Erasure, List, and Detection Zero-Error Capacities for Low Noise and a Relation to Identification

Rudolf Ahlswede, Ning Cai, and Zhen Zhang, *Senior Member, IEEE*

***Abstract*—For the discrete memoryless channel $(\mathcal{X}, \mathcal{Y}, W)$ we give characterizations of the zero-error erasure capacity C_{er} and the zero-error average list size capacity $C_{\text{a}l}$ in terms of limits of suitable information (respectively, divergence) quantities (Theorem 1). However, they do not "single-letterize." Next we assume that $\mathcal{X} \subset \mathcal{Y}$ and $W(x|x) > 0$ for all $x \in \mathcal{X}$, and we associate with W the low-noise channel W_ε, where for $\mathcal{Y}^+(x) = \{y : W(y|x) > 0\}$**

$$W_\varepsilon(y|x) = \begin{cases} 1, & \text{if } y = x \text{ and } |\mathcal{Y}^+(x)| = 1 \\ 1-\varepsilon, & \text{if } y = x \text{ and } |\mathcal{Y}^+(x)| > 1 \\ \frac{\varepsilon}{|\mathcal{Y}^+(x)|-1}, & \text{if } y \neq x. \end{cases}$$

Our Theorem 2 says that as ε tends to zero the capacities $C_{\text{er}}(W_\varepsilon)$ and $C_{\text{a}l}(W_\varepsilon)$ relate to the zero-error detection capacity $C_{\text{de}}(W)$.

Our third result is a seemingly basic contribution to the theory of identification via channels. We introduce the (second-order) identification capacity C_{oid} for identification codes with zero misrejection probability and misacceptance probability tending to zero. Our Theorem 3 says that C_{oid} equals the zero-error erasure capacity for transmission C_{er}.

***Index Terms*—Zero-error erasure capacity, zero-error average list size capacity, zero-error detection capacity, identification with zero misrejection probability, low-noise channels.**

I. Introduction

We study a discrete memoryless channel (DMC) with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and transmission matrix W. By adding letters, if necessary, we can always assume that $\mathcal{X} \subset \mathcal{Y}$. Recall that for two words $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$

$$W^n(y^n|x^n) = \prod_{t=1}^{n} W(y_t|x_t). \tag{1.1}$$

Our studies are devoted to cases with zero-error probabilities for decisions (see [1]). They concern the performance of this channel for transmission codes under two criteria, namely, the erasure probability and the average list size. We also introduce identification codes with zero-error probability for misrejection.

Manuscript received December 15, 1993; revised July 4, 1995. This research was supported in part by NSF Grant NCR-9205265 and by SFB 343, Diskrete Strukturen in der Mathematik, Bielefeld University. This paper was presented at the IEEE International Symposium on Information Theory, Trondheim, Norway, June 1994.

R. Ahlswede and N. Cai are with Universität Bielefeld, Fakultät für Mathematik, 33501 Bielefeld, Germany.

Z. Zhang is with Communication Sciences Institute, University of Southern California, Los Angeles, CA 90089-2565 USA.

Publisher Item Identifier S 0018-9448(96)00013-2.

Let us fix any blocklength n. A code $\mathcal{C}$ for the channel is simply a subset of $\mathcal{X}^n$. $M = |\mathcal{C}|$ is the size of the code. For $y^n \in \mathcal{Y}^n$

$$\mathcal{L}(y^n, \mathcal{C}) = \{c \in \mathcal{C} : W^n(y^n|c) > 0\} \tag{1.2}$$

are the lists associated with $\mathcal{C}$ and

$$\ell(y^n, \mathcal{C}) = |\mathcal{L}(y^n, \mathcal{C})| \tag{1.3}$$

are their sizes. We use the short-hands $\mathcal{L}(y^n)$ and $\ell(y^n)$, if it is clear which code $\mathcal{C}$ is used.

The set of erasures is

$$\mathcal{Y}_{\text{er}} = \{y^n \in \mathcal{Y}^n : \ell(y^n) > 1\}. \tag{1.4}$$

The associated erasure probability is

$$P_{\text{er}} = \frac{1}{M}\sum_{c \in \mathcal{C}} \sum_{y^n \in \mathcal{Y}_{\text{er}}} W^n(y^n|c) \tag{1.5}$$

and the associated average list size is

$$\overline{L} = \frac{1}{M}\sum_{c \in \mathcal{C}} \sum_{y^n \in \mathcal{Y}^n} W^n(y^n|c)\ell(y^n). \tag{1.6}$$

Define $M(n, \lambda)$ as the maximal size of a code of blocklength n with erasure probability at most λ and define the (zero-error) erasure capacity

$$C_{\text{er}} = \lim_{\lambda \to 0} \overline{\lim_{n \to \infty}} \frac{1}{n} \log M(n, \lambda). \tag{1.7}$$

Similarily, define $\widehat{M}(n, \mu)$ as the maximal size of a code of blocklength n with average list size at most μ and define the (zero-error) average-list size capacity

$$C_{\text{al}} = \lim_{\mu \to 1+} \overline{\lim_{n \to \infty}} \frac{1}{n} \log \widehat{M}(n, \mu). \tag{1.8}$$

Our first result, Theorem 1 in Section II, gives a characterisation of both quantities, C_{er} and $C_{\text{a}l}$, in terms of limits of suitable information (respectively, divergence) quantities.

However, they do not "single-letterize:" already for

$$W = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} & 0 \\ 0 & \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{3}{4} \end{pmatrix}$$

a two-letter optimization is better than the one-letter optimization: rate value $0.6156\ldots$ versus $0.6128\ldots$.

Next we analyze our formulas for C_{er} and C_{al} for low-noise channels W_ε. They are defined by the properties that for every $x \in \mathcal{X} \subset \mathcal{Y}$ there is a nonempty $S(x) \subset \mathcal{Y} \setminus \{x\}$ with

$$W_\varepsilon(x|x) = 1 - \varepsilon$$

and

$$\begin{aligned} W_\varepsilon(y|x) &= \varepsilon|S(x)|^{-1}, && \text{if } y \in S(x) \\ W_\varepsilon(x|x) &= 1, && \text{if } S(x) = \phi \end{aligned} \tag{1.9}$$

where ε is small.

We establish relations to the capacity C_{de} of zero-error detection codes for W_ε. Recall that a detection code for a channel W of blocklength n is simply a subset $\mathcal{C} \subset \mathcal{X}^n \subset \mathcal{Y}^n$. The associated probability of undetected errors is

$$P_{\text{de}} = \frac{1}{M} \sum_{c \in \mathcal{C}} \sum_{c' \in \mathcal{C} \setminus \{c\}} W^n(c'|c). \tag{1.10}$$

In the classical AWAC system the receiver asks for retransmission, if his received word is not in $\mathcal{C}$, that is, if he detects an error.

