On the Capacity of Computer Memory with Defects

CHRIS HEEGARD, MEMBER, IEEE, AND ABBAS A. EL GAMAL, MEMBER, IEEE

Abstract—**A computer memory with defects is modeled as a discrete memoryless channel with states that are statistically determined. The storage capacity is found when complete defect information is given to the encoder or to the decoder, and when the defect information is given completely to the decoder but only partially to the encoder. Achievable storage rates are established when partial defect information is provided at varying rates to both the encoder and the decoder. Arimoto–Blahut type algorithms are used to compute the storage capacity.**

I. INTRODUCTION

We model a computer memory with defects and noise as a discrete memoryless channel with states that are statistically determined [1]. For example, a binary memory with stuck-at faults and soft errors is modeled by the three discrete memoryless channels depicted in Fig. 1. Each memory cell has probability $p/2$ of being stuck at 0, probability $p/2$ of being stuck at 1, and probability $1-p$ of behaving as a binary symmetric channel (BSC) with parameter ϵ.

Manuscript received April 2, 1981; revised December 14, 1982. This work was supported in part by the National Science Foundation under Grant ECS 78-23334, in part by NSF under Grant ENG-79-08948, and in part under DARPA Contract MDA 903-79-C 0680. This work was presented in part at the IEEE International Symposium on Information Theory, Santa Monica, CA, February 1981.

C. Heegard is with Cornell University, School of Electrical Engineering, Ithaca, NY 14853.

A. A. El Gamal is with Stanford University, 137 Durand Hall, Stanford, CA 94305.

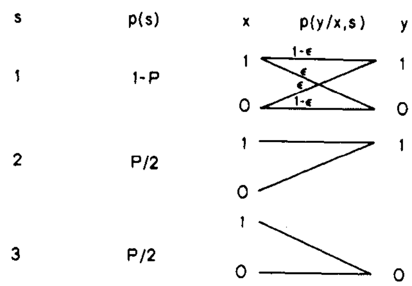

Fig. 1. Model for a binary memory cell with defects.

If neither the decoder nor the encoder knows the states of the memory the defects act as binary symmetric noise and the storage capacity of the memory (i.e., the maximum number of bits that can be reliably stored) is simply given by

$$C_{\min} = 1 - h\left((1-p)\epsilon + \frac{p}{2}\right) \text{ bits/cell,} \qquad (1.1)$$

where

$$h(x) = -x\log(x) - (1-x)\log(1-x).$$

On the other hand, if both the encoder and the decoder know the states of the memory the storage capacity can be easily shown to be

$$C_{\max} = (1-p)(1-h(\epsilon)) \text{ bits/cell.} \qquad (1.2)$$

0018-9448/83/0900-0731$01.00

When only the decoder knows the states of memory, the stuck-at cells can be treated as erasures, and the memory model reduces to a noisy binary erasure channel with erasure probability p and error probability $(1-p)\epsilon$. The storage capacity, in this case, is again given by (1.2).

Suppose only the encoder is given the defect information. The following argument shows that the capacity is again given by (1.2). Choose a set of $2^{nR'}$ Bernoulli 1/2 binary sequences for some $R' < 1 - h(\epsilon)$. With high probability this set will be a good code for the BSC with parameter ϵ. Randomly partition the set into 2^{nR} equal size subsets (or bins) and associate a different message with each bin. When the ith message is to be stored, search the ith bin for a sequence that is ϵ-compatible with the known defect. (A sequence is ϵ-compatible with the defect if it agrees with the stuck-at values in a fraction $1-\epsilon$ of these cells.) If an ϵ-compatible sequence is found in the ith bin it is stored, otherwise an error is declared. This codeword will now be correctly decoded with high probability, since each bit will experience an error with probability ϵ. For large n there will be approximately np stuck bits, thus there are approximately $2^{n(1-p+ph(\epsilon))}$ binary sequences that are ϵ-compatible with a given defect. The probability that an ϵ-compatible sequence is in the code is $2^{n(R'-1)}$, thus the expected number of ϵ-compatible sequences in the code book is the product $2^{n(R'-p+ph(\epsilon))}$. If the number of bins is much smaller than the number of ϵ-compatible sequences, then with high probability there will be an ϵ-compatible sequence in the ith bin. This is true for $R < R' - p + ph(\epsilon)$ and sufficiently large n. Since R' can be made arbitrarily close to $1 - h(\epsilon)$ we see that the capacity is given by (1.2).

Observe that for the case $\epsilon = 0$, the storage capacity is given by

$$C = (1-p) \text{ bits/cell.}$$

This is consistent with the work of Kusnetsov and Tsybakov [2]. Their work initiated a series of papers [3]–[8] concerning classes of binary codes for the case when the defects are known only to the encoder.

In this paper we investigate the problem of finding the storage capacity of a memory with arbitrary but finite storage and retrieval alphabets and arbitrary collection of states. In Section II, we give a general lower bound (Theorem 1) to the capacity when the states are partially known at arbitrary rates to the encoder and to the decoder. This lower bound is tight in the following special cases (Theorem 2): a) no state information to either the encoder or the decoder, b) complete state information at both encoder and decoder, c) complete state information to encoder and no information to the decoder, d) complete state information to decoder and arbitrary state information at the encoder. Results a), b) can be found in Wolfowitz [1], and c) has been established by Gel'fand and Pinsker [9]. In Section III, Arimoto–Blahut [10]–[12] algorithms are presented for determining the capacities in the cases of full state information. Examples are given. The proofs and derivations of theorems and algorithms are deferred to the appendixes.

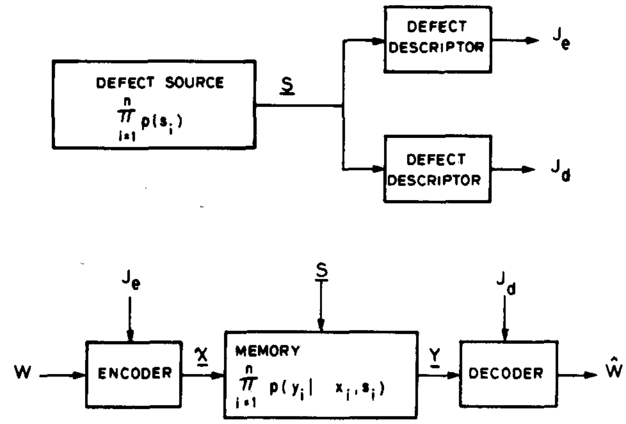

Fig. 2. General coding model for memory with defects.

II. Model and Results

A *discrete memoryless memory cell* $(S, p(s), X, p(y|x,s), Y)$ consists of three finite alphabets S, X, and Y, a probability mass function $p(s)$ on the alphabet S, and a probability transition matrix $p(y|x,s)$. The interpretation is that $s \in S$ is a state that the cell can assume with probability $p(s)$. In this state, if the letter $x \in X$ is stored then the letter $y \in Y$ is retrieved with probability $p(y|x,s)$.

