Information and Control: Matching Channels

Rudolf Ahlswede and Ning Cai

(Dedicated to Mark Pinsker on his 70th birthday)

***Abstract*— The transmission problem for noisy channels is usually studied under the condition that the decoding error probability λ is small and is sometimes studied under the condition that $\lambda = 0$. Here we just require that $\lambda < 1$ and obtain a problem which is equivalent to a coding problem with small λ for the "Deterministic Matching channel." In this new model, a cooperative person knows the codeword to be sent and can choose (match) the state sequence of the channel. There are interesting connections to combinatorial matching theory and extensions to the theory of identification as well as to multi-way channels. In particular, there is a surprising connection to Pinsker's coding theorem for the deterministic broadcast channel.**

***Index Terms*— Combinatorial matching, detection, feedback, identification, new channel models, zero-error problems;**

I. New Concepts and Results

A. The Matching Channel

Let $\mathcal{X}$ serve as input alphabet and let $\mathcal{Y}$ serve as output alphabet. By adding dummy letters we can always assume that $\mathcal{X} \subset \mathcal{Y}$. The transmission of letters is ruled by a class $\mathcal{W}$ of stochastic matrices with $|\mathcal{X}|$ rows and $|\mathcal{Y}|$ columns as follows. In addition to a sender and a receiver, there is a third person (or device) called a controller who decides which matrix $W \in \mathcal{W} = \{w(\cdot|\cdot|s): s \in \mathcal{S}\}$ shall govern the transmission of a letter by the sender. The controller knows which codeword the sender wants to transmit. The receiver has no knowledge about the actions of the controller. As code concept appropriate for this situation we introduce a matching code (MC). We call $\{(u_i, \mathcal{D}_i): 1 \leq i \leq M\}$ an (n, M, λ)-MC-code for $\mathcal{W}$, if

$$u_i \in \mathcal{X}^n, \mathcal{D}_i \subset \mathcal{Y}^n, \qquad \text{for } i = 1, 2, \cdots, M \tag{1.1}$$

$$u_i \neq u_j \text{ and } \mathcal{D}_i \cap \mathcal{D}_j = \varnothing, \qquad \text{for } i \neq j \tag{1.2}$$

and if for every i there is a sequence

$$s^n(i) = (s_1(i), \cdots, s_n(i)) \in \mathcal{S}^n = \Pi_1^n \mathcal{S}$$

with

$$W^n(\mathcal{D}_i|u_i|s^n(i)) \geq 1 - \lambda \tag{1.3}$$

if

$$W^n l(y^n|x^n|s^n(i)) = \prod_{t=1}^{n} W(y_t|x_t|s_t(i))$$

for

$$x^n = (x_1, \cdots, x_n) \in \mathcal{X}^n$$

and

$$y^n = (y_1, \cdots, y_n) \in \mathcal{Y}^n.$$

Let $C(\mathcal{W})$ be the capacity of the matching channel $\mathcal{W}$.

As usual we denote by X, S, Y random variables (RV's) with values in $\mathcal{X}, \mathcal{S}$, and $\mathcal{Y}$, respectively. Let P_{XS} be the joint distribution of (X, S) and

$$P_{XSY}(x, s, y) = P_{XS}(x, s)W(y|x, s)$$

for $x \in \mathcal{X}$, $s \in \mathcal{S}$, and $y \in \mathcal{Y}$.

Theorem 1: The capacity of the matching channel is given by

$$C(\mathcal{W}) = \max_{P_{XS}} \min(H(X), I(XS \wedge Y)).$$

Notice that the quantity $C = \max_{P_{XS}} I(XS \wedge Y)$ is the capacity of the corresponding discrete memoryless channel $(\mathcal{X} \times \mathcal{S}, \mathcal{Y}, W')$ (or for the model, where the controller knows not only the codeword but even the message to be sent) and that, therefore, $C(\mathcal{W}) \leq C$.

The minimization with $H(X)$ reflects the fact that only pairs (x^n, s^n), which are all different in the first component are permitted in the encoding. Obviously, choosing $M \sim \exp\{n \min(H(X), I(XS \wedge Y)\}$ of such pairs, say, $\{(x_i^n, s_i^n): 1 \leq i \leq M\}$, independently with distribution P_{XS}^n results with high probability in a code, for which most x_i^n's are different—and those we keep! This gives the direct part of Theorem 1 and the converse part is also obvious.

Remarks:

1) In case the controller is restricted to choose only state sequences $(s, s, \cdots, s), s \in S$, we are led to the "optimistic" channel of [25].

2) Massey conveyed the following interpretation to us:

   One can speak of "Coding with a Barrister" for the following reason. In the British system of law there are two kinds of lawyers, solicitors and barristers. The solicitor is the lawyer who prepares the case, but only the barrister is permitted to argue the case before the court. In the American system of law, the same lawyer usually performs both functions. Previously in coding theory, the "encoder" performed like an American lawyer, both mapping the message into a codeword then transmitting this codeword over the channel. The new feature of the present model is that the "encoder"

Manuscript received August 28, 1995; revised July 8, 1997.

The authors are with Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany.

Publisher Item Identifier S 0018-9448(98)00835-9.

acts like a "solicitor," only mapping the message into a codeword. It is then the "barrister" who transmits the codeword over the channel. Of course, if the barrister knew the message, there would be nothing new.

B. The Deterministic Matching Channel

It is instructive to consider the case where $\mathcal{W} = \mathcal{W}_0$ contains only 0–1 matrices.

Then Theorem 1 has the following specialization.

Theorem 2:

$$C(\mathcal{W}_0) = \max_{P_{XS}} \min(H(X), H(Y)).$$

Clearly, by the definition of an (n, M, λ)-MC-code we can assume now that $|\mathcal{D}_i| = 1$ and thus $\mathcal{D}_i = \{v_i\}$. Also if $\lambda < 1$, then it can actually be chosen to equal 0. So the distinct u_i's are matched with distinct v_i's. In determining the capacity $C(\mathcal{W}_0)$ we are thus led to a *new probabilistic coding theory*, whose mathematical structure is interesting and natural: *a novel combinatorial matching theory for products of bipartite graphs*.

It is convenient to work with an equivalent formulation of coding problems for $\mathcal{W}_0$ in terms of an *associated* DMC W

$$W(\cdot|x) = \sum_{s \in \mathcal{S}} Q(s)W(\cdot|x|s) \tag{1.4}$$

where Q is any probability distribution on $\mathcal{S}$ with $Q(s) > 0$ for $s \in \mathcal{S}$.

Since for the DMC

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

one notices that for any $(n, M, 0)$-MC-code $\{(u_i, \mathcal{D}_i): 1 \leq i \leq M\}$ for $\mathcal{W}_0$ the condition (1.3) can equivalently be described in the "dummy" formulation by

$$W^n(\mathcal{D}_i|u_i) > 0, \qquad \text{for } i = 1, 2, \cdots, M. \tag{1.5}$$

It is mathematically and esthetically quite appealing that by weakening the requirements on the error performance we are led from zero-error codes [15], where

$$W^n(\mathcal{D}_i|u_i) = 1, \qquad \text{for } i = 1, \cdots, M \tag{1.6}$$

to the most familiar λ-error codes $(\lambda > 0)$ with

$$W^n(\mathcal{D}_i|u_i) \geq 1 - \lambda, \qquad \text{for } i = 1, \cdots, M, \tag{1.7}$$

to MC-codes for $\mathcal{W}_0$ in the "dummy" formulation.

The corresponding capacities $C_0(W), C(W)$, and $C(\mathcal{W}_0)$ satisfy, of course,

$$C_0(W) \leq C(W) \leq C(\mathcal{W}_0). \tag{1.8}$$

There is a code concept between the first two, namely, that of an erasure code, for which in addition to (1.7) we also have

$$W^n(\mathcal{D}_j|u_i) = 0, \qquad \text{for } i \neq j. \tag{1.9}$$

The zero-error erasure capacity $C_{\text{er}}(W)$ has been studied in several papers and recently quite intensively by several authors (c.f. [9]). Until now, no "single-letter" formula exists.

Quite analogously we can also require (1.9) in conjunction with (1.5). This gives exactly the MDC-code defined in Section I-D in terms of $\mathcal{W}_0$.

It is interesting that coding for $\mathcal{W}_0$ is equivalent with matching in products of bipartite graphs (see Section III). This connection leads to a combinatorial version of Theorem 2, which is stated as Theorem 4 in Section IV. It has a nice direct proof with König's Minimax Theorem. Beyond this result on the asymptotic behavior of matching numbers under products, we give an exact result for two factors in Theorem 5 (Section V). This enables us to get also exact results for powers of certain bipartite graphs (Theorem 6 in Section VI). Finally, in Section VII we underline with two examples the significance of Theorems 5 and 6.

C. Multi-Way Matching Channels

The concept of a matching DMC has straightforward extensions to several sender and receiver models. A highlight in Section VIII is the solution of the *general* broadcast problem in this matching theory. As a special case of Theorem 9 we obtain Pinsker's [19] capacity region for the deterministic broadcast channel.

The corresponding Theorems 7 and 8 for compound and multiple-access channels are stated without their (routine) proofs.

D. The Controller Falls Asleep

It seems to us that the interplay between information transfer and controlling certain channels deserves more and deeper investigations. Channels with control aspects are the permuting relay channels of [22] and [24], as well as the outputwise varying channels of [18], which arose in the study of rewritable storage media. We also draw attention to [23] for still another philosophy: controlling by creating order.

Now we are more specific. In the model described in Section I-A, the controller is not only assumed to be cooperative, but he also acts perfectly. Next the communicators safeguard against mistakes of the controller and even against malicious operations (jamming) by using matching zero-error detection codes (MDC) $\{(u_i, \mathcal{D}_i): 1 \leq i \leq M\}$ which in addition to (1.3) (automatically with $\lambda = 0$) satisfy

$$W^n(\mathcal{D}_j|u_i|s^n) = 0, \qquad \text{for } i \neq j \text{ and all } s^n \in \mathcal{S}^n. \tag{1.10}$$

To determine its capacity, $C_{mde}(\mathcal{W})$, is a formidable task. For the deterministic matching channel $\mathcal{W}_0$ results and relations to other zero-error capacities are contained in Section IX. Appendix I contains instructive examples and Theorem I.1 as the main contribution on the relation. For one genuine channel, the $\binom{\alpha}{\beta}$-uniform hypergraph channel $W_{\alpha,\beta}$, we succeeded in determing the capacity in Section X.

As in the classical AWAC system we assume here that there is a noiseless feedback channel or just an active feedback channel on which the receiver can ask for retransmission. The frequency of such retransmissions depends on the error frequency (the sleeping habits) of the controller.

We emphasize again that in Section I-A it makes a big difference whether the controller knows the message (and thus the same word can be used in conjunction with different state sequences to represent the different messages) or only the codeword (and thus no word can represent different messages).

In the present situation it makes no sense to assume that the controller knows in addition to the codeword the messages, because the codewords have to be different, anyhow, to cope with a sleeping controller.

E. Matching Zero-Error Detection Codes with Feedback for $\mathcal{W}_0$ (MDCF)

We mention first that feedback or also randomization in the encoding increases the capacity of the matching channel $\mathcal{W}$ to the effect that the term $H(X)$ has to be dropped in the formula of Theorem 1. This is stated in (11.1) and proved in Section XI.

We turn now to the MDCF.

The feedback is now really used in the design of the code. There is given a finite set of messages $\mathcal{M} = \{1, 2, \cdots, M\}$. One of these messages is to be sent over the channel. Message $i \in \mathcal{M}$ is encoded by a (vector-valued) function

$$f_i^n = [f_{i1}, f_{i2}, \cdots, f_{in}]$$

where, for $t \in \{2, \cdots, n\}$, f_{it} is defined on $\mathcal{Y}^{t-1}$ and takes values in $\mathcal{X}$. f_{i1} is an element of $\mathcal{X}$. It is understood that after the received elements $Y_1, \cdots, Y_{t-1}$ have been made known to the sender by the feedback channel, the sender transmits $f_{it}(Y_1, \cdots, Y_{t-1})$. At $t = 1$ the sender transmits f_{i1}. Again, we assume that the controller knows only the encoding functions, but not the messages, and therefore $f_i^n \neq f_j^n$ if $i \neq j$. The distribution of the RV's Y_t $(t = 1, 2, \cdots, n)$ is determined by f_i^n and $W(\cdot|\cdot|s^n)$. We denote the probability of receiving $y^n = (y_1, \cdots, y_n) \in \mathcal{Y}^n$, if i has been encoded and the controller uses s^n, by

$$\begin{aligned} W^n(y^n|f_i^n|s^n) &= W(y_1|f_{i1}|s_1)W(y_2|f_{i2}(y_1)|s_2) \\ &\quad \cdots W(y_n|f_{in}(y_1, \cdots, y_{n-1})|s_n). \end{aligned}$$

In an (n, M, λ) matching zero-error detection feedback code $\{(f_i, \mathcal{D}_i, s_i^n)\colon 1 \le i \le M\}$ the $\mathcal{D}_i$ are disjoint subsets of $\mathcal{Y}^n$ and

$$\begin{aligned} &W^n(\mathcal{D}_i|f_i|s_i^n) \ge 1 - \lambda, && \text{for } i = 1, \cdots, N \\ &W^n(\mathcal{D}_j|f_i|s^n) = 0, && \text{for all } s^n \in \mathcal{S}^n,\ j \neq i. \quad (1.11) \end{aligned}$$

We are interested in the capacity $C_{m\,def}(\mathcal{W}_0)$. Here we can assume the $\mathcal{D}_i$'s to have one element, say the v_i's. Our optimism for finding a nice formula was originally just speculative: in case of feedback Shannon found also a nice formula for his zero-error capacity! Indeed, we have a surprising result, which is proved in Section XII.

Theorem 3:

a) $$C_{m\,de\,f}(\mathcal{W}_0) = \begin{cases} \max_{P_{XS}} I(XS \wedge Y) \\ \text{or} \\ 0 \end{cases}$$

b) $C_{m\,de\,f}(\mathcal{W}_0) = 0$ exactly if all columns have positive or zero entries only.

The astute reader may notice that Shannon's formula or the alternate formula of [15] (asked for by Shannon in [8]) has also a dichotomy relative to positivity. The formula in [15] describes the capacity of a jamming problem, namely, that of an arbitrarily varying channel with feedback. Our formula for $C_{m\,de\,f}(\mathcal{W}_0)$ also settles a feedback problem involving jamming.

Problem 1: Is there a common generalization of both jamming problems?

F. Identification

We emphasize that a systematic analysis of code concepts is still rewarding. By giving up the disjointness of the decoding sets and by requiring in addition to (1.7)

$$W^n(\mathcal{D}_j|u_i) \le \lambda, \qquad \text{for } i \neq j \qquad (1.12)$$

we get the concept of a (nonrandomized) identification code [12]. Randomization means here that instead of $u_i \in \mathcal{X}^n$ we allow $Q_i \in \mathcal{P}(\mathcal{X}^n)$, the set of PD's on $\mathcal{X}^n$.

In [9] we assumed (1.12) in conjunction with (1.6), that is, zero-error probability for misrejection and found that here the second-order identification capacity equals the (first-order) erasure capacity C_{er}.

Now we combine for instance (1.5) and (1.9), that is, identification with zero probability of misacceptance. Actually, we analyze all possible capacity concepts in Theorems 11–13 in Section XIII.

G. Further Code Concepts Leading to New Combinatorial Extremal Problems

Finally, we present and analyze in Definitions II.1–II.3 in Appendix II pairwise zero-error detection codes, component-pairwise zero-error detection codes, and pseudomatching zero-error detection codes (Theorem II.1). We comment also on other concepts. *With the only exception of Section XI, we consider from now on the deterministic $\mathcal{W}_0$ or an associated DMC.*

II. Definitions, Known Facts, and Abbreviations

We use essentially the terminology of [12].