$\mathcal{C}$ is a zero-error detection code, if $P_{\text{de}} = 0$, that is

$$W^n(c'|c) = 0 \text{ for all } c, c' \in \mathcal{C}, c \neq c'. \tag{1.11}$$

(More familiar are t-error detecting codes in algebraic coding theory.)

Our third result is a seemingly basic contribution to the theory of identification via channels ([11], [12]).

Recall that in identification the role of codewords is taken by probability distributions from $\mathcal{P}(\mathcal{X}^n)$, the set of all PD's on $\mathcal{X}^n$. Thus $\mathcal{C} = \{\mathcal{P}_i : 1 \leq i \leq \mathcal{N}\} \subset \mathcal{P}(\mathcal{X}^n)$ is an identification code.

We are now interested in a decoding rule $\{D_i : 1 \leq i \leq N\}$ which guarantees for all $i \in \{1, \cdots, N\}$ with probability one that i is accepted, if it is present. Therefore, necessarily

$$D_i \supset \left\{ y^n : \sum_{x^n} W^n(y^n|x^n) P_i(x^n) > 0 \right\}. \tag{1.12}$$

Furthermore, we are interested in having the maximal probability of misacceptance

$$P_{\text{ma}} = \max_i \max_{j \neq i} \sum_{y^n \in D_j} W^n(y^n|x^n) P_i(x^n)$$

small. Obviously, the best choice for the D_i's is with equality in (1.12). We call $\{(P_i, D_i) : 1 \leq i \leq N\}$ an identification code with zero misrejection probability and misacceptance probability P_{ma}. Let $N(n, \lambda)$ be the maximum size of such a code of length n and with $P_{\text{ma}} \leq \lambda$.

In short, we speak of the zero-error identification capacity C_{oid}, if

$$\begin{aligned} \inf_{\lambda > o} \underline{\lim}_{n \to \infty} \frac{1}{n} \log\log N(n, \lambda) &\geq C_{\text{oid}} \\ &\geq \inf_{\lambda > o} \overline{\lim}_{n \to \infty} \frac{1}{n} \log\log N(n, \lambda). \end{aligned} \tag{1.13}$$

Our Theorem 3 says that C_{oid} equals the zero-error erasure capacity for transmission C_{er}.

II. Non-Single-Letter Characterizations of C_{er} and C_{al}

We need some new definitions. For an input distribution P and a given channel W let the pair of RV's (X, Y) have the joint distribution $P \times W$. Y has the marginal distribution PW. We write for two matrices $\hat{W} \ll W$ and say that $\hat{W}$ is absolutely continuous with respect to W, if for all x, y $W(y|x) = 0$ implies $\hat{W}(y|x) = 0$. We call

$$\underline{I}(P, W) = \inf_{\hat{W} \ll W, P\hat{W} = PW} I(P, \hat{W}) \tag{2.1}$$

the "lower information" of X and Y (or for P and W). We write this quantity also as $\underline{I}(X \wedge Y)$ and introduce the "upper conditional entropy" by

$$\overline{H}(X|Y) = H(X) - \underline{I}(X \wedge Y). \tag{2.2}$$

Theorem 1: For every DMC with transmission matrix W

i) $C_{\text{er}} = \lim_{m \to \infty} \max_{P^{(m)}} \frac{1}{m} \underline{I}(P^{(m)}, W^m)$

ii)

$$C_{al} = \lim_{m \to \infty} \max_{P^{(m)}} \frac{1}{m} \min_{\substack{\hat{W}^m, \tilde{W}^m : \hat{W}^m \ll W^m \\ P^{(m)} \hat{W}^m = P^{(m)} \tilde{W}^m}} I(P^{(m)}, \hat{W}^m) + D(\tilde{W}^m \| W^m | P^{(m)})$$

where we use the conditional divergence

$$D(\tilde{W}^{(m)} \| W^m | P^{(m)}) = \sum_{x^m, y^m} P^{(m)}(x^m) \tilde{W}^{(m)}(y^m|x^m) \log \frac{\tilde{W}^{(m)}(y^m|x^m)}{W^m(y^m|x^m)}$$

iii) $C_{\text{er}} \geq C_{al}$.

Remarks:

1) We have been informed of independent work ([15]–[17]) by I. E. Telatar and Robert G. Gallager. The formula for C_{er} and the fact, that it does not "single-letterize," are also established in [17].
2) We are especially grateful to I. E. Telatar for drawing our attention to the fact that, quite amazingly, originally we used instead of our correct (2.8) a wrong formula for $|\mathcal{X}(y^n)|$ in (2.15).

Proof: i) We begin with the direct part. Select M codewords independently according to the uniform distribution on $\mathcal{T}_P^n$ (or $\mathcal{T}_X^n$), the set of words x^n of type $P_{x^n} = P$. Let this selection be described by the random variables $U_1, \cdots, U_M$. Its analysis requires a few auxiliary results. It proceeds via an upper bound on the mean value of $\ell(y^n)$.

Set first $Q = PW$ and consider $\mathcal{T}_{Q,\varepsilon}^n$, that is, the set of words $y^n \in \mathcal{Y}^n$, whose type P_{y^n} satisfies

$$|P_{y^n}(y) - Q(y)| \leq \varepsilon \text{ for } y \in \mathcal{Y}.$$

It is well known that for every $\hat{W}$ with $P\hat{W} = Q$

$$\hat{W}((\mathcal{T}_{Q,\varepsilon}^n)^c | x^n) \leq \exp\{-f(\varepsilon)n\} \quad \text{for some } f(\varepsilon) > 0, \text{ if } x^n \in \mathcal{T}_P^n. \tag{2.3}$$

It suffices therefore to consider any $y^n \in \mathcal{T}_{Q,\varepsilon}^n$ and to consider for it the set

$$\mathcal{X}(y^n) = \{x^n \in \mathcal{T}_P^n : W^n(y^n|x^n) > 0\}. \tag{2.4}$$

Let the joint type of (x^n, y^n) be denoted by $P \times \hat{W}_{x^n}$ and define

$$\mathcal{X}_{P,\hat{W}} = \{x^n : P_{x^n} = P, \hat{W}_{x^n} = \hat{W}\}. \tag{2.5}$$

It is well known that

$$|\mathcal{X}_{P,\hat{W}}| = \exp\{nH(X|\hat{Y}) + o(n)\} \tag{2.6}$$

if the pair of RV's $(X, \hat{Y})$ has distribution $P \times \hat{W}$.