It is assumed that the cells of a memory consisting of n DMMC's are identically distributed and statistically independent.

An (n, R, R_e, R_d, P_e) code for a memory composed of n DMMC's consists of four functions (see Fig. 2):

$$J_e: S^n \to \{1, 2^{nR_e}\}$$
$$J_d: S^n \to \{1, 2^{nR_d}\}$$
$$f_e: \{1, 2^{nR}\} \times \{1, 2^{nR_e}\} \to X^n$$

and

$$f_d: Y^n \times \{1, 2^{nR_d}\} \to \{1, 2^{nR}\}.$$

The map J_e provides a description of the state vector s for the encoder f_e at a rate of R_e bits/cell. The map J_d provides a similar service for the decoder f_d at a rate of R_d bits/cell. The encoder maps the message w and the state description J_e into an input vector x. Finally, the decoder maps the output sequence y and the J_d description into an estimate of the message $\hat{w}$.

The probability of error is defined as

$$P_e = 2^{-nR} \sum_{w=1}^{2^{nR}} P(w \neq \hat{w} | w \text{ is stored})$$
$$= 2^{-nR} \sum_{w=1}^{2^{nR}} \sum_{s \in S^n} \sum_{y \in Y^n} p(s) p(f_d(y, J_d) \neq w | x = f_e(w, J_e), s).$$

This is the probability that for a random message (and memory), the estimated message disagrees with the true message.

For a fixed R_e and R_d, a rate triple (R, R_e, R_d) is *achievable* if and only if for any $\epsilon > 0$ there exists a

(n, R, R_e, R_d, P_e) code with $P_e < \epsilon$ for some (possibly large) n. The *capacity function* $C(R_e, R_d)$ of the memory is defined as the supremum over all achievable rates R for fixed (R_e, R_d). We define the following four special values of $C(R_e, R_d)$:

$$C_{\max} = \sup_{R_e, R_d} C(R_e, R_d),$$

$$C_{\min} = \inf_{R_e, R_d} C(R_e, R_d),$$

$$C_{\text{enc}} = \sup_{R_e} C(R_e, 0),$$

$$C_{\text{dec}} = \sup_{R_d} C(0, R_d).$$

It can easily be seen that $C_{\max}$ is the capacity when full knowledge of the states is given to both the encoder and decoder. Similarly, $C_{\min}$ is the capacity when no state information is provided, C_{enc} is the capacity when only the encoder is given full information about the states, and C_{dec} is the capacity when full state information is given only to the decoder.

We would hope to find an explicit characterization of (R_e, R_d) for all rates (R_e, R_d). The complete solution to this problem has not been found. However, the following theorem gives lower bounds to $C(R_e, R_d)$ by describing a set of achievable rates (R, R_e, R_d).

Theorem 1: Fix $(\boldsymbol{S}, p(s), \boldsymbol{X}, p(y|x, s), \boldsymbol{Y})$ and alphabets U, S_0, S_e, and S_d. All rates (R, R_e, R_d) in the convex hull of the set

$$\begin{aligned}\{(R, R_e, R_d) | R_e &> I(S_0, S_e; S)\\ R_d &> I(S_0, S_d; S) - I(S_0, S_d; Y)\\ R_d &> I(S_d; S|S_0) - I(S_d; Y|S_0)\\ R_e + R_d &> I(S_0, S_e, S_d; S) - I(S_0, S_d; Y) + I(S_e; S_d|S_o)\\ R_e + R_d &> I(S_e, S_d; S|S_0) - I(S_d; Y|S_0) + I(S_e; S_d|S_0)\\ R &< I(U; Y, S_d|S_0) - I(U; S_e|S_0)\end{aligned}$$

for some probability mass function

$$p(s, s_0, s_e, s_d, u, x) = p(s)p(s_0, s_e, s_d|s)p(u, x|s_0, s_e)\}$$

are achievable.

The proof of Theorem 1 involves techniques similar to those found in [14]–[16], [19], [20] and is therefore deferred to Appendix I.

Remark: If we let $R_e = 0$, Theorem 1 reduces to the convex hull of the set

$$\{(R, R_d) | R_d > I(S_d; S|Y); R < I(X; Y|S_d),$$
for some probability mass function
$$p(s, s_d, x) = p(s)p(s_d|s)p(x)\}.$$

This result is identical to a result by Ahlswede and Han [13].

We now show that for several values of (R_e, R_d), Theorem 1 is optimal. These include $C_{\min}$, $C_{\max}$, C_{enc}, and C_{dec}.

Theorem 2: a) $R_e = R_d = 0$ (No description of defects)

$$C_{\min} = \max_{p(x)} I(X; Y).$$

b) $R_e > H(S)$, $R_d > H(S|Y)$. (Complete description of defects at encoder and decoder)

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

c) $R_e > H(S)$, $R_d = 0$. (Complete description of defects at encoder and no description at decoder)

$$C_{\text{enc}} = \max_{p(u, x|s)} I(U; Y) - I(U; S),$$

where

$$\|U\| \leqslant \min(\|X\|, \|Y\|) + \|S\| - 1.$$

d) $R_d > H(S|Y)$ (Complete description of defects at decoder)

$$C = \max_{\substack{p(s_0|s)\\ \text{subject to } R_e > I(S_0; S)}} \max_{p(x|s_0)} I(X; Y|S).$$

Furthermore, in the special case $R_e = 0$:

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

The achievability of this theorem follows from Theorem 1 by identifying the auxiliary random variables as follows:

a) $S_0 = S_e = S_d = \varphi$, $U = X$;
b) $S_0 = S$, $S_e = S_d = \varphi$, $U = X$;
c) $S_0 = S_d = \varphi$, $S_e = S$; and
d) $S_e = \varphi$, $(S_0, S_d) = S$, $U = X$.

The converses are proved in Appendix II.

The following corollary to Theorem 2 concerns the capacity of a memory with stuck-at type defects.

Corollary: Let $\boldsymbol{S} = \{1, 2, \cdots, m\}$, $p(y|x, s = 1)$ be arbitrary and for $i > 1$ let $p(y_i|x, s = i) = 1$ for some $y_i \in \boldsymbol{Y}$. Then

$$C_{\max} = C_{\text{enc}} = C_{\text{dec}} = p(s = 1)C,$$

where C is the capacity of the DMC With $p(y|x) = p(y|x, s = 1)$.

Proof: Let $p^*(x)$ be a probability mass function on $\boldsymbol{X}$ which achieves capacity for the DMC $(\boldsymbol{X}, p(y|x, s = 1), y)$, and let $p^*(x|y)$ be the (backward) probability transition matrix for the DMC induced by $p^*(x)$. To achieve C_{dec}, set $p(x) = p^*(x)$. To achieve C_{enc}, set $U = X$, $p(x|s = 1) = p^*(x)$, and for $i > 1$ set $p(x|s = i) = p^*(x|y_i)$. □

This corollary shows that for memories which have only stuck-at type defects, $C_{\max}$ can be achieved by providing the full defect information to only the encoder or only to the decoder. We also note that $C_{\max} = C_{\text{dec}} = p(s = 1)C$ if for $i > 1$ $\boldsymbol{Y}$ is independent of $\boldsymbol{X}$ (i.e., $p(y|x, s = i) = p(y|s = i)$). In this case, it need not be true that $C_{\text{enc}} = C_{\max}$, as some of the examples in the following section will demonstrate.

