Reusable Memories in the Light of the Old Arbitrarily Varying and a New Outputwise Varying Channel Theory

Rudolf Ahlswede and Gabor Simonyi

Abstract—**Arbitrarily varying (AV) channels were introduced as a model for transmission in cases of jamming. It is shown that their theory applies naturally to memories and yields, in a unified way, some new and old capacity theorems for write-unidirectional (and more general) memories with side information. If the encoder has no side information, it is still not understood what the optimal rates for many cycles are. More insight through a theory of outputwise varying (OV) channels is expected.**

Index Terms—**Memories capacity channels, side information, Markov channels.**

I. Introduction

The starting point of our investigations are capacity problems for a certain kind of reusable memories, called WUM, which is an abbreviation for write-unidirectional memory. They have been introduced by Borden [5] and Willems–Vinck [15]. These authors explain how the concept arose in modeling an optical disk as a storage device with updating constraints imposed by laser technology. The reader is advised to consult Simonyi [14] and Cohen [6] for an account of the known results relevant for us here. We start right away with some basic definitions and go on with our own work.

Actually the models of Borden and Willems–Vinck are slightly different. This led to different notions of codes. Here we discuss only the WUM code in Willems–Vinck's sense, which has later been called alternating WUM code by Simonyi [14].

Definition: A family $\mathscr{C}=\{S_i\colon 1\le i\le M\}$ of subsets of $\mathscr{Y}^n=\{0,1\}^n$ is an alternating WUM code if

$$S_i\cap S_j=\varnothing \qquad (i\neq j), \tag{1.1}$$

$$S_i=T_{i0}\cup T_{i1}, \qquad \text{for } i=1,\cdots,M. \tag{1.2}$$

Manuscript received April 26, 1990; revised January 3, 1991. This work was performed in Bielefeld, August–October 1989 and was supported by SFB-343. This work was presented at the IEEE Information Theory Workshop, Veldhoven, The Netherlands, June 10–15, 1990.

R. Ahlswede is with Universität Bielefeld, Fakultät für Mathematik, Postfach 8640, 4800 Bielefeld 1, Germany.

G. Simonyi is with the Mathematical Institute, Hungarian Academy of Sciences, Realtanoda u. 13–15, 1053 Budapest, Hungary.

IEEE Log Number 9144767.

For all i, all j, and $y^n\in T_{i0}$ (resp. T_{i1}), there exists a $y'^n\in T_{j1}$ (resp. T_{j0}) with

$$y'^n\ge y^n \qquad (\text{resp } y'^n\le y^n). \tag{1.3}$$

The partial order "$\ge$" is defined by $y'^n=(y'_1,\cdots,y'_n)\ge y^n=(y_1,\cdots,y_n)$, iff $y'_t\ge y_t$ for $t=1,2,\cdots,n$.

Such a code can be used as follows. At every time instant the memory is in a state $y^n\in\mathscr{Y}^n$. There are two persons (or devices): the encoder E and the decoder D. They use the memory in so called cycles. In odd (resp. even) cycles the encoder can print only 1's (resp. 0's) in some of the n positions, that is change y^n in those positions to 1 (resp. 0). This is an updating of the memory. The purpose of the encoder is to store a new message $i\in\{1,\cdots,M\}$, which can be decoded (read) by the decoder.

Any word in the set T_{i0} (resp. T_{i1}) can represent message i in an even (resp. odd) cycle. Having read this word, by (1.1) the decoder can recover the message. Knowing the state of the memory, by (1.3), the encoder can change this state in the next cycle in order to store a new message.

Clearly, the code concept presented is for a situation, where the encoder knows the state of the memory before changing it to the next state, but the decoder needs to know only the state he is actually reading.

The issue of side information has played an important role in multiuser source and channel coding. We use the notation E_+ (resp. D_+), if the encoder (resp. decoder) has a specified side information, and the notation E_- (resp. D_-), if the encoder (resp. decoder) does not have this side information. For memory cells the side information refers to the knowledge of the content of the memory before a new encoding. For WOM's, papers [12] and [9] consider the case (E_+,D_-) and [17] analyze the remaining cases.

The case for WUM described above is (E_+,D_-). In [13] the other cases were first investigated. Using ϵ-error or 0-error as performance criteria, results in eight cases with corresponding capacities $C(E_+,D_+,\epsilon)$, $C(E_+,D_-,0)$, etc. The known results are described in [6], (see Theorem 0).

We derive a lower bound for $C(E_-, D_-, \epsilon)$ (Theorem 1) and we give a complete characterization for $C(E_-, D_+, \epsilon)$ (Theorem 2). Perhaps more importantly, we derive *all* known capacities by a unified approach, that is, AV-channel theory.

Furthermore, we gain a first understanding of the role of cycles via OV-channels, a concept we introduce. The name refers to "outputwise varying channels." In particular we derive exact conditions for memories to have positive capacity "in the long run." Actually all our results are for much more general memories than the WUM.

II. Results

According to Cohen's survey [6], at least four capacities are known. We summarize the results.

Theorem 0:

a) $C(E_+, D_+, \epsilon) = C(E_+, D_-, \epsilon) = \log(1+\sqrt{5})/2 \sim 0.694$;

b) $C(E_+, D_+, 0) = C(E_+, D_-, 0) = \log(1+\sqrt{5})/2$.

Our first contribution is the following.

Theorem 1:

$$C(E_-, D_-, \epsilon) \geq 0.545.$$

This result has been first claimed by Godlewski according to [6]. He had the idea to relate the capacity of the WUM to the transmission capacity of the so called Z-channel. However, neither the orally presented sketch of a proof nor the sketch in [6] provide a complete set of ideas necessary for a proof. The gap can naturally be closed with the help of AV-channel theory.

Next we show that our approach gives the complete solution for another case.

Theorem 2:

$$C(E_-, D_+, \epsilon) = \log(1+\sqrt{5})/2.$$

Moreover, we show that all other known results (Theorem 0) can be derived in a unified way. Phenomena from the theory of AV-channels (see [3]) have complete analogs for WUM's as for instance the fact that

$$C(E_+, D_-, 0) = C(E_+, D_-, \epsilon). \tag{2.1}$$

Remarks: We draw attention to the fact that I. Csiszár has shown (lecture at the Information Theory Meeting in Oberwolfach, 1989) that certain storage problems for memories can be viewed as coding problems for AV-channels. For instance memories with some cells stuck at zero and one can be regarded as AVC's with three possible states (cells stuck at zero, stuck at one, or good).

