_j(x^n)\mu(y^n|x^n)\varphi(x^n, y^n) \\ &= \sum_{j=1}^{s} p(\Omega_j) \overline{D}_j, \end{aligned} \tag{1.13}$$

where $\overline{D}_j$ is the average correcting cost for class Ω_j.

Therefore best use of the code is made, if we delete all but one of the essential classes with minimal value $\overline{D}_j$. This means that the Markov chain corresponding to the WEM code should be irreducible. For such codes the average correcting cost rate $\overline{d}_{\mathcal{C}}$ is well defined. Also we can define the average correcting cost rate

$$\overline{R}(d) = \sup_{\overline{d}_{\mathcal{C}} \leq d} r_{\mathcal{C}}. \tag{1.14}$$

Our main result is a characterization of $R(d)$ in terms of entropies (Theroem 1). The next result is an even more explicit characterization (Theorem 2). It is also of interest to know that $\overline{R}(d)$ equals $R(d)$ (Theorem 3). In the special case of Hamming WEMs, $R(d)$ just equals the binary entropy function $h(d)$ (Theorem 4). Finally, we draw attention to a certain duality between WEM codes and channel codes.

At the end of the paper we announce further results obtainable by our approach and we also sketch some directions of further investigations.

2 Coding Theorems

For the description of our results we need some definitions. For a set $\mathcal{Z}$, $\mathcal{P}(\mathcal{Z})$ denotes the set of all probability distributions on $\mathcal{Z}$. Let (X, Y) be a pair of random variables with values in $\mathcal{X} \times \mathcal{X}$ and distributions P_{XY}. We denote the (marginal) distributions of X (resp. Y) by P_X (resp. P_Y). A set of distributions important to us is

$$\mathcal{P}_d = \big\{P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{X}) : P_X = P_Y, \mathbb{E}\varphi(X,Y) \leq d\big\}. \tag{2.1}$$

These are distributions with equal marginals and an expectation of costs not greater than d.

Furthermore, for $P \in \mathcal{P}(\mathcal{X})$ and $\mathcal{L} \subset \mathcal{P}(\mathcal{X})$ we define

$$\mathcal{P}(P, \mathcal{L}) = \big\{P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{X}) : P_X = P, P_Y \in \mathcal{L}\big\} \tag{2.2}$$

and

$$\mathcal{P}_d(P, \mathcal{L}) = \big\{P_{XY} \in \mathcal{P}(P, \mathcal{L}) : \mathbb{E}\varphi(X,Y) \leq d\big\}. \tag{2.3}$$

Finally, we need the quantities

$$\rho(d) = \sup_{P_{XY} \in \mathcal{P}_d} H(Y|X) \tag{2.4}$$

$$\rho'(d) = \sup_{\mathcal{L} \subset \mathcal{P}(\mathcal{X})} \inf_{P \in \mathcal{L}} \sup_{P_{XY} \in \mathcal{P}_d(P,\mathcal{L})} H(Y|X). \tag{2.5}$$

Here $H(Y|X)$ is the conditional entropy.

Our main result can now be stated.

Theorem 1. *For any $d \geq 0$ and $\varphi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+$, $R(d) = \rho'(d)$.*

Lemma 1. $R(d) = \rho'(d)$.

Theorem 2. $\rho(d) = \rho(d)$.

Theorem 3. $\overline{R}(d) = R(d)$.

One possibility to define the average updating correction of a Hamming WEM code is

$$D_{\text{ave}} = \frac{1}{M^2} \sum_{1 \leq i,j \leq M} \frac{1}{|C_i|} \sum_{c_i \in C_i} \min_{c_j \in C_j} d_H(c_i, c_j).$$

We call $R_{\text{ave}}(d)$ the maximal rate of codes (n, M, D_{ave}). Our last result concerns Hamming WEM codes.

Theorem 4. (a) *For an (n, M, D) Hamming* WEM *code $M \leq \sum_{i=0}^{D} \binom{n}{i}$.*

(b) *For an (n, M, D_{ave}) Hamming* WEM *code,*

$$M D_{ave} \geq (t+1)M - \sum_{i=0}^{t} (t+1-i) \binom{n}{i},$$

where $\sum_{i=0}^{t} \binom{n}{i} \leq M \leq \sum_{i=0}^{t+1} \binom{n}{i}$.