III. Algorithms and Examples

Determining $C_{\min}$, $C_{\max}$, C_{enc}, and C_{dec} for an arbitrary DMMC can be a difficult analytical problem. However, these quantities can be easily evaluated numerically on a computer. The Arimoto–Blahut [10]–[12] algorithm maximizes $I(X; Z)$ over $p(x)$ for a given transition probability $p(z|x)$. This algorithm can be used to compute $C_{\min}$, $C_{\max}$, and C_{dec}.

The value of $C_{\min}$ can be computed by setting $Z = Y$ and maximizing $I(X; Z)$ for

$$P(Z = y|x) = \sum_{s\in S} p(s)p(y|x, s).$$

The value of $C_{\max}$ can be determined by setting $Z = Y$ for each $s \in S$ and maximizing $I(X; Z) = I(X; Y|s)$ for $P(Z = y|x) = p(y|x, s)$. Then we average the results of each maximization

$$I(X; Y|S) = \sum_{s\in S} p(s)I(X; Y|s).$$

To compute C_{dec}, we use the Arimoto–Blahut (A–B) algorithm with $Z = (Y, S)$. We maximize $I(X; Z) = I(X; Y, S)$ with

$$P(Z = (y, s)|x) = p(s)p(y|x, s).$$

The result of this maximization is $C_{\text{dec}} = I(X; Y|S)$ since X and S are independent (i.e., $I(X; S) = 0$).

The value of C_{enc} cannot be computed directly by the A–B algorithm. A new algorithm is developed to compute C_{enc}. This algorithm is presented here. The detailed derivation is given in Appendix III.

To compute C_{enc}, we express the capacity as

$$C_{\text{enc}} = \max_{q'} \max_{q} \max_{Q} F(q', q, Q),$$

where

$$F = \sum_{s\in S} \sum_{u\in U} \sum_{x\in X} \sum_{y\in Y} p(s)q(u|s)q'(x|u, s) \cdot p(y|x, s) \log\left\{\frac{Q(u|y)}{q(u|s)}\right\}.$$

The parameters $q(u|s)$, $Q(u|y)$ are conditional probability mass functions on the alphabet U, and $q'(x|u, s)$ is a conditional probability mass function on the alphabet X.

A flowchart describing the algorithm is given in Fig. 3.

Initially, $q(u|s)$ is set equal to an arbitrary positive transition matrix (e.g., $q(u|s) = 1/\|U\|$) and $q'(x|u, s)$ is set equal to an arbitrary zero–one transition matrix. The main iteration cycles through each of the three arguments increasing F with respect to one argument while leaving the other two fixed.

F is maximized over Q for fixed q and q' by setting $Q(u|y)$ equal to the conditional probability of u given y

$$Q(u|y) = \frac{\sum_{s\in S} \sum_{x\in X} p(s)q(u|s)q'(x|u, s)p(y|x, s)}{\sum_{u\in U} \sum_{s\in S} \sum_{x\in X} p(s)q(u|s)q'(x|u, s)p(y|x, s)}. \tag{3.1}$$

For this value of Q, $F = I(U; Y) - I(U; S)$.

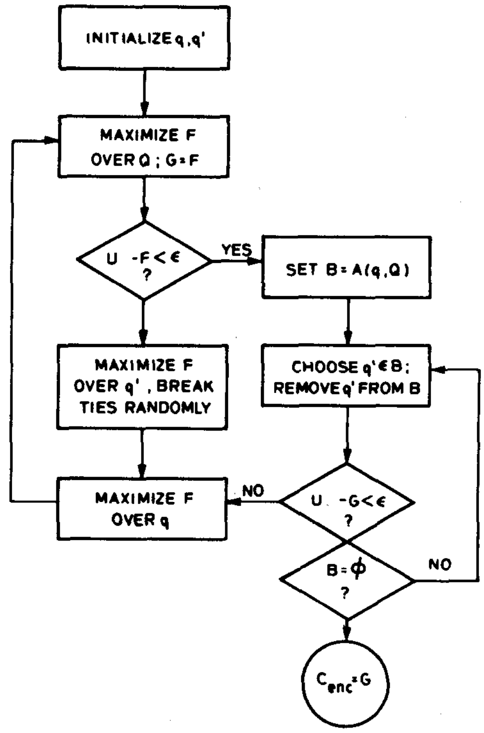

Fig. 3. Algorithm for computing C_{enc}.

To maximize F with respect to q' we set $q'(x|u, s) = 1$ for an $x \in X$ which maximizes

$$\prod_{y\in Y} Q(u|y)^{p(y|x, s)}$$

for each $u \in U$ and $s \in S$.

This implies that we can always assume that x is a deterministic function of u and s without loss of capacity. We define

$$A(q, Q) = \{q'|q' \text{ is a 0–1 matrix which maximizes } F\}.$$

Finally, F is maximized over q by

$$q(u|s) = \frac{\prod_{y\in Y} Q(u|y)^{(\sum_{x\in X} q'(x|u, s)p(y|x, s))}}{\sum_{u\in U} \prod_{y\in Y} Q(u|y)^{(\sum_{x\in X} q'(x|u, s)p(y|x, s))}}.$$

For fixed q'

$$U(q', q) = \sum_{s\in S} p(s) \max_{u\in U} \max_{x\in X} \sum_{y\in Y} p(y|x, s) \cdot \log\left(\frac{Q_0(u|y)}{q(u|s)}\right),$$

(where $Q_0(u|y)$ is given by (3.1)) forms an upper bound on $F(q', q, Q)$ which converges to F as F approaches a maximum. The algorithm terminates with $|C_{\text{enc}} - F| < \epsilon$, for any desired accuracy $\epsilon > 0$.

We now discuss three examples of a memory with defects. The first example is an extension of the simple example discussed in the introduction. For this example, it is found that $C_{\text{enc}} \leqslant C_{\text{dec}} \leqslant C_{\max}$. The second example con-

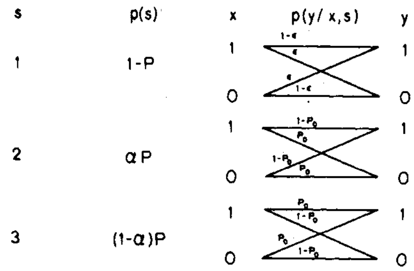

Fig. 4. Memory cell model for Example 1.