However we emphasize that this connection is entirely different from ours. In the present paper, methods from the theory of AVC's, in particular robustification techniques, serve as tools for solving memory problems. The memories are not modeled as AV-channels.

The second part of the paper was motivated by an attempt to prove a converse for Theorem 1. The converses of Willems [16] and Ozarow–Wyner (mentioned in [16]) address only 1-cyclic coding strategies. (This means that encoding and decoding strategies are the same for every cycle-modulo the change of the role of 0's and 1's.) Similarly "r-cyclic" means that they are the same for the block of the first r consecutive cycles, the block of the next r consecutive cycles, etc. In this sense "∞-cyclic" means, that all coding strategies are permitted. It is not clear whether the upper bound obtained holds for any number of cycles. An attempt to understand this question has convinced us that the "cycle problem" constitutes a new area of research in information theory. We try to understand it from first principles and ask a seemingly simple question such as "under which condition is for memories the ∞-cyclic capacity positive?" This naturally leads us to introduce and study OV-channels. Their definition follows.

Let $\mathscr{X}, \mathscr{Y}$ be finite sets, let $\mathscr{S}$ be a set of states, and let $\mathscr{W} = \{w(\cdot|\cdot|s): s \in \mathscr{S}\}$ be a set of stochastic $|\mathscr{X}| \times |\mathscr{Y}|$-matrices. It is assumed that

$$\mathscr{S} = \mathscr{Y}. \tag{2.2}$$

(This assumption describes the typical character of WUM's or of a more general memory concept, namely that the result of the previous writing determines the new state that governs what can happen at the next writing. For WUM's, cf. the remark in Section VII.)

Thus the sequence $(\mathscr{W}^n)_{n=1}^{\infty}$, with

$$\mathscr{W}^n = \{w(\cdot|\cdot|s^n): s^n \in \mathscr{S}^n\} \tag{2.3}$$

$$w(y^n|x^n|s^n) = \sum_{t=1}^{n} w(y_t|x_t|s_t), \tag{2.4}$$

for all $x^n = (x_1, \cdots, x_n) \in \mathscr{X}^n$, $y^n \in \mathscr{Y}^n$ and $s^n \in \mathscr{Y}^n$ defines a special AV-channel. We speak here of an OV-channel if the mode of operation is such that in the lth cycle of length n there is a probability distribution Q_l on the states $\mathscr{S}^n$ that equals the output distribution in the $(l-1)$st cycle. We say that $\{(u_i(l), D_i(l)): 1 \leq i \leq M_l;\ 1 \leq l \leq L\}$ is an $(n, (M_1, \cdots, M_L), \bar{\lambda})$-code for the OVC, if

$$u_i(l) \in \mathscr{X}^n, \qquad D_i(l) \subset \mathscr{Y}^n, \qquad \text{for } 1 \leq i \leq M_l,\ 1 \leq l \leq L, \tag{2.5}$$

$$D_i(l) \cap D_{i'}(l) = \varnothing, \qquad \text{for } i \neq i', \tag{2.6}$$

$$\frac{1}{M_l} \sum_{i=1}^{M_l} \sum_{s^n \in \mathscr{S}^n} w(D_i(l)|u_i(l)|s^n) Q_l(s^n) \geq 1 - \bar{\lambda} \qquad \text{for } 1 \leq l \leq L, \tag{2.7}$$

where

$$Q_l(s^n) = \frac{1}{M_{l-1}} \sum_{i=1}^{M_{l-1}} \sum_{s'^n \in \mathscr{S}^n} w(s^n|u_i(l-1)|s'^n) Q_{l-1}(s'^n) \tag{2.8}$$

and Q_0 is any initial distribution on $\mathscr{S}^n$. One way to define capacity is this.

R is $(n, \bar{\lambda})$ achievable if in the definition above $L = \infty$ and for any $\delta > 0$

$$\liminf_{L' \to \infty} \frac{1}{L'} \sum_{l=1}^{L'} \frac{1}{n} \log M_l \geq R - \delta. \tag{2.9}$$

R is $\bar{\lambda}$ achievable, if (2.9) holds for all $n \geq n_0(\bar{\lambda}, \delta)$. Finally R is achievable if it is $\bar{\lambda}$ achievable for all $\bar{\lambda} > 0$. The maximal achievable rate R is the capacity. The main point of this definition is that it expresses what is achievable in the "long run," that is, with repeated cycles. It will become clear from our proof of Theorem 1 that the problem reduces to an investigation of the behavior of information quantities. For this it suffices to consider the case $n = 1$, to fix an initial distribution Q_0 on $\mathscr{S}$ and to choose a sequence $P_0, P_1, \cdots, P_K$ of distributions on the input alphabet $\mathscr{X}$. They determine a state sequence $Q_0, Q_1, \cdots, Q_K$ and "average channels" $W(\cdot|\cdot|Q_i) = \sum_s w(\cdot|\cdot|s) Q_i(s)$.

Determine

$$C(Q_0) \triangleq \lim_{K \to \infty} \max_{P_0, P_1, \cdots, P_K} \frac{1}{K} \sum_{i=0}^{K} I(P_i, W(\cdot|\cdot|Q_i)).$$

In particular it is of interest to know when $C \triangleq \min_{Q_0} C(Q_0) > 0$. We give a complete answer to this question. We also give conditions for $C(Q_0)$ to be positive. A memory-like device can be used with positive rate if there is any Q_0 with $C(Q_0) > 0$, so in this sense $C^* \triangleq \max_{Q_0} C(Q_0)$ is the relevant quantity. Our analysis is based on well-known facts about the limiting behaviour of Markov chains with finite state space $\mathscr{Z} = \{1, 2, \cdots, c\}$. A good presentation can be found in [18].

We need some definitions. Let $W: \mathscr{Z} \rightsquigarrow \mathscr{Z}$ be a stochastic matrix defining a Markov chain.

For $Q = (Q_1, \cdots, Q_c) \in \mathscr{P}(\mathscr{Z})$, the set of PD's on $\mathscr{Z}$, we call

$$\operatorname{supp}(Q) \triangleq \{z | Q_z \neq 0\} \tag{2.10}$$

the support of Q and we denote the trace of Q under W by

$$\operatorname{tr}(Q) = \bigcup_{t=0}^{\infty} \operatorname{supp}(QW^t). \tag{2.11}$$

For $Q = \delta_z$ we also write $\operatorname{tr}(z) = \operatorname{tr}(\delta_z)$. The set $I \subseteq Z$ is called W-invariant, if $\operatorname{tr}(z) \subseteq I$ for all $z \in I$. An invariant $I \neq \varnothing$ is called minimal invariant, if no proper nonempty subset is invariant.