(c) *For Hamming* WEM *codes, $R(d) = R_{ave}(d) = h(d)$.*

3 Notation and Known Facts

We abbreviate $\mathcal{P}(\mathcal{X})$, the set of probability distribution on $\mathcal{X}$, as $\mathcal{P}$. For integers n we put

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

For $x^n \in \mathcal{X}^n$ we define for every $x \in \mathcal{X}$, $P_{x^n}(x) = \frac{1}{n}$ (number of occurences of x in x^n). P_{x^n} is a member of $\mathcal{P}_n$ by definition. It is called the type of x^n. Analogously we define the type $P_{x^n y^n}$ for pairs $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{X}^n$. For $P \in \mathcal{P}$ the set $\mathcal{T}_P^n$ of all P–typical sequences in $\mathcal{X}^n$ is given by $\mathcal{T}_P^n = \{x^n | P_{x^n} = P\}$.

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

$$G_{\mathcal{Q}}^n(x^n) = \{y^n | P_{x^n y^n} = \mathcal{Q}\}. \tag{3.1}$$

If for the random variables X, Y we have

$$P_{XY}(x,y) = P(x)W(x|y) \text{ for all } x,y,$$

then for the entropy $H(X)$ and the conditional entropy $H(Y|X)$ we also write $H(P)$ and $H(W|P)$, respectively.

We shall use the facts

$$|\mathcal{P}_n| \leq (n+1)^{|\mathcal{X}|} \tag{3.2}$$

$$(n+1)^{-|\mathcal{X}|^2} \exp\big\{nH(W|P_{x^n})\big\} \leq |G^n_{\mathcal{Q}}(x^n)|$$

$$\leq \exp\big\{n(W|P_{x^n})\big\}, \text{ if } \mathcal{Q} = P_{x^n} \cdot W. \tag{3.3}$$

4 Proof of the Converse Part in Theorem 1

We show that $R(d) \leq \rho'(d)$. For this let $\mathcal{C} = \{C_i | 1 \leq i \leq M\}$ be a WEM code with blocklength n and maximal per letter cost $d_{\mathcal{C}} \leq d$.

For any $x^n \in \Omega \triangleq \bigcup\limits_{i=1}^{M} C_i$ consider the set

$$S_d(x^n) = \big\{y^n \in \Omega | \varphi(x^n, y^n) \leq nd\big\}. \tag{4.1}$$

By the definition of a WEM code we have

$$M \leq \min_{x^n \in \Omega} |S_d(x^n)|. \tag{4.2}$$

We evaluate this bound by partitioning $S_d(x^n)$ according to joint types as follows. Consider the set of joint types

$$\mathcal{P}\big(x^n, S_d(x^n)\big) = \big\{P_{x^n y^n} | y^n \in S_d(x^n)\big\} \tag{4.3}$$

and write

$$S_d(x^n) = \bigcup_{\mathcal{Q} \in \mathcal{P}\left(x^n, S_d(x^n)\right)} \big(S_d(x^n) \cap G^n_{\mathcal{Q}}(x^n)\big). \tag{4.4}$$

By (3.2) and (3.3) we have

$$\frac{1}{n} \log |S_d(x^n)| \leq |\mathcal{X}|^2 \cdot \frac{\log n + 1}{n} + \max_{P_{x^n} \cdot W \in \mathcal{P}\left(X^n, S_d(x^n)\right)} H(W|P_{x^n})$$

and from (4.2), therefore,

$$\frac{1}{n}\log M \le |\mathcal{X}|^2 \frac{\log n+1}{n} + \min_{x^n\in\Omega} \max_{P_{x^n}\cdot W \in \mathcal{P}\left(x^n, S_d(x^n)\right)} H(W|\mathcal{P}_{x^n}).$$

The first summand tends to 0 as n tends to infinity. Since for $y^n \in S_d(x^n)$,

$$\begin{aligned}\sum_{t=1}^{n} \varphi(x_t, y_t) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{x^n y^n}(x,y)\varphi(x,y)\\ &= n\mathbb{E}_{P_{x^n y^n}}\varphi(X,Y) \le nd,\end{aligned}$$

by definitions (2.3) and (2.5) the second summand does not exceed $\rho'(d)$, because we can choose $\mathcal{L} \triangleq \{P_{z^n} : z^n \in \Omega\}$.