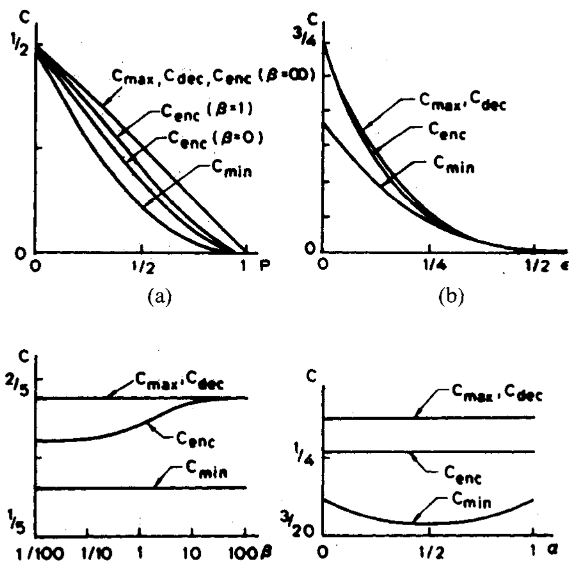

Fig. 5. Storage Capacity for Example 1: ($P_0 = ((1 + 1/\beta) - \sqrt{(1 + 1/\beta)^2 - 8\epsilon/\beta})/4$. (a) $a = 1/2$; $\epsilon = 0.11$. (b) $p = 1/4$; $a = 1/2$; $\beta = 1$. (c) $p = 1/4$; $\epsilon = 0.11$; $a = 1/2$. (d) $p = 0.4$; $\epsilon = 0.11$; $\beta = 1$.

cerns a binary memory cell which behaves as one of two binary symmetric channels. It is found that if the cross over probabilities are both less than 1/2, $C_{\min} = C_{\text{enc}} < C_{\text{dec}} = C_{\max}$. Thus no increase in storage capacity can result from providing the encoder with the defect state information. The last example shows that sometimes $C_{\text{enc}} > C_{\text{dec}}$.

Example 1: (Binary Cell with defects and symmetric noise) The memory cell model is depicted in Fig. 4. In Fig. 5, plots are given of $C_{\min}$, $C_{\max}$, C_{enc}, and C_{dec}. We note that the only capacity that varies with the parameter α is $C_{\min}$ and that only C_{enc} depends on the parameter β.

Example 2: (Two Binary Symmetric Channels (Fig. 6)) In Fig. 7, we give plots of $C_{\min}$, $C_{\max}$, C_{enc}, and C_{dec}. If ϵ_1 and ϵ_2 are both less than 1/2, then $C_{\min} = C_{\text{enc}} \leqslant C_{\text{dec}} = C_{\max}$. However, if $\epsilon_1 < 1/2$, and $\epsilon_2 > 1/2$, then the encoder can achieve rates higher than $C_{\min}$ by complementing the input on the $s = 2$ cells.

Example 3: The example depicted in Fig. 8 shows that there are cases where $C_{\text{dec}} < C_{\text{enc}}$. For this example, $C_{\min} = 0$, $C_{\text{dec}} = 2/3$, and $C_{\text{enc}} = C_{\max} = 1$ bit/cell.

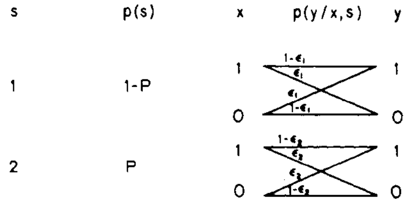

Fig. 6. Memory cell model for Example 2.

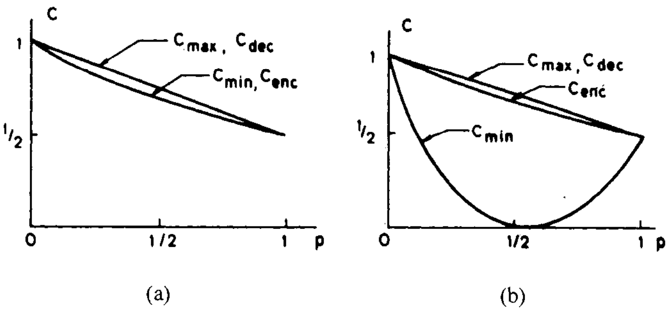

Fig. 7. Storage capacity for Example 2. (a) $\epsilon_1 = 0$; $\epsilon = 0.11$. (b) $\epsilon = 0.01$; $\epsilon_2 = 0.89$.

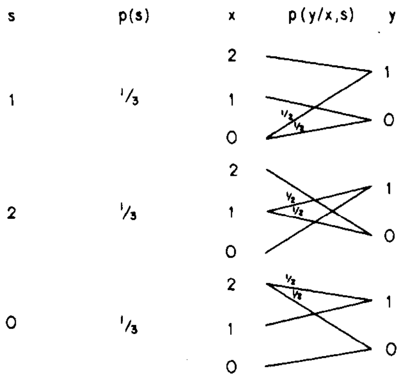

Fig. 8. Memory cell model for Example 3.

IV. Conclusion

We have seen that the storage capacity of a computer memory can be improved by providing the encoder or the decoder with information about the permanent defects. When the encoder or decoder is provided with an exact description of the locations and nature of all defects we have established the capacity. However, except for the case pertaining to part d) of Theorem 2, the storage capacity of the memory with a partial description of the defects is still an open problem.

Acknowledgment

The authors are thankful to Thomas Cover for the many valuable suggestions he has made to improve this paper. We also thank Dave Gluss and Max Costa for their help. Thanks are also due to Te Sun Han for his valuable comments on Appendix I.

Appendix I: Proof of Theorem 1

The outline consists of three parts. For this discussion, we fix $\epsilon > 0$, alphabets U, S_0, S_e, S_d and joint distribution

$$p(s, s_0, s_e, s_d, u, x, y) = p(s)p(s_0, s_e, s_d|s)p(u, x|s_0, s_e)p(y|x, s).$$

The first part concerns the existence of three maps

$$g_0: S^n \to S_0^n$$
$$g_1: S^n \to S_e^n$$
$$g_2: S^n \to S_d^n$$

such that if

$$\begin{aligned} R_0 &\triangleq \frac{1}{n}\log\|g_0\| > I(S_0; S) + 3\delta \\ R_1 &\triangleq \frac{1}{n}\log\|g_1\| > R_0 + I(S_e; S|S_0) + 4\delta \\ R_2 &\triangleq \frac{1}{n}\log\|g_2\| > R_0 + I(S_d; S|S_0) + 4\delta \\ R_1 + R_2 &> R_0 + I(S_e, S_d; S|S_0) + I(S_e; S_d|S_0) + 6\delta \end{aligned} \tag{A1}$$

and n is sufficiently large, then

$$P((g_0(S), g_1(S), g_2(S), S) \in T_\epsilon(S_0, S_e, S_d, S)) > 1 - \epsilon. \tag{A2}$$

Note that $\|g_i\|$ = cardinality of the range of g_i and $\delta(\epsilon) > 0$ is defined such that $\delta \to 0$ as $\epsilon \to 0$. (See [17] for a discussion of jointly typical sets $T_\epsilon(\cdot)$.)