Theorem (see [18]): For the stochastic matrix $W: \mathscr{Z} \rightsquigarrow \mathscr{Z}$, there is a partition

$$\mathscr{Z} = E \mathbin{\dot{\cup}} D_1 \cdots \mathbin{\dot{\cup}} D_r$$

with subpartitions

$$D_\rho = D_{\rho,0} \mathbin{\dot{\cup}} \cdots \mathbin{\dot{\cup}} D_{\rho, d_\rho - 1}, \qquad d_\rho \geq 1,$$

for $\rho = 1, \cdots, r$, which has the following properties.

a) D_ρ is minimal-invariant with respect to W.

b) For $Q \in \mathscr{P}(\mathscr{Z})$ with $\operatorname{supp}(Q) \subseteq D_{\rho,\nu}$ necessarily $\operatorname{supp}(QW^t) \subseteq D_{\rho,\nu+t}$ $(t = 0, 1, 2, \cdots)$ and equality holds here for t sufficiently large. The second index $\nu + t$ is taken mod d_ρ.

c) For every $D_{\rho,\nu}$ there is actually one $q_{\rho,\nu} \in \mathscr{P}(\mathscr{Z})$, $\operatorname{supp}(q_{\rho,\nu}) \subset D_{\rho,\nu}$ with $q_{\rho,\nu} W^t = q_{\rho,\nu+t}$ $(t = 0, 1, 2, \cdots)$. Furthermore, for all $Q \in \mathscr{P}(\mathscr{Z})$, $\operatorname{supp}(Q) \subset D_{\rho,\nu}$,

$$\lim_{t \to \infty} \|QW^t - q_{\rho,\nu+t}\| = 0 \qquad \text{(exponentially fast)}.$$

d) For every $Q \in \mathscr{P}(\mathscr{Z})$ there is exactly one $q \in \operatorname{conv}\{q_{\rho,\nu}: \rho = 1, \cdots, r;\ \nu = 0, \cdots, d_\rho - 1\}$ with

$$\lim_{t \to \infty} \|QW^t - qW^t\| = 0 \qquad \text{(exponentially fast)}.$$

We call $\vec{Q} = (Q_0, Q_1, \cdots, Q_{b-1})$ a b-orbit of W, if for $i = 0, \cdots, b-1$ and $j = 1, 2, \cdots$

$$Q_i W^j = Q_{(i+j) \bmod b}, \tag{2.12}$$

and denote the set of these b-orbits by $\mathscr{O}(W, b)$. Also, $\mathscr{O}(W) = \bigcup_{b=1}^{\infty} \mathscr{O}(W, b)$. Finally, we abbreviate the stochastic matrix $(W(\cdot|x, s))_{s \in \mathscr{S}}$ by W_x.

Theorem 3:

$$C > 0 \Leftrightarrow \bigcap_{x \in \mathscr{X}} \mathscr{O}(W_x) = \varnothing.$$

The structure of the sets $\mathscr{O}(W_x)$ is known as far as the $q^x_{\rho,\nu}$ corresponding to W_x in Theorem 2 are known. The positivity of $C(Q_0)$ is harder to analyze. The following result shows that for $|\mathscr{S}| = 2$ the positivity of $C(Q_0)$ is equivalent to that of C.

Theorem 4: If $|\mathscr{S}| = 2$, then for all Q_0 on $\mathscr{S}$ $C(Q_0) > 0$ exactly if $\bigcap_{x \in \mathscr{X}} \mathscr{O}(W_x) = \varnothing$. (An equivalent, more explicit condition is (7.13).)

III. Proof of Theorem 1

A. Heuristics

According to [6] Godlewski had the idea to look at the WUM in case (E_-, D_-) as follows: Two users U_1 and U_2 "communicate" over the WUM. User U_1 (odd cycles) transforms certain 0's in the memory to 1's and U_2 (even cycles) certain 1's to 0's. Let us look first at odd cycles and let us assume that "in average" there are $\sim Q(0)n$ 0's in the memory, that is, $s^n \in \mathscr{T}^n_{Q,\delta}$, the set of typical sequences with δ deviation. If U_1 encodes messages $1, \cdots, M$ by $u_1, u_2, \cdots, u_M \in \mathscr{X}^n$, and sends u_i, then U_2 receives $u_i \vee s^n$. Let us also assume that $u_i \in \mathscr{T}^n_P$ for $i = 1, \cdots, M$. Then "in the average" the received $y^n = u_i \vee s^n$ is in $\mathscr{T}^n_{Q',\delta'}$, where

$$Q' = P \cdot W(\cdot|\cdot|Q) \text{ with } W(\cdot|\cdot|Q) = \begin{array}{c c} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} Q(0) & Q(1) \\ 0 & 1 \end{pmatrix} \end{array},$$

and therefore this writing can be "visualized" as the operation of a Z-channel $W(\cdot|\cdot|Q)$.

We get a lower bound on its capacity by imposing symmetry: The density $Q'(0)$ of 0's before U_2 writes shall be equal to $Q(1)$, the density of 1's before U_1 writes. The situation in the cycles are then symmetrically the same.

The imposed condition is

$$Q'(0) = P(0)Q(0) = Q(1). \tag{3.1}$$

We get the mutual information

$$(1-P(0))h(Q(0)) = \left(2 - \frac{1}{Q(0)}\right)h(Q(0))$$

and the optimal rate

$$C_Z \triangleq \max_{\rho \geq \frac{1}{2}} \left(2 - \frac{1}{\rho}\right)h(\rho). \tag{3.2}$$

The maximum is assumed at $Q^*(0) \sim 0.776$ and has the value ~ 0.545. In "conclusion" $C(E_-, D_-, \epsilon) \geq 0.545$.

Apparently this is not a rigorous proof, because in the original model there are no probabilities ruling the states of the memory. The performance has to be good for every state of the memory and not only for some kind of an average expressed by the Z-channel. This is exactly the kind of situation for which AV-channel theory is appropriate!

Define $\mathscr{W} = \{w(\cdot|\cdot|0), w(\cdot|\cdot|1)\}$ with

$$w(\cdot|\cdot|0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad w(\cdot|\cdot|1) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \tag{3.3}$$

and a DMC with transmission matrix

$$W(\cdot|\cdot|Q) = w(\cdot|\cdot|0)Q(0) + w(\cdot|\cdot|1)Q(1). \tag{3.4}$$

Thus

$$W(y^n|x^n|Q^n) = \sum_{s^n} w(y^n|x^n|s^n)Q^n(s^n),$$

where

$$Q^n = \prod_1^n Q.$$

We need two methods from the theory of AV-channels, which were developed in [1] and [2]. The present formulations are essentially those from [3]. A few definitions are needed.