Since

$$\mathcal{X}(y^n) = \bigcup_{P\hat{W}=Q, \hat{W} \ll W} \mathcal{X}_{P,\hat{W}} \tag{2.7}$$

and since there are only polynomially many types, (2.6) and (2.7) imply

$$|\mathcal{X}(y^n)| = \exp\left\{n \max_{P\hat{W}=Q, \hat{W}\ll W} H(X|\hat{Y}) + o(n)\right\}. \tag{2.8}$$

Now, a codeword is selected from $\mathcal{X}(y^n)$ with a probability smaller than $|\mathcal{X}(y^n)||\mathcal{T}_P^n|^{-1}$, which in turn is smaller than $\exp\{-n\underline{I}(X \wedge Y) + o(n)\}$.

For the expected value of the random erasure probability $P_{\text{er}}(U_1, \cdots, U_N)$ we get with (1.5) by symmetry

$$\mathbb{E}P_{\text{er}}(U_1, \cdots, U_N) = \mathbb{E} \sum_{y^n \in \mathcal{Y}_{\text{er}}(U_1, \cdots, U_N)} W^n(y^n|U_1),$$

where $\mathcal{Y}_{\text{er}}(\mathcal{U}_1, \cdots, \mathcal{U}_N)$ is the random erasure set for the random code $(U_1, \cdots, U_N)$.

Therefore, with (2.3)

$$\begin{aligned}\mathbb{E}P_{\text{er}}(U_1, \cdots, U_N) &\le \sum_{x^n \in T_P^n} \Pr(U_1 = x^n) \\ &\cdot \sum_{y^n \in T_{Q,\varepsilon}^n} \Pr(\{U_2, \cdots, U_N\} \cap \mathcal{X}(y^n) \ne \phi) \\ &\cdot W(y^n|x^n) + \exp\{-f(\varepsilon)n\} \\ &\le (M-1)\exp\{-n\underline{I}(X \wedge Y) + o(n)\} \\ &\cdot \sum_{x^n \in T_P^n} \Pr(U_1 = x^n) \\ &\cdot \sum_{y^n \in T_{Q,\varepsilon}^n} W^n(y^n|x^n) + \exp\{-f(\varepsilon)n\}.\end{aligned}$$

If now $M < \exp\{n\underline{I}(X \wedge Y) - n\delta\}$, then

$$\mathbb{E}P_{\text{er}}(U_1, \cdots, U_N) \le \exp\left\{-n\frac{\delta}{2}\right\} + \exp\{-f(\varepsilon)n\}$$

for n large enough.

The direct part is proved for $m = 1$ and can be proved for general m in exactly the same way.

We continue with the converse part. If $\mathcal{C}$ is a code of blocklength n and erasure probability λ, then

$$\frac{1}{n}\log|\mathcal{C}| = \frac{1}{n}H(X^n) \tag{2.9}$$

where X^n has uniform distribution $P^{(n)}$ over $\mathcal{C}$, and since for any $\hat{W}^{(n)} \ll W^n$, $P^{(n)}\hat{W}^{(n)} = P^{(n)}W^n$, the erasure probability is not increasing, we obtain from Fano's Lemma

$$H(X^n|\hat{Y}^n) \le h(\lambda) + \lambda \log|\mathcal{C}|. \tag{2.10}$$

Finally, (2.9) and (2.10) yield

$$\frac{1}{n}\log|\mathcal{C}| \le \frac{1}{n}\underline{I}(X^n \wedge Y^n) + o(\lambda) \le C_{\text{er}} + o(\lambda).$$

We complete the proof by letting λ go to zero.

ii) For the direct part we select M codewords at random as before.

Let the chosen code be $\mathcal{C} = \{c_1, \cdots, c_M\}$. Its average list size $\overline{L} = \overline{L}(c_1, \cdots, c_M)$ is

$$\overline{L} = \frac{1}{M}\sum_{c \in \mathcal{C}} \sum_{y^n \in \mathcal{Y}^n} W^n(y^n|c)\ell(y^n).$$

This can be written in the form

$$\overline{L} = \frac{1}{M}\sum_{c\in\mathcal{C}} \overline{L}(c), \quad \text{with } \overline{L}(c) = \sum_{y^n \in \mathcal{Y}^n} W^n(y^n|c)\ell(y^n). \tag{2.11}$$

Therefore, for any $c \in \mathcal{T}_P^n$

$$\mathbb{E}\overline{L}(U_1, \cdots, U_M) = \mathbb{E}\overline{L}(c, U_2, \cdots, U_M). \tag{2.12}$$

We estimate now the last quantity from above.

Let $\mathcal{P} = \mathcal{P}(c, \mathcal{Y}, n)$ be the set of all joint types P_{c,y^n}. For every $P \times V \in \mathcal{P}$ define the generated set

$$G_V(c) = \{y^n \in \mathcal{Y}^n : P_{c,y^n} = P \times V\}. \tag{2.13}$$

Then we have

$$W^n(G_V(c)|c) = \exp\{-nD(V\|W|P) + o(n)\}. \tag{2.14}$$

We estimate now the average list size for $y^n \in G_V(c)$. We obtain for $y^n \in G_V(c)$ by (2.8)

$$|\mathcal{X}(y^n)| = \exp\left\{n \max_{P\hat{W}=PV, \hat{W}\ll W} H(X|\hat{Y}) + o(n)\right\}. \tag{2.15}$$

or in terms of distributions

$$|\mathcal{X}(y^n)| = \exp\{n \max_{P\hat{W}=PV, \hat{W}\ll W} H(\hat{W}|P) + o(n)\}.$$

Each element of this set is selected as a codeword with probability $\exp\{-nH(P) + o(n)\}$. Therefore, the average list size for $y^n \in G_V(c)$ is at most

$$\exp\left\{-n \min_{\hat{W}: P\hat{W}=PV, \hat{W}\ll W} I(P, \hat{W}) + o(n)\right\}M + 1. \tag{2.16}$$

This gives with (2.11) and (2.14)

$$\begin{aligned}\mathbb{E}\overline{L}(c, U_2, \cdots, U_M) = \sum_{P\times V \in \mathcal{P}} M\exp\{-nD(V\|W/P) \\ \cdot \min_{\hat{W}: P\hat{W}=PV, \hat{W}\ll W} I(P, \hat{W}) + o(n)\} + 1.\end{aligned}$$

If now

$$\begin{aligned}M \le \min_{P\times V\in\mathcal{P}} \exp\{n(D(V\|W|P) \\ + \min_{\hat{W}: P\hat{W}=PV, \hat{W}\ll W} I(P, \hat{W})) - \varepsilon n\}\end{aligned}$$

then

$$\mathbb{E}\overline{L}(c) \le 1 + \exp\{-2\varepsilon n\} \tag{2.17}$$

and the direct part is proved.

