On Storage Media with Aftereffects

H. S. WITSENHAUSEN AND A. D. WYNER

Bell Laboratories, Murray Hill, New Jersey 07974

Successive independent batches of data are written into a memory and then read out. The behavior of a memory cell in each read/write cycle is dependent (this is the "aftereffect") on its previous history. Each cell can be modelled as an automaton. Using a large number of identical cells and coding/decoding across the memory, what is the maximum throughput that can be achieved in $N$ cycles with negligible error probability?

This problem is equivalent (by a time/space interchange) to finding the total capacity of a certain multiuser interference channel.

Exact answers are obtained for the cell-model in which the output of a cell is the exclusive OR of the two most recent inputs. For the (more realistic) inclusive OR, lower and upper bounds are determined.

The increase in throughput obtainable by delaying some of the read cycles is also determined or bounded.

1. INTRODUCTION

We consider a medium made up of a large number $N$ of independent cells, each capable of storing one bit of information. The medium is used in the following way: For $t=1,2,...,T$, data from a source $S_t$, possibly encoded, is stored at time $t$ and read at time $t + \theta_t$ by a user who seeks to recover the data from $S_t$ with arbitrarily small error probability. This user does not care about the sources $S_t$ (for $t \neq t$). The sources are independent.

If the cells function perfectly and $0 < \theta_t < 1$, then $N$ bits of data can be transmitted this way from each source. This is a rate of 1 bit per cell at each usage cycle.

Now consider the case of imperfect cells, each cell being a copy of a certain stochastic or deterministic finite state machine. (In the stochastic case, the randomness in each cell is assumed independent of that in all other cells.) Then, the aftereffect on the cells of usage in prior time periods introduces errors in the current storage-retrieval cycle. As the data of previous cycles is only known in distribution, not in realization, one is faced with a noisy channel and error-correcting codes must be used. As $N$ is large, the capacity of this channel will indicate the maximum possible rate of transmission.
However, using the maximum rate may severely hamper the next few time cycles. How then is one to assign the rates between the successive cycles to maximize the long-range average throughput? This is a dynamic programming problem of substantial difficulty.

Our problem is related to the earlier work on so-called "write-once memories," such as punched paper tape. That case was investigated by Rivest and Shamir (1982), Heegard (to appear), and Wolf, Wyner, Ziv, and Körner (to appear). For punched paper, it is evident that the long-range average rate per cell and usage cycle is zero, as the medium cannot recover, the aftereffect never dies out. Note that for such a medium the state of a cell is the last output (readout). It is a special case of a medium in which the new output is a function of only the current input and the previous output.

Denote by \( X^i_t \) the input to cell \( i \) in cycle \( t \), and by \( Y^i_t \) the corresponding output. Then the above case is characterized by

\[
Y^i_t = f(X^i_t, Y^i_{t-1}).
\]

(For punched tape, \( f \) is the inclusive or.)

As opposed to this, we consider media where the state of a cell is the previous input, so that

\[
Y^i_t = f(X^i_t, Y^i_{t-1}).
\]

We will primarily consider \( f \) to be the inclusive or the exclusive or of its arguments.

The problem can be considered as part of multiuser information theory. However, we do not attempt to find a complete rate region, only the maximum total rate is investigated.

**2. Cross-Dependence**

At the first cycle, all cells are assumed to be in the same initial state, corresponding to a clear medium. Then the coding problem for the first cycle is that of a discrete memoryless channel. However, at later cycles, the codes previously used may have statistics that introduce dependence between the cells. Then the coding problem is equivalent to that for a channel with memory.

As will be seen, we cannot in all cases rule out the possibility that introducing cross-dependence may be optimal. This a major difficulty in this investigation.
More precisely, the information-theoretic upper bound on the rate over $T$ cycles is

$$
\frac{1}{TN} \sum_{t=1}^{T} I(X_t; Y_t),
$$

(3)

where $X_t = (X^1_t, X^2_t, ..., X^N_t)$ and $Y_t = (Y^1_t, Y^2_t, ..., Y^N_t)$. If cross-dependence is to be discarded, one has to establish the upper bound

$$
\frac{1}{TN} \sum_{t=1}^{T} \sum_{i=1}^{n} I(X^i_t; Y^i_t).
$$

(4)

The earlier work (Rivest and Shamir, 1982; Heegard, to appear) was based upon (1), and, in that case, the desired bound is readily established as follows.

