rev. Oct. 1982

STORAGE RATES FOR A MEMORY WITH A SELECTOR †

Jonathan W. Greene *

Abbas El Gamal **

Abstract

We consider a memory composed of N discrete cells, each characterized by a defect state s drawn independently according to $p(s)$. The probability of retrieving a symbol y given s and the stored symbol x is completely specified by $p(y|x,s)$. The selector identifies a subset of "good" cells, which alone are used to store data, in an effort to improve the reliable storage rate of the memory.

For some fixed $r \le 1$, the selector chooses a subset of rN cells based on the states $[s_1,...,s_N] = \mathbf{s}$. The subset is denoted by a binary vector $\mathbf{u}$, where $u_i = 1$ if and only if the i^{th} cell is used. The symbols $[x_1,...,x_{rN}]$ are stored in order in the selected cells. A storage rate R is achieved if there exists a sequence of $(2^{RN}, rN)$ codes, selection rules $p(\mathbf{u}|\mathbf{s})$ and decoding rules such that the probability of error tends to zero.

The storage capacity is established for independent selection rules $p(\mathbf{u}|\mathbf{s}) = \prod_{i=1}^{N} p(u_i|s_i)$. It is then shown that the capacity is higher for the more general class of causal rules $p(\mathbf{u}|\mathbf{s}) = \prod_{i=1}^{N} p(u_i|s_1, \ldots, s_i)$. However, for the cell consisting of two binary symmetric channels (BSCs), the capacity for causal rules is achieved by an independent rule. A similar result holds for any two-state cell when the state is known to the decoder.

For arbitrary selection rules, rates higher than those possible with causal rules are achievable, even for two-state BSC cells. The capacity for arbitrary rules is as yet unknown.

---

† To appear, IEEE Trans. Inform. Theory.

* Department of Electrical Engineering, Stanford University. Partially supported by NSF Graduate Fellowship.

** Department of Electrical Engineering, Stanford University. Work partially supported under Air Force contract F 49620-79-C-0058, DARPA contract MDA 903-79-C-0608 and NSF Grant 80-26102.

1. Introduction

In many computer memory systems, it is possible to identify totally or partially defective storage locations by testing. Examples are media imperfections on magnetic disc or drum memories which decrease the signal to noise ratio and "leaky" or "stuck" storage cells or columns in integrated circuit random access memories (RAMs). Such memory systems can be modeled as a discrete memoryless channel with statistically determined states [1]. Several researchers have studied the achievable rates for this channel when information about the states is available to the encoder, decoder or both [1]-[3]. Heegard and El Gamal [3] give a set of achievable combinations of storage rates and rates of defect description to the encoder and decoder. For the case of no defect description, and also the cases of complete description of defects to either the encoder or decoder or both, the given achievable storage rate is optimal.

Although defect state information is often exploited in current memory technology, it is used in a simpler way than state dependent coding. With magnetic disk memories, the user simply avoids writing data in sectors which have been identified as defective [4]. Many commercially available 64Kbit RAMs employ either laser or electrically blown fuses which permit the connection of on-chip spares to replace defective memory cells [5] [6]. These strategies amount to mere avoidance of defective locations, rather than the use of state information in coding. Ordinary coding is still employed, of course, to ensure reliable storage in the presence of noise and any remaining partially defective cells.

In this paper, we investigate the storage capacity when defect information is used to skip over some of the cells. The memory is again modeled as a discrete memoryless channel with statistically determined states with the addition of a *selector* as depicted in Figure 1. Based on the defect information, the selector chooses some fixed fraction r of the N cells and connects them in order to the encoder/decoder. The storage rate of the memory is defined to be $R=K/N$.

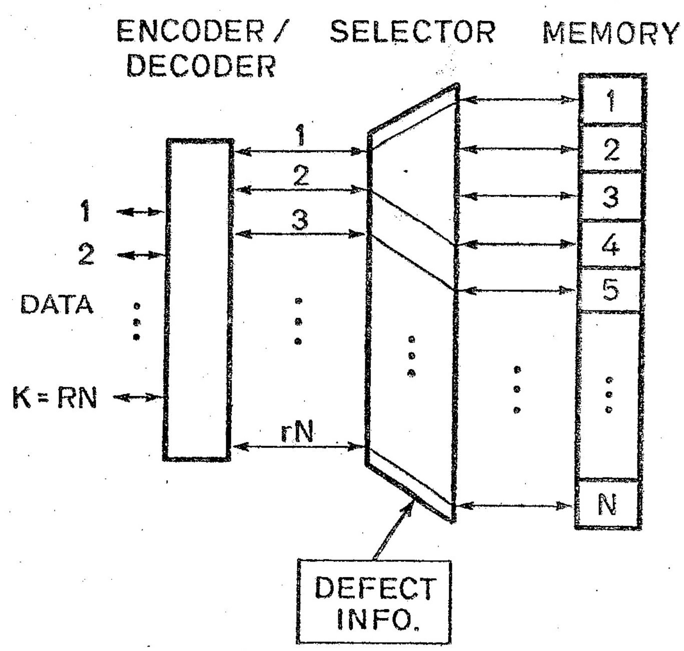

Fig. 1: Model for a memory with a selector.

The requirement that the selector preserve the order of the cells, though somewhat arbitrary, insures that the defect information is used only to skip cells, not as a basis for permuting them. Otherwise the cells could be reordered into groups by state and a separate encoder/decoder used for each group. This reduces the capacity problem to the case when both the encoder and decoder receive the defect state information. The complexity of such a permuting selector is certainly higher than that of the order-preserving one considered here.

Clearly, if a defective cell has zero storage capacity then the selector can only increase the overall capacity of the memory by skipping it. However, for a partially defective cell, the following tradeoff arises. If the selector is very demanding and selects only perfect cells, a small fraction of the cells will be used, limiting the storage capacity. If, on the other hand, the selector uses many defective cells, some mixing of defective and non-defective channels must occur due to the ordering restriction; this may lower the overall capacity of the memory. This tradeoff is examined in the following example.

Example A: Three state binary symmetric channel: clean, noisy and stuck.

Consider the channel shown in Figure 2, where X represents the storage symbol, Y the retrieval symbol and S one of three possible defect states.

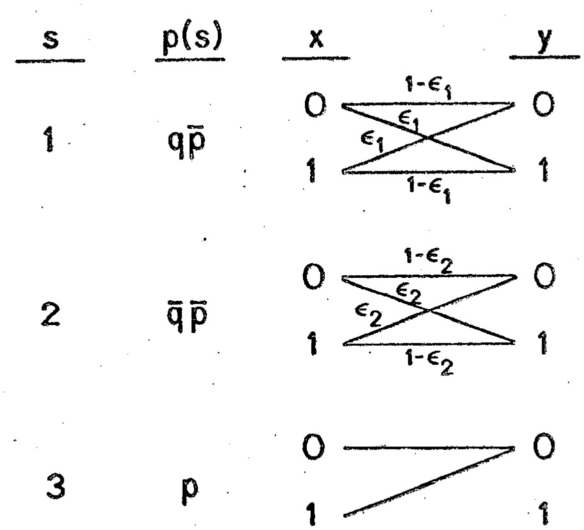

Fig. 2: Memory cell of Example A.

State 1, the "clean" state, is characterized by a binary symmetric channel (BSC) with crossover parameter ε_1. State 2, the "noisy" state, is characterized by a BSC with parameter ε_2, where $\varepsilon_1 \le \varepsilon_2 \le 1/2$. State 3 is stuck at zero. We have

$$p(y \mid x, s) = \begin{cases} \varepsilon_1 & y \neq x \quad s=1 \\ 1-\varepsilon_1 & y=x \quad s=1 \\ \varepsilon_2 & y \neq x \quad s=2 \\ 1-\varepsilon_2 & y=x \quad s=2. \\ 1 & y=0 \quad s=3 \\ 0 & y=1 \quad s=3 \end{cases}$$

Let $P(S=3)=p$, $P(S=1)=\bar{p}q$ and $P(S=2)=\bar{p}\bar{q}$, where $\bar{\vartheta}$ denotes $1-\vartheta$.