B. Some Notation

We use for any finite set $\mathscr{Z}$ the following notions: $\mathscr{P}(\mathscr{Z}) \triangleq$ set of all distributions on $\mathscr{Z}$, $\mathscr{P}(n, \mathscr{Z}) \triangleq \{p \in P(\mathscr{Z})$: $p(z)n$ is integral for all $z \in \mathscr{Z}\}$; $z^n = (z_1, \cdots, z_n)$ is said to be (q, δ)-typical (in the terminology of [2]) if

$$||\{t: z_t = z\}| - q(z)n| \leq \delta n, \quad \text{for all } z \in \mathscr{Z}, \text{ and } \mathscr{Z}^n(q, \delta)$$

is the set of those sequences in $\mathscr{Z}^n$. $(q,0)$-typical sequences are also said to be of type q. Often, if the reference set is clear, with a hint to typicality $\mathscr{Z}^n(q, \delta)$ is written as $\mathscr{T}^n_{q,\delta}$. For $\delta = 0$ we omit the δ. We use the symmetric group Σ_n, that is, the group of permutations acting on $\{1, 2, \cdots, n\}$. Every $\tau \in \Sigma_n$ induces a bijection π: $\mathscr{S}^n \to \mathscr{S}^n$ defined by $\pi s^n \triangleq (s_{\tau(1)}, \cdots, s_{\tau(n)})$ for $s^n = (s_1, \cdots, s_n) \in \mathscr{S}^n$. Π_n denotes the group of these bijections. Their restrictions to $\mathscr{S}^n(p)$, $p \in \mathscr{P}(n, \mathscr{S})$, are also bijective.

C. The Robustification Technique

Theorem RT: If g: $\mathscr{S}^n \to [0,1]$ satisfies for an $\alpha \in [0,1]$ the inequality

$$\sum_{s^n \in \mathscr{S}^n} g(s^n)p^n(s^n) > 1 - \alpha,$$

for $p^n = \prod_1^n p$ with $p \in \mathscr{P}(n, \mathscr{S})$, then it also satisfies the inequality

$$\frac{1}{n!} \sum_{\pi \in \Pi_n} g(\pi s^n) > 1 - \alpha_n, \qquad \text{for all } s^n \in \mathscr{S}^n(p),$$

where $\alpha_n = \alpha(n+1)^{|\mathscr{S}|}$.

Proof: The first inequality is equivalent to

$$\sum_{s^n} (1 - g(\pi s^n))p^n(\pi s^n) < \alpha, \qquad \text{for } \pi \in \Pi_n$$

because π is bijective. Since $p^n(\pi s^n) = p^n(s^n)$, it follows that

$$\sum_{s^n} \left(1 - \frac{1}{n!} \sum_{\pi} g(\pi s^n)\right) p^n(s^n) < \alpha.$$

Here $(1 - (1/n!)\Sigma_\pi g(\pi s^n)) \geq 0$ and therefore the left side is decreased when summing for $s^n \in \mathscr{S}^n(p)$ only. For these s^n $\Sigma_\pi g(\pi s^n)$ is constant and we get

$$\left(1 - \frac{1}{n!} \sum_{\pi} g(\pi s^n)\right) p^n(\mathscr{S}^n(p)) < \alpha, \quad \text{for } s^n \in \mathscr{S}^n(p).$$

With the well-known fact

$$p^n(\mathscr{S}^n(p)) \geq (n+1)^{-|\mathscr{S}|},$$

we finally conclude

$$\left(1 - \frac{1}{n!} \sum_{\pi} g(\pi s^n)\right)(n+1)^{-|\mathscr{S}|} < \alpha, \qquad \text{for } s^n \in \mathscr{S}^n(p)$$

and thus the result follows.

D. Application of Theorem RT

We need the following concepts. A correlated (n, M)-code is specified by a finite probability space (Γ, μ) and a collection $\{(\mu_i^\gamma, D_i^\gamma)$: $1 \leq i \leq M\}_{\gamma \in \Gamma}$ of (n, M)-codes. In using such a code, the index γ is chosen according to the random experiment (Γ, μ), and thus sender and receiver use the code indexed by γ. Since γ has to be made known to both of them, there must be a common knowledge or correlation in the system. It serves here only as a mathematical tool. For any $\mathscr{S}^{*n} \subset \mathscr{S}^n$ the average error is measured by

$$\max_{s^n \in \mathscr{S}^{*n}} \sum_{\gamma \in \Gamma} \mu(\gamma) \frac{1}{M} \sum_{i=1}^M w\left((D_i^\gamma)^c | u_i^\gamma | s^n\right).$$

Suppose now that for some (n, M)-code $\{(u_i, D_i)$: $1 \leq i \leq M\}$ and for $Q \in \mathscr{P}(n, \mathscr{S})$

$$\frac{1}{M} \sum_{i=1}^M \sum_{s^n \in \mathscr{S}^n} w(D_i|u_i|s^n)Q^n(s^n) > 1 - \alpha, \tag{3.5}$$

where α is exponentially small.

For $g: \mathcal{S}^n \to [0,1]$ defined by

$$\frac{1}{M}\sum_{i=1}^{M} w(D_i|u_i|s^n), \tag{3.6}$$

the inequality (3.5) ensures the validity of the hypothesis of Theorem RT and therefore

$$\frac{1}{n!}\sum_{\pi\in\Pi_n}\frac{1}{M}\sum_{i=1}^{M} w(D_i|u_i|\pi s^n) > 1-\alpha_n$$

or, equivalently, for $s^n\in\mathcal{S}^n(Q)$

$$\frac{1}{n!}\sum_{\pi\in\Pi_n}\frac{1}{M}\sum_{i=1}^{M} w(\pi^{-1}D_i|\pi^{-1}u_i|s^n) > 1-\alpha_n. \tag{3.7}$$

But this says that the correlated code specified by the collection of codes

$$\{(\pi^{-1}u_i,\pi^{-1}D_i): 1\le i\le M\}_{\pi\in\Pi_n}$$

and the uniform distribution on Π_n has an average error probability less than α_n for all $s^n\in\mathcal{S}^n(Q)$. Clearly, $a_n \le e^{-\epsilon n}$ for $\alpha=(n+1)^{-|\mathcal{S}|}e^{-\epsilon n}$.