The proof of this result follows by showing that on the average the following construction obtains the desired maps [14], [15]. Choose independently a set of vectors $\{S_0^i\}$, $1 \leq i \leq 2^{nR_0}$ according to a uniform distribution over the set $T_\epsilon(S_0)$. For each $1 \leq i \leq 2^{nR_0}$, independently choose two sets of vectors $\{S_e^{ij}\}$, $1 \leq j \leq 2^{n(R_1 - R_0)}$ and $\{S_d^{ik}\}$, $1 \leq k \leq 2^{n(R_2 - R_0)}$ according to a uniform distribution over the sets $T_\epsilon(S_\epsilon|S_0^i)$ and $T_\epsilon(S_d|S_0^i)$, respectively. For each $s \in T_\epsilon(S)$ look for an $1 \leq i \leq 2^{nR_0}$, $1 \leq j \leq 2^{n(R_1 - R_0)}$ and $1 \leq k \leq 2^{n(R_2 - R_0)}$ with $(S_0^i, S_e^{ij}; S_d^{ik}) \in T_\epsilon(S_0, S_e, S_d|s)$. Set $g_0(s) = S_0^i$, $g_1(s) = S_e^{ij}$ and $g_2(s) = S_d^{ik}$. For any $s \notin T_\epsilon(S)$ or if we cannot find such a triple, set $g_0(s) = S_0^1$, $g_1(s) = S_e^{11}$ and $g_2(S) = S_d^{11}$.

With this construction, if S is drawn according to $P(S = s) = \prod_{m=1}^n p(s_m)$ and (A1) is satisfied then (A2) will follow for sufficiently large n.

The second part of the proof concerns the existence of two maps

$$J_e: S_0^n \times S_e^n \to \{1, 2^{nR_e}\}$$

and

$$J_d: S_0^n \times S_d^n \to \{1, 2^{nR_d}\}$$

that can be used to label the ranges of three maps g_0, g_1, and g_2 constructed in part 1 of the proof. Specifically we argue that for sufficiently large n, when

$$\begin{aligned} R_e &> R_1 \\ R_d &> \max\{R_2 - I(S_0, S_d; Y), R_2 - R_0 - I(S_d; Y|S_0)\} \end{aligned} \tag{A3}$$

then given $J_e(g_0(S), g_1(S))$ we can find an estimate $(\hat{S}_0, \hat{S}_e)$ with

$$P((\hat{S}_0, \hat{S}_e) = (g_0(S), g_1(S))) > 1 - \epsilon \tag{A4}$$

and given $J_d(g_0(S), g_1(S))$ and the output vector Y we can find an estimate $(\tilde{S}_0, \tilde{S}_d)$ with

$$P((\tilde{S}_0, \tilde{S}_d) = (g_0(S), g_2(S))) > 1 - \epsilon. \tag{A5}$$

The first map J_e can be easily disposed of by recognizing that $\|g_0, g_1\| = \|g_1\| = 2^{nR_1}$. The existence of the second map J_d follows by averaging the following random construction [17], [19], [20]. For each unique pair $(g_0(s), g_2(s))$ independently choose $J_d(g_0(s), g_2(s))$ according to a uniform distribution over the set of integers $\{1, 2^{nR_d}\}$. Let (A3) be satisfied and choose (S, Y) according to the distribution $\prod_{m=1}^n p(s_m, y_m)$. Given $J_d(g_0(S), g_2(S))$ and Y choose the estimate $(\tilde{S}_0, \tilde{S}_2)$ as any pair $(g_0(s), g_2(s)) \in T\epsilon(S_0, S_d|Y)$ with $J_d(g_0(s), g_2(s)) = J_d(g_0(S), g_2(S))$. Then if n is sufficiently large, (A5) will follow. By combining (A1) and (A3) we arrive at the desired bounds

$$\begin{aligned} R_e &> I(S_0, S_e; S) + 7\delta \\ R_d &> \max\{I(S_0, S_d; S) - I(S_0, S_d; Y), I(S_d; S|S_0) \\ &\quad - I(S_d; Y|S_0)\} + 10\delta \\ R_e + R_d &> \max\{I(S_0, S_e, S_d; S) \\ &\quad - I(S_0, S_d; Y), I(S_e, S_d; S|S_0) \\ &\quad - I(S_d; Y|S_0)\} + I(S_e; S_d|S_0) + 12\delta. \end{aligned}$$

The final part of the proof concerns the last inequality of the theorem. This portion concerns the existence of maps

$$f_e: \{1, 2^{nR}\} \times S_0^n \times S_e^n \to X^n$$

and

$$f_d: Y^n \times S_0^n \times S_d^n \to \{1, 2^{nR}\}.$$

such that when

$$R \leq I(U; Y, S_d|S_0) - I(U; S_e|S_0) - 9\delta \tag{A6}$$

and n is sufficiently large than for a random message $W \in \{1, 2^{nR}\}$

$$P(f_d(Y, \tilde{S}_0, \tilde{S}_d) = W | X = f_e(W, \hat{S}_0, \hat{S}_e)) > 1 - \epsilon. \tag{A7}$$

The following random construction will prove the existence of these maps. Let R, R' satisfy

$$\begin{aligned} I(U; S_e|S_0) + 4\delta < R' &< I(U; S_e|S_0) + 5\delta \\ R + R' &< I(U; S_d, Y|S_0) - 4\delta. \end{aligned} \tag{A8}$$

For each $s_0 \in T_\epsilon(S_0)$, $1 \leq l \leq 2^{nR'}$ and $1 \leq w \leq 2^{nR}$ independently choose a vector $U^{s_0 l w}$ according to a uniform distribution over the the set $T_\epsilon(U|s_0)$. For each $(s_0, s_e) \in T_\epsilon(S_0, S_e)$ and $1 \leq w \leq 2^{nR}$ look for an $1 \leq l \leq 2^{nR'}$ such that $U^{s_0 l w} \in T_\epsilon(U|s_0, s_e)$. Choose X according to a uniform distribution over the set $T_\epsilon(X|U^{s_0 l w}, s_0, s_e)$ and define $f_e(w, s_0, s_e) = X$. For every other value of (w, s_0, s_e) set $f_e(w, s_0, s_e) = constant \in X^n$.

To construct f_d, take each $(y, s_0, s_d) \in T_\epsilon(Y, S_0, S_d)$ and look for a unique $1 \leq l \leq 2^{nR'}$ and $1 \leq w \leq 2^{nR}$ with $U^{s_0 l w} \in T_e(U|y, s_0, s_d)$ and set $f_d(y, s_0, s_e) = w$. For every other value of (y, s_0, s_d) set $f_d(y, s_0, s_d) = 1$.

Under this construction, for sufficiently large n and a random message W uniformly distributed over $\{1, 2^{nR}\}$ (A7) will be satisfied. Combining (A8) we arrive at the desired bound (A6).