The simplest type of selection procedure would decide whether to use each cell based only on the state of the cell itself. This is termed an *independent selection rule*. Obviously, state 3 cells should be skipped since they can store no information. However, it is not so clear whether state 2 cells should be used or not. If only state 3 cells are skipped, a fraction $\bar{p}$ of the cells will be used. Of these, a fraction q will be of state 1 and $\bar{q}$ of state 2. The rate achieved is

$$R = \bar{p}\ [1-h(q\,\varepsilon_1+\bar{q}\,\varepsilon_2)].$$

If this mixing of states 1 and 2 lowers the rate too much, we might consider skipping state 2 cells as well. This achieves a rate

$$R = \bar{p}q\ [1-h(\varepsilon_1)].$$

As shown in Figure 3, each strategy achieves the higher rate in certain circumstances.

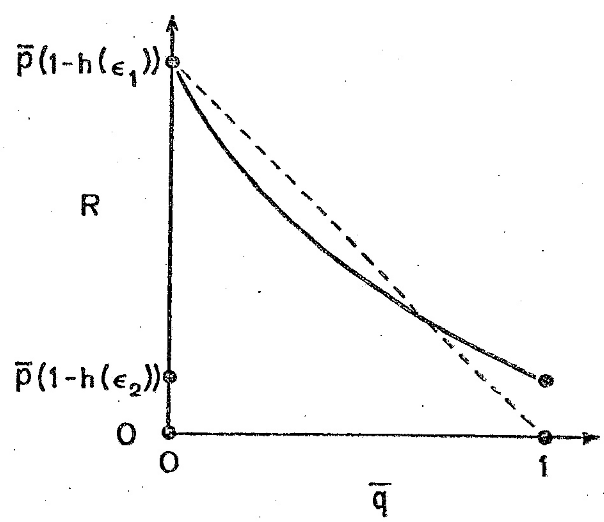

Fig. 3: Independent storage rates for Example A. Solid curve is $R=\bar{p}[1-h(q\,\varepsilon_1+\bar{q}\,\varepsilon_2)]$. Dashed line is $R=\bar{p}q\,[1-h(\varepsilon_1)]$. $\varepsilon_1=0.001$, $\varepsilon_2=0.27$.

Theorem 1 shows that the better of these two strategies achieves capacity for independent selection rules.

Can more complicated selection rules achieve higher rates? If we examine a *causal* selection rule, under which the decision to use the i^{th} cell is based on the states $s_1, \ldots, s_i$, the answer for this example is no. This is proved by Theorem 3.

The main reason for studying independent and causal rules is their simplicity, although there may be certain practical situations where these restrictions on selection rules apply.

If arbitrary selection rules are permitted, higher rates may be achieved for certain values of q, ε_1 and ε_2. This is done by skipping so as to bring as many state 1 cells as possible to odd numbered positions. Codes of different rates are then used for even and odd symbol positions. The coder has no knowledge of the state other than through the different statistics for the different positions.

The remainder of this paper focuses on storage rates for a memory with a selector. In Section 2, the problem is formally introduced. In Section 3, we establish that the storage capacity for independent selection rules is achieved by deterministic rules. We also present a simple extension of Shannon's result for causal state information at the encoder [1] to allow for the addition of an independent selection rule.

The example of Section 4 shows that the storage capacity for the more general class of causal rules is higher than for independent rules. However, for the cell consisting of two binary symmetric states, the capacity for causal rules is achieved by an independent rule. A similar result holds for any two-state cell when state information is known to the decoder.

The examples of Section 5 demonstrate that arbitrary rules can achieve higher rates than causal rules. The storage capacity for arbitrary selection rules is, as yet, unknown. We conclude with a tabular summary of the known capacities when the state information is available to one or a combination of the encoder, decoder or selector.

2. Definition of the Model

We now give a detailed definition of the model for a memory with a selector, shown in Figure 1. The memory itself is composed of N independent, discrete memory cells $[X, Y, S, p(s), p(y|x,s)]$. Each defect state $s \in S$ is chosen independently according to $p(s)$. The probability of retrieving a symbol $y \in Y$ given the stored symbol $x \in X$ and the state s is completely specified by a transition matrix $p(y|x,s)$. S, X and Y are assumed finite.

For some fixed *selection rate* $r \leq 1$, the selector chooses a subset of rN cells based on the defect states $[s_1,s_2,\ldots,s_N] = \mathbf{s}$. Thus the selection is completely specified by a vector $\mathbf{U}$, where for $1 \leq i \leq N$, $u_i = 1$ if the i^{th} cell is used and is zero otherwise. The conditional distribution $p(\mathbf{u}|\mathbf{s})$ is called a *selection rule*. Note that $\mathbf{U}$ and $\mathbf{X}$ are independent.

Storage symbols $[x_1,x_2,\ldots,x_{rN}] = \mathbf{x}$ are stored in order in the selected cells. That is, x_j, $1 \leq j \leq rN$, is stored in cell i if

$$U_i = 1 \text{ and } \sum_{k=1}^{i} U_k = j. \tag{2.1}$$

For notational convenience we define the random vector $\mathbf{T} \in S^{rN}$ by $t_j = s_i$ for i, j satisfying (2.1). Thus t_j is the state of the cell used to store x_j.

A code $[2^{RN}, rN]$ for the memory with a selector consists of a set of 2^{RN} equally likely messages $W = \{1,2,\ldots,2^{RN}\}$, an encoding function $f : W \to X^{rN}$, and a decoding function $g : Y^{rN} \to W$. The probability of decoding error, averaged over all messages, is given by

$$P_e = P(\, g(\mathbf{Y}) \neq W \,)$$

where $p(w,\mathbf{y})$ is evaluated based on the selection rule and memory cell statistics.

A storage rate R is achievable if there exists a sequence of $[2^{RN}, rN]$ codes and selection rules $p(\mathbf{u} \mid \mathbf{s})$ such that the probability of decoding error tends to zero with increasing N.

We define two subclasses of selection rules:

(i) independent selection rules, where $p(\mathbf{u} \mid \mathbf{s})$ is of the form

$$\prod_{i=1}^{N} p(u_i \mid s_i).$$

(ii) causal selection rules, where $p(\mathbf{u} \mid \mathbf{s})$ is of the form

$$\prod_{i=1}^{N} p(u_i \mid s_1, s_2, \ldots, s_i).$$

The storage capacity C for a given class of selection rules is the supremum over all rates achievable with rules in the class.

3. Independent Selection Rules

The independent selection rules $\prod_{i=1}^{N} P(u_i | s_i)$ allow for randomization in selecting the cells to be used. In the following theorem, we prove that the optimal independent rules are deterministic. Thus randomization is not necessary and any independent rule can be outperformed by a rule that selects a cell if and only if its state is in some subset v of S.

Theorem 1: For any memory cell $[X,Y,S,p(s),p(y|x,s)]$, independent selection rules achieve storage capacity

$$C_i = \max_{v \subseteq S} P(S \in v) \max_{p(x)} I(X;Y|S \in v). \tag{3.1}$$

Proof: Achievability is easily shown. Let $p(u_i|s_i) = 1(s_i \in v)$, where $1(\cdot)$ denotes the indicator function. By the law of large numbers, the fraction of cells used approaches $P(S \in v)$. Each used cell stores information at a rate given by the mutual information in (3.1).