For $t = 1$, starting with a clear medium, one has $I(X_1, Y_1) = H(X_1) = H(Y_1)$, while for $t > 1$ one has componentwise $Y_t = f(X_t, Y_{t-1})$. The $X_t (t = 1, ..., T)$ are independent since they are derived from independent sources. Thus

$$
H(Y_{t-1}, X_t, Y_t) = H(Y_{t-1}, X_t, Y_{t-1}) = H(Y_t) + H(X_t)
$$

(5)

because $Y_{t-1}$ is a function of $(X_1, X_2, ..., X_{t-1})$ hence independent of $X_t$.

Then if

$$
S = \sum_{t=1}^{T} I(X_t, Y_t),
$$

one has using (5),

$$
S = H(Y_1) + \sum_{t=2}^{T} [H(Y_t) + H(X_t) - H(X_t, Y_t)]
$$

$$
= H(Y_1) + \sum_{t=2}^{T} [H(X_t) + H(Y_t) - H(X_t, Y_t)]
$$

$$
+ H(Y_{t-1}, X_t, Y_t) - H(Y_{t-1}) - H(X_t)]
$$

$$
= H(Y_T) + \sum_{t=2}^{T} H(Y_{t-1} | X_t, Y_t).
$$

All terms can now be upperbounded by their bitwise versions

$$
S \leq \sum_{i=1}^{N} \left[ H(Y^i_T) + \sum_{t=2}^{T} H(Y^i_{t-1} | X^i_t, Y^i_t) \right]
$$

(6)

which shows that cross-dependence can be ignored.
This line of proof fails, except for \( T = 2 \), when the state of a cell is the previous input as in (2).

3. The Exclusive OR Case

Suppose that one has componentwise,

\[ Y_t = X_t \oplus X_{t-1}. \tag{7} \]

This case has the virtue that exact answers to the major questions are easily obtained.

3.1. The Optimal Throughput

One way to use the medium, starting from a clear (i.e., all zeros) position is to use at every odd \( t \), i.i.d. binary symmetric input, and at every even \( t \) an all-zero input. Thus, one transmits one bit per cell at odd times and nothing at even times. The result is an average rate per cell and cycle of

\[ \frac{1}{2} \text{ for } T \text{ even}, \]

\[ \frac{1}{2} + \frac{1}{T} \text{ for } T \text{ odd}, \tag{8} \]

so that the long run average is \( \frac{1}{2} \). Note that this implementation requires no coding and involves no cross-correlation.

**Theorem 1.** The above rates are optimal.

Indeed, for any policy of medium usage, the information-theoretic upper bound on the rate is

\[ \frac{1}{NT} \left( H(X_1) + \sum_{t=2}^{T} I(X_t; Y_t) \right). \tag{9} \]

Obviously

\[ H(X_1) \leq N. \tag{10} \]

Further,

\[ I(X_t; Y_t) = H(X_t) + H(Y_t) - H(X_t Y_t) \]

\[ = H(X_t) + H(Y_t) - H(X_t X_{t-1}) \tag{a} \]

\[ = H(Y_t) - H(X_{t-1}) \tag{b}, \]

(11)
where (a) holds because $X_t Y_t$ and $X_t X_{t-1}$ determine the same $\sigma$-field (each can be obtained from the other) and (b) holds because $X_t$ and $X_{t-1}$ are independent.

Thus, for the first two stages one has

$$H(X_1) + I(X_2 ; Y_2) = H(Y_2) \leq N, \quad (12)$$

and for any two consecutive stages one has

$$I(X_t, Y_t) + I(X_{t+1}, Y_{t+1}) = H(X_t) - H(X_t | Y_t) + H(Y_{t+1}) - H(X_t)
\leq H(Y_{t+1}) \leq N. \quad (13)$$

Now for $T$ even use (12) once and (13) $(T-2)/2$ times; for odd $T$ use (10) once and (13) $(T-1)/2$ times to conclude that (8) is the absolute maximum, proving the theorem.

3.2. Smoothing

It may be considered a disadvantage to have a rate of transmission oscillating between 0 and 1 on successive cycles. Then one can use half the cells (assume $N$ even) in the above manner, while using the other half on even cycles while applying zero to them on odd cycle. This gives a uniform rate of $\frac{1}{2}$ at each cycle, which is optimal for even $T$ and asymptotically.