Now we go a step further. Let

$$\mathcal{P}_{Q^*,\eta}(n,\mathcal{S}) = \left\{Q'\in\mathcal{P}(n,\mathcal{S}): \sum_s |Q^*(s)-Q'(s)|\le\eta\right\} \tag{3.8}$$

be the set of types in the η-neighbourhood of Q^*, which was defined in the line following (3.2). From the theory of compound channels ([19]) we know that an (n,M,α)-code $\{(u_i,D_i): 1\le i\le M\}$ exists such that (3.5) holds for *all* $Q\in\mathcal{P}_{Q^*,\eta}(n,\mathcal{S})$ and

$$\frac{1}{n}\log M \ge \max_{P\in\mathcal{P}(\mathcal{X})}\ \min_{Q\in\mathcal{P}_{Q^*,\eta}(n,\mathcal{S})} I(P,W(\cdot|\cdot|Q)) - r$$

$$\ge \max_P I(P,W(\cdot|\cdot|Q^*))f(\eta,r),$$

where

$$\lim_{\eta\to 0,r\to 0} f(\eta,r)=1 \text{ and } n\ge n_0(\eta,r).$$

Theorem RT implies now that (3.7) holds for all $s^n\in \bigcup_{Q\in\mathcal{P}_{Q^*,\eta}(n,\mathcal{S})}\mathcal{S}^n(Q)$.

E. Applications of the Elimination Technique (ET)

We use a result of [1], which says how the randomization in a correlated code can be reduced drastically.

Theorem ET: If for $\epsilon>0$, $\lambda\in(0,1)$ and n sufficiently large for any $\mathcal{S}'^n\subset\mathcal{S}^n$ the average error of the correlated code

$$(\Gamma,\mu;(\mu_i^\gamma,D_i^\gamma): 1\le i\le M)_{\gamma\in\Gamma}$$

satisfies for all $s^n\in\mathcal{S}'^n$

$$\sum_{\gamma\in\Gamma}\mu(\gamma)\frac{1}{M}\sum_{i=1}^{M} w((D_i^\gamma)^c|u_i^\gamma|s^n)\le e^{-\epsilon n},$$

then a $\Gamma^*\subset\Gamma$ with

$$|\Gamma^*|\le n^2 \tag{3.9}$$

and a μ^* exist such that the code

$$(\Gamma^*,\mu^*;(u_i^\gamma,D_i^\gamma): 1\le i\le M)_{\gamma\in\Gamma^*}$$

satisfies for its (even maximum) probability of error

$$\max_{1\le i\le M}\ \max_{s^n\in\mathcal{S}'^n}\sum_{\gamma\in\Gamma^*}\mu^*(\gamma)w((D_i^\gamma)^c|u_i^\gamma|s^n)\le\lambda.$$

Since U_1 and U_2 can "clean" the memory by sending 1's resp. 0's only, they obviously can communicate over the WUM at a positive rate. By sending the outcome of (Γ^*,μ^*), performed at the sender's side, the correlation can be eliminated totally without an essential loss in rate.

F. Updating in Many Cycles

We know, from the theory of compound channels, that the code in E can actually be chosen with the properties

$$u_i\in\mathcal{T}_P^n,\qquad D_i\subset\bigcup_{Q'\in\mathcal{P}_{Q^*,\eta}(n,\mathcal{S})}\mathcal{T}_\delta^n(W(\cdot|u_i|Q')), \tag{3.10}$$

that is, the decoding sets contain only sequences generated from u_i by some $W(\cdot|\cdot|Q')$ from the neighborhood of $W(\cdot|\cdot|Q^*)$. Since u_i is $(P,0)$-typical all sequences in $\bigcup_{i=1}^M D_i$ are $(Q,0)$-typical for some $Q\in\mathcal{P}_{Q^*,\eta^{(1)}}(n,\mathcal{S})$. Notice that $\eta^{(1)}\to 0$ as $(\delta,\eta)\to(0,0)$.

If the received sequence y^n is not in $\bigcup_{i=1}^M D_i$ an error is declared. Its probability is smaller than λ. For the second cycle, the first procedure will be iterated with $\mathcal{P}_{Q^*,\eta}(n,\mathcal{S})$ replaced by the slightly larger $\mathcal{P}_{Q^*,\eta^{(1)}}(n,\mathcal{S})$. Reiteration L times results in a total error probability less than $L\cdot\lambda$ and an increasing sequence $\eta^{(1)},\cdots,\eta^{(L)}$. The choice $\eta^{(l)}=l\cdot\eta$ would work: just make for a constant L, λ and η small enough. Finally, after L cycles, spend one cycle to clean the memory and in the next cycle use codewords $u\in\mathcal{T}_{Q^*}^n$ (or just one of them) to get into the proper states. Now repeat the L cycles as before. □

IV. Proof of Theorem 2

As in the proof of Theorem 1, we impose the symmetry condition yielding (3.1). We apply now random coding in the following form: $V_1,\cdots,V_M$ are i.i.d. RV's taking values in $\mathcal{T}_P^n$ with equal probabilities. The decoding sets $D_i(V_1,\cdots,V_M;s^n)$ are obtained by maximum likelihood decoding via the channel $w(\cdot|\cdot|s^n)$ and omission of those sequences y^n that are not generated by V_i via this channel. Then for the expected average error probability

$$\mathbb{E}\bar\lambda_{s^n}(V_1,\cdots,V_M;D_1(s^n),\cdots,D_M(s^n))\le\bar\lambda,$$

provided that for $\gamma>0$ and $n\ge n_0(\bar\lambda,\gamma)$

$$\log M\le I(P\times\cdots\times P,W(\cdot|\cdot|s^n))-\gamma n.$$

As before we consider only states

$$s^n\in\bigcup_{Q\in\mathcal{P}_{\bar Q,\eta}(n,\mathcal{S})}\mathcal{S}^n(Q),$$

where $\bar{Q}$ is determined below. Then the new states have the desired properties as in the previous proof. A random code is a correlated code and Theorem ET applies also, if codewords and/or decoding sets depend on s^n (see [3]).

Now, for $s^n \in \mathscr{S}^n(Q)$,

$$I(P \times \cdots \times P, w(\cdot|\cdot|s^n)) = Q(0)h(P(0)),$$

and by (3.1) this equals $Q(0)h((1-Q(0))/Q(0))$. Thus we get that

$$\bar{C} \triangleq \max_{Q(1) \le \frac{1}{2}} (1-Q(1))h\left(\frac{Q(1)}{1-Q(1)}\right)$$

is an achievable rate.