The converse is more involved. We are given a $(2^{RN}, rN)$ code for the cell $[X,Y,S,p(s),p(y|x,s)]$ and an independent selection rule $\prod_{i=1}^{N} p(u_i|s_i)$. In order to take account of the effect of the unused cells on the capacity, we define the random vector $\mathbf{Z} \in (Y \cup \{e\})^N$, $e \notin Y$, by the following one-to-one function of $\mathbf{Y}$ and $\mathbf{U}$. For $j(i,\mathbf{u}) = \sum_{k=1}^{i} u_k$, $1 \le i \le N$, let

$$z_i = \begin{cases} y_{j(i,\mathbf{u})} & \text{if } u_i = 1 \\ e & \text{if } u_i = 0 \end{cases} \tag{3.2}$$

The symbol e may be interpreted as an erasure. We define

$$p(w,\mathbf{x},\mathbf{z},\mathbf{s},\mathbf{u}) = p(w)p(\mathbf{x}|w)\prod_{i=1}^{N} p(s_i)p(u_i|s_i)\left[1(u_i=1)P(Y=z_i|x_{j(i,\mathbf{u})},s_i) + 1(u_i=0)1(z_i=e)\right].$$

Note that the induced conditional distribution on $\mathbf{Z}$ given $\mathbf{X}$ and $\mathbf{U}$ is of the form

$$p(\mathbf{z}|\mathbf{x},\mathbf{u}) = \prod_{i=1}^{N} p(z_i | x_{j(i,\mathbf{u})}, u_i) \tag{3.3}$$

where

$$p(z_i \mid x_{j(i,\mathbf{u})}, u_i) = \frac{\sum_{s_i} p(s_i) p(u_i \mid s_i) \left[1(u_i=1) P(Y=z_i \mid x_{j(i,\mathbf{u})}, s_i) + 1(u_i=0) 1(z_i=e) \right]}{\sum_{s_i} p(s_i) p(u_i \mid s_i)}.$$

By Fano's Inequality, for $\varepsilon_N \to 0$,

$$\begin{aligned}
R &\le \frac{1}{N} I(\mathbf{X};\mathbf{Y}) + \varepsilon_N \\
&\le \frac{1}{N} I(\mathbf{X};\mathbf{Y},\mathbf{U}) + \varepsilon_N \\
&= \frac{1}{N} I(\mathbf{X};\mathbf{Z},\mathbf{U}) + \varepsilon_N \qquad \text{((3.2) is one-to-one)} \\
&= \frac{1}{N} I(\mathbf{X};\mathbf{Z} \mid \mathbf{U}) + \varepsilon_N \\
&= \frac{1}{N} E\, I(\mathbf{X};\mathbf{Z} \mid \mathbf{u}) + \varepsilon_N \\
&\le \frac{1}{N} E \sum_{i=1}^{N} I(X_{j(i,\mathbf{u})}; Z_i \mid u_i) + \varepsilon_N
\end{aligned}$$

since $p(\mathbf{z} \mid \mathbf{x},\mathbf{u})$ is a product distribution (3.3). Note that the mutual information depends on u_k, $k \ne i$, only through $j(i,\mathbf{u})$ and thence through the corresponding distribution $p(x_{j(i,\mathbf{u})})$. We escape this dependence by maximizing over all distributions $p(x_i)$:

$$\begin{aligned}
R &\le \frac{1}{N} E \sum_{i=1}^{N} \max_{p(x_i)} I(X_i;Z_i \mid u_i) + \varepsilon_N \\
&\le \max_i E \max_{p(x_i)} I(X_i;Z_i \mid u_i) + \varepsilon_N \\
&= \max_i \max_{p(x_i)} I(X_i;Z_i \mid U_i) + \varepsilon_N && (3.4) \\
&\le \max_{p(u \mid s)} \max_{p(x)} I(X;Z \mid U) + \varepsilon_N . && (3.5)
\end{aligned}$$

Equation (3.4) follows by noting that $z=e$ (an erasure) occurs when $U_i=0$ and so $I(X_i;Z_i \mid U_i=0)=0$ regardless of $p(x_i)$. Thus the maximization over $p(x_i)$ can be moved outside the expectation over U_i.

Now any $p(u \mid s)$ can be represented as a mixture of deterministic distributions. More formally, for any $p(u \mid s)$ there exists a probability mass function $p(v)$ over $\{v : v \subseteq S\}$ such that $\sum_v p(v) p(u \mid s,v) = p(u \mid s)$ for every $u \in U$ and

$s \in S$, where

$$p(u \mid s,v) = \begin{cases} 1 & \text{if } u=1 \text{ and } s \in v \\ 1 & \text{if } u=0 \text{ and } s \in v \\ 0 & \text{otherwise.} \end{cases}$$

We then have the joint probability mass function

$$p(x,z,s,v,u) = \\ p(x)\,p(v)\,p(s)\,p(u \mid s,v)\left[1(u=1)P(Y=z \mid x,s)+1(u=0)1(z=e)\right] \quad (3.6)$$

It can easily be shown that under (3.6):

$$I(X;Z \mid U=1, V=v) = P(S \in v)\; I(X;Y \mid S \in v), \text{ and} \\ I(X;Z \mid U=0, V=v) = 0. \quad (3.7)$$

Returning to (3.5), we have

$$\begin{aligned} R &\le \max_{p(u|s)} \max_{p(x)} I(X;Z \mid U) + \varepsilon_N \\ &= \max_{p(v)} \max_{p(x)} I(X;Z \mid U) + \varepsilon_N \\ &\le \max_{p(v)} \max_{p(x)} I(X;Z \mid U,V) + \varepsilon_N \qquad \text{(independence)} \\ &\le \max_{v} \max_{p(x)} I(X;Z \mid U,v) + \varepsilon_N \\ &= \max_{v} P(S \in v) \max_{p(x)} I(X;Y \mid S \in v) + \varepsilon_N \end{aligned}$$

by (3.7) and the theorem is proved. ∎

Specializing Theorem 1 to two defect states yields the following corollary.

Corollary 1: For the case $S = \{1,2\}$ with $P(S=1) = q$ and $P(S=2) = \bar{q}$, independent selection rules achieve capacity

$$C_{i2} = \max\left\{ \max_{p(x)} I(X;Y),\; q \max_{p(x)} I(X;Y \mid S=1),\; \bar{q} \max_{p(x)} I(X;Y \mid S=2) \right\}.$$

The following example demonstrates the use of the above corollary. It will be considered again in Section 4.

Example B1: Binary memory with three retrieval symbols.

Consider the memory cell of Figure 4, with $X=\{0,1\}$, $Y=\{a,b,c\}$, $S=\{1,2\}$, and $P(S=1) = 1/2$. (This situation might arise if the signal retrieved from

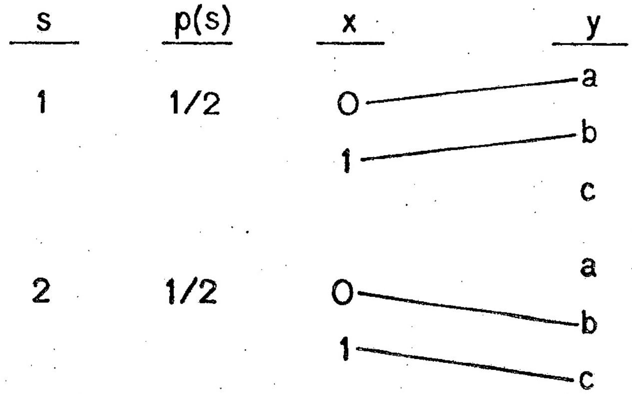

Fig. 4: Memory cell of Example B.

defective cells is offset from normal readout levels). Corollary 1 gives the following capacity for this memory under an independent selection rule.

$$C = C_{i2} = \max\left\{\max_{p(x)} \tfrac{1}{2} I(X;Y|S=1),\ \max_{p(x)} \tfrac{1}{2} I(X;Y|S=2),\ \max_{p(x)} I(X;Y)\right\}.$$

The first two terms ore obviously equal to ½ bit. The third term corresponds to a mixture of the two states identical to a binary erasure channel with parameter ½, which has a capacity of ½ bit. Thus $C = \tfrac{1}{2}$ bit.

The following corollary extends Theorem 1 to allow for the provision of defect information to the decoder.

Corollary 2: If **t**, the vector containing the states of the selected cells, is known to the decoder, the capacity for independent rules is

$$C_{id} = \max_{p(x)} I(X;Y|S)$$

Proof: This rate is obviously achievable by using all cells. The converse is a straightforward variation of the proof of Theorem 1 and is therefore omitted.

The following example demonstrates the use of Corollary 2. It will be considered again in Section 5.

Example C1: Channel with defect information at the decoder.

Consider the memory cell of Figure 5 with $X=\{a,b,c,d\}$, $Y=\{0,1\}$, $S=\{1,2\}$ and $P(S=1)= 1/2$. Assume in addition that the decoder is given t.