5 Proof of the Direct Part in Theorem 1

Recall the definition of $\mathcal{P}_n \subset \mathcal{P}(\mathcal{X})$ in Section 3 and define

$$\rho'(d) = \sup_{\mathcal{L}\subset\mathcal{P}_n} \inf_{P\in\mathcal{L}} \sup_{P_{XY}\in\mathcal{P}_d(P,\mathcal{L})} H(Y|X).$$

By continuity,

$$\lim_{n\to\infty} \rho'_n(d) = \rho'(d). \tag{5.1}$$

Fix now n and choose $\mathcal{L}_n \subset \mathcal{P}_n$ such that

$$\min_{P\in\mathcal{L}_n} \max_{P_{XY}\in\mathcal{P}_d(P,\mathcal{L}_n)} H(Y|X) \ge \rho'_n(d) - \varepsilon. \tag{5.2}$$

We shall show by random selection that for any n and ε $R(d) \ge \rho'_n(d) - 2\varepsilon$ and thus the result.

For the $\mathcal{L}_n$ chosen, define

$$\Omega \triangleq \bigcup_{P\in\mathcal{Q}_n} \mathcal{T}_P^n \subset \mathcal{X}^n. \tag{5.3}$$

Now label the elements of Ω independently and with probability $1/M$ with one of the numbers $1, 2, \ldots, M$. The elements labeled with i form the codeset C_i. M will be determined below. We first calculate the probability $\Pr(x^n, i)$ that for a fixed $x^n \in \Omega$ there does not exist a $y^n \in C_i \cap S_d(x^n, \Omega)$, where

$$S_d(x^n, \Omega) = \left\{z^n \in \Omega : \varphi(x^n, z^n) \le nd\right\}. \tag{5.4}$$

Clearly,

$$\Pr(x^n, i) = \left(1 - \frac{1}{M}\right)^{|S_d(x^n,\Omega)|}. \tag{5.5}$$

If now $\Gamma = \sum_{x^n\in\Omega}\sum_{i=1}^{n} \Pr(x^n, i) < 1$, then there exists a code of size M and maximal per letter cost less than d.

Now by (5.5)

$$\begin{aligned}\Gamma &\leq |\Omega| M \max_{x^n\in\Omega}\left(1 - \frac{1}{M}\right)^{|S_d(x^n,\Omega)|}\\ &= |\Omega| M \left(1 - \frac{1}{M}\right)^{\min_{x^n\in\Omega}|S_d(x^n,\Omega)|} < 1,\end{aligned}$$

if $M \leq (2\log|\Omega|)^{-1} \min_{x^n\in\Omega} |S_d(x^n\Omega)|$. By (3.3) and (5.4)

$|S_d(x^n, \Omega)| = |\bigcup_{\mathcal{Q}\in\mathcal{P}_d(P_{x^n},\mathcal{L}_n)} G^n_{\mathcal{Q}}(x^n)| \geq \max_{\mathcal{Q}=W\cdot P_{x^n}\in\mathcal{P}_d(P_{x^n},\mathcal{L}_n)} \exp\{nH(W|P_{x^n})\}\cdot(n+1)^{-|\mathcal{X}|^2}$

and we can choose M such that

$$\frac{1}{n}\log M \geq \min_{P\in\mathcal{L}_n}\max_{P_{XY}\in\mathcal{P}_d(P,\mathcal{L}_n)} H(Y|X) - O(\log n)\frac{1}{n}.$$

6 Proof of Lemma 1

Since obviously $\rho(d) \leq \rho'(d)$, it suffices to establish the opposite inequality. For $\varepsilon > 0$, let $\mathcal{L} \subset \mathcal{P}(\mathcal{X})$ be a set of distributions with

$$\inf_{P\in\mathcal{L}}\sup_{P_{XY}\in\mathcal{P}_d(P,\mathcal{L})} H(Y|X) \geq \rho'(d) - \varepsilon \tag{6.1}$$

and, for $P \in \mathcal{L}$, let $P_{XY}(P)$ be a distribution in $\mathcal{P}_d(P, \mathcal{L})$ with $H(Y|X) \geq \rho'(d) - \varepsilon$. This gives rise to two maps,

$$\begin{aligned}\Psi_1 &: \mathcal{L} \to \bigcup_{P\in\mathcal{L}} \mathcal{P}_d(P, \mathcal{L})\\ \Psi_2 &: \mathcal{L} \to \mathcal{L},\end{aligned}$$

with $\Psi_1(P) = P_{XY}(P)$ and $\Psi_2(P) = P_Y(P)$.