Notice that this is Borden's formula for $C(E_+, D_-, 0)$, which is shown to equal $C(E_+, D_-, \epsilon)$ in [6]. But this equals also $C(E_+, D_+, \epsilon)$ (Theorem 0, Part a)). Since obviously $C(E_+, D_+, \epsilon) \ge C(E_-, D_+, \epsilon)$, the proof is complete. □

V. AV-Channel Theory Gives also Theorem 0

If the sender and the receiver known s^n then they can use a code

$$\{(u_i(s^n), D_i(s^n)) : 1 \le i \le M\},$$

for $w(\cdot|\cdot|s^n)$ with

$$\log M \ge \sum_{t=1}^{n} \max_{P_t} I(P_t, w(\cdot|\cdot|s_t)) - \gamma n$$

and *error probability 0*.

Using symmetry and restricting s^n as usual, gives again the same optimal rate, because for this special AV-channel (WUM!) the maximizing P_t's can be chosen to be equal. This gives Theorem 0, Part a). In particular we understand why in the case (E_+, D_+) the capacity is the same for both error concepts. The same phenomenon occurs in case (E_+, D_-) exactly as for AV-channels in general (see [3]).

The formula in [3, p. 622],

$$C_Q = \max_{(U,S,X) \in \mathscr{R}_Q} [I(U \wedge Y) - I(U \wedge S)]$$

yields now

$$C(E^+, D^-, 0) \ge \max\{C_Q : Q \text{ satisfies } (3.1)\}.$$

As in the Appendix of [3] one can choose U such that $U = Y$ with probability 1. Then $I(U \wedge Y) - I(U \wedge S) = H(Y) - H(Y) + H(Y|S) = Q(0)h(P(0))$, a familiar expression.

VI. Proof of Theorem 3

For $\vec{Q} = (Q_0, \cdots, Q_{b-1}) \in \bigcap_{x \in \mathscr{X}} \mathscr{O}(W_x)$, we have for the initial distribution Q_0 that $Q_0 W_x^t$ is independent of x and therefore, for any $P_1, P_2, \cdots$,

$$\sum_x (Q_0 W_x^t) P_t(x) \text{ is independent of } P_t.$$

Thus, $I(P_t, W(\cdot|\cdot|Q_t)) = 0$ for all t, hence $C(Q_0) = 0$ and finally $C = 0$. Conversely, if $\bigcap_{x \in \mathscr{X}} \mathscr{O}(W_x) = \varnothing$ choose $P_1 = P_2 = \cdots = (1/|\mathscr{X}|, \cdots, 1/|\mathscr{X}|)$ and the associated stochastic matrix $\bar{W} = \Sigma_x (1/|\mathscr{X}|) W_x$. Determine $\mathscr{O}(\bar{W})$. For any initial Q_0, we have asymptotically the periodic behavior of $Q_0 \bar{W}^t$ described in Part d) of the Theorem in terms of the $\bar{q}_{\mathscr{S},\nu}$ corresponding to $\bar{W}$. Denote the limiting cycle by $(\bar{q}, \bar{q}\bar{W}, \cdots, \bar{q}\bar{W}^{\bar{b}-1})$. The cycles in $\mathscr{O}(W_x)$ have as a common period $\prod_{\rho=1}^{r(x)} d_\rho(x)$ and therefore the cycles in $(\bigcup_{x \in \mathscr{X}} \mathscr{O}(W_x)) \cup \mathscr{O}(\bar{W})$ have as common period

$$\left(\prod_{x \in \mathscr{X}} \prod_{\rho=1}^{r(x)} d_\rho(x)\right) \prod_{\rho=1}^{\bar{r}} \bar{d}_\rho \triangleq \Omega, \quad \text{say.}$$

Since $\mathscr{O}_1(W_x) = \{q : (q, \cdots,) \in \mathscr{O}(W_x)\}$ is compact, $\bigcap_{x \in \mathscr{X}} \mathscr{O}(W_x) = \varnothing$ implies

$$\sup_{x' \ne x} \inf_{\substack{q \in \mathscr{O}_1(W_x) \\ q' \in \mathscr{O}_1(W_{x'})}} \sum_{t=1}^{\Omega} \|qW_x^t - q'W_{x'}^t\| \ge \epsilon,$$

and therefore,

$$\lim_{T \to \infty} \max_x \sum_{T\Omega+1}^{(T+1)\Omega} \|\bar{q}\bar{W}^t - \bar{q}W_x^t\| \ge \frac{\epsilon}{2}.$$

This implies $C(Q_0) \ge f(\epsilon) > 0$ for all Q_0 and hence $C > 0$.

VII. Conditions for $C(Q_0)$ to be Positive: Proof of Theorem 4

By Theorem 3 the condition $\bigcap_x \mathscr{O}(W_x) = \varnothing$ is sufficient for $C(Q_0)$ to be positive. However, the condition may not be necessary, because starting in Q_0 it may be possible by suitable choice of $(P_1, P_2, \cdots)$ to avoid the common cycles. We shall show first that in case $\mathscr{X} = \mathscr{S} = \{0,1\}$ this is only possible at the price of having for $P_t = (p_t, \bar{p}_t)$

$$\lim_{t \to \infty} p_t \bar{p}_t = 0 \tag{7.1}$$

and thus,

$$\lim_{t \to \infty} I(P_t, \cdot) = 0. \tag{7.2}$$

Therefore, here still $C(Q_0)$ equals 0 if $\bigcap_x \mathscr{O}(W_x) \ne \varnothing$.

To see this, we investigate first the consequences of this condition for the two matrices

$$w(\cdot|\cdot|0) = \begin{pmatrix} \alpha & \bar{\alpha} \\ \beta & \bar{\beta} \end{pmatrix}, \quad w(\cdot|\cdot|1) = \begin{pmatrix} \gamma & \bar{\gamma} \\ \delta & \bar{\delta} \end{pmatrix}.$$

Suppose that $\vec{q} = (q_0, q_1, \cdots, q_{b-1})$ is a common orbit. Then necessarily

$$q_0 W_0 = q_0 W_1. \tag{7.3}$$

Therefore $QW_0 = QW_1$ must at least have a solution in $Q = (q, \bar{q})$, that is,

$$q\alpha + (1-q)\gamma = q\beta + (1-q)\delta$$

and thus the condition

$$(\alpha - \beta)q = (\delta - \gamma)\bar{q}. \tag{7.4}$$

We see that there is at most one such Q unless $\delta = \gamma$, $\alpha = \beta$.