1) *Sets, Channels, Types, Generated Sequences:* Script capitals $\mathcal{X}, \mathcal{Y}, \cdots$, denote finite sets. The cardinality of a set $\mathcal{A}$ is denoted by $|\mathcal{A}|$. $\binom{\mathcal{A}}{k}$ is the family of all k-element subsets of the set $\mathcal{A}$. The letters P, Q always stand for probability distributions on finite sets. $X, Y, \cdots$ denote RV's. The functions "log" and "exp" are understood to be to the base 2. For a stochastic $|\mathcal{X}| \times |\mathcal{Y}|$-matrix W we have already defined the transmission probabilities W^n of a DMC, and we have also introduced $\mathcal{P}(\mathcal{X}^n)$ as the set of PD's on $\mathcal{X}^n$. We abbreviate $\mathcal{P}(\mathcal{X})$ as $\mathcal{P}$. $\mathcal{V}$ denotes the set of all channels V with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. For positive integers n we set

$$\mathcal{P}_n = \{P \in \mathcal{P}\colon P(x) \in \{0, 1/n, 2/n, \cdots, 1\} \text{ for all } x \in \mathcal{X}\}.$$

For any $P \in \mathcal{P}_n$, called type or n-type, we define the set

$$\mathcal{V}_n(P)=\left\{V\in\mathcal{V}\colon V(y|x)\in\left\{0,\frac{1}{nP(x)},\frac{2}{nP(x)},\cdots 1\right\},\right.$$
$$\left. x\in\mathcal{X}, y\in\mathcal{Y}\right\}.$$

For $x^n \in \mathcal{X}^n$ we define for every $x \in \mathcal{X}$

$$P_{x^n}(x) = \frac{1}{n}\cdot(\text{number of occurrences of } x \text{ in } x^n).$$

P_{x^n} is a member of $\mathcal{P}_n$ by definition. P_{x^n} is called type of x^n. Similarly, we define the type $P_{x^n y^n}$ for pairs $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$. For $P \in \mathcal{P}$ the set $\mathcal{T}_P^n$ of all P-typical sequences in $\mathcal{X}^n$ is given by

$$\mathcal{T}_P^n = \{x^n\colon P_{x^n} = P\}.$$

For an RV Z with distribution P_Z we abbreviate $\mathcal{T}_{P_Z}^n$ as $\mathcal{T}_Z^n$, and when we emphasize that P is a distribution on $\mathcal{Z}$ and that $\mathcal{T}_Z^n \subset \mathcal{Z}^n$, then we write $\mathcal{Z}^n(P)$ instead of $\mathcal{T}_Z^n$.

For $V \in \mathcal{V}$, a sequence $y^n \in \mathcal{Y}^n$ is said to be V-generated by x^n if, for all $x \in \mathcal{X}, y \in \mathcal{Y}$

$$P_{x^n y^n}(x,y) = P_{x^n}(x)\cdot V(y|x).$$

The set of those sequences is denoted by $\mathcal{T}_V^n(x^n)$. Notice that $\mathcal{T}_P^n \neq \emptyset$ if and only if $P \in \mathcal{P}_n$ and $\mathcal{T}_V^n(x^n) \neq \emptyset$ if and only if $V \in \mathcal{V}_n(P_{x^n})$. For the pair of RV's (X, Y) with $\mathcal{P}r(Y = y|X = x) = V(y|x)$ we write also $\mathcal{T}_{Y|X}^n(x^n)$ instead of $\mathcal{T}_V^n(x^n)$. For $P \in \mathcal{P}$, $V \in \mathcal{V}$ we write PV for the PD on $\mathcal{Y}$ given by

$$PV(y) = \sum_x P(x)V(y|x), \qquad y \in \mathcal{Y}.$$

$\mathcal{T}_{PV}^n$ is the set of PV-typical sequences in $\mathcal{Y}^n$.

2) *Entropy and Information Quantities:* Let X be an RV with values in $\mathcal{X}$ and distribution $P \in \mathcal{P}$, and let Y be an RV with values in $\mathcal{Y}$ such that the joint distribution of (X,Y) on $\mathcal{X}\times\mathcal{Y}$ is given by

$$\Pr(X = x, Y = y) = P(x)\cdot V(y|x), \qquad V \in \mathcal{V}.$$

We write $H(P), H(V|P)$, and $I(P,V)$ for the entropy $H(X)$, the conditional entropy $H(Y|X)$, and the mutual information $I(X \wedge Y)$, respectively. For $P,\ \tilde{P} \in \mathcal{P}$

$$D(\tilde{P}\|P) = \sum_x \tilde{P}(x)\log\frac{\tilde{P}(x)}{P(x)}$$

denotes the I-divergence and for $V, \tilde{V} \in \mathcal{V}$ the quantity

$$D(\tilde{V}\|V|P) = \sum_x P(x)D(\tilde{V}(\cdot|x)\mathcal{V}\|V(\cdot|x))$$

for the conditional I-divergence.

3) *Elementary Properties of Typical Sequences and Generated Sequences:*

$$|\mathcal{P}_n| \le (n+1)^{|\mathcal{X}|}$$
$$|\mathcal{V}_n(P)| \le (n+1)^{|\mathcal{X}|\cdot|\mathcal{Y}|}$$
$$(n+1)^{-|\mathcal{X}|}\cdot\exp\{nH(P)\} \le |\mathcal{T}_P^n| \le \exp\{nH(P)\} \tag{2.1}$$

for $P \in \mathcal{P}_n$

$$|\mathcal{T}_V^n(x^n)| \ge (n+1)^{-|\mathcal{X}|\cdot|\mathcal{Y}|}\cdot\exp\{nH(V|P)\}$$
$$|\mathcal{T}_V^n(x^n)| \le \exp\{nH(V|P)\}$$

for $P \in \mathcal{P}_n, V \in \mathcal{V}_n(P), x^n \in \mathcal{T}_P^n$, and

$$W^n(\mathcal{T}_V^n(x^n)|x^n) \le \exp\{-nD(V\mathcal{V}\|W|P)\}$$
$$W^n(\mathcal{T}_V^n(x^n)|x^n) \ge (n+1)^{-|\mathcal{X}|\cdot|\mathcal{Y}|}\cdot\exp\{-nD(V\mathcal{V}\|W|P)\}$$

for

$$P \in \mathcal{P}_n, V \in \mathcal{W}_n(P), x^n \in \mathcal{T}_P^n, y^n \in \mathcal{T}_V^n(x^n)$$

and $W \in \mathcal{V}$.

Let

$$\mathcal{T}_{V,\delta}^n(x^n) = \{y^n \in \mathcal{T}_W^n(x^n)\colon \sum_{x,y}|W(y|x) - V(y|x)| < n\delta\}.$$

For $V \in \mathcal{V}_n(P_{x^n})$ one can always find a sequence $(\delta_n)_{n=1}^\infty$ with $\lim_{n\to\infty}\delta_n = 0$ and $\lim_{n\to\infty}\sqrt{n}\delta_n = \infty$ such that

$$V^n(\mathcal{T}_{V,\delta_n}^n(x^n)|x^n) \to 1, \qquad \text{as } n \to \infty.$$

Moreover, for any pair of RV's (X, Y) with $\mathcal{T}_{XY}^n \neq \emptyset$, we always have

$$\mathcal{T}_{XY}^n = \dot{\bigcup_{x^n \subset \mathcal{T}_X^n}} \{x^n\}\times\mathcal{T}_{Y|X}^n(x^n)$$

and, therefore,

$$|\mathcal{T}_{XY}^n| = |\mathcal{T}_X^n||\mathcal{T}_{Y|X}^n(x^n)|, \qquad \text{for all } x^n \in \mathcal{T}_{\mathcal{X}}{}^n. \tag{2.2}$$

This is used in Section VIII.

Table I provides a list of abbreviations and indicates where they are used.

III. The Deterministic Matching Channel and Matching in Products of Bipartite Graphs

We have shown in Section I that one can associate with the deterministic matching channel $\mathcal{W}_0$ a DMC W given by (1.4) so that any $(n, M, 0)$-MC-code $\{(u_i, D_i)\colon 1 \le i \le 1\}$ for $\mathcal{W}_0$ satisfies (1.5) for W. This condition means that the (correct) *decoding probability* is *positive.*

In codes with this property there are $v_i \in \mathcal{D}_i$ for $i = 1, 2, \cdots, M$ with

$$W^n(v_i|u_i) > 0. \tag{3.1}$$

So it suffices to study $\mathcal{D}_i'$ with one element or sets of codewords $\mathcal{U}$ with an injective map $f\colon \mathcal{U} \to \mathcal{Y}^n$ such that for $u \in \mathcal{U}$

$$W^n(f(u)|u) > 0. \tag{3.2}$$

Such an f is called a matching and $(\mathcal{U}, f)$ is a matching code. Their study obviously concerns only the support of W^n (i.e., the set of positive entries of W^n). This set can be viewed as the edge set $\mathcal{E}_n = \mathcal{E}_n(W)$ in the bipartite graph $\mathcal{G}(W^n) = (\mathcal{X}^n, \mathcal{Y}^n, \mathcal{E}_n(W))$, where $(x^n, y^n) \in \mathcal{E}_n$ iff $W(y^n|x^n) > 0$.

TABLE I

Abbreviation	Meaning	Section
$\mathcal{W}$	a class of channels called matching channel	I, II, XI
$\mathcal{W}_0$	deterministic matching channel	I, II, III, IX, XII, Appendix I
W	DMC associated with $\mathcal{W}_0$	I
$\mathcal{X}_W(\cdot), \mathcal{Z}_{V,c}(\cdot)$	column supports of matrices	IX, Appendix I, Appendix II
$\mathcal{Y}_W(\cdot), \mathcal{Z}_{V,r}(\cdot)$	row supports of matrices	IX, XIII, Appendix I, Appendix II
MC	matching codes	I, III
MDC	matching zero-error detection codes (for deterministic channels)	I, IX, X, Appendix I
MDCF	matching zero-error detection codes with feedback	I, XII
$M_{de}^n(W)$	largest size of zero-error detection codes for W^n	IX, Appendix I
$M_{m\,de}^n(W)$	largest size of MDC for W^n	IX, X, Appendix I
$C(\mathcal{W})$	capacity of matching channel $\mathcal{W}$	I
$C(\mathcal{W}_0)$	capacity of deterministic matching channels $\mathcal{W}_0$	I–VIII
$C_{m\,de}(\mathcal{W}_0)$ or $C_{m\,de}$	capacity of matching zero-error detection codes for $\mathcal{W}_0$	I, IX, X, Appendix I
$C_{m\,de\,f}(\mathcal{W}_0)$	capacity of matching zero-error detection codes with feedback for $\mathcal{W}_0$	I, XII
$C_f(\mathcal{W})$	capacity of matching codes with feedback for $\mathcal{W}$	I, XI, XII
$C(\geq \cdot, < \cdot)$	various second-order "identification capacities" for a DMC	XIII
$M_{-\,-}^n(W)$ $M_{-\,+}^n(W)$ $M_{+\,-}^n(W)$ $M_{+\,+}^n(W)$	the largest sizes of four kinds of pseudomatching 0-error detection codes for channel W^n	Appendix II
$M_0^n(W)$	the largest size of zero-error codes for W^n	Appendix II
$\mathcal{G}_1 \otimes \mathcal{G}_2$	product of graphs $\mathcal{G}_1$ and $\mathcal{G}_2$	III–VII
$\mathcal{G}^{\otimes n}$	nth power of graph $\mathcal{G}$	III–VII, Appendix II
$d_{\mathcal{G}}(v)$	degree of vertex v in graph $\mathcal{G}$	III–VIII
$\Gamma_{\mathcal{G}}(v)$	vertices connected with v	III-V
$\nu(\mathcal{G})$	matching number of graph $\mathcal{G}$	III–VI
$\tau(\mathcal{G})$	vertex covering number of $\mathcal{G}$	II, III, IV, VI
$\gamma(\mathcal{G})$	$\lim_{n\to\infty} (1/n) \log \nu(\mathcal{G}^{\otimes n})$	IV, V, VII
$\mathcal{K}(\mathcal{G})$	König–Hall pair of distributions	IV, V
$\mathcal{T}_P^n, \mathcal{T}_Z^n, \mathcal{Z}^n(P)$	the set of P-typical sequences	II, IV, VIII, XIII
$\mathcal{T}_{V,\delta}^n(x^n)$	(V, δ)-generated sequences of x^n	II, XIII

Clearly, for the f above $\{(u, f(u)): u \in \mathcal{U}\}$ is exactly a set of nonintersecting edges in $\mathcal{G}(W^n)$, that is, a matching in the terminology of graph theory. Conversely, such a matching is a matching in the corresponding matching code $(\mathcal{U}, f)$, where $\mathcal{U}$ is the set of vertices in $\mathcal{X}^n$, which are matched to vertices in $\mathcal{Y}^n$. So we have reduced the study of the matching channel $\mathcal{W}_0$ via W to the study of maximal matchings in the bipartite graph $\mathcal{G}(W^n)$.

At first we notice that this graph is an nth power $\mathcal{G}^{\otimes n} = \mathcal{G} \otimes \mathcal{G} \otimes \cdots \otimes \mathcal{G}$ of the graph $\mathcal{G}(W) = (\mathcal{X}, \mathcal{Y}, \mathcal{E})$, if the product $\mathcal{G}_1 \otimes \mathcal{G}_2$ of two bipartite graphs $\mathcal{G}_i = (\mathcal{X}_i, \mathcal{Y}_i, \mathcal{E}_i); i = 1, 2;$ is defined as $(\Pi_{i=1}^2\ \mathcal{X}_i, \Pi_{i=1}^2\ \mathcal{Y}_i, \mathcal{E})$ with

$$\mathcal{E} = \{(x^2, y^2): (x_i, y_i) \in \mathcal{E}_i \text{ for } i = 1, 2\}. \tag{3.3}$$

So we can write

$$\mathcal{G}(W^n) = \mathcal{G}^{\otimes n}(W). \tag{3.4}$$

Shannon looked at these graphs in his study of the zero-error capacity problem (i.e., the problem of determining the vertex-independence number).

For any graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ the size of a largest matching is called the matching number of $\mathcal{G}$ and is denoted by $\nu(\mathcal{G})$. A matching of the bipartite graphs $(\mathcal{V}_1, \mathcal{V}_2, \mathcal{E})$ is called a matching of $\mathcal{V}_1$ into $\mathcal{V}_2$ if every $v \in \mathcal{V}_1$ is an endpoint of an edge in the matching. A matching is perfect, if it is both; a matching of $\mathcal{V}_1$ into $\mathcal{V}_2$ and of $\mathcal{V}_2$ into $\mathcal{V}_1$.

A vertex cover of $\mathcal{G}$ is a subset $S \subset \mathcal{V}$ such that each edge from $\mathcal{E}$ has an endpoint in S. The cardinality of a smallest vertex cover of $\mathcal{G}$ is the vertex cover number $\tau(\mathcal{G})$.

We introduce for every $v \in \mathcal{V}$

$$\Gamma_{\mathcal{G}}(v) = \{v': (v, v') \in \mathcal{E}\} \tag{3.5}$$

and for every $S \subset \mathcal{V}$

$$\Gamma_{\mathcal{G}}(S) = \bigcup_{v \in S} \Gamma_{\mathcal{G}}(v). \tag{3.6}$$

The degree of v is

$$d_{\mathcal{G}}(v) = |\Gamma_{\mathcal{G}}(v)|. \tag{3.7}$$

Two of the first and most basic results in matching theory (see [7]) are as follows.