Choose now a sequence of distributions in $\mathcal{L}$, say $(P_i)_{i=1}^k$, where

$$P_{i+1} = \Psi_2(P_i) \quad \text{for} \quad 1 \leq i \leq k-1. \tag{6.2}$$

Define now

$$P_{\tilde{X}\tilde{Y}} = \frac{1}{k-1}\sum_{i=1}^{k-1}\Psi_1(P_i), \tag{6.3}$$

which has as marginals

$$P_{\tilde{X}} = \frac{1}{k-1}\sum_{i=1}^{k-1}P_i,\ P_{\tilde{Y}} = \frac{1}{k-1}\sum_{i=1}^{k-1}P_{i+1}. \tag{6.4}$$

Furthermore, let (X_i, Y_i) have distribution $\Psi_1(P_i)$ and let the random variable I take values in $\{1, \dots, k-1\}$ with equal probabilities. Then by our construction,

$$\rho'(d) - \varepsilon \leq \frac{1}{k-1}\sum_{i=1}^{k-1}H(X_i|Y_i) = H(Y_I|X_I, I) \leq H(\tilde{Y}|\tilde{X}) \tag{6.5}$$

and, from (6.4) we conclude that in the norm $\|\cdot\|$ of total variation

$$\|P_{\tilde{X}} - P_{\tilde{Y}}\| = \frac{1}{k-1}\|P_1 - P_k\| \leq \frac{2}{k-1}. \tag{6.6}$$

Therefore $P_{\tilde{X}\tilde{Y}}$ has almost identical marginal distributions. This, (6.5), and an elementary continuity argument complete the proof.

7 Proof of Theorem 3

Since $\overline{R}(d) \geq R(d)$, we have to prove only

$$\overline{R}(d) \leq R(d). \tag{7.1}$$

For this we make use of a known fact.

Lemma 2. *For any $\mathcal{A} \subset \mathcal{X}^n$ and distribution $\mathcal{Q}^n$ on $\mathcal{X}^n$ defined by*

$$\mathcal{Q}^n(x^n) = \begin{cases} 1/|\mathcal{A}| & \text{if } x^n \in \mathcal{A} \\ 0 & \text{otherwise,} \end{cases}$$

$$\log|\mathcal{A}| = H(\mathcal{Q}^n) \leq \sum_{t=1}^{n} H(\mathcal{Q}_t) \leq nH(\mathcal{Q}), \tag{7.2}$$

where $\mathcal{Q}_1, \dots, \mathcal{Q}_n$ are the 1–dimensional marginals of $\mathcal{Q}^n$ and $\mathcal{Q} = (1/n)\sum_{t=1}^{n}\mathcal{Q}_1$.

Proof: The two inequalities in (7.2) are basic properties of the entropy function. Notice that $\mathcal{Q}$ has the property

$$\mathcal{Q}(x) = \frac{1}{n|\mathcal{A}|}\left|\left\{(t, x^n)|1 \leq t \leq n, x^n \in \mathcal{A} \text{ with } x_t = x\right\}\right|. \tag{7.3}$$

Fix any Ψ as in (1.5), (1.6).

Let now $\mathcal{C} = \{C_i | 1 \leq i \leq M\}$ be a WEM code, which is irreducible under the Markov chain induced by Ψ. For $x^n \in \Omega = \bigcup_{i=1}^{M} C_i$ define

$$\Omega(x^n) = \big\{\Psi(x^n, i) | 1 \leq i \leq M\big\}. \tag{7.4}$$

We have assigned to every $\mathcal{A} \subset \mathcal{X}^n$ in (7.3) a probability distribution $\mathcal{Q}$ on $\mathcal{X}$. Here we carry this one step further by assigning a probability distribution $\mathcal{Q}_{x^n}(\cdot,\cdot)$ to $\Omega(x^n)$ by

$$\mathcal{Q}_{x^n}(x,y) = \frac{1}{n|\Omega(x^n)|} \big|\big\{(t, y^n) | 1 \leq t \leq n, y^n \in \Omega(x^n), x_t = x, y_t = y\big\}\big|. \tag{7.5}$$

We also use

$$P_{x^n}(x) = \frac{1}{n} \big|\big\{(t, x^n) | x_t = x\big\}\big|, \tag{7.6}$$

$$W_{x^n}(y|x) = \mathcal{Q}_{x^n}(x,y) P_{x^n}(x)^{-1}. \tag{7.7}$$

Now for $\mathcal{A} = \Omega(x^n)$ by Lemma 2,

$$\frac{1}{n} \log |\Omega(x^n)| \leq \frac{1}{n} \sum_{t=1}^{n} H(\mathcal{Q}_1) = \sum_x \frac{1}{n} \sum_{t: x_t = x} H(\mathcal{Q}_t)$$

and, by the convexity of the entropy function,

$$\begin{aligned} \sum_x \frac{1}{n} \sum_{t: x_t = x} H(\mathcal{Q}_t) &\leq \sum_x \frac{1}{n} n P_{x^n}(x) H\left(\frac{1}{n P_{x^n}(x)} \sum_{t: x_t = x} \mathcal{Q}_t\right) \\ &= \sum_x P_{x^n}(x) H\big(W_{x^n}(\cdot|x)\big) = H(W_{x^n}|P_{x^n}). \end{aligned}$$