In the latter $W(\cdot|\cdot|Q')$ has identical rows for all q' and thus $C(Q_0) = 0$ for all Q_0. We first simplify matters by assuming w.l.o.g. that $\alpha \geq \beta$. By (7.4), $\delta \geq \gamma$. Now by the assumed commonness of the orbits

$$q_1 W_0 = q_1 W_1$$

and therefore necessarily $q_1 = Q = q_0$. Hence,

$$QW_0 = QW_1 = Q, \tag{7.5}$$

and b can be chosen to equal 1.

We analyse this constraint next. Clearly $q\alpha + (1-q)\gamma = q$ or

$$q\bar{\alpha} = \bar{q}\gamma. \tag{7.6}$$

This and (7.4) yield

$$q\bar{\beta} = \bar{q}\delta \tag{7.7}$$

and thus

$$\bar{\alpha} : \bar{\beta} = \gamma : \delta. \tag{7.8}$$

Clearly $I(P, W(\cdot|\cdot|Q)) = 0$ for all $P \in \mathscr{P}(\mathscr{X})$.

But now we start in any Q_0 and try to choose $P_1, P_2, \cdots$ to do better! At step t we have $Q_t = (q_t, \bar{q}_t)$ with $q_t = q + \Delta q_t$. Then at step $t+1$

$$\begin{aligned} q_{t+1} &= p_t(q + \Delta q_t)\alpha + \bar{p}_t(q + \Delta q_t)\beta \\ &\quad + p_t(\bar{q} - \Delta q_t)\gamma + \bar{p}_t(\bar{q} - \Delta q_t)\delta \\ &= (p_t q\alpha + \bar{p}_t q\beta + p_t \bar{q}\gamma + \bar{p}_t \bar{q}\delta) \\ &\quad + (p_t\alpha + \bar{p}_t\beta - p_t\gamma - \bar{p}_t\delta)\Delta q_t. \end{aligned}$$

Since by (7.6) and (7.7) the term in the first bracket equals $p_t(q\alpha + q\bar{\alpha}) + \bar{p}_t(q\beta + q\bar{\beta}) = p_t q + \bar{p}_t q = q$ and since the term in the second bracket equals $(p_t(\alpha - \gamma) + \bar{p}_t(\beta - \delta))\Delta q_t$, we conclude that

$$|\Delta q_{t+1}| = |q_{t+1} - q| \leq |\Delta q_t| \, |p_t(\alpha - \gamma) + \bar{p}_t(\beta - \delta)|. \tag{7.9}$$

This is readily analysed by cases. Clearly, $\lim_{n\to\infty} q_n = q$, if

$$\max(|\alpha - \gamma|, |\beta - \delta|) < 1.$$

Otherwise, also if $m \triangleq \min(|\alpha - \gamma|, |\beta - \delta|) < 1$, we use a new argument. If $\epsilon \leq p_t \leq 1 - \epsilon$ for $\epsilon > 0$, then $|p_t(\alpha - \gamma) + \bar{p}_t(\beta - \delta)| \leq 1 - \epsilon + \epsilon \cdot m = 1 - \epsilon(1 - m)$.

However, if P_t does not satisfy the bounds, then $I(P_t, \cdot) \leq f(\epsilon) \to 0$ $(\epsilon \to 0)$ and we cannot maintain positive rate. We are left with the cases:

$$|\alpha - \gamma| = |\beta - \delta| = 1. \tag{7.10}$$

In case $\alpha = 1$, $\beta = 0$, $\gamma = 0$, $\delta = 1$, we get directly from (7.9)

$$|\Delta q_{t+1}| \leq \Delta q_t |2p_t - 1|$$

and, for $\epsilon \leq p_t \leq 1 - \epsilon$, we have $|2p_t - 1| \leq 1 - 2\epsilon$. The previous argument applies again.

By the restriction to $\alpha \geq \beta$, we are finally left with the cases

$$\left\{\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}\right\} \text{ and } \left\{\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}\right\}$$

as sets of matrices. This results in identical rows for every $Q' \in \mathscr{P}(\mathscr{S})$ and thus in $I(P, \cdot) = 0$.

Remark: For the WUM

$$\left\{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}\right\},$$

(7.8) does not hold, therefore $C > 0$.

Case: $|\mathscr{X}| = a \geq 2$. We are given

$$w(\cdot|\cdot|0) = \begin{pmatrix} \alpha(1) & \bar{\alpha}(1) \\ \vdots & \vdots \\ \alpha(a) & \bar{\alpha}(a) \end{pmatrix},$$

$$w(\cdot|\cdot|1) = \begin{pmatrix} \beta(1) & \bar{\beta}(1) \\ \vdots & \vdots \\ \beta(a) & \bar{\beta}(a) \end{pmatrix}.$$

If $\bigcap_{x \in \mathscr{X}} \mathscr{O}(W_x) \neq \varnothing$, then (7.3) to (7.8) take the form

$$QW_x = QW_{x'}, \qquad \text{for } x, x' \in \mathscr{X} \tag{7.11}$$

$$q\bar{\alpha}(x) = \bar{q}\beta(x), \qquad \text{for } x \in \mathscr{X} \tag{7.12}$$

$$\overline{\alpha(x)} : \overline{\alpha(x')} = \beta(x) : \beta(x'), \qquad \text{for } x, x' \in \mathscr{X}. \tag{7.13}$$

Furthermore,

$$\begin{aligned} q_{t+1} &= \sum_{x \in \mathscr{X}} P_t(x)(q + \Delta q_t)\alpha(x) \\ &\quad + \sum_{x \in \mathscr{X}} P_t(x)(\bar{q} - \Delta q_t)\beta(x) \\ &= \sum_{x \in \mathscr{X}} P_t(x)(q\alpha(x) + \bar{q}\beta(x)) \\ &\quad + \sum_{x \in \mathscr{X}} P_t(x)(\alpha(x) - \beta(x))\Delta q_t \\ &= q + \Delta q_t \sum_{x \in \mathscr{X}} P_t(x)(\alpha(x) - \beta(x)), \qquad \text{by (7.12)}. \end{aligned}$$

In the case $M \triangleq \max_x |\alpha(x) - \beta(x)| < 1$, we get $|\Delta q_{t+1}| = |q_{t+1} - q| \leq M|\Delta q_t|$ and the proof works as in the case $|\mathscr{X}| = 2$. Otherwise we have to go through cases again. First w.l.o.g. we can choose the ordering

$$\alpha(1) \geq \alpha(2) \geq \cdots \geq \alpha(a)$$

and then by (7.13) necessarily

$$\beta(1) \leq \beta(2) \leq \cdots \leq \beta(a).$$

Further reduction consists of omission up to one of those W_x that equals another $W_{x'}$. Furthermore, there are the following four candidates for a W_x with 0 and 1 as entries only:

		q
(10)	(01)	arbitrary
(10)	(10)	1
(01)	(01)	0
(01)	(10)	$\frac{1}{2}$.