For the converse part, let $\mathcal{C} \subset \mathcal{X}^n$ be a code with average list size $1+\lambda$ and size $|\mathcal{C}| = M$

$$\frac{1}{M}\sum_{c\in\mathcal{C}}\sum_{y^n\in\mathcal{Y}^{(n)}} W^n(y^n|c)\ell(y^n) = 1+\lambda \tag{2.18}$$

where $\mathcal{Y}^{(n)} = \{y^n \in \mathcal{Y}^n : \ell(y^n) \geq 1\}$.

Define X^n as in the previous converse proof and denote its distribution by $P^{(n)}$. Then

$$\frac{1}{n}\log M = \frac{1}{n}H(X^n) = \frac{1}{n}H(P^{(n)})$$

and

$$\sum_{x^n\in\mathcal{X}^n} P^{(n)}(x^n) \sum_{y^n\in\mathcal{Y}^{(n)}} W^n(y^n|x^n)\ell(y^n) = 1+\lambda. \tag{2.19}$$

We establish a connection to information quantities by showing first that for any $W^{(n)}$

$$\sum_{x^n} P^{(n)}(x^n) \sum_{y^n\in\mathcal{Y}^{(n)}} W^{(n)}(y^n|x^n)\log \ell(y^n) \leq D(W^{(n)}\|W^n|P^{(n)}) + \log(1+\lambda). \tag{2.20}$$

Clearly, by Jensen's inequality $\left(\mathbb{E}\exp\{Z\} \geq \exp\{\mathbb{E}Z\}\right)$

$$\begin{aligned}1+\lambda &= \sum_{x^n} P^{(n)}(x^n) \sum_{y^n\in\mathcal{Y}^{(n)}} W^n(y^n|x^n)\ell(y^n)\\
&= \sum_{x^n} P^{(n)}(x^n) \sum_{y^n\in\mathcal{Y}^{(n)}} W^{(n)}(y^n|x^n)\\
&\quad\cdot \exp\left\{-\log\frac{W^{(n)}(y^n|x^n)}{W^n(y^n|x^n)}\frac{1}{\ell(y^n)}\right\}\\
&\geq \exp\left\{-\sum_{x^n} P^{(n)}(x^n)\sum_{y^n\in\mathcal{Y}^{(n)}} W^{(n)}(y^n|x^n)\right.\\
&\quad\left.\cdot \log\frac{W^{(n)}(y^n|x^n)}{W^n(y^n|x^n)}\frac{1}{\ell(y^n)}\right\}\\
&= \exp\left\{-D(W^{(n)}\|W^n|P^{(n)}) + \sum_{x^n} P^{(n)}(x^n)\right.\\
&\quad\left.\cdot \sum_{y^n\in\mathcal{Y}^{(n)}} W^{(n)}(y^n|x^n)\log\ell(y^n)\right\}\end{aligned}$$

and thus (2.20) holds.

Next observe that for every $\hat{W}^n$ with $P^{(n)}\hat{W}^n = P^{(n)}W^{(n)}$

$$\begin{aligned}\sum_{x^n} P^{(n)}(x^n) &\sum_{y^n\in\mathcal{Y}^{(n)}} W^{(n)}(y^n|x^n)\log\ell(y^n)\\
&= \sum_{x^n} P^n(x^n)\sum_{y^n\in Y^n}\hat{W}^n(y^n|x^n)\log\ell(y^n)\\
&\geq H(X^n|\hat{Y}^n).\end{aligned} \tag{2.21}$$

The inequalities (2.20) and (2.21) imply

$$D(W^{(n)}\|W^n|P^{(n)}) + \log(1+\lambda) - H(X^n|\hat{Y}^n) \geq 0$$

and this yields

$$\begin{aligned}\frac{1}{n}\log M &= \frac{1}{n}H(X^n)\\
&\leq \frac{1}{n}(H(X^n) + D(W^{(n)}\|W^n|P^{(n)})\\
&\quad + \log(1+\lambda) - H(X^n|\hat{Y}^n)\\
&= \frac{1}{n}(I(X^n \wedge \hat{Y}^{(n)}) + D(W^{(n)}\|W^n|P^{(n)}))\\
&\quad + \frac{1}{n}\log(1+\lambda).\end{aligned} \tag{2.22}$$

Minimization over $W^{(n)}$ (which corresponds to $\tilde{W}^n$) and $\hat{W}^n$ completes the proof.

iii) This follows directly from the two definitions of the kinds of codes. Namely, if $\mathcal{C}$ has average list size $1+\lambda$ then $\mathcal{Y}_{\text{er}}$ has probability at most λ.

Remark 3: Notice that (2.19) is the substitute for Fano's inequality.

III. Capacities for Low-Noise Channels

For small ε, W_ε, defined in (1.9), is the prototype of a low-noise channel. We know that for its erasure capacity $C_{\text{er}}(\varepsilon)$ and for its average list capacity $C_{al}(\varepsilon)$ we have only the characterizations in terms of "non-single letter" information quantities of Theorem 1 in Section II.

However, if we know the limits

$$K_{\text{er}} = \lim_{\varepsilon\to 0} C_{\text{er}}(\varepsilon) \text{ and } K_{al} = \lim_{\varepsilon\to 0} C_{al}(\varepsilon) \tag{3.1}$$

then we have a certain knowledge also about the unknown quantities.