Therefore, Lemma 2 generalizes to

Lemma 3. $(1/n) \log |\Omega(x^n)| \leq H(W_{x^n}|P_{x^n})$.

This implies

$$R = \frac{1}{n} \log |\Omega(x^n)| \leq H(W_{x^n}|P_{x^n}). \tag{7.8}$$

This is true for every $x^n \in \Omega$.

Let Π be the stationary initial distribution and define

$$\mathcal{Q}(x,y) = \sum_{x^n \in \Omega} \Pi(x^n) \mathcal{Q}_{x^n}(x,y). \tag{7.9}$$

Let (X,Y) have distribution $\mathcal{Q}$. Then as in (6.5) we conclude that

$$H(Y|X) \geq \sum_{x^n \in \Omega} \Pi(x^n) H(W_{x^n}|P_{x^n}) \geq R.$$

We complete the proof of Theorem 3 by showing that $P_{XY} = \mathcal{Q} \in \mathcal{P}_d$, which is defined in (2.1). Now

$$\begin{aligned}
\sum_{x,y} \mathcal{Q}(x,y)\varphi(x,y) &= \sum_{x^n \in \Omega} \Pi(x^n) \sum_{i=1}^{M} \frac{1}{nM} \varphi\big(x^n, \Psi(x^n, i)\big) \\
&= \frac{1}{n} \sum_{x^n} \sum_{y^n} \Pi(x^n) \mu(y^n|x^n) \varphi(x^n, y^n) \\
&= \overline{d} \leq d.
\end{aligned}$$

It remains to be seen that $P_X = P_Y$,

$$\begin{aligned}
P_X(x) &= \sum_{y} \mathcal{Q}(x,y) = \sum_{y} \sum_{x^n \in \Omega} \Pi(x^n) \mathcal{Q}_{x^n}(x,y) \\
&= \sum_{y} \sum_{x^n \in \Omega} \Pi(x^n) \frac{1}{n|\Omega(x^n)|} \big|\big\{(t, y^n) | 1 \leq t \leq n, y^n \in \Omega(x^n), y_t = y, x_t = x\big\}\big| \\
&= \sum_{x^n \in \Omega} \Pi(x^n) \frac{1}{n} \big|\big\{(t | 1 \leq t \leq n, x_t = x\big\}\big|. \qquad (7.10)
\end{aligned}$$

$$\begin{aligned}
P_Y(y) &= \sum_{x} \mathcal{Q}(y,x) = \sum_{x} \sum_{x^n \in \Omega} \Pi(x^n) \mathcal{Q}_{x^n}(x,y) \\
&= \sum_{x} \sum_{x^n \in \Omega} \Pi(x^n) \frac{1}{n|\mathcal{Q}(x^n)|} \big|\big\{(t, y^n) | 1 \leq t \leq n, y^n \in \Omega(x^n), x_t = x, y_t = y\big\}\big| \\
&= \sum_{x} \sum_{x^n \in \Omega} \Pi(x^n) \frac{1}{n} \sum_{y^n \in \Omega} \mu(y^n|x^n) |\{t | 1 \leq t \leq n, x_t = x, y_t = y\}| \\
&= \sum_{x^n \in \Omega} \Pi(x^n) \frac{1}{n} \sum_{y^n \in \Omega} \mu(y^n|x^n) \times |\{t | 1 \leq t \leq n, y_t = y\}| \\
&= \sum_{y^n \in \Omega} \frac{1}{n} \Pi(y^n) |\{t | 1 \leq t \leq n, y_t = y\}|,
\end{aligned}$$

because Π is stationary. By (7.10) the last quantity equals $P_X(y)$.

8 Hamming WEM Codes

Proof of Theorem 4.

(a) For any $x^n \in \{0,1\}^n$ the total number of elements in $\{0,1\}^n$ with a distance not greater than D from x^n is $\sum_{i=0}^{D} \binom{n}{i}$. Therefore, necessarily, $M \leq \sum_{t=0}^{D} \binom{n}{i}$.