Theorem H (Hall's Marriage Theorem, [6], [7]):

A bipartite graph $\mathcal{G} = (\mathcal{V}_1, \mathcal{V}_2, \mathcal{E})$ has a matching of $\mathcal{V}_1$ into $\mathcal{V}_2$ iff

$$|\Gamma(S)| \geq S, \qquad \text{for all } S \subset \mathcal{V}_1 \tag{3.8}$$

and

Theorem K (König's Minimax Theorem [7], [10]):

For every bipartite graph $\mathcal{G}$

$$\tau(\mathcal{G}) = \nu(\mathcal{G}).$$

These theorems can easily be derived from each other. We need here a consequence of Theorem K.

Corollary 1: If $\mathcal{G} = (\mathcal{V}_1, \mathcal{V}_2, \mathcal{E})$ satisfies for two numbers $d_{\mathcal{V}_1}, d_{\mathcal{V}_2}$ and for $i = 1, 2$

$$d_{\mathcal{G}}(v) = d_{\mathcal{V}_i}, \qquad \text{for all } v \in \mathcal{V}_i$$

then

$$\tau(\mathcal{G}) = \nu(\mathcal{G}) = \min_{i=1,2} |\mathcal{V}_i|.$$

Proof: Without loss of generality we can assume $|\mathcal{V}_1| \leq |\mathcal{V}_2|$. By the hypothesis we have also $|\mathcal{E}| = d_{\mathcal{V}_i} |\mathcal{V}_i|$ for $i = 1, 2$ and thus $d_{\mathcal{V}_1} \geq d_{\mathcal{V}_2}$.

Hence, each vertex of $\mathcal{G}$ covers at most $d_{\mathcal{V}_1}$ edges and, therefore,

$$\tau(\mathcal{G}) d_{\mathcal{V}_1} \geq |\mathcal{E}| = d_{\mathcal{V}_1} |\mathcal{V}_1|$$

which gives the result.

Remark:

3) We draw attention to the fact that matching and covering in $\mathcal{G}(W^n)$ are different from packing and covering by edges for Cartesian products of hypergraphs (c.f. [1]).

IV. Main Results on Matchcing in Products of Bipartite Graphs

For any bipartite graph $\mathcal{G} = (\mathcal{X}, \mathcal{Y}, \mathcal{E})$ we study the asymptotic behavior of the matching number $\nu(\mathcal{G}^{\otimes n})$. A key idea is to extend the König–Hall condition (3.8), which is in terms of cardinalities as measure of sets, to pairs of PD's associated with $\mathcal{G}$. The matching capacity of $\mathcal{G}$ is

$$\gamma(\mathcal{G}) = \lim_{n \to \infty} \frac{1}{n} \log \nu(\mathcal{G}^{\otimes n}). \tag{4.1}$$

We define the set

$$\mathcal{K}(\mathcal{G}) = \{(P, Q) \colon P \in \mathcal{P}(\mathcal{X}), Q \in \mathcal{P}(\mathcal{Y}), P(S) \leq Q(\Gamma_{\mathcal{G}}(S))\ \forall\, S \subset \mathcal{X}\} \tag{4.2}$$

and call its members König–Hall pairs of distributions.

Moreover, in the sequel we assume that all graphs have no *isolated vertices.*

Theorem 4: For every bipartite graph $\mathcal{G} = (\mathcal{X}, \mathcal{Y}, \mathcal{E})$

$$\gamma(\mathcal{G}) = \max_{(P,Q) \in \mathcal{K}(\mathcal{G})} \min(H(P), H(Q)).$$

Proof: Recall the definitions of typical sequences and types in Section II. We shall decompose $\mathcal{G}^{\otimes n}$ into subgraphs

$$\mathcal{G}_n(P, Q) = (\mathcal{X}^n(P), \mathcal{Y}^n(Q), \mathcal{E}_n(P, Q))$$

where $P \in \mathcal{P}_n(\mathcal{X})$, $Q \in \mathcal{P}_n(\mathcal{Y})$, and

$$\mathcal{E}_n(P, Q) = \mathcal{E}_n \cap (\mathcal{X}^n(P) \times \mathcal{Y}^n(Q)).$$

Clearly, since $\mathcal{G}_n(P, Q)$ is a subgraph of $\mathcal{G}^{\otimes n}$

$$\tau(\mathcal{G}_n(P, Q)) \leq \tau(\mathcal{G}^{\otimes n}). \tag{4.3}$$

On the other hand, if $C_n(P, Q)$ is a cover of $\mathcal{G}_n(P, Q)$ of smallest size, then

$$\bigcup_{(P,Q) \in \mathcal{P}_n(\mathcal{X}) \times \mathcal{P}_n(\mathcal{Y})} C_n(P, Q)$$

is a cover of $\mathcal{G}^{\otimes n}$ and thus

$$\tau(G^{\otimes n}) \leq \sum_{(P,Q) \in \mathcal{P}_n(\mathcal{X}) \times \mathcal{P}_n(\mathcal{Y})} \tau C_n(P, Q)). \tag{4.4}$$

Now, since $|\mathcal{P}_n(\mathcal{X})|$ and $|\mathcal{P}_n(\mathcal{Y})|$ grow only polynomially in n, (4.3) and (4.4) imply

$$\lim_{n \to \infty} \frac{1}{n} \log \tau(\mathcal{G}^{\otimes n}) = \lim_{n \to \infty} \frac{1}{n} \log \max_{(P,Q) \subset \mathcal{P}_n(\mathcal{X}) \times \mathcal{P}_n(\mathcal{Y})} \tau(\mathcal{G}_n(P, Q)). \tag{4.5}$$

Next observe that for $(P, Q) \in \mathcal{P}_n(\mathcal{X}) \times \mathcal{P}_n(\mathcal{Y})$ with $\mathcal{E}_n(P, Q) \neq \emptyset$ $\mathcal{G}_n(P, Q)$ satisfies the hypothesis of Corollary 1, because for any $x^n, x'^n \in \mathcal{X}^n(P)$ there is a permutation π on $\{1, 2, \cdots, n\}$ with $\pi x^n = (x_{\pi(1)}, \cdots, x_{\pi(n)}) = x'^n$ and by the invariance of $\mathcal{Y}^n(Q)$ under π

$$\begin{aligned} |\Gamma_{\mathcal{G}_n(P,Q)}(x^n)| &= \left| \prod_{i=1}^{n} \Gamma_{\mathcal{G}}(x_t) \cap \mathcal{Y}^n(Q) \right| \\ &= \left| \prod_{t=1}^{n} \Gamma_{\mathcal{G}}(x_{\pi(t)}) \cap \mathcal{Y}^n(Q) \right| \\ &= |\Gamma_{\mathcal{G}_n(P,Q)}(x'^n)| \end{aligned}$$

(and, symmetrically this holds for $y^n, y'^n \in \mathcal{Y}^n(Q)$). We conclude with Corollary 1 that for these P, Q

$$\nu(\mathcal{G}_n(P, Q)) = \tau(\mathcal{G}_n(P, Q)) = \min(|\mathcal{X}^n(P)|, |\mathcal{Y}^n(Q)|). \tag{4.6}$$

By Theorem K also

$$\nu(\mathcal{G}^{\otimes n}) = \tau(\mathcal{G}^{\otimes n}). \tag{4.7}$$

Now, from (4.5) we conclude with (4.6), (4.7), and (2.1)

$$\lim_{n \to \infty} \frac{1}{n} \log \nu(\mathcal{G}^{\otimes n}) = \lim_{n \to \infty} \max_{P, Q \colon \mathcal{E}_n(P,Q) \neq \emptyset} \min(H(P), H(Q)). \tag{4.8}$$

The final step is based on a result of interest on its own.

Lemma 1:

i) For all n

$$(P,Q) \in \mathcal{K}(\mathcal{G}) \cap (\mathcal{P}_n(\mathcal{X}) \times \mathcal{P}_n(\mathcal{Y})) \Leftrightarrow \mathcal{E}_n(P,Q) \neq \emptyset. \tag{4.9}$$

ii) If $P(x) > 0$ for all $x \in \mathcal{X}$, then for all $\varepsilon > 0, (P,Q) \in \mathcal{K}(\mathcal{G})$ and sufficiently large n, there exists

$$(P',Q') \in \mathcal{K}(\mathcal{G}) \cap (\mathcal{P}_n(\mathcal{X}) \times \mathcal{P}_n(\mathcal{Y}))$$

such that

$$\sum_{x\in\mathcal{X}} |P(x) - P'(x)|, \sum_{x\in\mathcal{Y}} |Q(y) - Q'(y)| < \varepsilon.$$

Proof of Lemma 1:

i) Fix $x^n \in \mathcal{X}^n(P)$. By symmetry it will not matter which one. Clearly,

$$\mathcal{E}_n(P,Q) \neq \emptyset \Leftrightarrow d_{\mathcal{G}_n(P,Q)}(x^n) > 0. \tag{4.10}$$

We give first another characterization for $\mathcal{E}_n(P,Q) \neq \emptyset$ in terms of a matching property of another bipartite graph $\mathcal{G}_{(n)}(P,Q)$.

This graph has the vertex sets $\mathcal{X}_{(n)} = \{x_1, x_2, \cdots, x_n\}$ and $\mathcal{Y}_{(n)} = \{y_1, \cdots, y_n\}$, where $\mathcal{X}_{(n)}$ contains $n\ P(x)$ "copies" of each $x \in \mathcal{X}$ and $\mathcal{Y}_{(n)}$ contains $n\ Q(y)$ "copies" of each $y \in \mathcal{Y}$.

It has the edge set

$$\begin{aligned}\mathcal{E}_{(n)} = \{(x^*, y^*)\colon & x^* \text{ is copy of } x \in \mathcal{X},\\ & y^* \text{ is copy of } y \in \mathcal{Y}, \text{ and } (x,y) \in \mathcal{E}\}.\end{aligned}$$

By definitions of $\mathcal{G}^{\otimes n}$, $\mathcal{G}_n(P,Q)$, and $\mathcal{G}_{(n)}(P,Q)$, $x^n = (x_1, \cdots, x_n)$, is adjacent with at least one vertex in $\mathcal{G}_n(P,Q)$ iff $\mathcal{G}_{(n)}(P,Q)$ has a perfect matching or by (4.10)

$$\mathcal{E}_n(P,Q) \neq \emptyset \Leftrightarrow \mathcal{G}_{(n)}(P,Q) \text{ has a perfect matching.} \tag{4.11}$$

A fortiori (4.9) is equivalent to

$$\begin{aligned}&(P,Q) \in \mathcal{K}(\mathcal{G}) \cap \mathcal{P}_n(\mathcal{X}) \times \mathcal{P}_n(\mathcal{Y})\\ &\qquad \Leftrightarrow \mathcal{G}_{(n)}(P,Q) \text{ has a perfect matching.}\end{aligned} \tag{4.12}$$

To show this, let us start with a

$$(P,Q) \in \mathcal{K}(\mathcal{G}) \cap \mathcal{P}_n(\mathcal{X}) \times \mathcal{P}_n(\mathcal{Y}).$$

Now, every $S^* \subset \mathcal{X}_{(n)}$ is associated with a subset S of $\mathcal{X}$, where

$$x \in S \Leftrightarrow x \text{ has a copy } x_i \in S^*. \tag{4.13}$$

By the definitions of $\mathcal{G}_{(n)}, \mathcal{X}^n(P)$ and $\mathcal{K}(\mathcal{G})$ we have now

$$\begin{aligned}|S^*| &\le \sum_{x\in S} n\ P(x) = n\ P(S) \le n\ Q(\Gamma_{\mathcal{G}}(S))\\ &= \sum_{y\in\Gamma_{\mathcal{G}}(S)} n\ Q(y) = |\Gamma_{\mathcal{G}_{(n)}}(S^*)|.\end{aligned} \tag{4.14}$$

This Hall condition and Theorem H imply that $\mathcal{G}_{(n)}(P,Q)$ has a perfect matching. Conversely, let us assume now that $\mathcal{G}_{(n)}(P,Q)$ has a perfect matching. For any $S \subset \mathcal{X}$ define

$$S^{**} = \{x_i\colon x_i \text{ is copy of some } x \in S\} \tag{4.15}$$

a subset of $\mathcal{X}_{(n)}$. Then

$$n\ P(S) = \sum_{x\in S} n\ P(x) = |S^{**}| \tag{4.16}$$

and since $\mathcal{G}_{(n)}$ has a perfect matching, by Theorem H

$$|S^{**}| \le |\Gamma_{\mathcal{G}_{(n)}}(S^{**})|. \tag{4.17}$$

Also, by (4.15)

$$|\Gamma_{\mathcal{G}_{(n)}}(S^{**})| = \sum_{y\in\Gamma_{\mathcal{G}}(S)} n\ Q(y) = n\ Q(\Gamma_{\mathcal{G}}(S))$$

and finally, this and (4.16), (4.17) imply $P(S) \le Q(\Gamma_{\mathcal{G}}(S))$ and so $(P,Q) \in \mathcal{K}(\mathcal{G})$.

ii) We proceed by induction on $|\mathcal{X}|$. For $|\mathcal{X}| = 1$ the statement is trivial.

$\underline{|\mathcal{X}| > 1}$:

We say that a distribution P^* on $\mathcal{Z}$ is δ-approximated by $\hat{P}$ if

$$\sum_{z\in\mathcal{Z}} |P^*(z) - \hat{P}(z)| < \delta.$$

Case 1: For all $\phi \neq S \subsetneq \mathcal{X}$

$$P(S) < Q(\Gamma_{\mathcal{G}} S). \tag{4.18}$$

Let

$$\delta' = \frac{1}{4} \min_{\phi\neq S\subsetneq\mathcal{X}} (Q(\Gamma_{\mathcal{G}}(S)) - P(S))$$

and let $\delta = \min(\delta', \varepsilon)$. Then by (4.18), $\delta' > 0$. When n is sufficiently large, we always can choose $P' \in \mathcal{P}_n(\mathcal{X})$ and $Q' \in \mathcal{P}_n(\mathcal{Z})$ δ-approximating P and Q, respectively. Moreover, $(P',Q') \in \mathcal{K}(\mathcal{G})$ because for all $S \subset \mathcal{X}$, $|P(S) - P'(S)| < \delta$ and $|Q(\Gamma_{\mathcal{G}}(S)) - Q'(\Gamma_{\mathcal{G}}(S))| < \delta$. This completes the proof in this case.

Case 2: There exists an $\mathcal{X}_0 \subset \mathcal{X}$ with $0 < |\mathcal{X}_0| < |\mathcal{X}|$ (and so, by assumption, $0 < P(\mathcal{X}_0) < 1$), such that

$$P(\mathcal{X}_0) = Q(\Gamma_{\mathcal{G}}\mathcal{X}_0). \tag{4.19}$$

Let $\mathcal{Y}_0 = \Gamma_{\mathcal{G}}(\mathcal{X}_0)$, $\mathcal{X}_1 = \mathcal{X}\backslash\mathcal{X}_0$, and $\mathcal{Y}_1 = \mathcal{Y}\backslash\mathcal{Y}_0$, and introduce two subbipartite graphs $\mathcal{G}_0 = (\mathcal{X}_0, \mathcal{Y}_0, \mathcal{E}_0)$ and $\mathcal{G}_1 = (\mathcal{X}_1, \mathcal{Y}_1, \mathcal{E}_1)$ of $\mathcal{G}$, where

$$\mathcal{E}_i = \{(x,y)\colon x \in \mathcal{X}_i, y \in \mathcal{Y}_i, (x,y) \in \mathcal{E}\}$$

for $i = 1, 2$.