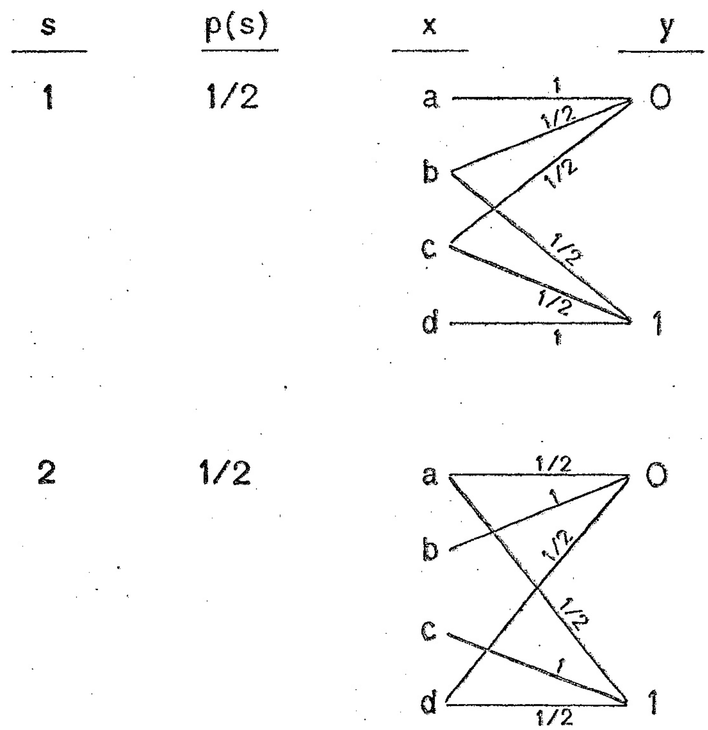

Fig. 5: Memory cell of Example C.

The capacity is given by Corollary 2 as $\max_{p(x)} I(X;Y|S)$. With $p(a)$, $p(b)$, $p(c)$, and $p(d)$ denoting the storage symbol probabilities, it may be shown that

$$I(X;Y|S=1) = E \log_2 \frac{p(Y|X,S=1)}{p(Y|S=1)}$$
$$= h\left[p(a)+ \tfrac{1}{2}p(b)+ \tfrac{1}{2}p(c)\right]-\left[p(b)+p(c)\right].$$

We can determine $I(X;Y|S=2)$ similarly, and thus

$$I(X;Y|S) = \tfrac{1}{2}h\left[p(a)+ \tfrac{1}{2}p(b)+ \tfrac{1}{2}p(c)\right] + \tfrac{1}{2}h\left[p(b)+ \tfrac{1}{2}p(a)+ \tfrac{1}{2}p(d)\right] - \tfrac{1}{2}$$
$$\leq \tfrac{1}{2}1 + \tfrac{1}{2}1 - \tfrac{1}{2} = \tfrac{1}{2}$$

But a mutual information of 1/2 bit is achieved by a uniform distribution on the storage symbols. Thus the storage capacity is 1/2 bit.

In general, the addition of a selector may permit higher rates of storage even if defect information is already provided to the encoder. This was the case in Example A. In the particular case that the encoder receives causal information only and the selection rule is independent, the storage capacity can be found by an application of Theorem 1.

Provision of causal state information to the encoder was considered by Shannon [1]. The input symbol $x_j \in X$ is allowed to depend on the current and previous states $s_1, s_2, \ldots, s_j$. As usual, the states are assumed to be independent. Without loss of generality, we take $S = \{1,2,\ldots,||S||\}$. Shannon proved that this channel is equivalent to the derived channel where the input symbol $\mathbf{x}_j \in X^{||S||}$ is independent of the states and represents a mapping $[x_1, \ldots, x_{||S||}]$ from the state alphabet S to the actual channel input alphabet X. Furthermore, if the states are identically distributed, the capacity of the channels is given by

$$C = \max_{p(x_1,\ldots,x_{||S||})} I(X_1, \ldots, X_{||S||}; Y),$$

where the mutual information is computed under the conditional distribution of the derived channel

$$p(y \mid x_1, \ldots, x_{||S||}) = \sum_s p(s)\, p(y \mid x_s, s).$$

This result can be extended to give the capacity of a memory with causal state information at the encoder and a selector operating under an independent selection rule. First, observe that under an independent selection rule the selected states $\{t_j\}$ are independent. Therefore, Shannon's equivalence result applies to the channel formed by the selected cells: simply substitute t_j for s_j and the induced distribution $p(t_j)$ for $p(s_j)$. A storage rate can be achieved with a given selection rule and state-dependent encoding algorithm if and only if the same rate can be achieved with the same selection rule for the derived channel. Since the derived channel does not provide state information to the encoder, its capacity is given by Theorem 1. We have the following corollary.

Corollary 3: When the encoder is provided with causal state information $t_1, t_2, \ldots t_j$ before symbol x_j is stored, the capacity for independent selection rules is

$$C = \max_{v \subseteq S} P(S \in v) \max_{p(x_1,\ldots,x_{||S||})} I(X_1, \ldots, X_{||S||}; Y | S \in v).$$

4. Causal Selection Rules

In general, causal rules achieve higher rates than independent rules. This is demonstrated by reconsidering Example B.

Example B2.

Consider the following causal rule for the cell of Example B1 (Figure 4):

$$P(U_i=1\,|\,s_1,s_2,\ldots,s_i)=\begin{cases}1 & j=\sum_{k=1}^{i-1}u_k+1 \text{ is odd.}\\ 1 & j \text{ is even and } s_i \text{ matches state of previous used cell.}\\ 0 & \text{otherwise.}\end{cases}$$

Breaking **t**, the vector of selected states, into pairs, we find that each pair of states are either both 1 or both 2, with equal probability. For example, **s**=[11211222121] would yield **u**=[11100111101] and **t**=[11222211].

Let L_e be the number of cells skipped prior to a cell selected for an even position in the **t** vector and after the previous selected cell. Let L_o be defined similarly for odd positions. Then

$$E\ L_o = 0 \text{ and}$$
$$E\ L_e = 0\,\frac{1}{2}+1\,\frac{1}{4}+2\,\frac{1}{8}+3\,\frac{1}{16}\cdots = 1.$$

Thus the expected number of unused cells per pair of selected cells is 1, for a selection rate $r=\frac{2}{3}$.

We now compute the capacity of the channel consisting of one pair of used cells with identical but unknown states, shown in Figure 6. A maximum mutual information $I(X_1,X_2;Y_1,Y_2)\approx 1.77155$ bits is achieved by input distribution $p(00)=p(11)=.207$, $p(10)=p(01)=.293$. The storage rate achieved is

$$R = r \max_{p(x_1,x_2)} \tfrac{1}{2} I(X_1,X_2;Y_1,Y_2)$$
$$\approx \frac{2}{3}\,\frac{1}{2}(1.77155) = .5905 \text{ bits}$$

which exceeds the independent rule capacity $C_{i2}=\tfrac{1}{2}$.

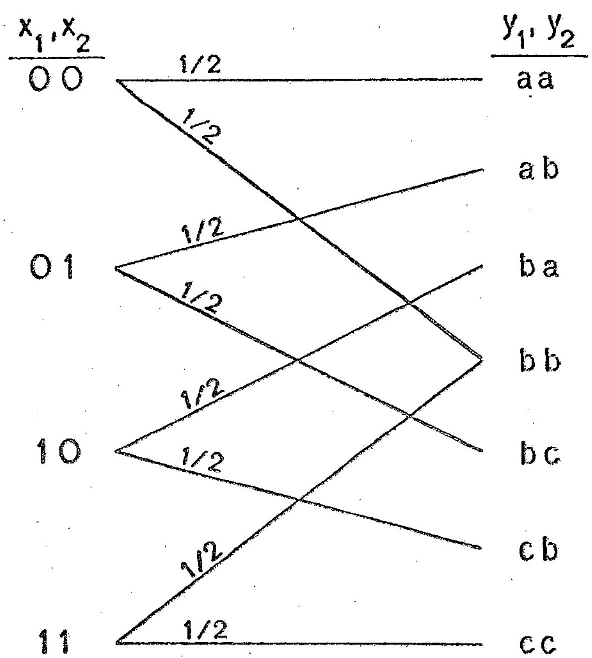

Fig. 6: Pairwise channel for Example B2.