(b) For any $x^n \in \Omega = \bigcup_{i=1}^{M} C_i$ the contribution to D_{ave} is at least

$$\frac{1}{M}\left[\sum_{i=0}^{t} i\binom{n}{i} + (t+1)\left(M - \sum_{i=0}^{t}\binom{n}{i}\right)\right],$$

if $\sum_{i=0}^{t} \binom{n}{i} \leq M \leq \sum_{i=0}^{t+1} \binom{n}{i}$.

(c) We show first that $R_{\text{ave}}(d) \leq h(d)$ for $d < \frac{1}{2}$, the case $d \geq \frac{1}{2}$ being trivial.

From (b) we know that

$$D_{\text{ave}} \geq \frac{1}{M}\sum_{i=0}^{t} i\binom{n}{i} \quad \text{and} \quad \frac{1}{M} \geq \left(\sum_{i=0}^{t+1}\binom{n}{i}\right)^{-1},$$

and thus $D_{\text{ave}} \geq \left(\sum_{i=0}^{t+1}\binom{n}{i}\right)^{-1} \sum_{i=0}^{t} i\binom{n}{i} \geq t - O\left(\sqrt{t}\right)$. Therefore we have also

$$M \leq \sum_{i=0}^{t+1}\binom{n}{i} \leq \sum_{i=0}^{D_{\text{ave}}+1}\binom{n}{i} \leq \exp\left\{nh\left(\frac{D_{\text{ave}}}{n}\right) + o(n)\right\}$$

and thus the claim.

Since $R_{\text{ave}} \geq R(d)$ and by Theorem 2, $R(d) = \rho(d)$, it suffices to show that $\rho(d) \geq h(d)$. In the formula for $\rho(d)$ choose now (X, Y) such that $P_X = P_Y$, $P_X(0) = P_X(1) = \frac{1}{2}$, and $P_{Y|X}(0|0) = P_{Y|X}(1|1) = \beta$. Then $\rho(d) \geq \max_{1-\beta \leq d} h(1-\beta) = h(d)$, if $d \leq \frac{1}{2}$, and hence the result.

Remarks.

(1) Statements (a) and (b) in Theorem 4 suggest two definitions. A WEM code is called *perfect*, if the equality holds in (a), and it is called quasi–perfect, if equality holds in (b).

A closer look at the definitions of perfectness and quasi–perfectness shows that the collection of cosets of a perfect *linear* channel code is a perfect WEM code; also the collection of the cosets of a quasi–perfect channel code is a quasi–perfect WEM code. On the other hand, we can prove very easily that any coset of a perfect linear WEM code is a perfect channel code and any coset of a quasi–perfect linear WEM code is a quasi–perfect channel code. By this duality the results (van Lint, 1975) for perfect linear codes lead to corresponding results for perfect linear WEM codes.

(2) The answer $R(d) = h(d)$ for the storage capacity of a Hamming WEM suggest a close connection to multi–user source coding. This can best be understood from an abstract point of view. In Ahlswede (1979, 1980) it was demonstrated that the essence of many multiuser source coding problems is a statement about vertex colorings of hypergraphs which assign different colors to almost all vertices in every edge.

The essence of Theorem 1 is a stetement of a similar kind. This can best be seen from the proof in Section 5. Indeed, let $(\Omega, \mathcal{E})$ be a hypergraph with vertex set Ω and edge set $\mathcal{E}$ (that is, $\big\{S_d(x^n, \Omega) | x^n \in \Omega\big\}$ in Section 5). We say that $(\Omega, \mathcal{E})$ carries M colors, if there is a vertex coloring with M colors such that *all* these colors occur in *every* edge. Let $M(\Omega, \mathcal{E})$ be the maximal number of colors carried by $(\Omega, \mathcal{E})$.

The derivation leading to (5.6) carried out in abstract gives the following result.

Coloring Lemma.

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

Since in typical applications the quantities $|\Omega|$ and $|\mathcal{E}|$ grow exponentially in the blocklength n, we have $M(\Omega, \mathcal{E}) \sim \min_{E \in \mathcal{E}} |E|$.

(3) We are grateful to N. Cai for having drawn out attention to the fact that the coloring problem, which consists in determining $M(\Omega, \mathcal{E})$ for any hypergraph $(\Omega, \mathcal{E})$, includes problems of Ramsey type.