3.3. Variable Read-out Delay

Note that input $X_t$ is independent of all outputs other than $Y_t$ and $Y_{t+1}$. With $0 < \theta_t < 1$, the readout is $Y_t$. Alternatively, if $1 < \theta_t < 2$, then $Y_{t+1}$ is used. As observed by Vyssotsky (1982), the total throughput can be increased by alternating between these two modes.

This is done as follows: if $t \equiv 0 \pmod{3}$, then let $X_t$ be all zeros, if $t \equiv 1$ then use all possible values of $X_t$ and readout $Y_t$ with $0 < \theta_t < 1$, if $t \equiv 2$ also use $X_t$ fully but readout $Y_{t+1}$ with $1 < \theta_t < 2$ for the reconstruction. Again, no coding or decoding is required. Taking the interpretation that no readout is possible after the final cycle $T$, this gives average rates

$$\frac{2}{3} \quad \text{if} \quad T \equiv 0 \pmod{3},
\frac{2}{3} + \frac{1}{3T} \quad \text{if} \quad T \equiv 1,
\frac{2}{3} - \frac{1}{3T} \quad \text{if} \quad T \equiv 2. \quad (14)$$

Theorem 2. The above rates are optimal.
The information transmitted successfully from source $S_t$ is upper bounded, for $t < T$, by

$$\max_{1 \leq t < T} I(X_t; Y_t) = \max[I(X_t; Y_t), I(X_t; Y_{t+1})].$$

(15)

For $t = 1$ this applies with $Y_1 = X_1$, i.e., $X_0 = 0$.

For the last stage, the bound is $I(X_T; Y_T)$ as a delay would cause loss of the information. Anyhow the bound of $N$ bits holds, as for any single $t$.

For the last two stages, a bound of $N$ also holds for if there is no delay, the argument of Theorem 1 applies, if the last stage is delayed it is lost and if the first stage is delayed then all information from $S_{T-1}$ and $S_T$ is recovered from $Y_T$ and $H(Y_T) \leq N$.

Now note that

$$I(X_t; Y_{t+1}) = H(Y_t) - H(Y_t | X_t) = H(Y_t) - H(X_{t-1}),$$

(16)

and

$$I(X_t; Y_{t+1}) = H(Y_{t+1}) - H(Y_{t+1} | X_t) = H(Y_{t+1}) - H(X_{t+1}),$$

(17)

and

$$H(Y_{t+1}) = H(X_t \oplus X_{t+1}) \leq H(X_t, X_{t+1}) = H(X_t) + H(X_{t+1}).$$

(18)

For simplicity, shift the indexing so that the 3 cycles under consideration have indices $t = 1, 2, 3$, while $t = 0$ refers to the previous cycle and $t = 4$ to the one following. Then the bound (15) on the total throughput for these 3 stages can be written, using (16) and (17)

$$\max(H(Y_1) - H(X_0), H(Y_2) - H(X_2))$$

$$+ \max(H(Y_3) - H(X_1), H(Y_3) - H(X_3))$$

$$+ \max(H(Y_3) - H(X_2), H(Y_4) - H(X_4)).$$

(19)

To see that this is at most $2N$ consider all eight combinations of choices, denoted L (left) or R (right) in the three maxima and use the following

| LLL | 1 |
| LLR | 1 |
| LRL | 3 |
| LRR | 4 |
| RLL | 2 |
| RLR | 2 |
| RRL | 3 |
| RRR | 3 |

(20)
Each choice gives 3 positive and 3 negative terms. By (18), one of the positive terms $H(Y_i)$ is overwhelmed by the negative terms $-H(X_{i-1}) - H(X_i)$, where $i$ is given by the table. Dropping the remaining negative term leaves the sum of two entropies, each of a binary $N$ vector. Thus (19) is bounded by $2N$.

### 3.4. Smoothing with Delay

Assume $N$ is a multiple of 3. The memory is divided into 3 equal parts, used on the following schedule where 0 means all zero input, 1 means unit rate input read immediately and D means unit rate input with delayed readout.

\[ 1 \text{ D } 0 \text{ 1 D 0 1 D 0 ...} \]
\[ 1 \text{ 0 1 D 0 1 D 0 1 ...} \]
\[ 0 \text{ 1 D 0 1 D 0 1 D ...} \]

This gives a rate of $\frac{2}{3}$ at each stage, except the last where the information in the D sector is lost. The average is thus $\frac{2}{3} - (1/3T)$ which is optimal for $T \equiv 2 \pmod{3}$ and asymptotically.