Let us use the abbreviations

$$C_{\text{er}}^n(\varepsilon) = \max_{P^{(n)}} \frac{1}{n}\underline{I}(P^{(n)}, W_\varepsilon^n) \tag{3.2}$$

$$\begin{aligned}C_{\text{al}}^n(\varepsilon) = \max_{P^{(n)}} \min_{W^{(n)}\ll W_2^n\, P^{(n)}W^{(n)} = P^{(n)}\widetilde{W}^{(n)}} &\frac{1}{n}(I(P^{(n)}, W^{(n)})\\
&+ D(\widetilde{W}^{(n)}\|W_\varepsilon^n|P^{(n)})).\end{aligned} \tag{3.3}$$

Then Theorem 1 says that

$$C_{\text{er}}(\varepsilon) = \lim_{n\to\infty} C_{\text{er}}^n(\varepsilon) \text{ and } C_{al}(\varepsilon) = \lim_{n\to\infty} C_{\text{al}}^n(\varepsilon) \tag{3.4}$$

and by (3.1) we have

$$K_{\text{er}} = \lim_{\varepsilon\to 0}\lim_{n\to\infty} C_{\text{er}}^n(\varepsilon), K_{al} = \lim_{\varepsilon\to 0}\lim_{n\to\infty} C_{al}^n(\varepsilon). \tag{3.5}$$

We study also the auxiliary quantities

$$\tilde{K}_{\text{er}} = \lim_{n\to\infty}\lim_{\varepsilon\to 0} C_{\text{er}}^n(\varepsilon), \tilde{K}_{al} = \lim_{n\to\infty}\lim_{\varepsilon\to 0} C_{\text{al}}^n(\varepsilon). \tag{3.6}$$

These two quantities exist, because by the definitions (3.2) and (3.3)

$$(n+m)\lim_{\varepsilon\to 0} C_{\text{er}}^{n+m}(\varepsilon) \geq n\lim_{\varepsilon\to 0} C_{\text{er}}^n(\varepsilon) + m\lim_{\varepsilon\to 0} C_{\text{er}}^m(\varepsilon)$$

and

$$(n+m)\lim_{\varepsilon\to 0} C_{al}^{n+m}(\varepsilon) \geq n\lim_{\varepsilon\to 0} C_{\text{al}}^n(\varepsilon) + m\lim_{\varepsilon\to 0} C_{al}^m(q).$$

However, the existence of the limits in (3.1) or (3.5) is not at all obvious. We introduce therefore the lower limits

$$\underline{K}_{\rm er} = \underline{\lim}_{\varepsilon\to 0} C_{\rm er}(\varepsilon) \text{ and } \underline{K}_{al} = \underline{\lim}_{\varepsilon\to 0} C_{al}(\varepsilon)$$

and the corresponding upper limits $\overline{K}_{\rm er}$ and $\overline{K}_{al}$.

Finally, let $C_{\rm de}(\varepsilon)$ be the (zero-error) detection capacity of W_ε. Since it is independent of ε for $\varepsilon \in (0,1)$ we simply write $C_{\rm de}$. It is the key quantity for our limits.

Theorem 2:

i) $K_{al} = \tilde{K}_{al} = C_{\rm de}$.
ii) $\underline{K}_{\rm er} \geq \tilde{K}_{\rm er} = C_{\rm de}$.

Remarks:

3) We conjecture that $K_{\rm er}$ exists and equals $C_{\rm de}$. Sufficient for this would be the continuity of $C_{\rm er}$ in ε or that $C_{\rm er}^n(\varepsilon)$ is nonincreasing in ε, because then $\overline{K}_{\rm er} \leq \tilde{K}_{\rm er}$.
4) Inspection of the proofs below shows that all lower bounds by $C_{\rm de}$ remain valid, if we replace W_ε by any matrix V_ε with $V_\varepsilon \ll W_\varepsilon$ and $V_\varepsilon(x|x) \geq 1-\varepsilon$ for $x \in \mathcal{X}$.

Proof: We conclude with iii) in Theorem 1 that

$$C_{\rm er}^n(\varepsilon) \geq C_{al}^n(\varepsilon) \qquad C_{\rm er}(\varepsilon) \geq C_{al}(\varepsilon) \tag{3.7}$$

and therefore also that

$$\underline{K}_{\rm er} \geq \underline{K}_{al},\ \overline{K}_{\rm er} \geq \overline{K}_{al}, \text{ and } \tilde{K}_{\rm er} \geq \tilde{K}_{al}. \tag{3.8}$$

We have by (3.2) and (3.3) the monotonicity properties

$$C_{\rm er}^{2^i}(\varepsilon) \text{ is nondecreasing in } i \tag{3.9}$$

and

$$C_{\rm al}^{2^i}(\varepsilon) \text{ is nondecreasing in } i. \tag{3.10}$$

These properties imply that

$$\underline{K}_{\rm er} \geq \tilde{K}_{\rm er} \text{ and } \underline{K}_{al} \geq \tilde{K}_{al}. \tag{3.11}$$

In the light of (3.11) the proof of i) in Theorem 2 is complete after we have shown that

1) $\tilde{K}_{al} \geq C_{\rm de}$ and 2) $\overline{K}_{al} \leq C_{\rm de}$.

After we have established i), by (3.8) and (3.11) it suffices for the proof of ii) to show that

3) $\tilde{K}_{\rm er} \leq C_{\rm de}$.

Proof of 1): Recall that

$$\tilde{K}_{al} = \lim_{n\to\infty} \lim_{\varepsilon\to 0} C_{al}^n(\varepsilon).$$

For any null sequence $(\delta_i)_{i=1}^\infty$, $\delta_i > 0$, there is a sequence $(n_i)_{i=1}^\infty$ of positive integers with $C_{\rm de}^{n_i} \geq C_{\rm de} - \delta_i$. There is a corresponding detection code $\mathcal{C}^{(n_i)}$ of rate $C_{\rm de}^{n_i}$ $(i = 1, 2, \cdots)$. Its average list size $\overline{L}_\varepsilon(\mathcal{C}^{(n_i)})$ under $W_\varepsilon^{n_i}$ satisfies

$$\overline{L}_\varepsilon(\mathcal{C}^{(n_i)}) \leq (1-\varepsilon)^n \cdot 1 + \big(1 - (1-\varepsilon)^n\big)|\mathcal{C}^{(n_i)}| \tag{3.12}$$

and

$$\lim_{\varepsilon\to 0} \overline{L}_\varepsilon(\mathcal{C}^{(n_i)}) = 1.$$

Using (2.22) in the converse proof for average list size codes we obtain for every ε and i

$$C_{\rm de} - \delta_i \leq \frac{1}{n_i}\log|\mathcal{C}^{(n_i)}| \leq C_{\rm al}^{n_i}(\varepsilon) + \frac{1}{n_i}\log\big(1 + \lambda(\varepsilon, n_i)\big)$$

where

$$\lambda(\varepsilon, n_i) = (1-\varepsilon)^{n_i} + \big(1 - (1-\varepsilon)^{n_i}\big)|\mathcal{C}^{(n_i)}| - 1.$$

Since

$$\lim_{\varepsilon\to 0} \lambda(\varepsilon, n_i) = 0$$

this yields

$$C_{\rm de} - \delta_i \leq \lim_{\varepsilon\to 0} C_{\rm al}^{n_i}(\varepsilon)$$

and thus

$$C_{\rm de} = \lim_{i\to\infty}(C_{\rm de} - \delta_i) \leq \lim_{i\to\infty}\lim_{\varepsilon\to 0} C_{\rm al}^{n_i}(\varepsilon) = \tilde{K}_{al}.$$