To see this, let us choose for integers n, k, l with $n > k > l$, the l–element subsets $\binom{\mathcal{N}}{l}$ and the k–element subsets $\binom{\mathcal{N}}{k}$ of $\mathcal{N} = \{1, 2, \dots, n\}$ as vertex set (resp. edge set) of a hypergraph $\left(\binom{\mathcal{N}}{l}, \binom{\mathcal{N}}{k}\right)$, whose vertex–edge incidence structure is defined by set–theoretic containment. Now the classical Ramsey number $n(k, l)$ is the smallest integer such that for $n \geq n(k, l)$ the hypergraph $\left(\binom{\mathcal{N}}{l}, \binom{\mathcal{N}}{k}\right)$ satisfies $M\left(\binom{\mathcal{N}}{l}, \binom{\mathcal{N}}{k}\right) = 1$.

9 Suggestions for Further Research

Here we mention problems and directions which deserve further study. In some cases we already know the answers. Since they are obtained by combinations of known proofs we just state the results.

Code Constructions

Recently Zhang has found new WUM codes, which exceed in rate the ones known previously (Willems and Vinck, 1986; Simonyi, 1989). In our much more general class of

WEMs there are several interesting cases such as the Hamming case discussed in Section 8, in which constructive results should be obtainable. It is of course very challenging to find out whether the capacities can be achieved.

Perfect WEM Codes

One may try to classify the perfect and quasi–perfect non–linear WEM codes.

Error–Correcting WEM Codes

There may be errors in writing, printing, and in reading. The familiar concept of the error–correcting code can be combined with that of a WEM code. The following definition seems to us to be the most natural: $\{C_i | 1 \leq i \leq M\}$ is an (n, M, d_1, d_2) WEM code, if

(i) for all i, j $(i \neq j)$ and all $x^n \in C_i$, $y^n \in C_j$, $d_H(x^n, y^n) \geq d_1$,

(ii) for all $x^n \in \Omega_{d_1/2} = \big\{z^n : \exists y^n \in \Omega, d_H(z^n, y^n) < d_1/2\big\}$ and all j there is a $y^n \in C_j$ with $\varphi^n(x^n, y^n) \leq d_2$.

Example: $\{000\ 000\ 000\}$,$\{111\ 000\ 000, 000\ 111\ 111\}$, $\{000\ 111\ 000, 111\ 000\ 111\}$, $\{000\,000\,111, 111\,111\,000\}$ is a $(9, 4, 1, 4)$ WEM code with d_H as the cost function. Several interesting combinatorial problems arise already for special choices of the parameters d_1 and d_2.

Side Information

The issue of side information has played an important role in multi–user source and channel coding. We use the notation E_+ (resp. D_+), if the encoder (resp. decoder) has side information, and the notation E_- (resp. D_-), if the encoder (resp. decoder) does not have side information. For memory cells the side information refers to the knowledge of the contents of the memory before a new action (encoding or decoding) is taken. For WOM's the papers of Rivest and Shamir (1982) and Heegard (1983) consider the case (E_+, D_-), and in Wolf, Wyner, Ziv, and Körner (1984), the remaining cases are analysed.

Our results for WEM codes concern the case (E_+, D_-). Additional knowledge of the decoder has no effect; that is, our results also hold in case (E_+, D_+). We have the same independence of decoder's knowledge in case E_-.

There is an interesting case E_* between E_- and E_+, where the encoder *knows time.*

In a forthcoming paper "Multi–user WEM Codes" we prove that for *average costs* in obvious notation

$$\overline{R}_-(d) = \sup_{P : E_{P \times P}\varphi \leq d} H(P) \tag{9.1}$$

$$\overline{R}^*(d) = \lim_{n \to \infty} \frac{1}{k} \sup \left\{ \sum_{i=1}^{k} H(P_i) | P_1, \dots, P_k : \mathbb{E}_{P_i \times P_{i+1}} \varphi \leq d \right\}. \tag{9.2}$$

For *maximal costs* we have no final answers. They depend on the solution of two seemingly basic extremal problems, which in the Hamming case with $\mathcal{X} = \{0, 1\}$ are solved and known as diametric and isoperimetric problem (see Ahlswede and Katona, 1979):

Maximize $|\mathcal{A}|$ subject to the constraints $\mathcal{A} \subset \mathcal{X}^n$ and

$$\varphi^n(x^n, y^n) \leq d \;\text{ for all }\; x^n, y^n \in \mathcal{A}. \tag{9.3}$$

Maximize $|\mathcal{A}|\ |\mathcal{B}|$ subject to the constraints $\mathcal{A}, \mathcal{B} \subset \mathcal{X}^n$ and

$$\varphi^n(x^n, y^n) \leq d \;\text{ for all }\; x^n \in \mathcal{A}, y^n \in B. \tag{9.4}$$

Defects

Work on memory cells with defects was initiated in Kusnetsov and Tsybakov (1974). The papers of Ahlswede (1986), Ahlswede and Han (1983), Costa (1983), El Gamal and Greene (1983), Gelfand and Pinsker (1980), Heegard and El Gamal (1983), and Tsybakov (1977) either continue this line of investigation or relate to it.