Then by (4.19)

$$P(\mathcal{X}_i) = Q(\mathcal{Y}_i), \qquad \text{for } i = 0, 1 \tag{4.20}$$

and, therefore, since $(P,Q) \in \mathcal{K}(\mathcal{G})$,

$$(P(\cdot|\mathcal{X}_0), Q(\cdot|\mathcal{Y}_0)) \in \mathcal{K}(\mathcal{G}_0).$$

Since for any $\delta > 0$, $\overline{P} = (P(\mathcal{X}_0), P(\mathcal{X}_1))$, which equals $\overline{Q} = (Q(\mathcal{Y}_o), Q(\mathcal{Y}_1))$, can be δ-approximated by m-types for sufficiently large m, ii) follows from the induction hypothesis, if we can show that $(P(\cdot|\mathcal{X}_1), Q(\cdot|\mathcal{Y}_1)) \in \mathcal{K}(\mathcal{G}_1)$.

Indeed, this must be true, since otherwise one could find $S \subset \mathcal{X}_1$ such that $P(S|\mathcal{X}_1) < Q(\Gamma_{\mathcal{G}_1}(S)|\mathcal{Y}_1)$, and therefore by (4.20), such that

$$P(S) < Q(\Gamma_{\mathcal{G}_1}(S)) \le Q(\Gamma_{\mathcal{G}}(S))$$

which contradicts $(P,Q) \in \mathcal{K}(\mathcal{G})$.

Remark:

4) Lemma 1 shows that the definition of $\mathcal{K}(\mathcal{G})$ is symmetrical in the vertex sets, that is, we have also

$$(P,Q) \in \mathcal{K}(\mathcal{G}) \Leftrightarrow \forall\, T \subset \mathcal{Y} \qquad Q(T) \le P(\Gamma_{\mathcal{G}}(T)).$$

Lemma 1 has an immediate consequence:

Corollary 2: For all $P \in \mathcal{P}_n(\mathcal{X})$, $Q \in \mathcal{P}_n(\mathcal{Y})$, $W \in \mathcal{V}$, and $x^n \in \mathcal{T}_P^n$: $W^n(\mathcal{T}_Q^n(x^n)|x^n) > 0$ iff

$$P(S) \le Q(\{y\colon w(y|x) > 0 \text{ for some } x \in S\}$$

for all $S \subset \mathcal{X}$.

V. Matching in Products of Nonidentical Bipartite Graphs

The result of this section answers a natural combinatorial question, but its main motivation was to extend our coding theorems for the deterministic matching channel to the nonstationary situation. In terms of the associated discrete memoryless channel this means that we are given a sequence $(W_t)_{t=1}^{\infty}$ of $|\mathcal{X}| \times |\mathcal{Y}|$-stochastic matrices and the transmission for words of length n is governed by $W^n = \Pi_{t=1}^n W_t$. In this situation an approach with typical sequences is very clumsy, but our approach via König–Hall pairs of distributions goes rather smoothly.

The heart of the matter is the case of two factors: $\mathcal{G}_i = (\mathcal{X}_i, \mathcal{Y}_i, \mathcal{E}_i)$ $(i = 1,2)$. For $P_i \in \mathcal{P}(\mathcal{Z}_i)$, where $\mathcal{Z}_i$ is finite and $i = 1,2$, we define the product distribution $P_1 \times P_2$ by

$$P_1 \times P_2(z_1,z_2) = P_1(z_1)P_2(z_2), \qquad \text{for } z_i \in \mathcal{Z}_i \text{ and } i = 1,2. \tag{5.1}$$

We introduce a product of König–Hall sets, namely,

$$\mathcal{K}(\mathcal{G}_1) \times \mathcal{K}(\mathcal{G}_2) = \{(P_1 \times P_2, Q_1 \times Q_2)\colon (P_i,Q_i) \in \mathcal{K}(\mathcal{G}_i) \text{ for } i = 1,2\}. \tag{5.2}$$

Theorem 5: For bipartite graphs $\mathcal{G}_i = (\mathcal{X}_i, \mathcal{Y}_i, \mathcal{E}_i)$ $(i=1,2)$

$$\begin{aligned}\gamma(\mathcal{G}_1 \otimes \mathcal{G}_2) &= \max_{(P_1 \times P_2, Q_1 \times Q_2) \in \mathcal{K}(\mathcal{G}_1) \times \mathcal{K}(\mathcal{G}_2)} \cdot \min(H(P_1) + H(P_2), H(Q_1) + H(Q_2)) \\ &= \max_{\substack{(P_i,Q_i) \subset \mathcal{K}(\mathcal{G}_i) \\ i=1,2}} \cdot \min(H(P_1) + H(P_2), H(Q_1) + H(Q_2)).\end{aligned} \tag{5.3}$$

Proof: Obviously, the second equation follows immediately from (5.2). We show now the first equation.

By Theorem 4 and Lemma 1 it suffices to prove that for

$$\mathcal{K}_n(\mathcal{G}_i) = \mathcal{K}(\mathcal{G}_i) \cap (\mathcal{P}_n(\mathcal{X}_i) \times \mathcal{P}_n(\mathcal{Y}_i)), \qquad i = 1,2 \tag{5.4}$$

and

$$\mathcal{K}_n(\mathcal{G}_1 \otimes \mathcal{G}_2) = \mathcal{K}(\mathcal{G}_1 \otimes \mathcal{G}_2) \cap (\mathcal{P}_n(\mathcal{X}_1 \times \mathcal{X}_2) \times \mathcal{P}_n(\mathcal{Y}_1 \times \mathcal{Y}_2)) \tag{5.5}$$

we have for all n

$$\begin{aligned}&\max_{(P,Q) \subset \mathcal{K}_n(\mathcal{G}_1 \otimes \mathcal{G}_2)} \min(H(P), H(Q)) \\ &\quad\le \max_{(P_1 \times P_2, Q_1 \times Q_2) \in \mathcal{K}_n(\mathcal{G}_1) \times \mathcal{K}_n(\mathcal{G}_2)} \cdot \min(H(P_1) + H(P_2), H(Q_1) + H(Q_2)) \\ &\quad\le \max_{(P,Q) \in \mathcal{K}_{n^2}(\mathcal{G}_1 \otimes \mathcal{G}_2)} \min(H(P), H(Q)).\end{aligned} \tag{5.6}$$

We need

$$\mathcal{K}_n(\mathcal{G}_1) \times \mathcal{K}_n(\mathcal{G}_2) \subset \mathcal{K}_{n^2}(\mathcal{G}_1 \otimes \mathcal{G}_2). \tag{5.7}$$

To verify this, by (4.9), we have to show that for all

$$(P_1 \times P_2, Q_1 \times Q_2) \in \mathcal{K}_n(\mathcal{G}_1) \times K_n(\mathcal{G}_2)$$
$$\mathcal{E}_{n^2}(P_1 \times P_2, Q_1 \times Q_2) \ne \emptyset.$$

Actually, for all

$$(v_i, v_i') \in \mathcal{E}_n(P_i, Q_i) \ne \emptyset, \qquad i = 1,2$$

$$(v_1 v_2, v_1' v_2') \in \mathcal{E}_{n^2}(P_1 \times P_2, Q_1 \times Q_2).$$

Therefore, (5.7) holds and the second inequality in (5.6) follows.

Finally, we have to prove the first inequality in (5.6). Suppose that $(\hat{P}, \hat{Q})$ achieve the maximum in the left-hand side of (5.6) and that $\hat{P}_1, \hat{P}_2$ (respectively, $\hat{Q}_1, \hat{Q}_2$) are the marginal distributions of $\hat{P}$ (respectively, $\hat{Q}$).

Since clearly

$$H(\hat{P}) \le H(\hat{P}_1) + H(\hat{P}_2)$$

and

$$H(\hat{Q}) \le H(\hat{Q}_1) + H(\hat{Q}_2) \tag{5.8}$$

it suffices to prove that $(\hat{P}_i, \hat{Q}_i) \in \mathcal{K}(\mathcal{G}_i)$ for $i = 1,2$.

Actually, one readily verifies that for all $S \subset \mathcal{X}_1$

$$\Gamma_{\mathcal{G}_1 \otimes \mathcal{G}_2}(S \times \mathcal{X}_2) = \Gamma_{\mathcal{G}_1}(S) \times \mathcal{Y}_2 \tag{5.9}$$

and, therefore, $(\hat{P}, \hat{Q}) \in \mathcal{K}(\mathcal{G}_1 \otimes \mathcal{G}_2)$ implies

$$\hat{P}_1(S) = \hat{P}(S \times \mathcal{X}_2) \le \hat{Q}(\Gamma_{\mathcal{G}_1}(S) \times \mathcal{Y}_2) = \hat{Q}_1\,(\Gamma_{\mathcal{G}_1}(S))$$

and hence $(\hat{P}_1, \hat{Q}_1) \in \mathcal{K}(\mathcal{G}_1)$. By the same reasons also $(\hat{P}_2, \hat{Q}_2) \in \mathcal{K}(\mathcal{G}_2)$.

VI. An Exact Formula for the Matching Number of Powers of "Starred" Bipartite Graphs

We consider here bipartite graphs $\mathcal{G} = (\mathcal{X}, \mathcal{Y}, \mathcal{E})$, which can be presented in the following form.

There are sets of vertices $\mathcal{J} \subset \mathcal{X}$ and $\mathcal{K} \subset \mathcal{Y}$ such that

i) every vertex in $\mathcal{J}$ (respectively, $\mathcal{K}$) is adjacent with at least one vertex in $\mathcal{Y} \setminus \mathcal{K}$ (respectively, $\mathcal{X} \setminus \mathcal{J}$);

ii) every vertex in $\mathcal{X} \setminus \mathcal{J}$ (respectively, $\mathcal{Y} \setminus \mathcal{K}$) is adjacent with exactly one vertex in $\mathcal{K}$ (respectively, $\mathcal{J}$), and there is no edge between $\mathcal{X} \setminus \mathcal{J}$ and $\mathcal{U} \setminus \mathcal{K}$.

We speak of a *starred* bipartite graph. We also introduce the abbreviation

$$\mathcal{Z} = \mathcal{J} \cup \mathcal{K} \tag{6.1}$$

and for every $z \in \mathcal{Z}$ we define a *star with center* z as $S_z = \{\{z\}, \mathcal{V}_z, \mathcal{F}_z\}$, where

$$\mathcal{V}_z = \{v \in (\mathcal{X} \setminus \mathcal{J}) \cup (\mathcal{Y} \setminus \mathcal{K}) \colon (z,v) \in \mathcal{E}\} \tag{6.2}$$

$$\mathcal{F}_z = \{(z,v) \colon v \in \mathcal{V}_z\}. \tag{6.3}$$

Of course, since $\mathcal{G}$ is bipartite, for $z = j \in \mathcal{J}$ (respectively, $z = k \in \mathcal{K}$) necessarily $\mathcal{V}_j \subset \mathcal{Y} \setminus \mathcal{K}$ (respectively, $\mathcal{V}_k \subset \mathcal{X} \setminus \mathcal{J}$). By conditions i) and ii), obviously $\{\{z\} \cup \mathcal{V}_z \colon z \in \mathcal{Z}\}$ is a partition of $\mathcal{X} \cup \mathcal{Y}$.

Now we associate with every $z^n \in \mathcal{Z}^n = \Pi_1^n\, \mathcal{Z}$ the complete bipartite graph

$$S_{z^n} = S_{z_1} \otimes S_{z_2} \otimes \cdots \otimes S_{z_n} \tag{6.4}$$

where S_{z_t}'s are stars defined by (6.2) and (6.3) for $z_t = z$.

Denote its vertex set by $\mathcal{V}^*_{z^n}$ and its edge set by $\mathcal{F}_{z^n} = \Pi_{t=1}^n\, \mathcal{F}_{z_t}$. Notice that

$$|\mathcal{V}^*_{z^n} \cap \mathcal{X}^n| = \prod_{z_t \notin \mathcal{J}} |\mathcal{V}_{z_t}|. \tag{6.5}$$

This is the number of vertices of S_{z^n} falling into $\mathcal{X}^n$ and will be denoted by $\omega_{\mathcal{X}}(z^n)$. Similarly, we define

$$\omega_{\mathcal{Y}}(z^n) = \prod_{z_t \notin \mathcal{K}} |\mathcal{V}_{z_t}|. \tag{6.6}$$

We speak of the $\mathcal{X}$-weight and of the $\mathcal{Y}$-weight of S_{z^n}.

Theorem 6: For every stared bipartite graph $\mathcal{G}$ the matching number of its nth power is given by

$$\nu(\mathcal{G}^{\otimes n}) = \sum_{z^n \subset \mathcal{Z}^n} \min(\omega_{\mathcal{X}}(z^n), \omega_{\mathcal{Y}}(z^n)). \tag{6.7}$$

Proof: Since S_{z^n} is a complete bipartite graph with vertex sets of sizes $\omega_{\mathcal{X}}(z^n)$ and $\omega_{\mathcal{Y}}(z^n)$, it has obviously a matching of size $\min(\omega_{\mathcal{X}}(z^n), \omega_{\mathcal{Y}}(z^n))$. Furthermore, by our definitions, for $z^n \neq z'^n$, an edge in $\mathcal{F}_{z^n}$ and an edge in $\mathcal{F}_{z'^n}$ have never a common vertex.

Therefore, the matching corresponding to the different z^n's can be taken together to form one matching. Thus

$$\nu(\mathcal{G}^{\otimes n}) \geq \sum_{z^n \subset \mathcal{Z}^n} \min(\omega_{\mathcal{X}}(z^n), \omega_{\mathcal{Y}}(z^n)). \tag{6.8}$$

To show the opposite inequality, by Theorem K it suffices to find a vertex cover of $\mathcal{G}^{\otimes n}$ of size

$$\sum_{z^n \in \mathcal{Z}^n} \min(\omega_{\mathcal{X}}(z^n), \omega_{\mathcal{Y}}(z^n)).$$

Our candidate is the set of vertices

$$\left(\bigcup_{z^n \in \mathcal{Z}^n : \omega_{\mathcal{X}}(z^n) \leq \omega_{\mathcal{Y}}(z^n)} (\mathcal{V}^*_{z^n} \cap \mathcal{X}^n) \right) \cup \left(\bigcup_{z^n \in \mathcal{Z}^n : \omega_{\mathcal{X}}(z^n) > \omega_{\mathcal{Y}}(z^n)} (\mathcal{V}^*_{z^n} \cap \mathcal{Y}^n) \right). \tag{6.9}$$

Clearly, it has the desired cardinality. It remains to be seen that it is a vertex cover for $\mathcal{G}^{\otimes n}$.

Suppose that $(x^n, y^n) \in \mathcal{E}_n$ is not covered. Then necessarily $T = \{t\colon \text{both, } x_t \text{ and } y_t, \text{ are centers of a star}\} \neq \emptyset$, because all edges in $\mathcal{F}_{z^n}, z^n \in \mathcal{Z}^n$, are covered for our candidate. Next we observe that for every $t \in \{1, 2, \cdots, n\} \setminus T$ x_t and y_t are in the same star. Therefore, if $x^n \in \mathcal{V}_{z^n}, y^n \in \mathcal{V}_{z'^n}$, then $z_t = z'_t$ for $t \in \{1, 2, \cdots, n\} \setminus T$ and thus

$$\omega_{\mathcal{X}}(z^n) = \omega_{\mathcal{X}}(z'^n) \left(\prod_{t \subset T} |\mathcal{V}_{z'_t}| \right)^{-1} \tag{6.10}$$

$$\omega_{\mathcal{Y}}(z^n) = \omega_{\mathcal{Y}}(z'^n) \left(\prod_{t \in T} |\mathcal{V}_{z_t}| \right). \tag{6.11}$$

(Since for all $t \in T$ x_t and y_t are centers of S_{z_t} and $S_{z'_t}$, respectively.)