Proof of 2): We know from the proof of Theorem 1 (see (2.16) and (2.17)) that there are codes $\mathcal{C}^{(n)}(\varepsilon)$ with average list size $1 + \alpha(n)$ and rate

$$\frac{1}{n}\log|\mathcal{C}^{(n)}(\varepsilon)| \geq \overline{K}_{al} - \delta_n(\varepsilon)$$

where

$$\lim_{n\to\infty} \alpha(n) = 0$$

and

$$\lim_{n\to\infty} \delta_n(\varepsilon) = 0.$$

The probability of the output set $\mathcal{C}^{(n)}(\varepsilon)$ is

$$Q\big(\mathcal{C}^{(n)}(\varepsilon)\big) = \sum_{c\in\mathcal{C}^{(n)}(\varepsilon)} \sum_{y^n\in\mathcal{C}^{(n)}(\varepsilon)} \frac{1}{|\mathcal{C}^{(n)}(\varepsilon)|} W^n(y^n|c) \geq (1-\varepsilon)^n.$$

Therefore, the average list size over this set is

$$\hat{L} = \frac{1}{Q\big(\mathcal{C}^{(n)}(\varepsilon)\big)} \sum_{c\in\mathcal{C}^{(n)}(\varepsilon)} \sum_{y^n\in\mathcal{C}^{(n)}(\varepsilon)} \frac{1}{|\mathcal{C}^{(n)}(\varepsilon)|} W^n(y^n|c)\ell(y^n) \leq (1-\varepsilon)^{-n}\big(1 + \alpha(n)\big) = \Delta.$$

Let $\mathcal{C}_1$ be the subset of $\mathcal{C}^{(n)}(\varepsilon)$, which has list size at most 2Δ. The cardinality of $\mathcal{C}_1$ is at least $\frac{1}{2}|\mathcal{C}^{(n)}(\varepsilon)|$. Randomly select a subcode of $\mathcal{C}_1$ of cardinality

$$\frac{1}{2}|\mathcal{C}^{(n)}(\varepsilon)|(1-\varepsilon)^n(1+\alpha(n))^{-1}\exp\{-\varepsilon n\}.$$

The list size of a codeword in this subcode is not 1 with probability $\exp\{-\varepsilon n\}$. Deleting those codewords whose list size is greater than 1 results in a code of cardinality in average at least

$$\frac{1}{2}|\mathcal{C}^{(n)}(\varepsilon)|(1-\varepsilon)^n(1+\alpha(n))^{-1}\exp\{-\varepsilon n\}\big(1 - \exp\{-\varepsilon n\}\big).$$

This is a detection code and this leads to

$$C_{\rm de} \geq \overline{K}_{al} - \delta_n(\varepsilon) + \log(1-\varepsilon) - frac1n\log(1+\alpha(n)) - \varepsilon - o(1).$$

Letting n go to infinity and then ε go to zero gives

$$C_{\rm de} \geq \overline{K}_{al}.$$

Proof of 3): For fixed n let $P_\varepsilon^{(n)}$ be the optimal distributions and let $W_\varepsilon^{(n)}$ be the optimal stochastic matrices in the definition of $C^{(n)}(\varepsilon)$. By compactness there exists a null-sequence $(\varepsilon_k)_{k=1}^\infty$ such that

$$\lim_{k\to\infty} P_{\varepsilon_k}^{(n)} = P^{(n)} \text{ and } \lim_{k\to\infty} W_{\varepsilon_k}^{(n)} = W^{(n)}. \tag{3.13}$$

By the continuity of the mutual information function I we obtain

$$\lim_{k\to\infty} I\big(P_{\varepsilon_k}^{(n)}, W_{\varepsilon_k}^{(n)}\big) = I(P^{(n)}, W^{(n)})$$

and since for fixed n, $C_{\text{er}}^{(n)}$ is continuous in ε

$$\lim_{\varepsilon\to 0} C_{\text{er}}^{(n)}(\varepsilon) = I(P^{(n)}, W^{(n)}).$$

It is also easy to see that

$$P^{(n)}W^{(n)} = P^{(n)}. \tag{3.14}$$

Also, any $\overline{W}^{(n)}$ with $P^{(n)}\overline{W}^{(n)} = P^{(n)}$ satisfies

$$I(P^{(n)}, W^{(n)}) \le I(P^{(n)}, \overline{W}^{(n)}).$$

We find now a blocklength nN detection code by randomly and independently selecting M codewords in $\mathcal{X}^{nN}$ according to the PD $(P^{(n)})^N$.

We choose

$$M = \exp\{NI(P^{(n)}, W^{(n)}) - \delta nN\}.$$

The list size for a codeword selected is in average

$$M\exp\{-NI(P^{(n)}, W^{(n)}) + o(Nn)\} + 1 = 1 + \exp\{-\delta nN\}$$

because (2.16) holds. By deleting codewords with list size at least 2 we obtain a detection code of size at least $M(1 - \exp\{-\delta nN\})$. This concludes the proof.

IV. The Identification Capacity for Zero-Error Probability of Misrejection

We recall the definitions given at the end of the Introduction.

Theorem 3: For every DMC the zero-error second-order identification capacity C_{oid} equals the first-order zero-error erasure capacity for transmission C_{er}.

Remark 6): The results about C_{er} in the previous sections are now also of interest for identification.

Proof: Let $\mathcal{C}^{(n)}$ be an optimal erasure code of length n with maximal erasure probability P_{er} of the order $1/n$. We know that

$$\lim_{n\to\infty} \frac{1}{n}\log|\mathcal{C}^{(n)}| = C_{\text{er}}.$$

Let $\{\mathcal{C}_i : 1 \le i \le N\}$ be a collection of subcodes of $\mathcal{C}^{(n)}$ with the following properties:

1) $|\mathcal{C}_i| = \frac{|\mathcal{C}^{(n)}|}{n^2}$ for $i = 1, \cdots, N$.
2) $|\mathcal{C}_i \cap \mathcal{C}_j| \le \frac{|\mathcal{C}^{(n)}|}{2n^3}$ for $i \ne j$.

By the same reasoning as in [11] one can show that N can be made as big as $\exp\{\exp\{\log|\mathcal{C}^{(n)}| - o(n)\}\}$.