Gelfand and Pinsker (1980) gives the most general result in case (E^+, D^-) and Ahlswede and Han (1983) gives the most general result in case (E^+, D^+), with partial side information. Both results are incorporated in a more general theorem with partial side information for encoder and decoder. Optimality is far from having been accomplished. The model is that of a memoryless channel with state $\mathcal{S}$, where the states are selected by an i.i.d. process. The side information concerns knowledge about the outcome of this process. By a combination of the methods, our results for WEM codes can mostly be extended to include cases of defects. One can also include aftereffects in the sense of Witsenhausen and Wyner (1983).

10 References

[1] R. Ahlswede, Coloring hypergraphs: A new approach to multi–user source coding, Part I, J. Combin. Inform. and System Sci., Vol. 4, No. 1, 76–115, 1979; Part II, J. Combin. Inform. and System Sci., Vol. 5, No. 3, 220–268, 1980.

[2] R. Ahlswede, Arbitrarily varying channels with state sequence known to the sender, IEEE Trans. Inform. Theory, IT–32, No. 5, 621–629, 1986.

[3] R. Ahlswede and T.S. Han, On source coding with side information via a multiple–access channel and related problems in multi–user information theory, IEEE Trans. Inform. Theory, IT–29, 396–412, 1983.

[4] R. Ahlswede and G. Katona, Contributions to the geometry of Hamming spaces, Discrete Math., 17, 1–22, 1979.

[5] J.M. Borden, Coding for write–unidirectional memories, IEEE Trans. Inform. Theory, in press.

[6] K.L. Chung, "Markov Chains", Springer, New York, 1967.

[7] G.D. Cohen and G. Simonyi, Coding for write–unidirectional memories and conflict resolution, Discrete Applied Mathematics, in press.

[8] M.H.M. Costa, Writing on dirty paper, IEEE Trans. Inform. Theory, IT–29, No. 3, 439–441, 1983.

[9] A.A. El Gamal and J.W. Greene, Storage rates for a restructurable memory, preprint, 1983.

[10] S.I. Gelfand and M.S. Pinsker, Coding for channel with random parameters, Problems Control and Inform. Theory, 9, No. 1, 19–31, 1980.

[11] C. Heegard, On the capacity of a permanent memory, in "1983 Conf. on Inform. Sciences and Systems, John Hopkins University", 1983.

[12] C. Heegard and A.A. El Gamal, On the capacity of computer memory with defects, IEEE Trans. Inform. Theory, IT–29, No. 5, 731–739, 1983.

[13] A.V. Kusnetsov and B.S. Tsybakov, Coding a memory with defective cells, transl. from Problemy Peredachi Informatsii, 10, No. 2, 52–60, 1974.

[14] R.L. Rivest and A. Shamir, How to reuse a write–once memory, Inform. and Control, 55, 1–19, 1982.

[15] G. Simonyi, On write–unidirectional memories, "Report Interne Enst 86 Doo 7", 1986.

[16] G. Simonyi, On write–unidirectional memory codes, IEEE Trans. Inform. Theory, 663–667, 1989.

[17] B.S. Tsybakov, Bounds for the codes correcting errors and defects, transl. from Problemy Perdachi Informatsii, 13, No. 2, 11–22, 1977.

[18] J.H. van Lint, Recent results on perfect codes and related topics, in "Combinatorics" (M. Hall, Jr. and J.H. van Lint, Eds.), Reidel Dordrecht, 1975.

[19] F.M.J. Willems and A.J. Vinck, Repeated recording for an optical disk, in "Proceedings, Seventh Symposium on Information Theory in the Benelux, May 1986", 49–53, Delft Univ. Press, 1986.

[20] H.S. Witsenhausen and A.D. Wyner, On storage media with aftereffects, Inform. and Control, 56, No. 3, 1983.

[21] J.K. Wolf, A.D. Wyner, J. Ziv, and J. Körner, Coding for write–once memory, AT& T Tech. J. 63, No. 6, 1098–1112, 1984.