In order to study causal rules, we must find a way to relate the probability distribution on **T**, the states of the selected cells, and the expected number of cells which have to be skipped to achieve that distribution. The next theorem is a step in this direction. It gives the minimum expected number of unused cells for any causal selection rule and two-state memory cell in terms of the distribution of T_j conditioned on any function of the past states. This bound will be used to prove Theorems 3 and 4.

Theorem 2: Take $S=\{1,2\}$ with $P\{S=1\} = q$. Let L_j be the number of unused cells between the $j-1^{st}$ and j^{th} used cells, as shown in Figure 7. (L_1 is the number of unused cells prior to the first used cell). Let Z be any random variable independent of the set of "future" states, $\{s_i, s_{i+1}, \ldots, s_N\}$, where i is the smallest integer such that $j = \sum_{k=1}^{i-1} u_k + 1$. Finally, define $L_j(\alpha)$ as the smallest $E(L_j \mid z)$ under any causal selection rule for which

$$\alpha = P\{T_j=1 \mid Z=z\}.$$

Then for all j, $1 \le j \le rN$,

$$L_j(\alpha) = \max\left\{\frac{q-\alpha}{1-q},\ \frac{\alpha-q}{q}\right\}.$$

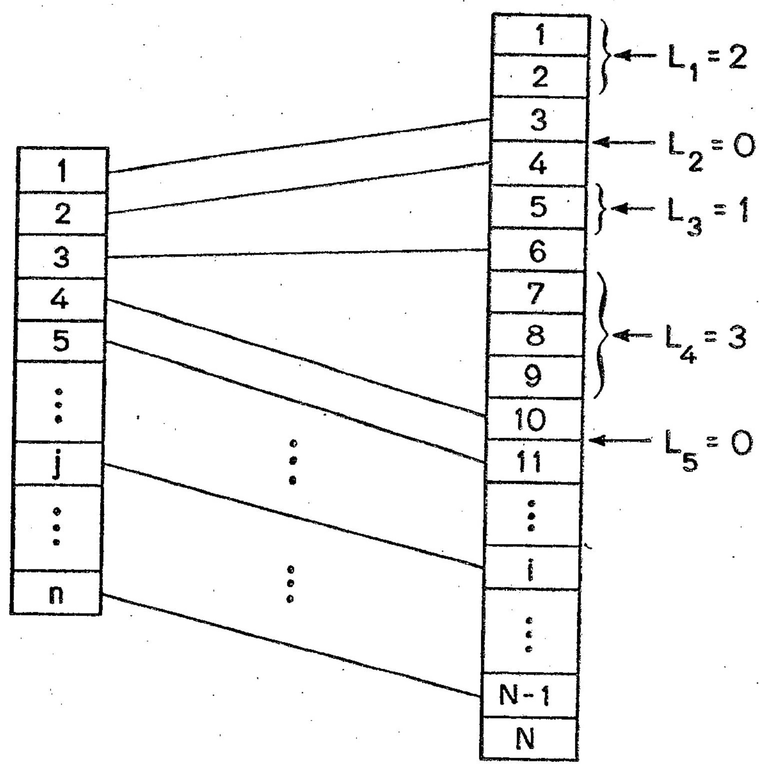

Fig. 7: A selection of cells.

Proof: Since the future states are independent of each other and of Z,

$$p(s_i, s_{i+1}, \ldots s_N \mid z) = \prod_{k=i}^{N} p(s_k).$$

Consider any particular $L_j(\alpha)$ and drop the subscript. Begin with the case $q \le \alpha \le 1$. For $\alpha=q$ it is obvious that $L(q)=0$ since we can simply use s_i without skipping at all. For $\alpha=1$,

$$L(1) = 0\,q + 1\,\bar{q}q + 2\,\bar{q}^2 q + \cdots = \bar{q}/q.$$

It is easily shown that $L(\alpha) \leq (\alpha - q)/q$ for $q \leq \alpha \leq 1$. Given any z and a desired value of $\alpha = P(T=1|z)$, construct the following rule: with probability $(\alpha - q)/\bar{q}$ we keep skipping until a state 1 cell is reached and with probability $1-(\alpha-q)/\bar{q}$ we do not skip at all. Under this rule, $E(L|z) = (\alpha - q)/q$, and by definition, $L(\alpha) \leq E(L|z)$.

Now we show that this bound on $L(\alpha)$ is tight. Suppose $L(\alpha) < (\alpha - q)/q$ for some $q < \alpha < 1$. Then by definition there must exist a z and a selection rule under which $P(T=1|z) = \alpha$ and $E(L|z) = L(\alpha) < (\alpha - q)/q$. Modify the rule as follows: whenever the rule would use a state 2 cell, which happens with probability $\bar{\alpha}$, skip it and start again from that cell applying the same rule. Continue to start over whenever a state 2 cell would be selected, so that the selection of a state 1 cell is assured.

The modified rule achieves a new value of:

$$P(T=1|z) = \alpha + \bar{\alpha}\left[\alpha + \bar{\alpha}\left[\cdots\right]\cdots\right] = 1$$

and

$$\begin{aligned} E(L|z) &= L(\alpha) + \bar{\alpha}\left[L(\alpha)+1\right] + \bar{\alpha}^2\left[L(\alpha)+1\right]\cdots \\ &= \left[L(\alpha)+1\right]/\alpha - 1. \end{aligned}$$

Then

$$\begin{aligned} L(1) &\leq \left[L(\alpha)+1\right]/\alpha - 1 \\ &< \left[\frac{\alpha - q}{q} + 1\right]/\alpha - 1 \\ &= \frac{\bar{q}}{q} \end{aligned}$$

which is a contradiction since $L(1) = \bar{q}/q$. Thus $L(\alpha) = (\alpha - q)/q$ for $q \leq \alpha \leq 1$. The result for $0 \leq \alpha \leq q$ follows by symmetry, concluding the proof. ∎

We now apply Theorem 2 to demonstrate that for certain two-state memory cells causal selection rules do no better than independent rules.

Theorem 3: For a memory cell consisting of two binary symmetric channels with crossover probabilities ε_1 and ε_2, and with $P(S=1)=q$, the storage capacity for causal selection rules is equal to the capacity for independent selection rules and is thus given by

$$C = C_{i2} = \max\left\{\bar{q}[1-h(\varepsilon_2)],\ [1-h(q\,\varepsilon_1+\bar{q}\,\varepsilon_2)],\ q[1-h(\varepsilon_1)]\right\}.$$

Proof: Since an independent rule is also a causal rule, achievability follows from Corollary 1. To prove the converse, we first reduce consideration to rules with only two or fewer distinct values of $P(T_j=1)$, $1\le j\le rN$.

Given any causal selection rule $\prod_{i=1}^{N} p(u_i\,|\,s_1,\ldots,s_i)$ with selection rate r, let $z_j=x_j\oplus y_j$ and $\mathbf{z}_j^-=\{z_1,\ldots z_{j-1}\}$. Define the following probability density functions for $0\le\alpha\le1$ and $1\le j\le rN$:

$$f_j(\alpha\,|\,\mathbf{z}_j^-)=\delta\left[\alpha-P(T_j=1\,|\,\mathbf{z}_j^-)\right] \quad \text{and}$$
$$f(\alpha)=\frac{1}{rN}\sum_{j=1}^{rN} E\, f_j(\alpha\,|\,\mathbf{z}_j^-),$$

where $\delta(x)$ is a unit area impulse at $x=0$, the Dirac delta function.