Let P_i be the uniform distribution over $\mathcal{C}_i$ and set

$$D_i = \{y^n : \exists x^n \in \mathcal{C}_i \text{ such that } W(y^n|x^n) > 0\}.$$

Apparently

$$\sum_{x^n} P_i(x^n) \sum_{y^n\in D_i} W^n(y^n|x^n) = 1.$$

By the properties of $\mathcal{C}^{(n)}$ and the $\mathcal{C}_i$'s, one gets for the second kind of error probability

$$P_{\text{ma}} \le \max_{i\ne j} \sum_{x^n} P_j(x^n) \sum_{y^n\in D_i} W^n(y^n|x^n) \le P_{\text{er}} + \frac{1}{2n}.$$

To prove the converse part, we consider again

$$\mathcal{X}(y^n) = \{x^n : W^n(y^n|x^n) > 0\}.$$

We have for any $P \in \mathcal{P}(\mathcal{X}^n)$ and any V with $PV = PW^n$ and $V \ll W^n$

$$\begin{aligned}
&\sum_{x^n,y^n} P(x^n)V(y^n|x^n)\log\frac{V(y^n|x^n)}{PV(y^n)}\\
&\quad= \sum_{y^n,x^n\in\mathcal{X}(y^n)} P(x^n)V(y^n|x^n)\log\frac{V(y^n|x^n)P(x^n)}{PV(y^n)P(x^n)}\\
&\quad\ge \sum_{y^n} PV(y^n)\log\frac{PV(y^n)}{PV(y^n)P(\mathcal{X}(y^n))}\\
&\quad\text{(by the log–sum inequality)}\\
&\quad= \sum_{y^n} PW^n(y^n)\log\frac{1}{P(\mathcal{X}(y^n))}.
\end{aligned}$$

Therefore

$$\begin{aligned}
\underline{I}(P, W^n) &= \min_{V:PV=PW^n, V\ll W^n} \sum_{x^n,y^n} P(x^n)V(y^n|x^n)\\
&\qquad\cdot \log\frac{V(y^n|x^n)}{PV(y^n)}\\
&\ge \sum_{y^n} PW^n(y^n)\log\frac{1}{P(\mathcal{X}(y^n))}.
\end{aligned} \tag{4.1}$$

By Chebychev's inequality and (4.1)

$$\begin{aligned}
PW^n(\{y^n : P(\mathcal{X}(y^n)) < \exp\{-\underline{I}(P, W^n) - n\varepsilon\}\})&\\
\le \frac{\underline{I}(P, W^n)}{\underline{I}(P, W^n) + n\varepsilon}.&
\end{aligned} \tag{4.2}$$

Define

$$\mathcal{Y}^* = \{y^n : P(\mathcal{X}(y^n)) \ge \exp\{-\underline{I}(P, W^n) - n\varepsilon\}\}$$

and notice that by (4.2)

$$PW^n(\mathcal{Y}^*) \ge \frac{n\varepsilon}{\underline{I}(P, W^n) + n\varepsilon} \triangleq \delta, \text{ say.}$$

Now randomly select a code C^* of cardinality

$$\exp\{\underline{I}(P, W^n) + 2n\varepsilon\}$$

according to the PD P such that different codewords are selected independently. Associate with the random set $\mathcal{C}^*$

$$\mathcal{Y}(C^*) = \{y^n : \exists x^n \in C^* \text{ with } W^n(y^n|x^n) > 0\}.$$

Notice that for any $y^n \in \mathcal{Y}^*$

$$\Pr(y^n \in \mathcal{Y}(C^*)) \geq 1-(1-\exp\{-\underline{I}(P,W^n)-n\varepsilon\})^{\exp\{\underline{I}(P,W^n)+2n\varepsilon\}} \geq 1-\exp\{-\exp\{n\varepsilon\}\}.$$

Therefore, there exists a code $\mathcal{C}^*$ such that

$$\mathcal{Y}^* \subset \mathcal{Y}(\mathcal{C}^*).$$

We can always assume that

$$D_i = \mathcal{Y}(\text{supp}(P_i))$$

where $\text{supp}(P) = \{x^n : P(x^n) > 0\}$. Since for any i

$$\underline{I}(P_i, W^n) \leq nC_{\text{er}} + o(n)$$

we get for

$$\mathcal{Y}_i^* = \{y^n : P_i(\mathcal{X}(y^n)) \geq \exp\{-nC_{\text{er}} - n\varepsilon\}\}$$

$$P_i W^n(\mathcal{Y}_i^*) \geq \frac{\varepsilon}{C_{\text{er}} + \varepsilon} = \delta, \quad \text{say.}$$

Since for every i we can find a subcode $\mathcal{C}_i$ of $\text{supp}(P_i)$ with

$$\mathcal{Y}_i^* \subset \mathcal{Y}(\mathcal{C}_i)$$

we conclude that

$$P_i W^n(\mathcal{Y}(\mathcal{C}_i)) \geq \delta.$$

We see that for $i \neq j$ also $\mathcal{C}_i \neq \mathcal{C}_j$, because otherwise

$$P_i W^n(D_j) \geq P_i W^n(\mathcal{Y}(\mathcal{C}_j)) = P_i W^n(\mathcal{Y}(\mathcal{C}_i)) \geq \delta$$

and this contradicts the fact that

$$P_i W^n(D_j) \leq \frac{1}{n}.$$

The total number of codes of cardinality $\exp\{n(C_{\text{er}}+2\varepsilon)\}$ is at most $|\mathcal{X}^n|^{\exp\{n(C_{\text{er}}+2\varepsilon)\}}$. Since

$$\lim_{n\to\infty} \frac{1}{n} \log\log |\mathcal{X}^n|^{\exp\{n(C_{\text{er}}+2\varepsilon)\}} = C_{\text{er}} + 2\varepsilon$$

letting ε go to zero proves the converse.

V. Concluding Remarks

We mention here some connections to other work and also further directions of research.

1) It is clear from (4.1) that our characterization of $C_{\text{er}}(W)$, in particular its "direct part," is better than Forney's [2] bound

$$C_{\text{er}}(W) \geq \max_{P\in\mathcal{P}(\mathcal{X}^n)} -\sum_{y^n} PW^n(y^n) \log \sum_{x^n : W^n(y^n|x^n)>0} P(x^n).$$

It should be noted, however, that Forney's bound is tight in the limit ($n \to \infty$). A rigorous and simple proof of the converse was shown to us by I. Telatar.