Since by assymption (x^n, y^n) is not covered, neither x^n nor y^n is in our candidate subset. By the construction of this subset

$$\omega_{\mathcal{X}}(z^n) > \omega_{\mathcal{Y}}(z^n) \quad \text{and} \quad \omega_{\mathcal{X}}(z'^n) \leq \omega_{\mathcal{Y}}(z'^n) \tag{6.12}$$

and (6.10)–(6.12) imply

$$\begin{aligned} \omega_{\mathcal{Y}}(z^n) &\geq \omega_{\mathcal{X}}(z'^n) \left(\prod_{t \subset T} |\mathcal{V}_{z_t}| \right) \\ &= \omega_{\mathcal{X}}(z^n) \left(\prod_{t \in T} |\mathcal{V}_{z'_t}| \right) \left(\prod_{t \in T} |\mathcal{V}_{z_t}| \right) \\ &> \omega_{\mathcal{Y}}(z^n) \left(\prod_{t \in T} |\mathcal{V}_{z'_t}| \right) \left(\prod_{t \in T} |\mathcal{V}_{z_t}| \right). \end{aligned}$$

This contradicts the definition of the graph, in which

$$|\mathcal{V}_z| \geq 1, \qquad \text{for all } v \in \mathcal{Z} = \mathcal{J} \cup \mathcal{K}.$$

VII. Two Examples Illustrating the Significance of Theorems 5 and 6

Example 1: $\gamma(\mathcal{G}_1 \otimes \mathcal{G}_2) > \gamma(\mathcal{G}_1) + \gamma(\mathcal{G}_2)$.

Consider two complete bipartite graphs

$$\mathcal{G}_i = (\mathcal{X}_i, \mathcal{Y}_i, \mathcal{E}_i), \qquad i = 1, 2$$

with parameters

$$|\mathcal{X}_1| = |\mathcal{Y}_2| = \alpha < \beta = |\mathcal{X}_2| = |\mathcal{Y}_1|.$$

Obviously,

$$\nu(\mathcal{G}_i^{\otimes n}) = \alpha^n, \qquad \text{for } i = 1, 2 \tag{7.1}$$

and, therefore,

$$\gamma(\mathcal{G}_i) = \log \alpha, \qquad \text{for } i = 1, 2. \tag{7.2}$$

However, by Theorem 5 or by direct reasoning

$$\nu((\mathcal{G}_1 \otimes \mathcal{G}_2)^{\otimes n}) = (\alpha\beta)^n \tag{7.3}$$

because $\mathcal{G}_1 \otimes \mathcal{G}_2$ is a complete bipartite graph.

Thus we have

$$\gamma(\mathcal{G}_1 \otimes \mathcal{G}_2) = \log \alpha + \log \beta > 2 \log \alpha = \gamma(\mathcal{G}_1) + \gamma(\mathcal{G}_2).$$

Example 2:

$$\nu(\mathcal{G}^{\otimes(m_1+m_2)}) > \prod_{i=1}^{2} \nu(\mathcal{G}^{\otimes m_i}) \quad \text{and} \quad 2^{n\gamma(\mathcal{G})} > \nu(\mathcal{G}^{\otimes n})$$

can occur infinitely often and for arbitrarily large m_1, m_2, and n.

Consider the starred bipartite graph $\mathcal{G} = (\mathcal{X}, \mathcal{Y}, \mathcal{E})$ with $\mathcal{X} = \{x_i\colon i = 0, 1, \cdots, \alpha\}, \mathcal{Y} = \{y_j\colon j = 0, 1, \cdots, \alpha\}$, and $\mathcal{E} = \{(x_i, y_j)\colon i = 0 \text{ or } j = 0\}$.

By Theorem 6

$$\nu(\mathcal{G}^{\otimes n}) = \begin{cases} 2\left[\sum_{i=0}^{n/2} \binom{n}{i}\alpha^i + \frac{1}{2}\binom{n}{\frac{n}{2}}\alpha^{n/2}\right], & n \text{ even} \\ 2\sum_{i=0}^{(n-1)/2} \binom{n}{i}\alpha^i, & n \text{ odd.} \end{cases} \tag{7.4}$$

By Theorem 4 ($P(x_0) = Q(y_0) = \frac{1}{2}$, and $Q(y_j) = P(x_i) = 1/2\alpha$ for $i = 1, \cdots, \alpha$) or directly by (7.4)

$$\gamma(\mathcal{G}) = 1 + \tfrac{1}{2}\log\alpha. \tag{7.5}$$

Therefore, for *all* n

$$2^{\gamma(\mathcal{G})n} = 2^n\alpha^{n/2} > \nu(\mathcal{G}^{\otimes n}).$$

Moreover, (7.4) also shows that

$$\nu(\mathcal{G}^{\otimes(m_1+m_2)}) > \prod_{i=1}^{2} \nu(\mathcal{G}^{\otimes m_i}).$$

VIII. Multi-Way Deterministic Matching Channels

We take first another look at the (one-way) deterministic matching channel in order to get a certain understanding of its structure, which helps us when dealing with more complex channels such as compound, multiple-access and broadcast deterministic matching channels.

A. Another Look at Theorem 2

A straightforward proof of its direct part by random coding is sketched in Section I. Actually, this approach gives even a more general Theorem 1. Here we deal only with *deterministic* channels. Our first and detailed proof of Theorem 2 via an *extension* of *combinatorial matching theory* is contained in Sections III and IV. It arose in an analysis in the "control" model of the "dummy" model (see (1.4) and (1.5)).

Actually, there is a very simple direct path to Theorem 2 using König's Theorem K in Section III. Indeed, just consider the bipartite graph

$$\mathcal{G}_n = \mathcal{G}_n(P_{XS}, W) = (\mathcal{T}_X^n, \mathcal{T}_Y^n, \mathcal{E}_{P_{XS,W}}^n) \tag{8.1}$$

where $P_Y = P_{XS} \cdot W$ and $(x^n, y^n) \in \mathcal{E}_{P_{XS,W}}^n$ iff there exists an $(x^n, s^n) \in \mathcal{T}_{XS}^n$ with $W(y^n|x^n|s^n) = 1$. We know (see Section II) that

$$\begin{aligned} |\mathcal{T}_X^n| &= \exp\{H(X)n + o(n)\} \\ |\mathcal{T}_Y^n| &= \exp\{(H(Y)n + o(n)\} \\ d_{\mathcal{G}_n}(x^n) &= \exp\{H(Y|X)n + o(n)\} \\ d_{\mathcal{G}_n}(y^n) &= \exp\{H(X|Y)n + o(n)\} \end{aligned}$$

and $d_{\mathcal{G}_n}(x^n)$ (respectively, $d_{\mathcal{G}_n}(y^n)$) has the same value for all $x^n \in \mathcal{T}_X^n$ (respectively, $y^n \in \mathcal{T}_Y^n$).

The vertex covering number

$$\tau(\mathcal{G}_n(P_{XS}, W))$$

obviously equals $\min(H(X), H(Y))$ and the proof of Theorem 2 follows with Theorem K.

Problem 2: Can one give explicit constructions of matchings achieving the capacity in Theorem 2? Can it be done as an orbit of a group of permutations?

B. Compound Channels

We are given now c deterministic matching channels

$$W_j\colon \mathcal{X} \times \mathcal{S} \to \mathcal{Y}_j \qquad (j = 1, 2, \cdots, c)$$

and ask for a simultaneous code, that is, one set of codewords $\mathcal{U} \subset \mathcal{X}^n$ and decoding sets $\{\mathcal{D}_{ji}\colon 1 \le i \le |\mathcal{U}|\}$ for $j = 1, 2, \cdots, c$.

Here the random choice described above works again. Details are left to the reader.

Theorem 7: For the compound deterministic matching channel the capacity equals

$$\max_{XS} \min_{j=1,\cdots,c} \min(H(X), H(Y_j)).$$

Here it was implicitly assumed that the encoder and the controller do not know the individual (implicitly) channel, but that the receiver does. However, it is easy to show that (as for classical compound channels) the capacity is not affected by the receiver's knowledge.

More generally, one can also describe the capacity region of the compound multiple-access deterministic matching channel

$$W\colon \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_b \times \mathcal{S} \to \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_c$$

where we consider codes $\mathcal{U}_j \subset \mathcal{X}_j^n$ for $j = 1, 2, \cdots, b$ and decoding sets $\{\mathcal{D}_{jk}(i)\colon 1 \le i \le |\mathcal{U}_j|\}$ for $j = 1, 2, \cdots, b$ and $k = 1, 2, \cdots, c$.

C. Multiple-Access Channels (MAC's)

A matching MAC is given by a stochastic

$$W\colon (\mathcal{X} \times \mathcal{S}) \times (\mathcal{Y} \times \mathcal{T}) \to \mathcal{Z}.$$

It is understood that there are two controllers, K_S and K_T. Observing the input word x^n controller K_S can choose $s^n = s^n(x^n)$. Similarly, controller K_T responds to the input y^n. K_S does not observe y^n and K_T does not observe x^n.

Combining the sketch of proof of Theorem 1 with standard proofs for the MAC coding theorem gives the following result.

Theorem 8: The capacity region of the matching MAC contains $conv\{(R_{\mathcal{X}}, R_{\mathcal{Y}})\colon 0 \le R_{\mathcal{X}}, R_{\mathcal{Y}}, \exists P_{XS}, P_{YT}$ with $R_{\mathcal{X}} \le \min(I(XS \wedge Z|YT), H(X)), R_{\mathcal{Y}} \le \min\,(I(YT \wedge Z|XS), H(Y)), R_{\mathcal{X}} + R_{\mathcal{Y}} \le I(XSYT \wedge Z)\}$.

Adding a time-sharing parameter (if necessary) gives the exact region.

Problem 3: In another model there is only one controller $K_{\mathcal{X}\mathcal{Y}}$ who acts upon independent inputs x^n, y^n. The pair $(R_{\mathcal{X}}, R_{\mathcal{Y}})$ is achievable, if for some

$$\begin{aligned} P_{XYS} &= P_{S|XY} \cdot P_X \cdot P_Y \\ R_{\mathcal{X}} &\le \min(I(XS \wedge Z|Y), H(X)) \\ R_{\mathcal{Y}} &\le \min(I(YS \wedge Z|X), H(Y)) \\ R_{\mathcal{X}} + R_{\mathcal{Y}} &\le I(XYS \wedge Z). \end{aligned}$$

Establish the capacity region!

Problem 4: What are the capacity regions for matching zero-error detection codes for deterministic channels in both models?

D. Broadcast Channels

It is surprising that we can also determine here the capacity region! This means in the noisy channel terminology that under the (not realistic) condition that the error probability is strictly smaller than 1 we moved the rock!

Here is the result.

For the broadcast deterministic matching channel

$$W\colon \mathcal{X} \times \mathcal{S} \to \mathcal{Y} \times \mathcal{Z}$$

an (n, M, N) matching code is a family

$$\{(u_{ij}, \mathcal{D}_i(\mathcal{Y}), \mathcal{D}_j(\mathcal{Z}))\colon 1 \le i \le M, 1 \le j \le N\}$$

where $u_{ij} \in \mathcal{X}^n$, $\mathcal{D}_i(\mathcal{Y}) \subset \mathcal{Y}^n$, $\mathcal{D}_j(\mathcal{Z}) \subset \mathcal{Z}^n$, $\mathcal{D}_i(\mathcal{Y}) \cap \mathcal{D}_{i'}(\mathcal{Y}) = \emptyset (i \ne i')$, $\mathcal{D}_j(\mathcal{Z}) \cap \mathcal{D}_{j'}(\mathcal{Z}) = \emptyset (j \ne j')$, and for every pair of messages (i, j) there is a sequence $s^n(i, j) = (s_1(i, j), \cdots, s_n(i, j)) \in \mathcal{S}^n$ with

$$W^n(\mathcal{D}_i(\mathcal{X})|u_{ij}|s^n(i, j)) = 1$$

and

$$W^n(\mathcal{D}_j(\mathcal{Y})|u_{ij}|s^n(i, j)) = 1$$

for $1 \le i \le M, 1 \le j \le N$, if

$$W^n(y^n, z^n|x^n|s^n) = \prod_{t=1}^{n} W(y_t, z_t|x_t|s_t).$$

Theorem 9: The broadcast deterministic matching channel has the set of all achievable pairs of rates $(R_{\mathcal{Y}}, R_{\mathcal{Z}})$ defined by the convex hull of the sets

$$0 \le R_{\mathcal{Y}} \le H(Y) \qquad 0 \le R_{\mathcal{Z}} \le H(Z)$$

$$0 \le R_{\mathcal{Y}} + R_{\mathcal{Z}} \le \min(H(X), H(YZ))$$

where all RV's are induced by distributions P_{XS} and the channel.

Remarks:

5) Actually, more generally we also have a solution for the case with a common message set [21].

6) In case $|\mathcal{S}| = 1$ this yields Pinsker's characterization [19] of the capacity region for deterministic broadcast channels.

We use in our proof a certain bining idea, namely the

Color Carrier Lemma [26]: For each hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ there is a coloring $\varphi\colon \mathcal{V} \to \mathcal{L} = \{1, \cdots, L\}$ such that every color in $\mathcal{L}$ appears in every edge from $\mathcal{E}$ whenever

$$L \le \left(\ell n|\mathcal{E}| \min_{E \in \mathcal{E}} |E|\right)^{-1} \min_{E \in \mathcal{E}} |E|.$$

Proof of Theorem 9: The converse follows by a standard decomposition into subcodes corresponding to sets of typical sequences. The issue is the direct part. We give its flavor first in the case $|\mathcal{S}| = 1$, which is Pinsker's celebrated result.

1) Embedding of $\mathcal{T}^n_{YZ}$ into $\mathcal{T}^n_{XS}$: Every pair of typical sequences $(y^n, z^n) \in \mathcal{T}^n_{YZ}$ is the image of at least one pair $(x^n, s^n) \in \mathcal{T}^n_{XS}$ under W. We select any such pair $(\widetilde{y^n, z^n})$.

The matrix

$$\Omega = (\widetilde{y^n, z^n})_{y^n \in \mathcal{T}^n_Y, z^n \in \mathcal{T}^n_Z}$$

shall have entries "0" in all positions $(y^n, z^n) \notin \mathcal{T}^n_{YZ}$.

2) A Code Based on the Color Carrier Lemma: We can replace XS by X. Now we have $(\widetilde{y^n, z^n}) \in \mathcal{T}^n_{XZ}$ and Ω has

$$\sim \exp\{H(Y, Z)n\} \le \exp\{H(X)n\}$$

nontrivial entries, $r \sim \exp\{H(Y)n\}$ rows and $c \sim \exp\{H(Z)n\}$ columns. Consider the hypergraph

$$(\{1, \cdots, c\}, \{E_\rho\colon 1 \le \rho \le r\})$$

where E_ρ is the set of nontrivial positions in row ρ, so $E_\rho \subset \{1, 2, \cdots, c\}$.

Now message $\rho \in \{1, 2, \cdots, r\}$ can be associated with the row index ρ and message $\ell \in \{1, 2, \cdots, L\}$ can be associated with any $\varphi^{-1}(\ell) \in E_\rho$, where $\varphi\colon \{1, 2, \cdots, c\} \to \{1, 2, \cdots, L\}$ has by the Color Carrier Lemma the property that $\{\ell\colon \varphi^{-1}(\ell) \in E_\rho\} = \{1, 2, \cdots, L\}$. Moreover,

$$L \sim (H(Y)n)^{-1} \exp\{H(Z|Y)n\} \sim \exp\{H(Z|Y)n\}. \tag{8.2}$$

Thus with (r, L) we have achieved the rates $(H(Y), H(Z|Y))$ and the direct part is completed by time-sharing.