### 3.5. Time Invariant Policy

The optimal policies derived above or their smoothed versions require the time-dependent selection of the rate or, in the smoothed version, of the storage sector. If $t$ is not known, even modulo 2 or 3, these policies are not implementable. What then is the best time invariant policy?

Assuming $\theta_t < 1$ for all $t$, this amounts to maximizing, over all distributions $P$ on $\{0, 1\}^N$, the quantity

\[ \frac{1}{N} I(X_1; X_0) = \frac{1}{N} (H(X_1 | X_0) - H(X_0)), \]

where $X_0, X_1$ are independently drawn under $P$.

For $N = 1$ the maximum is 0.214418 attained for the distribution $p_0 = 0.1211, p_1 = 0.8789$ (or vice versa).

If there were no cross-dependence, this would be valid for all $N$ with $P$ the product of the bit probabilities. However, already for $N = 2$ the distribution $p_{00} = 0, p_{01} = p_{10} = 0.14, p_1 = 0.72$ yields the rate 0.224593, so that cross-dependence is shown to matter.

The limit for $N \to \infty$ is not known.
4. An Inequality

Note the identity, most easily verified by expanding all terms into unconditional entropies, for 3 arbitrary discrete random variables \( X, Y, Z \),

\[
H(Z \mid X) + H(Z \mid Y) - H(Z) = I(X; Y \mid Z) - I(X; Y) + H(Z \mid XY). \tag{23}
\]

One has \( I(X; Y \mid Z) \geq 0 \), as for \( H(Z \mid XY) \), it vanishes when \( Z \) is a function of \( X, Y \); otherwise, it is nonnegative, so that

\[
H(Z) \leq H(Z \mid X) + H(Z \mid Y) + I(X; Y). \tag{24}
\]

If \( X \) and \( Y \) are independent, then

\[
H(Z) \leq H(Z \mid X) + H(Z \mid Y). \tag{25}
\]

In particular, applying (25) to independent binary \( X, Y \) with \( \Pr\{X = 0\} = x \), \( \Pr\{Y = 0\} = y \) and letting \( Z = X \lor Y \) be the inclusive or of \( X \) and \( Y \), one has the "subderivation" inequality for the function \( h(p) = -p \log p - (1 - p) \log(1 - p) \),

\[
h(xy) \leq xh(y) + yh(x). \tag{26}
\]

By induction, this generalizes to

\[
h \left( \prod_{i=1}^{n} p_i \right) \leq \sum_{i=1}^{n} h(p_i) \prod_{j \neq i} p_j, \tag{27}
\]

5. Inclusive OR, No Delays

For the inclusive OR case the results are mostly bounds and required extensive machine computation.

5.1. Two Stage Case

For \( T = 2 \) the proof of section 2 applies because, starting with a clear medium one has \( Y_1 = X_1 \). Hence the problem reduces to finding the maximum over all \( p, q \) in \([0, 1]\) of

\[
(1 - q) h(p) + h(pq), \tag{28}
\]

for this is \( H(X_1) + I(X_2; X_1 \lor X_2) \) when \( \Pr\{X_1 = 0\} = p \) and \( \Pr\{X_2 = 0\} = q \). Of course, the maximum over \( q \) is \( h(p) + C(p) \), where \( C(p) \) is the capacity of the \( Z \)-channel with crossover probability \( 1 - p \).
This maximum is 1.388085713..., which is attained for \( p = 0.66489... \), and \( q = 0.41668... \).

Thus the throughput is 38\% larger than with the exclusive or.

5.2. **Single Cell Optima**

For \( T > 2 \) we cannot, a priori, rule out cross-dependence. We can, however, investigate the optimum subject to the single-cell restriction. This will at least be a lower bound on the unconstrained optimum. An upper bound will be considered later.

Thus, for a single cell, let

\[
p_t = \Pr\{X_t = 0\}.
\]  

Then the problem is to maximize over \((p_1, ..., p_T) \in [0, 1]^T\) the expression

\[
H(X_t) + \sum_{t=2}^{T} I(X_t; Y_t) = h(p_1) + \sum_{t=2}^{T} h(p_{t-1} - p_t) - p_t h(p_{t-1}).
\]  

As each term of the sum involves only two successive \(p_t\), the maximum can be efficiently determined by dynamic programming, with the following results.