2) The quantity $\underline{I}(P, W)$ defined in Section II is not convex in P, whereas $I(P, W)$ is. We therefore alternatively suggest to take the upper envelope

$$I_L(P,W) = \max\left\{\sum_{j\in J} \alpha_j \underline{I}(P_j, W) : P = \sum_{j\in J} \alpha_j P_j, 0 \leq \alpha_j, \sum_{j\in J} \alpha_j = 1\right\} \tag{5.1}$$

and call it "lower information." It is a quantity of some operational significance, which naturally arises in time-sharing arguments. In terms of random variables X, Y we write for it also $I_L(X \wedge Y)$. It can be shown to be symmetric in X and Y.

$$H_L(X|Y) = H(X) - I_L(X \wedge Y)$$

is then the "upper-conditional entropy." For an extension of our work to multiuser models one can use a calculus of these quantities.

3) It seems that the study of low-noise channels should be rewarding also, if the usual probabilistic error criteria are used. In some instances non-single-letter characterization may become computable in the limit $\varepsilon \to 0$.

4) In [3] it was shown that C_{er} equals the ordinary channel capacity C, if the following condition holds: For $\ell \geq 2$ there do not exist $x_1, x_2, \cdots, x_\ell \in \mathcal{X}$, $x_{\ell+1} = x_1$, and $y_1, \cdots, y_\ell \in \mathcal{Y}$ with

$$W(y_i|x_i) > 0, W(y_i|x_{i+1}) > 0 \quad \text{for } i = 1, \cdots, \ell. \tag{5.2}$$

This condition is not necessary for $C_{\text{er}} = C$ to hold. We have a complete characterization of this equality for the case $\min(|\mathcal{X}|, |\mathcal{Y}|) = 2$.

5) Since in [4] the zero-error capacity of a DMC has been shown to equal the maximal error capacity of an associated arbitrarily varying channel (AVC) with 0–1-matrices only, there have developed more connections between zero-error problems and AVC theory. One line of investigations, starting with the discoveries of the "worst channel" for binary-output AVC's in [5] and the "maximum probability decoder" in [6], studies the performance of seemingly simple decoding rules such as minimum-distance decoding in [7]–[9]. There the "distance" is actually a distortion function $d_n : \mathcal{X}^n \times \mathcal{Y}^n \to \mathbb{R}_+$ with

$$d_n(x^n, y^n) = \sum_{t=1}^{n} d(x_t, y_t) \text{ and } d : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+. \tag{5.3}$$

In [8] the maximal rate of codes for the DMC W with an error probability tending to zero under d-distance decoding is called d-capacity and denoted as C_d. The most known example of such a decoding rule is the maximum-likelihood decoder $d(x, y) = -\log W(y|x)$. Another one is the "mismatch decoder" $d(x, y) = -\log V(y|x)$, where V corresponds to another DMC. It is on another line of investigations. Furthermore, for suitable $d : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ one obtains problems equivalent to the classical zero-error problem and the zero-error problem for erasures.

The lower bound for C_d stated in [8] is not tight for Shannons zero-error capacity, but it is also not tight in case of erasures (Example in the Introduction).

Other partial results mentioned in [9] concern an extension of the question of [3] to general d: "Under which conditions is C_d equal to C?" and positivity conditions for C_d in the familiar line of AVC theory.

6) Zero-error detection capacity has been called Sperner capacity in [10]. The name is suggestive to combinatorialists familiar with an extremal problem solved by Sperner and with related work. It seems to us that in information theory names like t-error detecting and zero-error detection codes are almost self-explainatory and therefore preferable. Actually, C_{de} is also a special case of d-capacity.
7) One can extend our work to zero-error detection codes with bounds on the erasure probabilities.

Acknowledgment

The authors wish to thank A. Flammenkamp for his dedicated computer analysis of the example mentioned in the Introduction.

References

[1] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inform. Theory*, vol. IT-2, pp. 8–19, 1956.
[2] G. D. Forney, "Exponential error bounds for erasure, list and decision feedback schemes," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 206–220, 1968.
[3] M. S. Pinsker and A. J. Ševerdjaev, "Zero error capacity with erasure" (in Russian), *Probl. Pered. Inform.*, vol. 6, no. 1, pp. 20–24, 1970.
[4] R. Ahlswede, "A note on the existence of the weak capacity for channels with arbitrarily varying channel probability functions and its relation to Shannon's zero error capacity," *Annals Math. Statist.*, vol. 41, pp. 1027–1033, 1970.
[5] R. Ahlswede and J. Wolfowitz, "The capacity of a channel with arbitrarily varying cpt's and binary output alphabet," *Z. Wahrscheinlichkeitsth. u. verw. Geb.*, vol. 15, pp. 186–194, 1970.
[6] R. Ahlswede, "A method of coding and an application to arbitrarily varying channels," *J. Combin., Inform. Syst. Sci.*, vol. 5, pp. 10–35, 1980.
[7] I. Csiszár and P. Narayan, "Capacity and decoding rules for classes of arbitrarily varying channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 752–769, 1989.
[8] I. Csiszár, "Channel capacity for minimum distance decoding," in *Proc. Swedish–Russian Workshop on Information Theory* (Visby, Sweden, 1989), pp. 8–10.
[9] I. Csiszár and P. Narayan, "Channel capacity with a given decoding metric," in *Proc. Swedish–Russian Workshop on Information Theory* (Mölle, Sweden, 1993), p. 461.
[10] L. Gargano, J. Körner, and U. Vaccaro, "Sperner capacities," to appear in *Graphs Combinat.*
[11] R. Ahlswede and G. Dueck, "Identification via channels," *IEEE Trans. Inform. Theory*, vol. 35, no. 1, pp. 15–29, 1989.
[12] ——, "Identification in the presence of feedback—A discovery of new capacity formulas," *IEEE Trans. Inform. Theory*, vol. 35, no. 1, pp. 30–39, 1989.
[13] R. Ahlswede and I. Csiszár, "Hypothesis testing under communication constraints," *IEEE Trans. Inform. Theory*, vol. IT-32, no. 4, pp. 543–549, 1986.
[14] R. Ahlswede, "Channel capacities for list codes," *J. Appl. Prob.*, vol. 10, pp. 824–826, 1973.
[15] I. E. Telatar and R. G. Gallager, "Zero error decision feedback capacity of discrete memoryless channels," in *Proc. 1990 Bilkent Int. Conf. on New Trends in Communication, Control, and Signal Processing* (Bilkent University, Ankara, Turkey), July 1990, pp. 228–233.
[16] I. E. Telatar, "Multi-access communications with decision feedback decoding," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, May 1992.
[17] I. E. Telatar and R. G. Gallager, "New exponential upper bounds to error and erasure probabilities," in *Proc. 1994 IEEE Int. Symp. on Information Theory* (Trondheim, Norway, June 27–July 1, 1994).