Appendix II: Proof of Theorem 2

The achievability part of the theorem follows immediately by application of Theorem 1. The converse of the theorem is easily derived for parts a) and b) and has already been presented for part c) [9]. The converse for part d) is derived here. The proof shows that for any (n, R, R_e, R_d, P_e) code with $R_d > H(S|Y)$, there exists an $\epsilon \geqslant 0$ and a random variable S_0 such that $S \to S_0 \to X$ as a Markov chain and

$$\begin{aligned} R_e &\geqslant I(S_0; S) \\ R &\leqslant I(X; Y|S) + \epsilon. \end{aligned} \tag{A9}$$

Furthermore, $\epsilon(P_e)$ is defined such that $\epsilon \to 0$ as $P_e \to 0$.

First we need prove that the region of all rate triples $(R, R_e, H(S|Y))$ achieved by Theorem 1 is convex in (R, R_e) for fixed $p(s)$, $p(y|x, s)$. Let

$$A = \{(R, R_e)|R_e > I(S_0; S), R < I(X; Y|S),$$

where

$$p(s, s_0, x, y) = p(s)p(s_0|s)p(x|s_0)p(y|x, s)\},$$

and

$$B = \{(R, R_e)R_e > I(S_0; S|Z), R < I(X; Y|S, Z),$$

where

$$p(z, s, s_0, x, y) = p(z)p(s)p(s_0|s, z)p(x|s_0, z)p(y|x, s)\}.$$

Apparently, B = convex hull (A). We show $A = B$. Obviously $A \subseteq B$; we need to show $B \subseteq A$. Fix $p(z)p(s_0|s, z)p(x|s_0, z)$ then $(R, R_e) \in B$ for

$$\begin{aligned} R_e > I(S_0; S|Z) &= I(S_0, Z; S) \quad \text{(independence)} \\ &\geqslant I(S_0; S) \\ R < I(X; Y|S, Z) &\leqslant I(X, Z; Y|S) \\ &= I(X; Y|S) \end{aligned}$$

since $I(Z; Y|X, S) = 0$. Thus $(R, R_e) \in A$ and convexity is proved.

The entropy of the message random variable is

$$H(W) = \log\|W\| \triangleq nR$$

by assumption. If we take

$$\epsilon(P_e) = P_e R + \frac{1}{n} h(p_e),$$

then we obtain the Fano's inequality

$$H(W|\boldsymbol{Y}, \boldsymbol{S}) \leqslant n\epsilon. \tag{A10}$$

Note $\epsilon \geqslant 0$ and $\epsilon \to 0$ as $P_e \to 0$ as required.

Now, let $J_e = J_e(\boldsymbol{S})$ be the encoder's description of the defect vector $\boldsymbol{S}$,

$$\begin{aligned} nR_e &\triangleq H(J_e) \\ &\geqslant I(J_e; \boldsymbol{S}) \\ &= \sum_{i=1} I(J_e; S_i|\boldsymbol{S}_i^+) \\ &= \sum_{i=1}^{n} I(J_e, \boldsymbol{S}_i^+; S_i) \quad \text{(independence)}, \end{aligned} \tag{A11}$$

where $\boldsymbol{S}_n^+ = \varphi$, $\boldsymbol{S}_i^+ = [S_{i+1}, S_{i+2}, \cdots, S_n]$. Then

$$\begin{aligned} nR &= H(W) \\ &= H(W|J_e, \boldsymbol{S}) \quad \text{(independence)} \\ &\leqslant I(W; \boldsymbol{Y}|J_e, \boldsymbol{S}) = n\epsilon \quad \text{(from A10)} \\ &= \sum_{i=1}^{n} I(W; Y_i|J_e, S_i Y_i^-) + \epsilon \end{aligned}$$

(where $Y_1^- = \phi$, $Y_i^- = [Y_1, Y_2, \cdots, Y_{i-1}]$)

$$\begin{aligned} &\leqslant \sum_{i=1}^{n} I(W, J_e, \boldsymbol{S}_i^-, \boldsymbol{S}_i^+, \boldsymbol{Y}_i^-; Y_i|S_i) + \epsilon \\ &\leqslant \sum_{i=1}^{n} I(X_i; Y_i|S_i) + \epsilon \end{aligned} \tag{A12}$$

since $W, J_e, \boldsymbol{S}_i^-, \boldsymbol{S}_i^+; Y_i^- \to X_i, S_i \to Y_i$ form a Markov chain. Finally, since $S_i \to J_e, \boldsymbol{S}_i^+ \to X_i$ form a Markov chain and by convexity of the achievable region, there exists a distribution $p(s_0|s)p(x|s_0)$ such that (A11) and (A12) imply (A9).

Appendix III: Derivation of Algorithm for C_{enc}

Fix $(\boldsymbol{S}, p(s), \boldsymbol{X}, p(y|x, s), \boldsymbol{Y}), \boldsymbol{U}$ and let

$$p_0(s, u, x, y) = p(s)q(u|s)q'(x|u, s)p(y|x, s) \tag{A13}$$

be a probability mass function on $\boldsymbol{S} \times \boldsymbol{U} \times \boldsymbol{X} \times \boldsymbol{Y}$. Then

$$Q_0(u|y) = \frac{\sum_{s \in \boldsymbol{S}} \sum_{x \in \boldsymbol{X}} p_0(s, u, x, y)}{\sum_{s \in \boldsymbol{s}} \sum_{u \in \boldsymbol{U}} \sum_{x \in \boldsymbol{X}} p_0(s, u, x, y)} \tag{A14}$$

is the conditional probability mass function of U given Y under (A13). Let $Q(u|y)$ be an arbitrary conditional probability mass function, and define

$$F(q', q, Q) = \sum_{s \in \boldsymbol{S}} \sum_{u \in \boldsymbol{U}} \sum_{x \in \boldsymbol{X}} \sum_{y \in \boldsymbol{Y}} p_0(s, u, x, y) \ln\left(\frac{Q(u|u)}{q(y|s)}\right), \tag{A15}$$

and

$$U(q', q) = \sum_{s \in \boldsymbol{S}} p(s) \max_{x \in \boldsymbol{X}} \max_{u \in \boldsymbol{U}} \sum_{y \in \boldsymbol{Y}} p(y|x, s) \ln\left(\frac{Q_0(u|y)}{q(u|s)}\right). \tag{A16}$$

The derivation of the algorithm consists of four parts. First, we show $C_{\text{enc}} = \max_{q'} \max_q \max_Q F(q', q, Q)$. Second, we show how to maximize F with respect to one argument when the other two arguments are fixed. Next, we show for fixed $q'(x|u, s)$ and arbitrary $\tilde{q}(u|s)$ that $\max_q \max_Q F(q', q, Q) \leqslant U(q', \tilde{q})$. Finally, we show that F_i, the value of F after the ith iteration of the algorithm converges montonically to $F^* = C_{\text{enc}}$ and $F^* = U^* =$ the limit of U_i.