For odd \( T \) the maximum is attained by taking \( p_t = \frac{1}{2} \) for \( t \) odd and \( p_t = 1 \) for \( t \) even. This gives (30) the value \( 1 + (T - 1)/2 \).

For even \( T \) the optimal policy differs from the previous one only in the occurrence of a single "parity correction sequence" which can be located, with equal effect, at the beginning, the end or in between. This sequence is the one obtained for \( T = 2 \), that is \( p_t = 0.66489, p_{t+1} = 0.41668 \) surrounded by zeros (interpreting \( p_0 \) and \( p_{T+1} \) as zeros). Then (30) has the value \( 1.388... + (T - 2)/2 \).

The rates so obtained are

\[
\frac{1}{2} + \frac{1}{2T} \quad \text{for } T \text{ odd,}
\]

\[
\frac{1}{2} + \frac{0.388085713}{T} \quad \text{for } T \text{ even.}
\]  

This gives the same asymptotic rate of \( \frac{1}{4} \) as for the exclusive or. Smoothing is possible in the same way by alternately using \( p_t = \frac{1}{2} \) on half of the memory and \( p_t = 1 \) on the other half and reversing the halves at each step.

5.3. **An Upper Bound**

For coding using pairs of cells and \( T < 7 \), no improvement by cross-dependence is obtained. In general, the question is open.
Thus we turn to an upper bound. To this end observe that by (25)

\[ I(X_t; Y_t) = H(Y_t) - H(Y_t | X_t) \leq H(Y_t | X_{t-1}). \]  

(32)

Thus defining \( X_0 \) to be 0 with probability 1, (30) is bounded by

\[ \sum_{t=1}^{T} H(Y_t | X_{t-1}) \leq \sum_{t=1}^{N} \sum_{t=1}^{T} H(Y_t | X_{t-1}). \]  

(33)

Hence, an upper bound is the maximum of

\[ \sum_{t=1}^{T} H(Y_t | X_{t-1}) = \sum_{t=1}^{T} p_{t-1} h(p_t), \]  

(34)

with \( p_0 = 1 \).

This maximum can again be determined by dynamic programming. One finds that as \( T \) gets larger, except near the beginning and end, the optimal \( p_t \) tend to a common value. From this it follows that the average rate for large \( T \) tends to

\[ \max_{0 < p < 1} ph(p) = 0.6169476434..., \]  

(35)

which is attained for \( p = 0.703506... \).

This bound is a rough one and it is entirely possible that 0.5 is the largest possible asymptotic rate.

5.4. The Time Invariant Policy

If one makes it a requirement that the coding be the same at all stages (as in subsection 3.5), then on a single cell basis one has to maximize (refer to (30)) \( h(p^2) - ph(p) \) over \( p \). This maximum is 0.3831859... attained for \( p = 0.71356 \).

As in subsection 3.5 one can compare this to time invariant policies allowing cross-dependence between just pairs of cells. One such is \( p_{00} = 0.46, \ p_{01} = p_{10} = 0.27, \ p_{11} = 0 \) which gives a rate of 0.385525, showing that cross-dependence is beneficial.

6. INCLUSIVE OR, VARIABLE DELAY

6.1. Lower Bound

As in the exclusive OR case we allow \( X_t \) to be recovered from either \( Y_t \) or \( Y_{t-1} \) with the choice made at each \( t \). For \( T = 2 \) delay cannot help, for \( T > 2 \) we cannot rule out cross-dependence. Again, a lower bound, possibly sharp, is obtained on a single-cell basis. Then an upper bound is derived.
Interestingly, the upper bound is much closer to the lower bound than in the case without delay.

As in the exclusive OR case, one may clear the medium at every third cycle then use a stage with no delay followed by a stage with delay whose output is read while the input is clearing. The rates so obtained are the ones of Theorem 2. Asymptotically, the rate is $\frac{3}{4}$.

6.2. Upper Bound

Letting $X_0$ consist of $N$ zeros and $X_{T-1}$ of $N$ ones, the exact answer requires maximizing

$$\sum_{t=1}^{T} \max(I(X_t; Y_t), I(X_t; Y_{t-1}))$$

over all sequences $P_1, \ldots, P_T$ of distributions on $\{0, 1\}^N$ for the $X_t$.