E. General Case $|\mathcal{S}| > 1$

Now we have to cope with the fact that now $\mathcal{T}^n_{YZ}$ has to be an embedding in $\mathcal{T}^n_{XS}$ such that every $x^n \in \mathcal{T}^n_X$ is used only at most once in conjunction with an s^n (the controller has to choose always the same s^n for x^n. Otherwise he illegally transmits information.)

So let us consider any quadruple (X, S, Y, Z) of RV's with joint distribution

$$P_{XSYZ}(x, s, y, z) = P_{XS}(x, s)W(y, z|x|s) \tag{8.3}$$

for $(x, s, y, z) \in \mathcal{X} \times \mathcal{S} \times \mathcal{Y} \times \mathcal{Z}$.

Clearly, for $(x^n, y^n, z^n) \in \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n$

$$(y^n, z^n) \in \mathcal{T}^n_{YZ|X}(x^n) \text{ implies } W^n(y^n, z^n | x^n | s^n) > 0 \text{ for some } s^n \in \mathcal{S}^n. \tag{8.4}$$

Moreover, by symmetry and time sharing it is sufficient to show that

$$(R_\mathcal{Y}, R_\mathcal{Z}) = ((\min(H(X), H(Y)), \min(H(X), H(YZ)) - \min(H(X), H(Y)))$$

is achievable.

Case $H(X) \le H(Y)$: Here $R_\mathcal{Y} = H(X), R_\mathcal{Z} = 0$. For instance, by Corollary 1 in Section III one can match $\mathcal{T}^n_X$ into $\mathcal{T}^n_Y$ in the bipartite graph $(\mathcal{T}^n_X, \mathcal{T}^n_Y, \mathcal{T}^n_{XY})$ (where $\mathcal{T}^n_{XY}$ serves as the edge set). Thus we get

$$\{(u_{i,1}, v_i) \in \mathcal{T}^n_{XY}\colon 1 \le i \le |\mathcal{T}^n_X|\}$$

where for any $i \ne j$ $u_{i,1} \ne u_{j,1}$ and $v_i \ne v_j$. With the decoding sets $\mathcal{D}_i(\mathcal{Y}) = \{v_i\}$ for $i = 1, 2, \cdots, |\mathcal{T}^n_X|$ and $\mathcal{D}_1(\mathcal{Z}) = \mathcal{T}^n_Z$ we see from (8.4) that

$$\{(u_{i,1}, \mathcal{D}_i(\mathcal{Y}), \mathcal{D}_1(\mathcal{Z}))\colon 1 \le i \le |\mathcal{T}^n_X|\}$$

is a matching code. It has the pair of rates $(H(X), 0)$.

Case $H(Y) < H(X)$: We proceed in two steps.

First, we find a subset $\hat{\mathcal{X}}_n \subset \mathcal{T}^n_X$ and an injective mapping $g\colon \hat{\mathcal{X}}_n \to \mathcal{T}^n_{YZ}$ such that for all $x^n \in \hat{\mathcal{X}}_n$

$$g(x^n) \in \mathcal{T}^n_{YZ|X}(x^n) \tag{8.5}$$

and for all $y^n \in \mathcal{T}^n_Y$ and any fixed $\theta \in (0,1)$ the set

$$E_{y^n} \triangleq \{z^n\colon g(x^n) = (y^n, z^n) \text{ for some } x^n \in \hat{\mathcal{X}}_n\} \tag{8.6}$$

satisfies

$$|E_{y^n}| \ge \exp\{n(R_\mathcal{Z} - \theta)\}. \tag{8.7}$$

This describes an embedding of a subset of $\mathcal{T}^n_{YZ}$ into $\mathcal{T}^n_X$.

Subcase $H(Y) \le H(YZ) \le H(X)$: Here $R_\mathcal{Y} = H(Y)$ and $R_\mathcal{Z} = H(Z|Y)$. Obviously, the matching of $\mathcal{T}^n_{YZ}$ into $\mathcal{T}^n_X$ for the bipartite graph $(\mathcal{T}^n_X, \mathcal{T}^n_{YZ}, \mathcal{T}^n_{XYZ})$ given by Corollary 1 provides a g with all desired properties.

We are left with the crucial

Subcase $H(Y) < H(X) < H(YZ)$: Our concern are the rates

$$R_\mathcal{Y} = H(Y) \quad \text{and} \quad R_\mathcal{Z} = H(X) - H(Y) > 0. \tag{8.8}$$

Originally we achieved them by a fairly lengthy (due to complications caused by the exponential sizes of $\mathcal{T}^n_X$ and $\mathcal{T}^n_{YZ}$) counting argument.

Now we use a large deviational argument based on Bernstein's version of Chebyshev's inequality. Its power lies in double exponential bounds, which proved to be very useful in Information Theory already in [27].

Let $U(x^n), x^n \in \mathcal{T}^n_X$, be a family of independent RV's and for each x^n $U(x^n)$ have uniform distribution over $\mathcal{T}^n_{YZ|X}(x^n)$.

For fixed $y^n \in \mathcal{T}^n_Y$ and $(y^n, z^n) \in \mathcal{T}^n_{YZ}$, we introduce the events

$$E_1(y^n) \triangleq \left\{ \sum_{x^n \in \mathcal{T}^n_X} \delta^*(y^n, U(x^n)) \le \frac{1 - e^{-1}}{2} \frac{|\mathcal{T}^n_X|}{|\mathcal{T}^n_Y|} \right\} \tag{8.9}$$

and

$$E_2((y^n, z^n)) \triangleq \left\{ \sum_{x^n \in \mathcal{T}^n_X} \delta((y^n, z^n), U(x^n)) \ge \exp\left\{ n\frac{\theta}{2} \right\} \right\} \tag{8.10}$$

where

$$\delta^*(y^n, (y'^n, z'^n)) = \begin{cases} 0, & \text{if } y^n \ne y'^n \\ 1, & \text{if } y^n = y'^n \end{cases}$$

and δ is Kronecker's delta, i.e.,

$$\delta((y^n, z^n), (y'^n, z'^n)) = \begin{cases} 0, & \text{if } (y^n, z^n) \ne (y'^n, z'^n) \\ 1, & \text{if } (y^n, z^n) = (y'^n, z'^n). \end{cases}$$

The definitions of $U(x^n), E_1(y^n), E_2((y^n, z^n))$, and the Bernstein version of Chebyshev's inequality imply for all $y^n \in \mathcal{T}^n_Y$

$$\begin{aligned}
&\Pr(E_1(y^n)) \\
&\quad\le e^{(1-e^{-1}/2)(|\mathcal{T}^n_X|/|\mathcal{T}^n_Y|)} \mathbb{E}\, e^{-\Sigma_{x^n \in \mathcal{T}^n_X} \delta^*(y^n, U(x^n))} \\
&\quad\overset{1)}{=} e^{(1/2)(1-e^{-1})(|\mathcal{T}^n_X|/|\mathcal{T}^n_Y|)} \mathbb{E}\, e^{-\Sigma^n_{x^n \in \mathcal{T}^n_{X|Y}(y^n)} \delta^*(y^n, U(x^n))} \\
&\quad\overset{2)}{=} e^{(1/2)(1-e^{-1})(|\mathcal{T}^n_X|/|\mathcal{T}^n_Y|)} \prod_{x^n \in \mathcal{T}^n_{X|Y}(y^n)} \mathbb{E}\, e^{-\delta^*\left(y^n, U(x^n)\right)} \\
&\quad= e^{(1/2)(1-e^{-1})(|\mathcal{T}^n_X|/|\mathcal{T}^n_Y|)} \prod_{x^n \in \mathcal{T}^n_{X|Y}(y^n)} \\
&\qquad\cdot \left(\frac{|\mathcal{T}^n_{Z|XY}(x^n, y^n)|}{|\mathcal{T}^n_{YZ|X}(x^n)|} e^{-1} + 1 - \frac{|\mathcal{T}^n_{Z|XY}(x^n, y^n)|}{|\mathcal{T}^n_{YZ|X}(x^n)|} \right) \\
&\quad\overset{3)}{=} e^{(1/2)(1-e^{-1})(|\mathcal{T}^n_X|/|\mathcal{T}^n_Y|)} \\
&\qquad\cdot \left(1 - (1 - e^{-1}) \frac{|\mathcal{T}^n_X|}{|\mathcal{T}^n_{XY}|} \right)^{|\mathcal{T}^n_{X|Y}(y^n)|} \\
&\quad\le e^{(1-e^{-1})[(1/2)(|\mathcal{T}^n_X|/|\mathcal{T}^n_Y|) - (|\mathcal{T}^n_X|/|\mathcal{T}^n_{XY}|)|\mathcal{T}^n_{X|Y}(y^n)|]} \\
&\quad\overset{4)}{=} \exp\left\{ -\frac{1}{2}(1 - e^{-1}) \frac{|\mathcal{T}^n_X|}{|\mathcal{T}^n_Y|} \log e \right\} \\
&\quad\overset{5)}{=} \exp\{-\exp(n(R_z + o(1))\}
\end{aligned} \tag{8.11}$$

if n is big enough, and for all $(y^n, z^n) \in \mathcal{T}^n_{YZ}$

$$\begin{aligned}
&\Pr(E_2(y^n, z^n)) \\
&\quad\le e^{-\exp\{(1/2)n\theta\}} E\, e^{\Sigma_{x^n \in \mathcal{T}^n_X} \delta(y^n, z^n), U(x^n))} \\
&\quad\overset{6)}{=} e^{-\exp\{(1/2)n\theta\}} \mathbb{E}\, e^{\Sigma_{x^n \in \mathcal{T}^n_{X|YZ}(y^n, z^n)} \delta((y^n, z^n), U(x^n))} \\
&\quad\overset{2)}{=} e^{-\exp\{(1/2)n\theta\}} \prod_{x^n \in \mathcal{T}^n_{X|YZ}(y^n, z^n)} \mathbb{E}\, e^{\delta((y^n, z^n), U(x^n))} \\
&\quad= e^{-\exp\{(1/2)n\theta\}} \prod_{x^n \in \mathcal{T}^n_{X|YZ}(y^n, z^n)} \\
&\qquad\cdot \left(\frac{e}{|\mathcal{T}^n_{YZ|X}(x^n)|} + 1 - \frac{1}{|\mathcal{T}_{YZ|X}(x^n)|} \right)
\end{aligned}$$

$$\overset{7)}{=} e^{-\exp\{(1/2)n\theta\}} \left(1 + (e-1)\frac{|\mathcal{T}_X^n|}{|\mathcal{T}_{XYZ}^n|}\right)^{|\mathcal{T}_{X|YZ}^n(y^n,z^n)|}$$
$$\leq e^{-\exp\{(1/2)n\theta\}+(e-1)(|\mathcal{T}_X^n||\mathcal{T}_{X|YZ}^n(y^n,z^n)|/|\mathcal{T}_{XYZ}^n|)}$$
$$\overset{8)}{\leq} \left\{-\exp\left\{\frac{1}{2}n\theta\right\}\log e\right\} \tag{8.12}$$

for n large enough, where the steps are justified as follows:

1) for all $x^n \notin \mathcal{T}_{X|Y}^n(y^n)$
$$(y'^n, z'^n) \in \mathcal{T}_{YZ|X}^n(x^n), \qquad y'^n \neq y^n;$$
2) $U(x^n), x^n \in \mathcal{T}_X^n$, are independent;
3) for all $(x^n, y^n) \in \mathcal{T}_{XZ}^n$
$$|\mathcal{T}_{Z|XY}^n(x^n, y^n)||\mathcal{T}_{XY}^n| = |\mathcal{T}_{XYZ}^n|$$
and for all $x^n \in \mathcal{T}_X^n$
$$|\mathcal{T}_{YZ|X}^n(x^n)||\mathcal{T}_X^n| = |\mathcal{T}_{XYZ}^n|$$
(c.f. (2.2));
4) for all $y^n \in \mathcal{T}_Y^n$
$$|\mathcal{T}_Y^n||\mathcal{T}_{X|Y}^n(y^n)| = |\mathcal{T}_{XY}^n|$$
(c.f. (2.2));
5) by (8.8);
6) for all $x^n \notin \mathcal{T}_{X|YZ}^n(y^n, z^n)$
$$(y'^n, z'^n) \in \mathcal{T}_{YZ|X}^n(x^n), \quad (y'^n, z'^n) \neq (y^n, z^n);$$
7) for all $x^n \in \mathcal{T}_X^n$
$$|\mathcal{T}_{YZ|X}^n(x^n)||\mathcal{T}_X^n| = |\mathcal{T}_{XYZ}^n|$$
(c.f. (2.2));
8) by Section II and the inequalities characterizing this subcase
$$(e-1)\frac{|\mathcal{T}_X^n||\mathcal{T}_{X|YZ}^n(y^n, z^n)|}{|\mathcal{T}_{XYZ}^n|}$$
$$= \exp\{n(H(X) + H(X|YZ) - H(XYZ) + o(1))\}$$
$$= \exp\{n(H(X) - H(YZ) + o(1))\} < 1.$$

Thus (8.11) and (8.12) imply for n large enough

$$\Pr\left(\left[\bigcup_{y^n \in \mathcal{T}_Y^n} E_1(y^n)\right] \cup \left[\bigcup_{(y^n,z^n) \in \mathcal{T}_{YZ}^n} E_2(y^n, z^n)\right]\right) < \frac{1}{2} \tag{8.13}$$

So one can find a mapping $U\colon \mathcal{T}_X^n \to \mathcal{T}_{YZ}^n$, such that

$$U(x^n) \in \mathcal{T}_{YZ|X}^n(x^n) \tag{8.14}$$

and for all $y^n \in \mathcal{T}_Y^n$

$$|\{x^n\colon U(x^n) = (y^n, z^n) \text{ for some } z^n\}| > \frac{1-e^{-1}}{2}\,\frac{|\mathcal{T}_X^n|}{|\mathcal{T}_Y^n|} \tag{8.15}$$

and for all $(y^n, z^n) \in \mathcal{T}_{YZ}^n$

$$|U^{-1}((y^n, z^n))| = |\{x^n\colon U(x^n) = (y^n, z^n)\}| < \exp\{\tfrac{1}{2}n\theta\}. \tag{8.16}$$

Now for each (y^n, z^n) with $U^{-1}((y^n, z^n)) \neq \emptyset$, we take one $x^n \in U^{-1}((y^n, z^n))$ into our $\hat{\mathcal{X}}_n$ and define $U(x^n) = g(x^n)$ for this x^n. Then g is injective and by (8.8), (8.14)–(8.16), the relations (8.5) and (8.7) are satisfied.

Next we move to the second step and apply the Color Carrier Lemma to the hypergraph $\{\mathcal{T}_Z^n, \{E_{y^n}\}_{y^n \in \mathcal{T}_Y^n}\}$, to obtain a coloring $\varphi\colon \mathcal{T}_Z^n \to \mathcal{L} = \{1, \cdots, L\}$ for $L = \lceil \exp\{n(R_Z - 2\theta)\}\rceil$, and n big enough.

Label the elements of $\mathcal{T}_Y^n$ as $v_i, 1 \leq i \leq M = |\mathcal{T}_Y^n|$ (in any order), and for each $1 \leq i \leq M$ and $1 \leq \ell \leq L$ we choose a $z^n \in E_{v_i}$ with $\varphi(z^n) = \ell$ as $v'_{i,\ell}$.

Now the definition of E_{y^n} and injectivity of g allow us to choose $g^{-1}((v_i, v'_{i,\ell}))$, the inverse image of $(v_i, v'_{v,\ell})$ under g as our $u_{i,\ell}$. Finally, set $\mathcal{D}_i(\mathcal{Y}) = \{v_i\}$ for $1 \leq i \leq M$ and $\mathcal{D}_\ell(Z) = \{v_{i,\ell}\colon 1 \leq i \leq M\}$. We get our matching code

$$\{(u_{i,\ell}\mathcal{D}_i(\mathcal{Y}), \mathcal{D}_\ell(Z))\colon 1 \leq i \leq M, 1 \leq \ell \leq L\}$$

with rate $(R_{\mathcal{Y}} + o(1), R_{\mathcal{Z}} - 2\theta)$ (by (8.8)).