By Fano's Inequality and the symmetry of the channel,

$$\begin{aligned}
R &\le \frac{1}{N} I(\mathbf{X};\mathbf{Y})+\varepsilon_N\\
&\le \frac{1}{N}\sum_{j=1}^{rN}\left[1-H(Z_j\,|\,\mathbf{Z}_j^-)\right]+\varepsilon_N\\
&= \frac{1}{N}\sum_{j=1}^{rN} E\left[1-h\left[\varepsilon_1 P(T_j=1\,|\,\mathbf{z}_j^-)+\varepsilon_2 P(T_j=2\,|\,\mathbf{z}_j^-)\right]\right]+\varepsilon_N\\
&= \frac{1}{N}\sum_{j=1}^{rN} E\int_0^1 f_j(\alpha\,|\,\mathbf{z}_j^-)\left[1-h(\alpha\varepsilon_1+\bar{\alpha}\varepsilon_2)\right]d\alpha+\varepsilon_N\\
&\qquad\qquad\qquad \text{(by definition of } f_j(\alpha\,|\,\mathbf{z}_j^-)\text{)}\\
&= r\int_0^1 \frac{1}{rN}\sum_{j=1}^{rN} E\, f_j(\alpha\,|\,\mathbf{z}_j^-)\left[1-h(\alpha\varepsilon_1+\bar{\alpha}\varepsilon_2)\right]d\alpha+\varepsilon_N\\
&= r\int_0^1 f(\alpha)\left[1-h(\alpha\varepsilon_1+\bar{\alpha}\varepsilon_2)\right]d\alpha+\varepsilon_N
\end{aligned} \tag{4.1}$$

by the definition of $f(\alpha)$. We denote the quantity in brackets by $I(\alpha)$.

Note also that with L_j and $L(\alpha)$ defined as in Theorem 2, we have the following condition if the scheme is to skip few enough cells to achieve the given selection rate r:

$$
\begin{aligned}
N(1-r) &\geq E \sum_{j=1}^{rN} L_j \\
&= \sum_{j=1}^{rN} E\, E(L_j \mid \mathbf{z}_j^-) \\
&\geq \sum_{j=1}^{rN} E\, L(\, P(T_j=1 \mid \mathbf{z}_j^-)\,) \qquad \text{(by Theorem 2)} \\
&= \sum_{j=1}^{rN} E \int_0^1 f_j(\alpha_j \mid \mathbf{z}_j^-)\, L(\alpha_j)\, d\alpha_j
\end{aligned}
$$

by the definition of $f_j(\alpha_j \mid \mathbf{z}_j^-)$. Dividing by rN and expressing the integral in terms of $f(\alpha)$ we have

$$\frac{1-r}{r} \geq \int_0^1 f(\alpha)\, L(\alpha)\, d\alpha. \qquad (4.2)$$

We now employ the extended Ahlswede-Korner lemma as stated in [7].

Lemma: Let R be any subset of $\mathbf{R}^n$ consisting of at most k connected subsets. Let F_m, $m=1,2,...k$ be real valued continuous functions on R. Then for any probability measure p on R there exist k elements α_i of R and constants $p_i \geq 0$, $\sum_{i=1}^{k} p_i = 1$, such that for all m

$$\int_R F_m(\alpha)\, dp(\alpha) = \sum_{i=1}^{k} p_i F_m(\alpha_i) .$$

Returning to the proof of Theorem 3, we know from the lemma that there exist p_1, $p_2 = 1-p_1$, α_1 and α_2, all between zero and one, such that

$$
\begin{aligned}
p_1 I(\alpha_1) + p_2 I(\alpha_2) &= \int_0^1 f(\alpha)\, I(\alpha)\, d\alpha \\
&\geq (R-\varepsilon_N)/r \qquad (4.3)
\end{aligned}
$$

and

$$p_1 L(\alpha_1) + p_2 L(\alpha_2) = \int_0^1 f(\alpha)\, L(\alpha)\, d\alpha$$

$$\leq \frac{1-r}{r}. \tag{4.4}$$

The inequalities follow from (4.1) and (4.2), respectively.

We next show that consideration may be further reduced from two to only one distinct value of α. Solving (4.3) for R and using (4.4) to substitute for r yields

$$R \leq \left[p_1 L(\alpha_1) + p_2 L(\alpha_2) + 1\right]^{-1} \left[p_1 L(\alpha_1) + p_2 L(\alpha_2)\right] + \varepsilon_N$$
$$\leq \max_{\alpha \in \{\alpha_1, \alpha_2\}} \left[L(\alpha) + 1\right]^{-1} I(\alpha) + \varepsilon_N$$

by some simple calculus (see Appendix 1). Expanding the set over which the maximization is taken to include $0 \leq \alpha \leq 1$ we have

$$R \leq \max_{\alpha} \left[L(\alpha)+1\right]^{-1} I(\alpha) + \varepsilon_N$$
$$\leq \max_{\alpha \in \{0, q, 1\}} \left[L(\alpha)+1\right]^{-1} I(\alpha) + \varepsilon_N$$
$$= \max \left\{\bar{q}\left[1-h(\varepsilon_2)\right], \left[1-h(q\varepsilon_1+\bar{q}\varepsilon_2)\right], q\left[1-h(\varepsilon_1)\right]\right\} + \varepsilon_N.$$

The second inequality follows from the differentiability of $L(\alpha)$ in the intervals $(0,q)$ and $(q,1)$, and from the convexity of $I(\alpha)$ (see Appendix 2). The equality follows from the definitions of $L(\alpha)$ and $I(\alpha)$, and Theorem 3 is proved. ▪

We now prove a similar result for any two-state memory cell if the state information is available at the decoder.

Theorem 4: For any two-state memory cell if the decoder is given **t**, the state information for the selected cells, the storage capacity for causal selection rules is equal to the capacity for independent rules and is thus given by

$$C = C_{id} = \max_{p(x)} I(X;Y|S).$$

Proof: Achievability follows from Corollary 2, so we deal only with the converse.

Given a $(2^{RN}, rN)$ code for the two state cell $[X, S, Y, p(s), p(y|x,s)]$ with $S=\{1,2\}$ and $P(S=1) = q$ and a causal selection rule $\prod_{i=1}^{N} p(u_i|s_1, \ldots, s_i)$, define the joint probability mass function

$$p(w,\mathbf{x},\mathbf{y},\mathbf{s},\mathbf{u},\mathbf{t}) = p(w)p(\mathbf{x}\,|\,w)\prod_{i=1}^{N}[p(s_i)p(u_i\,|\,s_1,\ldots,s_i)]\,p(\mathbf{t}\,|\,\mathbf{s},\mathbf{u})\prod_{j=1}^{rN}p(y_j\,|\,x_j,S=t_j).$$

Note that the induced conditional distribution on $\mathbf{Y}$ given $\mathbf{X}$ and $\mathbf{T}$ is of the form

$$p(\mathbf{y}\,|\,\mathbf{x},\mathbf{t}) = \prod_{j=1}^{rN} p(y_j\,|\,x_j,S=t_j). \tag{4.5}$$

Let $p_j(t_j)$ and $p_j(x_j)$ be the induced marginal distributions on T_j and X_j. Then the joint marginal distribution is

$$p_j(x_j,y_j,t_j) = p_j(x_j)p_j(t_j)p(y_j\,|\,x_j,\ S=t_j).$$

We wish to make use of Lemma 1 to show that only two distinct joint marginal distributions $p_j(x_j,y_j,t_j)$ need be considered. To do this we need to define functions $I(\cdot)$ and $L(\cdot)$ which are continuous over some connected set of distributions that includes all the given marginal distributions. We are led to define a random variable Z, $1 \le Z \le rN$, with probability density function

$$f(z) = \frac{1}{rN}\sum_{j=1}^{rN}\delta(z-j).$$

We define the conditional distribution $p(x,y,t\,|\,z)$ by mixing the two distributions whose indices are the floor and ceiling of z:

$$p(x,y,t\,|\,z) =$$
$$\Big[(z-\lfloor z\rfloor)p_{\lceil z\rceil}(t) + (\lfloor z\rfloor+1-z)p_{\lfloor z\rfloor}(t)\Big]\Big[(z-\lfloor z\rfloor)p_{\lceil z\rceil}(x) + (\lfloor z\rfloor+1-z)p_{\lfloor z\rfloor}(x)\Big]\,p(y\,|\,x,S=t).$$

Note that $p(x,y,t\,|\,z)$ is continuous in z and that

$$p(x,y,t\,|\,z) = p_z(x,y,t), \qquad z\in\{1,2,\ldots rN\}. \tag{4.6}$$

Proceeding in the same fashion as in Theorem 3 we define

$$I(z) = I(X;Y\,|\,T,z),$$
$$\alpha(z) = P(T=1\,|\,z) = \sum_x\sum_y p(x,y,T=1\,|\,z) \text{ and}$$
$$L(\alpha(z)) = \max\left\{\frac{q-\alpha(z)}{1-q},\ \frac{\alpha(z)-q}{q}\right\}.$$

By Fano's Inequality, for $\varepsilon_N \to 0$,

$$R \le \frac{1}{N}I(\mathbf{X};\mathbf{Y}\,|\,\mathbf{T}) + \varepsilon_N$$

$$\leq \frac{1}{N}\sum_{j=1}^{rN} I(X_j;Y_j\,|\,T_j) + \varepsilon_N \qquad \text{(under distribution (4.5))}$$

$$= \frac{1}{N}\sum_{j=1}^{rN} I(X;Y\,|\,T,Z=j) + \varepsilon_N \qquad \text{(by (4.6))}$$

$$= r\int_1^{rN} f(z)I(z)dz + \varepsilon_N \qquad (4.7)$$

by the definitions of $f(z)$ and $I(z)$.