Part 1:

From Theorem 2c) and the identity

$$F(q', q, Q_0) = H(U|S) - H(U|Y) = I(U; Y) - I(U; S),$$

we see that $C_{\text{enc}} = \max_{q'} \max_q F(q', q, Q_0)$. Thus, we need only show that $F(q', q, Q_0) \geqslant F(q', q, Q)$:

$$\begin{aligned} &F(q', q, Q) - F(q', q, Q_0) \\ &\quad = \sum_{s \in \boldsymbol{S}} \sum_{u \in \boldsymbol{U}} \sum_{x \in \boldsymbol{X}} \sum_{y \in \boldsymbol{Y}} p_0(s, u, x, y) \ln \frac{Q(u|y)}{Q_0(u|y)} \\ &\quad \leqslant \sum_{s \in \boldsymbol{S}} \sum_{u \in \boldsymbol{U}} \sum_{x \in \boldsymbol{X}} \sum_{y \in \boldsymbol{Y}} p_0(s, i, x, y) \left[\frac{Q(u|y)}{Q_0(u|y)} - 1\right] = 0. \end{aligned}$$

The last step follows from the inequality $\ln x \leqslant x - 1$.

Part 2:

The inequality $F(q', q, Q) \leqslant F(q', q, Q_0)$ from the previous section satisfies equality if and only if $Q(u|y) = Q_0(u|y)$ for all u, y with $p_0(u, y) > 0$. Thus F is maximized over Q if and only if Q is given by (A14).

To maximize F over q we form the functional

$$G(q) = F(q', q, Q) + \sum_{s \in S} \lambda_s \Big(1 - \sum_{u \in U} q(u|s)\Big),$$

where the λ_s's are Lagrange multipliers. Since G is concave in q, the maximum is found by setting the derivative of G to zero. Since

$$\frac{\partial G}{\partial q(u_0|s_0)} = p(s_0) \sum_{x \in X} \sum_{y \in Y} q'(x|u_0, s_0) p(y|x, s_0) \cdot \left[\ln \frac{Q(u_0|y)}{q(u_0|s_0)} - 1\right] - \lambda_{s_0} \quad \text{(A17)}$$

G is maximized for

$$q(u_0|s_0) = \frac{\prod_{y \in Y} Q(u_0|y)^{\sum_{x \in X} q'(x|u_0, s_0) p(y|x, s_0)}}{\sum_{u \in U} \prod_{y \in Y} Q(u|y)^{\sum_{x \in X} q'(x|u, s_0) p(y|x, s_0)}}.$$

Note $q \geqslant 0$ as required.

To maximize F over q' we rewrite F as follows

$$F(q', q, Q) = H(U|S) + \sum_{s \in S} \sum_{u \in U} \sum_{x \in X} p(s) q(u \leqslant s) q'(x|u, s) \cdot \sum_{y \in Y} p(y|x, s) \ln Q(u|y).$$

We see that F is linear in q'. Thus F is maximized for each u and s by setting $q'(x|u, s) = 1$ for an x which maximizes $\sum_{y \in Y} p(y|x, s) \ln Q(u|y)$. This shows that to achieve C_{enc}, we can always take x as a deterministic function of u and s. We restrict $q'(x|u, s)$ to be a zero-one transition matrix and define

$$A(q, Q) = \{q' | q' \text{ is a zero-one matrix which maximizes } F\}.$$

Note that $1 \leqslant \|A\| \leqslant \|X\|^{\|U \times S\|}$.

Part 3:

Fix $q'(x|u, s)$, $q(u|s)$, $\tilde{q}(u|s)$ and define

$$U'(q', q, \tilde{q}) = \sum_{s \in S} \sum_{u \in U} \sum_{y \in Y} p(s) q(u|s) q'(x|u, s) \cdot p(y|x, s) \ln \frac{\tilde{Q}_0(u|y)}{\tilde{q}(u|s)},$$

where $\tilde{Q}_0(u|y)$ is the conditional probability mass function under $\tilde{q}(u \leqslant s)$.

First, it is clear by the definitions that $U'(q', q, \tilde{q}) \leqslant U(q', \tilde{q})$. We show that $F(q', q, Q) \leqslant U'(q', q, \tilde{q})$.

$$F(q', q, Q) - U'(q', q, \tilde{q})$$
$$= \sum_{s \in S} \sum_{u \in U} \sum_{x \in X} \sum_{y \in Y} p_0(s, u, x, y) \ln \frac{Q(u|y) \tilde{q}(u|s)}{q(u|s) Q_0(u|y)}$$
$$\leqslant \sum_{s \in S} \sum_{u \in U} \sum_{x \in X} \sum_{y \in Y} p_0(s, u, x, y) \left[\frac{Q(u|y) \tilde{q}(u|s)}{q(u|s) Q_0(u|y)} - 1\right] = 0.$$

We derive necessary and sufficient conditions for $F(q', q, Q) = U(q', q)$. First, $F(q', q, Q) = U'(q', q, q)$ if and only if for every u and y with $p_0(u, y) > 0$, we have $Q(u|y) = Q_0(u|y)$. Thus Q must maximize F for fixed q' and q. Second, F is maximized over q if and only if

$$\sum_{x \in X} \sum_{y \in Y} q'(x|u, s) p(y|x, s) \ln \frac{Q(u|y)}{q(u|s)} \begin{cases} = Cs & q(u|s) > 0 \\ \leqslant Cs & q(u|s) = 0 \end{cases}$$

where Cs depends only on s. This follows from (A17) and the Kuhn–Tucker conditions. Thus

$$F(q', q, Q) \leqslant \sum_{s \in S} p(s) \max_{u \in U} \sum_{x \in X} \sum_{y \in Y} q'(x|u, s) p(y|x, s) \ln \frac{Q(u|y)}{q(u|s)},$$

with equality if and only if q maximizes F.

Finally, from Part 2,

$$F(q', q, Q) \leqslant \sum_{s \in S} \sum_{u \in U} p(s) q(u|s) \max_{x \in X} \sum_{y \in Y} p(y|x, s) \ln \frac{Q(u|y)}{q(u|s)}$$

with equality if and only if q' maximizes F. Thus, we have $F(q', q, Q) \leqslant U(q', q)$ with equality if and only if F is maximized over q and Q with q' fixed and F is maximized over q' and q and Q fixed. Note, this does not guarantee that $F = C_{\text{enc}}$, since the q' which maximizes F may not be unique. However, if $F(q', q, Q) = U(q', q)$ for every $q' \in A(q, Q)$, then $F = C_{\text{enc}}$.