IX. The Controller Falls Asleep—On Matching Zero-Error Detection Codes

We consider now again the one-way deterministic matching channel $\mathcal{W}_0$. Now the communicators, sender and receiver, safeguard against mistakes of the controller and even against malicious operations (jamming) by using *matching zero-error detection codes* (MDC) $\{(u_i, v_i)\colon 1 \leq i \leq M\}$, which satisfy for some $s^n(i)(1 \leq i \leq M)$

$$W^n(v_i|u_i|s^n(i)) = 1 \tag{9.1}$$

and

$$W^n(v_j|u_i|s^n) = 0, \qquad \text{for } i \neq j \text{ and all } s^n \in \mathcal{S}^n. \tag{9.2}$$

$C_{m\ de}(\mathcal{W}_0)$ denotes the capacity of this channel.

Let $W(\cdot|\cdot)$ be as in (1.4) of Section I-B, associated with $\mathcal{W}_0$. Then (9.1) takes the form

$$W(v_i|u_i) > 0, \qquad \text{for } i = 1, 2, \cdots, M \tag{9.3}$$

and (9.2) takes the form

$$W(v_j|u_i) = 0, \qquad \text{for } i \neq j. \tag{9.4}$$

These conditions are quite similar to those defining a zero-error detection code $\{u_i\colon 1 \leq i \leq M\}$

$$\begin{aligned} W(u_i|u_i) &> 0, &&\text{for } 1 \leq i \leq M\\ W(u_j|u_i) &= 0, &&\text{for } i \neq j. \end{aligned} \tag{9.5}$$

In [9] its capacity was denoted by C_{de}. In another terminology this is called Sperner capacity [11].

Remark:

7) Just formally, one can skip the condition $W^n(u_i|u_i) > 0$ and arrive at another mathematically meaningful notion: zero-error pseudodetection codes. Related concepts can be found in Appendix II.

The connection is the following. Relevant for (9.5) are rows and columns of W with indices x,y in $\mathcal{X} \cap \mathcal{Y}$ and $W(y|x) > 0$, if $x = y$. This gives a square matrix $\tilde{W}$ as restriction of W to alphabets $\tilde{\mathcal{X}} = \tilde{\mathcal{Y}} \subset \mathcal{X} \cap \mathcal{Y}$, for which (9.5) can equivalently be stated. This matrix $\tilde{W}$ corresponds to a directed graph $\mathcal{G} = (\mathcal{V}, \vec{\mathcal{E}}$ with

$$\mathcal{V} = \tilde{\mathcal{X}} = \tilde{\mathcal{Y}} \text{ and } (v,v') \in \mathcal{E} \Leftrightarrow W(v'|v) > 0. \tag{9.6}$$

In the directed product graph $\mathcal{G}^n(\mathcal{V}^n, \vec{\mathcal{E}}^n)$ we have

$$(v^n, v'^n) \in \vec{\mathcal{E}}^n \Leftrightarrow W(v'^n|v^n) > 0$$

and therefore (9.5) is equivalent to an independent set in $\mathcal{Y}^n$. The rate of the independence numbers equals C_{de}. Notice also that all loops are included in $\mathcal{G}^n$.

Remark:

8) In [11], zero-error detection codes are described in terms of the dual graph $(\mathcal{G}^n)^* = (\mathcal{V}^n, (\vec{\mathcal{E}}^n)^*)$, which contains exactly the directed edges which are not in $\vec{\mathcal{E}}^n$. Then for any u_i, u_j in the code, there exists t with $(u_{i_t}, u_{j_t}) \in \vec{\mathcal{E}}^*$.

Now we discuss MDC (see (9.3) and (9.4)). For matrix W consider support sets

$$\mathcal{Y}_W(x) = \{y \in \mathcal{Y}\colon W(y|x) > 0\} \tag{9.7}$$

and define an associated directed graph $\mathcal{G}(W) = (\mathcal{X}, \mathcal{E}(W))$ by

$$(x,x') \in \mathcal{E}(W) \Leftrightarrow \mathcal{Y}_W(x) \supset \mathcal{Y}_W(x'). \tag{9.8}$$

So all loops are included and the MDC $\{(u_i^n, v_i^n)\colon 1 \le i \le M\}$ have the property: $\{u_i^n\colon 1 \le i \le M\}$ is an independent set in the directed product graph $\mathcal{G}^n(W) = \mathcal{G}(W^n)$. The converse is not true.

By (9.8), the associated graph has no directed cycles (if $\mathcal{Y}_W(x) \neq \mathcal{Y}_W(x')$ for $x \neq x'$) and, therefore, the class of these graphs is, again by (9.8), smaller than the class of graphs associated via (9.6).

Denote by $M^n_{m\,de}(W), M^n_{de}(W)$ and $M^n_0(W)$ the largest sizes of n-length MDC, zero-error detection codes, and zero-error codes for W^n, respectively. When $n = 1$ we write them as $M_{m\,de}(W), M_{de}(W)$, and $M_0(W)$.

Example 3: For

$$W_1 = \begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline 0 & + & + & 0 \\ 1 & 0 & + & + \\ 2 & + & 0 & + \end{array}$$

we have the independence number $I(\mathcal{G}(W_1)) = 3$, but $M_{m\,de}(W_1) = 2$. (Here $M_{de}(W_1) = 1$ and for zero-error codes $M_0(W_1) = 1$). One can identify x's with equal supports without loss in code length.

Example 4: For

$$W_2 = \begin{pmatrix} + & + & 0 & 0 & 0 \\ 0 & + & + & 0 & 0 \\ 0 & 0 & + & + & 0 \\ 0 & 0 & 0 & + & + \\ + & 0 & 0 & 0 & + \end{pmatrix}$$

$$M_0(W_2) = M_{de} = 2$$

and

$$M_{m\,de}(W_2) = |\{(1,1),(2,3),(4,4)\}| = 3.$$

However, by considering all matrices, we can show that the matching zero-error detection coding problem is a proper special case of the zero-error detection coding problem.

We discuss it in Appendix I.

X. The Matching Zero-Error Detection Capacity $C_{m\,de}$ in a Genuine Example

We introduce the $\binom{\alpha}{\beta}$-uniform complete hypergraph channel $W_{\alpha,\beta}$ as a channel with input alphabet $\mathcal{X} = \{1, 2, \cdots, \alpha\}$, output alphabet $\mathcal{Y} = \binom{\mathcal{X}}{\beta}$, that is, the letters of $\mathcal{Y}$ are the β-element subsets of $\mathcal{X}$, and for $x \in \mathcal{X}, E \in \mathcal{Y}$

$$W_{\alpha,\beta}(E|x) > 0 \Leftrightarrow x \in E. \tag{10.1}$$

Therefore, an MDC is a system $\{(u_i^n, E_i^n)\colon 1 \le i \le M\}$ with $u_i^n \in \mathcal{X}^n, E_i^n \in \binom{\mathcal{X}}{\beta}^n$, and $u_i^n \in E_j^n$ exactly if $i = j$. Its maximal size is $M_{\alpha,\beta}(n)$ and its capacity is

$$C_{\alpha,\beta} = \lim_{n \infty} \frac{1}{n} \log M_{\alpha,\beta}(n) \tag{10.2}$$

because MDC's can be concatenated. We are going to determine $M_{\alpha,\beta}(n)$ with an elegant method used by Blokhuis in [5]. Its main idea is that all polynomials in indeterminates $\xi_1, \cdots, \xi_n$ over any field, with degree $(\xi_i) \le d_i$ for $i = 1, \cdots, n$, form a linear space $\mathcal{L}(d_1, \cdots, d_n)$ of a dimension

$$\dim(\mathcal{L}(d_1, \cdots, d_n)) = \prod_{i=1}^{n} (d_i + 1). \tag{10.3}$$

Theorem 10: For $\binom{\alpha}{\beta}$-channels

$$M_{\alpha,\beta}(n) = (\alpha - \beta + 1)^n$$

and

$$C_{\alpha,\beta} = \log(\alpha - \beta + 1).$$

Proof: Since we can take products of codes, it suffices to show that $M_{\alpha,\beta}(1) \ge \alpha - \beta + 1$ in order to establish the lower bound. Define for $i = 1, 2, \cdots, \alpha - \beta + 1$

$$u_i = i \quad \text{and} \quad E_i = \{i, \alpha - \beta + 2, \cdots, \alpha\}. \tag{10.4}$$

Then $\{(u_i, E_i)\colon 1 \le i \le \alpha - \beta + 1\}$ is an MDC for the $\binom{\alpha}{\beta}$-channel. Conversely, let now $\{(u_i^n, E_i^n){:}1 \le i \le M\}$ be an MDC for the $\binom{\alpha}{\beta}$-channel. Consider the polynomials in $X^n = (X_1, \cdots, X_n)$

$$f_i(\xi^n) = \prod_{t=1}^{n} \prod_{x \notin E_{it}} (\xi_t - x), \qquad 1 \le i \le M, \tag{10.5}$$

where $(E_{i1}, \cdots, E_{in}) = E_i^n$.

Clearly, $f_i \in \mathcal{L}\ (\alpha - \beta, \cdots, \alpha - \beta)$ and for all i, i'

$$f_i(u_{i'}^n) \neq 0 \Leftrightarrow u_{i'}^n \in E_i^n \Leftrightarrow i = i'. \tag{10.6}$$

Therefore, $f_1, \cdots, f_n$ are linearly independent and hence

$$M \le \dim(\mathcal{L}(\alpha - \beta, \cdots, \alpha - \beta)) = (\alpha - \beta + 1)^n.$$

Remark 9

9) Comparison with [5] shows that our result is slightly stronger, because the algebraic method is better exploited.

XI. Feedback and Also Randomization Increase the Capacity of the Matching Channel

We have already seen in Section I that in the formula for $C(\mathcal{W})$ in Theorem 1 $H(X)$ enters only, because words from $\mathcal{X}^n$ can be used at most once as a codeword (in conjunction with an s^n) and if we drop this restriction, then we achieve the capacity of the DMC W': $\mathcal{X} \times \mathcal{S} \to \mathcal{Y}$

$$C = \max_{P_{XS}} I(XS \wedge Y). \tag{11.1}$$

This leaves room to enlarge $C(\mathcal{W})$ in the following two ways.

α) Suppose now that the controller knows in case of feedback the *encoding function for a message like earlier the codeword for a message*, then the C in (11.1) can still be achieved and is the capacity!

Clearly, C cannot be superceded, because feedback does not increase the capacity of a DMC. For the direct part, let us start with a set of codewords $\mathcal{U} = \{u_1, \cdots, u_M\}$ and decoding sets $\{\mathcal{D}_1, \cdots, \mathcal{D}_M\}$ for W: $\mathcal{X} \times \mathcal{S} \to \mathcal{Y}$, where

$$u_i = (x_i^n, s_i^n), \qquad 1 \le i \le M. \tag{11.2}$$

Let $\mathcal{S}(x^n)$ be the s^n's appearing with x^n. This set grows at most exponentially in n. Its elements serve as "first names" for x^n.

Recalling that encoding with feedback is described by a family of functions $\{f_i^n\}_{i=1}^M$ with $f_i^n = [f_{i1}, \cdots, f_{in}]$ and f_{it}: $\mathcal{Y}^{t-1} \to \mathcal{X}$, we use now a simple trick.

We add m positions before the word of length n. For these positions there are

$$|\mathcal{X}||\mathcal{X}|^{|\mathcal{Y}|}|\mathcal{X}|^{|\mathcal{Y}|^2} \cdots |\mathcal{X}|^{|\mathcal{Y}|^m}$$

many encoding functions g^m, which can be used as the beginnings of the encoding functions of length $n + m$. Since there are double exponentially in m many g^m's, they can be used as "nicknames" for the first names.

The controller knows by assumption the nicknames and therefore the first name s_i^n in (x_i^n, s_i^n), which he then uses in the set of positions $\{m+1, \cdots, m+n\}$. Obviously m's contribution to the loss in rate is negligible.

We summarize our findings for the matching capacity in case of feedback.

Proposition 1: $C_f(\mathcal{W}) = \max_{P_{XS}} I(XS \wedge Y)$.

β) Now we have no feedback, but the encoder can use randomization, that is, he encodes message i as $Q_i \in \mathcal{P}(\mathcal{X}^n)$. Q_i is known to the controller before he chooses s_i^n. Then again the capacity of the matching channel equals C. Again we use the previous idea, which can now be realized even simpler. We just use one additional position $(m = 1)$. This gives infinitely many nicknames, because $|\mathcal{P}(\mathcal{X})| = \infty$.

Formally, we choose an n-length code $\{(u_i, \mathcal{D}_i): 1 \le i \le M\}$ as in α) and M different distributions P_i on $\mathcal{P}(\mathcal{X})$. For message i the sender chooses and sends an $n+1$-length codeword xx_i^n, if this is the outcome of random experiment $(Q_i, \mathcal{X}^{n+1})$, where $Q_i(xx_i^n) = P_i(x)$. The controller, who knows i, chooses s_i^n. Finally, the sets $\{\mathcal{X} \times \mathcal{D}_1, \cdots, \mathcal{X} \times \mathcal{D}_M\}$ are used as decoding sets.

XII. The Capacity for Matching Zero-Error Detection Codes with Feedback (MDCF) for $\mathcal{W}_0$

A matching code with feedback $\{(f_i^n, \mathcal{D}_i): 1 \le i \le M\}$ for $\mathcal{W}_0$ or its associated W (in the sense of (1.4)) is specified by

$$f_i^n = [f_{i1}, \cdots, f_{in}], f_{it}: \mathcal{Y}^{t-1} \to \mathcal{X} \tag{12.1}$$

and for all $i \in \mathcal{M} = \{1, 2, \cdots, M\}$ exists a $v_i \in \mathcal{D}_i$ with

$$W(v_{i1}|f_{i1}) \prod_{t=2}^{n} W(v_{it}|f_{it}(v_{i1}, \cdots, v_{i(t-1)})) > 0 \tag{12.2}$$

$$v_i \ne v_{i'} (i \ne i'). \tag{12.3}$$

This code is zero-error detecting, if for $i \ne i'$ exists a $t \ge 1$ with $W(v_{it}|f_{i't}(v_{i1}, \cdots, v_{i(t-1)})) = 0$. Obviously, we can assume that $\mathcal{D}_i$ contains only one element.

We can assume again that W is nontrivial: not all row supports are identical and for all $y \in \mathcal{Y}$ $W(y|x) > 0$ for some $x \in \mathcal{X}$. Further, a necessary condition for $C_{m\,de\,f}$ to be positive is the existence of two input letters x_1, x_2 and an output letter y_0 with

$$W(y_0|x_1) > 0 \quad \text{and} \quad W(y_0|x_2) = 0. \tag{12.4}$$

Proposition 2:

$$C_{m\,de\,f} \begin{cases} \ge C(\mathcal{W}_0), & \text{if (12.4) holds} \\ = 0, & \text{otherwise.} \end{cases}$$

Proof: We can assume that (12.4) holds. We start with a matching code $\{(u_i^n, v_i^n): 1 \le i \le M\}$ and define a feedback encoding $f_i^{n+1} (1 \le i \le M)$ as follows:

For $u_i^n = (u_{i1}, \cdots, u_{in})$

$$f_{it}(y^{t-1}) = u_{it}, \quad \text{for } t = 1, 2, \cdots, n \text{ and all } y^n \tag{12.5}$$

$$f_{i\,n+1}(y^n) = \begin{cases} x_1, & \text{if } y^n = v_i^n \\ x_2, & \text{otherwise.} \end{cases} \tag{12.6}$$

Define now for all i

$$v_i^{n+1} = v_i^n y_0 \tag{12.7}$$

and verify that

$$\{(f_i^{n+1}, v_i^{n+1}): 1 \le i \le M\}$$

where $f_i^{n+1} = (f_{i1}, \cdots, f_{in}, f_{i\,n+1})$ is by (12.4) a matching zero-error detection code with feedback.