Now, with L_j defined as in Theorem 2 we have

$$N(1-r) \geq E\sum_{j=1}^{rN} L_j$$

$$= \sum_{j=1}^{rN} E\ L_j$$

$$\geq \sum_{j=1}^{rN} L(\alpha(j)) \qquad \text{(by Theorem 2)}$$

$$= rN\int_1^{rN} f(z)L(\alpha(z))dz. \qquad (4.8)$$

Applying Lemma 1 to (4.7) and (4.8) we obtain

$$(R-\varepsilon_N)/r \leq p_1 I(z_1) + p_2 I(z_2) \quad \text{and}$$

$$\frac{1-r}{r} \geq p_1 L(\alpha(z_1)) + p_2 L(\alpha(z_2))$$

for some p_1, $p_2=1-p_1$, z_1 and z_2. Then

$$R \leq \left[p_1 L(\alpha(z_1)) + p_2 L(\alpha(z_2)) + 1\right]^{-1}\left[p_1 I(z_1) + p_2 I(z_2)\right] + \varepsilon_N$$

$$\leq \max_{z\in\{z_1,z_2\}}\left[L(\alpha(z))+1\right]^{-1} I(z) + \varepsilon_N \qquad \text{(see Appendix 1)}$$

$$\leq \max_{p(z)}\max_{0\leq\alpha\leq 1}\left[\max\left\{\frac{q-\alpha}{1-q},\ \frac{\alpha-q}{q}\right\} + 1\right]^{-1}$$

$$\left[\alpha I(X;Y\,|\,S=1) + (1-\alpha)I(X;Y\,|\,S=2)\right] + \varepsilon_N$$

by the definitions of $L(\alpha)$ and $I(z)$ and (4.6).

We now perform the maximization of this expression over α. For $\alpha>q$,

$$\frac{d}{d\alpha}\left[\frac{\alpha-q}{q} + 1\right]^{-1}\left[\alpha I(X;Y\,|\,S=1) + (1-\alpha)I(X;Y\,|\,S=2)\right]$$

$$= \frac{-q}{\alpha^2} I(X;Y\,|\,S=2) \leq 0.$$

For $\alpha<q$,

$$\frac{d}{d\alpha}\left[\frac{q-\alpha}{1-q}+1\right]^{-1}\left[\alpha I(X;Y|S=1)+(1-\alpha)I(X;Y|S=2)\right]$$
$$=\frac{1-q}{(1-\alpha)^2}I(X;Y|S=2)\geq 0.$$

Therefore $\alpha=q$ must give the maximum value and

$$R\leq \max_{p(x)} I(X;Y|S)+\varepsilon_N$$

which completes the proof of Theorem 4. ▪

5. Arbitrary (Non-causal) Selection Rules

In this section we present two examples which demonstrate that non-causal selection rules can achieve greater rates than causal selection rules.

Example D: Two BSCs.

Consider the cell consisting of two binary symmetric states, 1 and 2, obtained by eliminating the stuck state (setting $p=0$) in Example A. Theorem 4 states that the capacity for causal rules is

$$C = \max\left\{q[1-h(\varepsilon_1)],\ [1-h(q\varepsilon_1+\bar{q}\varepsilon_2)],\ \bar{q}[1-h(\varepsilon_2)]\right\}.$$

We compare this with the rate achieved by the following non-causal rule for the case $q=\bar{q}=1/2$. For $j=\sum_{k=1}^{i-1} u_k+1$, let

$$P(U_i=1\,|\,s_1, s_2, \ldots, s_N) = \begin{cases} 0 & \text{if } j \text{ odd},\ s_i=2, \text{ and } s_{i+1}=1 \\ 1 & \text{otherwise.} \end{cases}$$

Essentially, the rule works sequentially by pairs of used cells. A state 2 cell is not selected for an odd numbered position in the $\mathbf{t}$ vector if there is a state 1 cell following it, which would be preferred. However, any state is acceptable for an even numbered position. The rule thus seeks to maintain a high proportion of state 1 cells in the odd numbered $\mathbf{t}$ positions. For example, $\mathbf{s}$=[111221121222] would yield $\mathbf{u}$=[111101101111] and $\mathbf{t}$=[1112111222].

Since the rule functions independently on pairs of selected cells, $p(\mathbf{t}) = \prod_{j \text{ odd}} p(t_j, t_{j+1})$. We can determine $p(t_j, t_{j+1})$ with the aid of the following table for j odd:

s_i	s_{i+1}	s_{i+2}	u_i	u_{i+1}	u_{i+2}	t_j	t_{j+1}
1	1	–	1	1	–	1	1
1	2	–	1	1	–	1	2
2	1	1	0	1	1	1	1
2	1	2	0	1	1	1	2
2	2	–	1	1	–	2	2

Since each triplet of states s_i, s_{i+1}, s_{i+2} occurs with probability 1/8, we have the following distribution on t_j, t_{j+1}:

t_j	t_{j+1}	$p(t_j, t_{j+1})$
1	1	3/8
1	2	3/8
2	1	0
2	2	2/8

Also, the expected number of cells skipped per used cell is $(1/8 + 1/8)/2 = 1/8$. The selection rate is thus $r = 8/9$.

We now compute the capacity of the symmetric channel consisting of a pair of used cells with states t_j and t_{j+1}. There are four possible error pairs $[z_j, z_{j+1}]$ whose probabilities are determined by ε_1, ε_2 and $p(t_j, t_{j+1})$.

z_j	z_{j+1}	$p(z_j, z_{j+1})$
0	0	$(3/8)\bar{\varepsilon}_1\bar{\varepsilon}_1+(3/8)\bar{\varepsilon}_1\bar{\varepsilon}_2+(2/8)\bar{\varepsilon}_2\bar{\varepsilon}_2$
0	1	$(3/8)\bar{\varepsilon}_1\varepsilon_1+(3/8)\bar{\varepsilon}_1\varepsilon_2+(2/8)\bar{\varepsilon}_2\varepsilon_2$
1	0	$(3/8)\varepsilon_1\bar{\varepsilon}_1+(3/8)\varepsilon_1\bar{\varepsilon}_2+(2/8)\varepsilon_2\bar{\varepsilon}_2$
1	1	$(3/8)\varepsilon_1\varepsilon_1+(3/8)\varepsilon_1\varepsilon_2+(2/8)\varepsilon_2\varepsilon_2$

Since the channel is symmetric, $I(X_j, X_{j+1}; Y_j, Y_{j+1}) = 2-H(Z_j, Z_{j+1})$ and we can achieve

$$R = \frac{1}{N}\sum_{\substack{j=1 \\ j \text{ odd}}}^{rN} I(X_j, X_{j+1}; Y_j, Y_{j+1})$$
$$= \frac{8}{9}\,\frac{1}{2}\Big[2-H(Z_j, Z_{j+1})\Big].$$

For many values of ε_1 and ε_2, the above rate is greater than the independent (or causal) rule capacity. For instance, ε_1=.120 and ε_2=.325 yield

$$\frac{1}{2}[1-h(\varepsilon_1)] \approx .235320 \quad \text{bits}$$
$$1-h(\tfrac{1}{2}\,\varepsilon_1+\tfrac{1}{2}\,\varepsilon_2) \approx .235294$$
$$\frac{8}{9}\,\frac{1}{2}[2-H(Z_j, Z_{j+1})] \approx .236252$$

so that the above non-causal rule rate exceeds the independent rule capacity by .000932 bits.