Again, by (25), expression (36) is bounded by

$$\sum_{t=1}^{T} \max(H(Y_t | X_{t-1}), H(Y_{t-1} | X_{t-1})),$$

and as each vector conditional entropy can be bounded by the sum of the bitwise conditional entropies, this is bounded by

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \max(H(Y_t^i | X_{t-1}^i), H(Y_{t+1}^i | X_{t+1}^i)).$$

With $p_t = \Pr(X_t = 0)$, this reduces to maximizing over all sequences $p_1, \ldots, p_T$ the expression

$$\sum_{t=1}^{T} h(p_t) \max(p_{t-1}, p_{t+1}),$$

where $p_0 = 1$, $p_{T+1} = 0$ and all other $p_t \in [0, 1]$.

As each term in (39) only involves 3 consecutive $p_t$, dynamic programming can be used. The computations show that as $T$ increases, apart from end effects, the optimal $p_t$ sequence has period 4 with every third and fourth $p_t$ equaling $\frac{1}{2}$. (It can be shown that for an even period, the maximum is attained with one probability $\frac{1}{2}$ in each parity class.) For the long run average one need only find the maximum over probabilities $\alpha, \beta$ of a periodic pattern $\frac{1}{2}, \frac{1}{2}, \alpha, \beta, \frac{1}{2}, \frac{1}{2}, \alpha, \beta, \ldots$, that is the maximum of

$$\frac{1}{4}(2 + h(\alpha) + h(\beta)).$$

It is attained for $\alpha = \beta = 0.81060\ldots$, and the corresponding upper bound on
the maximum achievable rate of $\frac{3}{4}$ by less than 3.5%. Strictly (Witsenhausen, to appear), from the value functions in the dynamic program, one can derive an upper bound of 0.7001 subject to the restriction of the input probabilities to multiples of 0.01.

7. A Time–Space Interchange

Consider a discrete memoryless $k$-input $k$-output channel. The inputs $X_1, \ldots, X_k$ and outputs $Y_1, \ldots, Y_k$ take values in $2k$ given alphabets. The channel is characterized by the transition probabilities

$$
\Pr\{Y_1 = y_1, \ldots, Y_k = y_k \mid X_1 = x_1, \ldots, X_k = x_k\}.
$$

Suppose this channel is used in the following way. There are $k$ independent sources $S_1, \ldots, S_k$. For each $i$ a stream of successive values for $X_i$ is derived from source $S_i$ by an encoder seeing only the output of $S_i$. The source outputs of $S_i$ are to be reconstructed in the Shannon sense by a decoder using only the stream of successive outputs $Y_i$. What $k$-tuples of source rates $R_i$ can be achieved in this way? This is an unsolved problem of the general interference channel.

Suppose now that for $j = 1, \ldots, k - 1$ $Y_1, \ldots, Y_j$ is independent of $X_{j+1}, \ldots, X_k$ given $X_1, \ldots, X_j$. This means that the interference propagates from lower to higher index outputs, but not in the opposite direction.

More specifically, suppose the channel can be described by the equations

$$
Y_i = g(X_i, S_{i-1}, N_i), \quad S_i = r(X_i, S_{i-1}, U_i),
$$

(40)

where the $S_i$ all take their values in the same finite alphabet, $S_0$ is a given character of that alphabet and the $N_i$ are independent random variables.

Then we have the situation described in Section 1 with time and space reversed: the input indices $i = 1, \ldots, k$, correspond to cycles of usage $t = 1, \ldots, T$, while the time streams of inputs and outputs correspond to the data read in and out of a large memory.

In this paper, we have concentrated on the case where the $N_i$ are constant (deterministic interference channel), all alphabets are $\{0, 1\}$, $S_i = r(X_i, S_{i-1}) = X_i$, $Y_i = g(X_i, S_{i-1}) = g(X_i, X_{i-1})$ with $g$ exclusive (inclusive) OR and $S_0 = X_0 = 0$.

However, we have only considered the problem of maximum average rate

$$
\frac{1}{k} \sum_{i=1}^{k} R_i
$$

and primarily the limit of this maximum as $k \to \infty$.

Received: September 22, 1983; Accepted: October 15, 1983
REFERENCES