Part 4:

First, we show that we can maximize F over q and Q for fixed q'. Define

$$p(y|u, s) = \sum_{x \in X} q'(x|u, s) p(y|x, s),$$

$$Q_i(u|y) = \frac{\sum_{s \in S} p(s) q_i(u|s) p(y|u, s)}{\sum_{s \in S} \sum_{u \in U} p(s) q_i(u|s) p(y|u, s)},$$

$$r_{i+1}(u|s) = \prod_{y \in Y} Q_i(u|s)^{p(y|u, s)},$$

$$q_{i+1}(u|s) = \frac{r_{i+1}(u|s)}{\sum_{u \in U} r_{i+1}(u|s)},$$

and

$$F_{i+1} = F(q', q_{i+1}, Q_i).$$

We show that if $q_0(u|s) > 0$ then

$$\lim_{i \to \infty} F_i = C \triangleq \max_{q(u|s)} I(U; Y) - I(U; S).$$

Let

$$p_{i+1}(s, u, y) = p(s) q_{i+1}(u|s) p(y|u, s)$$

and

$$p^*(s, u, y) = p(s) q^*(u|s) p(y|u, s)$$

be probability mass functions on $S \times U \times Y$ and assume $q^*(u|s)$ achieves C. Consider

$$F_{i+1} = \sum_{s \in S} \sum_{u \in U} \sum_{y \in Y} p_{i+1}(s, u, y) \ln \frac{Q_i(u|y) \sum_{u \in U} r_{i+1}(u|s)}{r_{i+1}(u|s)}$$
$$= \sum_{s \in S} p(s) \ln \sum_{u \in U} r_{i+1}(u|s).$$

Now,

$$\sum_{s\in S}\sum_{u\in U} p(s)q^*(u|s)\ln\frac{q_{i+1}(u|s)}{q_i(u|s)}$$
$$=\sum_{s\in S}\sum_{u\in U} p(s)q^*(u|s)\ln\frac{r_{i+1}(u|s)}{q_i(u|s)\sum_{u\in U} r_{i+1}(u|s)}$$
$$=-F_{i+1}+\sum_{s\in S}\sum_{u\in U}\sum_{y\in Y} p^*(s,u,y)\ln\frac{Q_i(u|y)}{q_i(u|s)}$$
$$=-F_{i+1}+\sum_{s\in S}\sum_{u\in U}\sum_{y\in Y} p^*(s,u,y)\ln\frac{p(y|u,s)p(s)}{p_i(y)p_i(s|u,y)}.$$

The last step follows from the identity $p_i(s,u,y)=p_i(s|u,y)p_i(u|y)p_i(y)$. We now see that

$$\sum_{s\in S}\sum_{u\in U} p(s)q^*(u|s)\ln\frac{q_{i+1}(u|s)}{q_i(u|s)}$$
$$=-F_{i+1}+\sum_{s\in S}\sum_{u\in U}\sum_{y\in Y} p^*(s,u,y)$$
$$\cdot\ln\frac{p(y|u,s)p(s)p^*(y)p^*(s|u,y)}{p^*(y)p^*(s|u,y)p_i(y)p_i(s|u,y)}$$
$$=-F_{i+1}+C+\sum_{s\in S}\sum_{u\in U}\sum_{y\in Y} p^*(s,u,y)\ln\frac{p^*(y)p^*(s|u,y)}{p_i(y)p_i(s|u,y)}$$
$$\leqslant -F_{i+1}+C,$$

where the last inequality follows from the nonnegativity of the dropped term. Finally, we see

$$\sum_{i=1}^{N-1}(C-F_{i+1})\leqslant\sum_{i=1}^{N-1}\sum_{s\in S}\sum_{u\in U} p(s)q^*(u|s)\ln\frac{q_{i+1}(u|s)}{q_i(u|s)}$$
$$=\sum_{s\in S}\sum_{u\in U} p(s)q^*(u|s)\ln\frac{q_N(u|s)}{q_0(u|s)}$$
$$\leqslant\sum_{s\in S}\sum_{u\in U} p(s)q^*(u|s)\ln\frac{q^*(u|s)}{q_0(u|s)}.$$

If $q_0(u|s)>0$, the last term is finite and $\sum_{i=1}^{\infty}C+F_{i+1}<\infty$. Since $C\geqslant F_{i+1}$, we conclude that $F_{i+1}\to C$ and this sequence converges. This shows that the algorithm for C_{enc} will maximize F over q and Q. The upper bound U described in Part 3 will determine if $C_{\text{enc}}-F<\epsilon$. If not, a new q' is chosen and the algorithm continues. Finally, since the set $A(q,Q)$ of admissible q''s is finite and the algorithm monotonically increases F, it will eventually find a q' which achieves C_{enc}. This establishes the convergence of the algorithm.

References

[1] J. Wolfowitz, *Coding Theorems of Information Theory*, 3rd ed. New York: Springer Verlag, 1978.

[2] A. V. Kusnetsov and B. S. Tsybakov, "Coding in a memory with defective cells," translated from *Prob. Peredach. Inform.*, vol. 10, no. 2, pp. 52–60, April–June 1974.

[3] B. S. Tsybakov, "Additive group codes for defect correction," translated from *Prob. Peredach. Inform.*, vol. II, no. 1, pp. 111–113, January–March, 1975.

[4] B. S. Tsybakov, S. I. Gel'fand, A. V. Kusnetsov, and S. I. Ortyukov, "Reliable Computation and reliable storage of information," in *Proc. 1975 IEEE–USSR Joint Workshop Inform. Theory*, pp. 215–226, Dec. 15–19 1975.

[5] I. B. Belov and A. M. Shashin, "Codes that correct triple defects in memory," translated from *Prob. Peredach. Inform.*, vol. 13, no. 4, pp. 62–65, October–December 1977.

[6] V. V. Losev, V. K. Konopel'ko, and Yu. D. Daryakin, "Double-and-triple-defect-correcting codes," translated from *Prob. Peredach. Inform.*, vol. 14, no. 4, pp. 98–101, October–December 1978.

[7] A. V. Kusnetsov, "Masking triple fixed defects in memory," preprint, Institute for Problems of Information Transmission, USSR Academy of Sciences.

[8] A. V. Kusnetsov, T. Kasami, and S. Tamamura, "An error correcting scheme for defective memory," *IEEE Trans. Inform. Theory*, vol. IT-24, no. 6, pp. 712–718, Nov. 1978.

[9] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Prob. Contr. and Inform. Theory*, vol. 9(1), pp. 19–31, 1980.

[10] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. IT-18, no. 1, pp. 14–20, Jan. 1972.

[11] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Infor. Theory*, vol. IT-18, no. 4, pp. 460–472, July 1972.

[12] A. Viterbi and J. Omura, *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.

[13] 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*, vol. IT-29, pp. 396–412, May 1983.

[14] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 306–311, May 1979.

[15] A. A. El Gamal and E. van der Meulen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 120–122, Jan. 1981.

[16] A. A. El Gamal and T. M. Cover, "Information theory of joint descriptions," *Trans. Inform. Theory*, pp. 851–857 Nov. 1982.

[17] T. Berger, "Multi-terminal source coding," CISM Courses and Lectures no. 229, *The Information Theory Approach to Communications*, G. Longo, Ed., 1977.

[18] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[19] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 4, pp. 471–480, July 1973.

[20] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inform. Theory*, vol. IT-21, no. 2, pp. 226–228, Mar. 1975.