Example C2: Defect Information at the Decoder, Revisited.

Consider the following non-causal selection rule for the cell of Example C1 (Figure 5). For $j = \sum_{k=1}^{i-1} u_k + 1$, let

$$P(U_i=1 \mid s_1, \ldots, s_N) = \begin{cases} 0 & \text{if } j \text{ odd}, s_i=2 \text{ and } s_{i+1}=1 \\ 0 & \text{if } j \text{ even}, s_i=1 \text{ and } s_{i+1}=2 \\ 1 & \text{otherwise.} \end{cases}$$

The rule seeks to increase the probability that a state 1 cell is selected for an odd numbered position in the **t** vector and the probability that a state 2 cell is selected for an even numbered position. These events are called matches.

For example, **s**=[111221121222] would yield **u**=[111101011111] and **t**=[1112121222].

Because of the symmetry of this rule, we only consider the case of odd j. We can break the rule into independent blocks by determining u_{i+1} as well as u_i if s_{i+1} is considered in the determination of u_i. This is shown in the following table:

s_i	s_{i+1}	u_i	u_{i+1}	t_j	t_{j+1}
1	–	1	–	1	–
2	1	0	1	1	–
2	2	1	1	2	2

For every execution of this tabulated procedure, we have

expected number of cells skipped = 1/4
expected number of cells matched = 1/2 + 1/4 + 1/4 = 1
expected number of cells mismatched = 1/4
expected total number of cells = 3/2

Thus the selection rate $r=(1+\frac{1}{4})/(3/2) = 5/6$ and $\frac{1}{4}/(1+\frac{1}{4}) = 4/5$ of the selected cells will be matches.

Since the decoder knows the state, a match will result in a clear channel and a mismatch in an erasure. The following rate is achieved.

$$R = r\ I(X;Y|S) = \frac{5}{6}\frac{4}{5} = \frac{2}{3},$$

which is greater than the independent (or causal) rule capacity C=1/2 calculated in Example C1.

6. Conclusions

In this paper we have studied the storage rates achievable when defect state information is provided to one or more of the encoder, decoder and selector. Information about a cell's state can be used by the encoder or decoder in coding for either the current and all subsequent cells, or all cells. These two cases are referred to as causal or arbitrary use of state information, respectively. The selection rule can be independent, causal or arbitrary, as defined in Section 2. The table lists the known results according to what type of state information is provided to each of the three components. Note that the result of Theorem 4 applies only for memory cells having two states, and the result of Theorem 3 applies only to cells consisting of two BSC states, as noted.

When information is provided arbitrarily to both encoder and decoder, the capacity is given by [3] as $\max_{p(x|s)} I(X;Y|S)$. In achieving this rate, only s_j is needed to encode and decode x_j. Therefore the same result applies to the causal case. Also note that a selector can be of no further use in these cases since its functions can be subsumed by the encoder and decoder.

TABLE: Summary of Known Results

n=no state information, i=independent, c=causal, a=arbitrary.

Enc.	Dec.	Selector	Capacity	Ref.
n	n	n	$\max_{p(x)} I(X;Y)$	-
		i	$\max_{v \subseteq S} P(S \in v) \max_{p(x)} I(X;Y \mid S \in v)$	Thm. 1
		c	$\max\{q[1-h(\varepsilon_1)],1-h(q\varepsilon_1+\bar{q}\varepsilon_2),\bar{q}[1-h(\varepsilon_2)]\}^{\dagger}$	Thm. 3
	c,a	n	$\max_{p(x)} I(X;Y \mid S)$	[3]
		i	$\max_{p(x)} I(X;Y \mid S)$	Cor. 2
		c	$\max_{p(x)} I(X;Y \mid S)^{\dagger\dagger}$	Thm. 4
c	n	n	$\max_{p(x_1 \ldots x_{\|\|S\|\|})} I(X_1 \ldots X_{\|\|S\|\|};Y)$	[1]
		i	$\max_{v \subseteq S} P(S \in v) \max_{p(x_1 \ldots x_{\|\|S\|\|})} I(X_1 \ldots X_{\|\|S\|\|};Y \mid S \in v)$	Cor. 3
c,a	c,a	n,i,c,a	$\max_{p(x \mid s)} I(X;Y \mid S)$	[3]
a	n	n	$\max_{p(v,x \mid s)} I(V;Y)-I(V;S)$	[2][3]

†for a cell consisting of two BSCs.

††for $||S||=2$.

Acknowledgement

The authors are grateful to Chris Heegard, Rudolph Ahlswede and Thomas Cover for several valuable remarks during the preparation of this paper.

Appendix 1

We show that for any given $L_1, L_2, I_1, I_2 \geq 0$,

$$R(p) = \left[pL_1+(1-p)L_2+1\right]^{-1} \left[pI_1+(1-p)I_2\right]$$

attains its maximum value over $0 \leq p \leq 1$ at $p=0$ or $p=1$.

Differentiating, we have

$$\frac{dR}{dp} = -\left[pL_1+(1-p)L_2+1\right]^{-2} \left[I_1-I_2+I_1L_2-I_2L_1\right] + \left[pL_1+(1-p)L_2+1\right]^{-1} \left[I_1-I_2\right]$$

which is either identically zero or never zero, regardless of p. In the first case, both $p=0$ and $p=1$ attain the maximum R. In the latter case, there is no interior maximum and so the maximum must occur at either endpoint.

Appendix 2

We have $I(\alpha) = 1-h(\alpha\varepsilon_1+\bar{\alpha}\varepsilon_2)$. Note that $I(\alpha)$ is convex $\cup$. For $1>\alpha>q$, let

$$R(\alpha) = \left[L(\alpha)+1\right]^{-1} I(\alpha) = (q/\alpha)I(\alpha).$$

We set the derivative equal to zero and find the second derivative under this condition.

$$\frac{dR}{d\alpha} = (-q/\alpha^2)I(\alpha)+(q/\alpha)\,\frac{d}{d\alpha}I(\alpha) = 0$$

$$\frac{d^2R}{d\alpha^2} = (2q/\alpha^3)I(\alpha) - (2q/\alpha^2)\,\frac{d}{d\alpha}I(\alpha) + (q/\alpha)\,\frac{d^2}{d\alpha^2}I(\alpha)$$

$$= (q/\alpha)\,\frac{d^2}{d\alpha^2}I(\alpha) > 0.$$

Since the second derivative is positive when the first derivative is zero, there can be no interior maximum.

For $0<\alpha<q$, let

$$R(\alpha) = \left[\frac{q-\alpha}{1-q}+1\right]^{-1} I(\alpha) = (\bar{q}/\bar{\alpha})I(\alpha).$$

Differentiating by $\bar{\alpha}$ we find as above that when the first derivative is zero, the second is positive. Thus there is no interior maximum in this region either. The maximizing α must therefore be 0, q, or 1.

References

[1] C. Shannon, "Channels with Side Information at the Transmitter," IBM J. Res. Develop., Vol. 2(4) Oct. 1958, pp. 289-293.

[2] S. Gel'fand, M. Pinsker, "Coding for Channel with Random Parameters," Problems of Control and Information Theory, Vol. 9(1), 1980, pp. 19-31.

[3] C. Heegard, A. El Gamal, "On the Capacity of a Computer Memory with Defects," to appear in IEEE Trans. Information Theory.

[4] "SA1000 Fixed Disk Drive OEM Manual," Shugart Assoc., Sunnyvale, CA, 1979, pp. 28-29.

[5] J. Posa, "What To Do When the Bits Go Out," Electronics, July 28, 1981, pp. 117-120.

[6] B.F. Fitzgerald and E.P. Thoma, "Circuit Implementation of Fusible Redundant Addresses on RAMs for Productivity Enhancement," IBM J. Res. Develop., Vol. 24(3), May 1980, pp. 291-298.

[7] M. Salehi, "Cardinality Bounds on Auxiliary Variables in Multiple-user Theory via the Method of Ahlswede and Korner," Department of Statistics, Stanford University, Technical Report No. 33, August 1978.