Here the first column corresponds to state 0 and the second column to state 1. The values for q associated with them allow only to combine the first row with any of the three other rows.

$$\begin{pmatrix}1 & 0\\1 & 0\end{pmatrix}, \begin{pmatrix}0 & 1\\1 & 0\end{pmatrix}; \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}, \begin{pmatrix}0 & 1\\0 & 1\end{pmatrix}$$

$$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}, \begin{pmatrix}0 & 1\\1 & 0\end{pmatrix}.$$

Thus we arrived at the situation

$$\mathscr{X}^- = \{x: |\alpha(x) - \beta(x)| < 1\},$$
$$\mathscr{X}^+ = \{x: |\alpha(x) - \beta(x)| = 1\},$$
$$\mathscr{X} = \mathscr{X}^- \dot{\cup} \mathscr{X}^+, \qquad |\mathscr{X}^+| \le 2.$$

We write $\mathscr{X}^+ = \{x_1, \cdots\}$, $1 - \delta \triangleq \max_{x \in \mathscr{X}} |\alpha(x) - \beta(x)|$. We analyze (7.12) for the first case:

$$\begin{aligned}\left|\sum_{x \in \mathscr{X}} P_t(x)(\alpha(x) - \beta(x))\right| &\le \sum_{x \in \mathscr{X}^-} P_t(x)(1-\delta) + P_t(x_1)\\ &= \epsilon(1-\delta) + 1 - \epsilon - P_t(x_2),\\ &\qquad \text{if } \epsilon = \sum_{x \in \mathscr{X}^-} P_t(x),\\ &= 1 - \epsilon\delta - P_t(x_2).\end{aligned}$$

This gives bad contraction only if $\epsilon \to 0$, $P_t(x_2) \to 0$ $(t \to \infty)$, but then $P_t(x_1) \to 1$ $(t \to \infty)$ and $I(P_t, \cdot) \to 0$ $(t \to \infty)$. The second case is similar and the third case is also easily settled.

Remark: The positivity for $C(Q_0)$ can be formulated without any reference to information theory. For Q_0, $\{W_x: x \in \mathscr{X}\}$ and a sequence $(P_t)_{t=1}^\infty$, define $Q_t = Q_0 W_{P_1} \cdots W_{P_{t-1}}$ and consider

$$C'(Q_0) = \sup_{(P_t)_{t=1}^\infty} \varliminf_{n \to \infty} \frac{1}{n} \cdot \sum_{t=1}^{n} \min\left(\max_{x \in \mathscr{X}} \|Q_t W_x - Q_t W_{P_t}\|, \min_{x \in \mathscr{X}}(1 - P_t(x))\right).$$

Now $C'(Q_0) > 0 \Leftrightarrow C(Q_0) > 0$.

References

[1] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Z. Wahrscheinlichkeitstheorie verw. Geb.*, vol. 44, pp. 159–175, 1978.

[2] ____, "Coloring hypergraphs: A new approach to multi-user source coding," *J. Comb. Inf. and Syst. Sci.*, vol. 4, pt. I, pp. 76–115, 1979; vol. 5, pt. II, pp. 220–268, 1980.

[3] ____, "Arbitrarily varying channels with state sequence known to the sender," *IEEE Trans. Inform. Theory*, vol. IT-32, no. 5, pp. 621–629, Sept. 1986.

[4] R. Ahlswede and Z. Zhang, "Coding for write-efficient memory," *Inform. Computat.*, vol. 83, no. 1, pp. 80–97, 1989.

[5] J. M. Borden, "Coding for write-unidirectional memories," submitted to *IEEE Trans. Inform. Theory*.

[6] G. Cohen, "On the capacity of write-unidirectional memories," *Bull. Insti. Math. Academia Sinica*, vol. 16, no. 4, Dec. 1988.

[7] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.

[8] S. I. Gelfand and M. S. Pinsker, "Coding for a channel with random parameters," *Probl. Control and Inform. Theory*, vol. 9, no. 1, pp. 19–31, 1981.

[9] C. D. Heegard, "On the capacity of a permanent memory," in *1983 Conf. Inform. Sci. Syst.*, Johns Hopkins Univ., 1983.

[10] C. D. Heegard and A. ElGamal, "On the capacity of a computer memory with defects," *IEEE Trans. Inform. Theory*, vol. IT-29, no. 5, pp. 731–739, Sept. 1983.

[11] A. V. Kusnetsov and B. S. Tsybsakov, "Coding a memory with defective cells," transl. *Prob. Peredach. Inform.*, vol. 10, no. 2, pp. 52–60, 1974.

[12] R. L. Rivest and A. Shamir, "How to reuse a write-once memory," *Inform. Contr.*, vol. 55, pp. 1–19, 1982.

[13] G. Simonyi, "On write-unidirectional memories," Rep. Interne ENST 86 Doo 7, 1986.

[14] ____, "On write-unidirectional memory codes," *IEEE Trans. Inform. Theory*, vol. 35, no. 3, pp. 663–667, May 1989.

[15] F. M. J. Willems and A. Vinck, "Repeated recording for an optical disk," in *Proc. Seventh Symp. Inform. Theory in the Benelux*, Delft, The Netherlands, Delft Univ. Press, pp. 49–59, 1986.

[16] F. M. J. Willems, "Converses for write-unidirectional memories," EUT Rep. 89-E-220, Eindhoven, The Netherlands, May 1989.

[17] J. K. Wolf, A. D. Wyner, J. Ziv, and J. Körner, "Coding for a write-once memory," *AT&T Bell Labs. Tech. J.*, vol. 63, no. 6, pp. 1089–1112, 1984.

[18] K. Jacobs, "Selecta mathematica IV," in *Heidelberger Taschenbücher Band 98*. Berlin: Springer-Verlag, 1972.

[19] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Stat.*, vol. 30, pp. 1229–1241, 1959.