Theroem 3:

$$C_{m\,de\,f} = \begin{cases} C_f(\mathcal{W}_0) = \max_{P_{XS}} I(XS \wedge Y), & \text{if (12.4) holds} \\ \text{or} & \\ 0, & \text{otherwise.} \end{cases}$$

Proof: Clearly, if (12.4) does not hold, then the nontrivial matrix has only positive entries and therefore $C_{m\,de\,f} = 0$.

Now, if (12.4) holds, then we start with a matching feedback code $\{(f_i^n, v_i^n): 1 \le i \le M\}$ and extend these encoding functions as in the proof of Proposition 2

$$f_{i\,n+1}(y^n) = \begin{cases} x_1, & \text{if } y^n = v_i^n \\ x_0, & \text{otherwise.} \end{cases}$$

Correspondingly, we set $v_i^{n+1} = (v_i^n y_0)$. Finally, we apply Proposition 1.

XIII. Identification for Matching Channels

For our DMC $W\colon \mathcal{X} \times \mathcal{S} \to \mathcal{Y}$ there are several models of identification with randomized encoding. Fix any code $\{(Q_i, \mathcal{D}_i)\colon 1 \le i \le N\}$. The Q_i's and $\mathcal{D}_i$'s are assumed to be different. One can consider the following performance criteria:

$$\text{I)}\quad \sum_{x^n} Q_i(x^n) W^n(\mathcal{D}_i|x^n) > 1 - \lambda, \qquad \text{for all } i$$

$$\text{II)}\quad \sum_{x^n} Q_i(x^n) W^n(\mathcal{D}_i|x^n) = 1, \qquad \text{for all } i$$

$$\text{III)}\quad \sum_{x^n} Q_i(x^n) W^n(\mathcal{D}_i|x^n) > 0, \qquad \text{for all } i$$

$$\text{a)}\quad \sum_{x^n} Q_j(x^n) W^n(\mathcal{D}_i|x^n) < \lambda, \qquad \text{for } j \ne i$$

$$\text{b)}\quad \sum_{x^n} Q_j(x^n) W^n(\mathcal{D}_i|x^n) = 0, \qquad \text{for } j \ne i$$

$$\text{c)}\quad \sum_{x^n} Q_j(x^n) W^n(\mathcal{D}_i|x^n) < 1, \qquad \text{for } j \ne i.$$

The classical work [12] concerns the combination (I, a) and in [9] the combination (II, a) was analyzed.

We settle (III, a), *the matching identification problem*, and also (III, c), in Theorem 11. Furthermore, the capacities for the cases (II, c) and (I, c) are characterized in Theorems 12 and 13, respectively. Finally, condition b). implies that $\mathcal{D}_i$ can be replaced by $\mathcal{D}_i \smallsetminus \bigcup_{j \ne i} \mathcal{D}_j$ and so we have here disjoint decoding sets: Actually, identification is reduced to transmission. (II, b) gives Shannon's classical zero-error capacity problem. (I, b) gives the erasure problem (see [9]) for transmission, and (III, b), reduces to the matching zero-error detection problem discussed in Sections IX and X. The discussion of all cases is complete.

Remark:

10) Some combinations are meaningful also in case of feedback.

We consider first the cases (III, a) and (III, c). Their capacities are denoted by $C(>0, <\lambda)$ and $C(>0, <1)$.

Theorem 11: We have for the second-order identification capacities

$$C(>0, <\lambda) = C(>0, <1) = \begin{cases} \log|\mathcal{Y}|, & \text{if } W \text{ is nontrivial} \\ 0, & \text{otherwise.} \end{cases}$$

W is nontrivial, if not all rows are identical and no column has only 0's as entries.

Proof: Let $\overline{Y}$ take values in $\mathcal{Y}$ according to the uniform distribution and let X take values in $\mathcal{X}$ such that for some small $\delta \in (0,1)$ and $P_Y = P_X W$

$$\|P_Y - P_{\overline{Y}}\| \ge \delta. \tag{13.1}$$

This is possible, because W is nontrivial.

Let

$$Q(x^n) = \begin{cases} \dfrac{n-1}{n} \cdot \dfrac{1}{|\mathcal{T}_X^n|}, & \text{if } x^n \in \mathcal{T}_X^n \\ \dfrac{1}{n} \dfrac{1}{|\mathcal{X}^n| - |\mathcal{T}_X^n|}, & \text{otherwise.} \end{cases} \tag{13.2}$$

Then

$$\sum_{x^n \subset \mathcal{T}_X^n} Q(x^n) W^n(y^n|x^n) > 0, \qquad \text{for all } y^n \in \mathcal{Y}^n$$

and for every $\mathcal{D} \subset \mathcal{T}_{\overline{Y}}^n$, by (13.1) (and the properties of typical and generated sequences stated in Section II)

$$\sum_{x^n} Q(x^n) W^n(\mathcal{D}|x^n) < \varepsilon_n(\delta)$$

and $\lim_{n \to \infty} \varepsilon_n(\delta) = 0$.

Now just choose the $2^{|\mathcal{T}_{\overline{Y}}^n|} - 1$ nonempty $\mathcal{D}$'s as decoding sets and choose the same number of arbitrary Q's on $\mathcal{T}_X^n$ as corresponding encoding distributions. Since

$$\log \log 2^{|\mathcal{T}_{\overline{Y}}^n|} \sim n \log |\mathcal{Y}|$$

the proof of the direct parts is complete. The converses follow from the fact that there are at most $2^{|\mathcal{Y}^n|}$ decoding sets.

We are left with the case (II, c) and its capacity $C(=1, <1)$.

Theorem 12: The second-order capacity $C(=1, <1)$ equals the first order $C_{m\,de}$.

Proof of Direct Part: Start with an MDC $\{(u_i, v_i)\colon 1 \le i \le M\}$

$$W^n(v_i|u_i) > 0 \text{ and } W^n(v_j|u_i) = 0 \text{ for all } i \ne j.$$

Define the supports

$$\mathcal{Y}^n(u) = \{v \in \mathcal{Y}^n\colon W^n(v|u) > 0\}, \qquad u \in \mathcal{X}^n \tag{13.3}$$

and notice that for any

$$\mathcal{U} = \{u_i\colon 1 \le i \le M\} \subset \mathcal{X}^n$$

$\{(u_i, v_i)\colon 1 \le i \le M\}$ is a set of codewords for some v_i if and only if

$$\mathcal{Y}^n(u) \not\subset \bigcup_{u' \subset \mathcal{U} \smallsetminus \{u\}} \mathcal{Y}^n(u'). \tag{13.4}$$

Let M be even and consider the $\frac{M}{2}$-element subsets of $\mathcal{U}$

$$\binom{\mathcal{U}}{\frac{M}{2}} = \left\{ A_\ell\colon 1 \le \ell \le \binom{M}{\frac{M}{2}} \right\}. \tag{13.5}$$

Let Q_ℓ be any distribution with support A_ℓ and define

$$\mathcal{D}_\ell = \bigcup_{u \in A_\ell} \mathcal{Y}^n(u). \tag{13.6}$$

Then obviously

$$\sum_{x^n} Q_\ell(x^n) W(\mathcal{D}_\ell|x^n) = 1, \qquad \text{for } 1 \le \ell \le \binom{M}{\frac{M}{2}} \tag{13.7}$$

and, moreover, for all $\ell \ne \ell'$

$$A_\ell \smallsetminus A_{\ell'} \ne \varnothing$$

and so by (13.4) and (13.6) for $u_j \in A_\ell \smallsetminus A_{\ell'}$

$$\mathcal{Y}^n(u_j) \not\subset \bigcup_{i \ne j} \mathcal{Y}^n(u_i) \supset \bigcup_{u \in A'_\ell} \mathcal{Y}^n(u) = \mathcal{D}'_\ell. \tag{13.8}$$

Since $Q_\ell(u_j) > 0$, we have finally that

$$\sum_{u \in A_\ell} Q_\ell(u) W^n(\mathcal{D}'_\ell | u) < 1.$$

Converse Part: Let $\{(Q_i, \mathcal{D}_i)\colon 1 \le i \le M\}$ be a (II, c) identification code. Define the support set of Q_i

$$S_i = \{u \in \mathcal{X}^n\colon Q_i(u) > 0\}.$$

Without loss of generality

$$\mathcal{D}_i = \bigcup_{u \in S_i} \mathcal{Y}^n(u) \tag{13.9}$$

and by (II, c)

$$\mathcal{D}_i \not\subset \mathcal{D}_j, \qquad \text{for } i \ne j. \tag{13.10}$$

Let now $S'_i \subset S_i$ be minimal with the properties

$$\bigcup_{u \in S'_i} \mathcal{Y}^n(u) = \bigcup_{u \in S_i} \mathcal{Y}^n(u) = \mathcal{D}_i. \tag{13.11}$$

The minimality property implies that for all i S'_i satisfies (13.4) and therefore corresponds to an MDC. Consequently,

$$|S'_i| \le \exp\{n\, C_{mde}\}. \tag{13.12}$$

Furthermore, the sets S'_i are different, because the $\mathcal{D}_i$'s are different and (13.11) holds. Therefore,

$$M \le |\mathcal{X}|^{n \cdot \exp\{n\, C_{mde}\}}$$

and, finally,

$$\frac{1}{n} \log \log M \le C_{mde} + \frac{1}{n} \log n + \frac{1}{n} \log \log |\mathcal{X}|.$$

Finally, we consider the combination (I, c) and determine its second-order identification capacity $C(> 1 - \lambda, < 1)$. We call W degenerate iff for some $y_0 \in \mathcal{Y}$

$$W(y_0|x) = 1, \qquad \text{for all } x \in \mathcal{X}. \tag{13.13}$$

Theorem 13:

$$C(> 1 - \lambda, < 1) = \begin{cases} \log|\mathcal{Y}|, & \text{if } W \text{ is nondegenerate and } \mathcal{X}| >| 1 \\ 0, & \text{if } W \text{ is degenerate or } |\mathcal{X}| = 1. \end{cases}$$

Proof: For a degenerate W we have for all $\mathcal{D} \subset \mathcal{Y}^n$ and all $x^n \in \mathcal{X}^n$

$$W^n(\mathcal{D}|x^n) = \begin{cases} 1, & \text{if } (y_0, \cdots, y_0) \in \mathcal{D} \\ 0, & \text{if } (y_0, \cdots, y_0) \notin \mathcal{D} \end{cases}$$

and so cannot have two decoding sets with the required conditions. For $|\mathcal{X}| = 1$ there is only one input distribution.

For nondegenerate W and $|\mathcal{X}| > 1$ it is easy to see that there are distinct $x_1, x_2 \in \mathcal{X}$ and $y_0 \in \mathcal{Y}$ with

$$W(y_0|x_1) > 0 \quad \text{and} \quad W(y_0|x_2) < 1. \tag{13.14}$$

Indeed, if there is no y_0 satisfying (13.14), for all y, $W(y|x_1) > 0$ implies that $W(y|x_2) = 1$. However, because there is at most one $y \in \mathcal{Y}$ with $W(y|x_2) = 1$, this means that for some y $W(y|x_1) = W(y|x_2) = 1$.

Now define

$$\begin{aligned} u_1 &= (x_1, \cdots, x_1) \\ u_2 &= (x_2, \cdots, x_2) \in \mathcal{X}^n \\ v_0 &= (y_0, \cdots, y_0) \in \mathcal{Y}^n \end{aligned}$$

and let n be so large that

$$W^n(v_0|u_2) < \frac{\lambda}{4} \tag{13.15}$$

and

$$W^n(\mathcal{T}^n_{W,\delta}(u_2)|u_2) > 1 - \frac{\lambda}{4} \tag{13.16}$$

(in the terminology of Section II). Set

$$\mathcal{E} = \mathcal{T}^n_{W,\delta}(u_2) \setminus \{v_0\}, \mathcal{F} = \mathcal{Y}^n \setminus (\{v_0\} \cup \mathcal{T}^n_{W,\delta}(u_2)).$$

Now, $|\mathcal{T}^n_{W,\delta}(u_2)|$ is much smaller than $|\mathcal{Y}|^n$ and $\mathcal{F}$ has the same rate as $|\mathcal{Y}|^n$, that is, for $\mu \in (0,1)$

$$\log|\mathcal{F}| > n(\log|\mathcal{Y}| - \mu) \tag{13.17}$$

for n large enough.

Now we choose as decoding sets the family

$$\{\mathcal{D}_i\colon \mathcal{D}_i = \mathcal{E} \cup \mathcal{F}'_i, \mathcal{F}'_i \subset \mathcal{F}, 1 \le i \le 2^{|\mathcal{F}|}\}.$$

It has cardinality $2^{|\mathcal{F}|}$ and $\log\log 2^{|\mathcal{F}|} > n(\log|\mathcal{Y}| - \mu)$. Finally, define Q_i with support $\{u_1, u_2\}$ and

$$0 < Q_i(u_1) < \frac{\lambda}{2}. \tag{13.18}$$

Since $W^n(v_0|u_1) > 0$ we get for all i, j

$$\sum_{x^n} Q_j(x^n) W^n(\mathcal{D}_i|x^n) < 1$$

and (13.16) and (13.18) imply

$$\sum_{x^n} Q_i(x^n) W^n(\mathcal{D}_i|x^n) > \left(1 - \frac{\lambda}{2}\right)^2 > 1 - \lambda$$

for all i.

The converse obviously holds.

Appendix I

Lemma AI.1: An MDC $\{(u_i, v_i)\colon 1 \le i \le M\}$ for $W^n\colon \mathcal{X}^n \to \mathcal{Y}^n (n \ge 1)$ can be viewed as zero-error detection code for V^n, where V has input and output alphabet

$$\mathcal{Z} = \{(x, y)\colon x \in \mathcal{X}, y \in \mathcal{Y}, W(y|x) > 0\} \tag{AI.1}$$

and

$$V((x', y')|(x, y)) > 0 \Leftrightarrow W(y'|x) > 0. \tag{AI.2}$$

Conversely, for every zero-error detection code

$$\{((u_{i1}, v_{i1}), \cdots, (u_{in}, v_{in}))\colon 1 \le i \le M\}$$

for V^n, $\{(u_i^n, v_i^n)\colon 1 \le i \le M\}$ is an MDC for W^n.

Proof: These are immediate consequences of the definitions.

Next we identify MDC problems as a certain class of zero-error detection problems. For this we procede as follows:

a) We observe that, while considering MDC for W^n, letters $x,\ x' \in \mathcal{X}$ with equal row support sets, that is,

$$\mathcal{Y}_W(x) = \mathcal{Y}_W(x') \quad \text{(defined in (9.7))} \tag{AI.3}$$

can be contracted to one letter.

b) We also observe that letters $y,\ y' \in \mathcal{Y}$ with equal columns support sets

$$\mathcal{X}_W(y) \triangleq \{x \in \mathcal{X}\colon W(y|x) > 0\} = \mathcal{X}_W(y') \triangleq \{x \in \mathcal{X}\colon W(y'|x) > 0\}$$

can be contracted to one.

We call a matrix W *irreducible*, if no contractions as decribed in a) and b) are possible.

It is clear from the definitions of $\mathcal{Z}, V$ in (I.1) and (I.2) that

$$V(z|z) > 0, \qquad \text{for all } z \in \mathcal{Z}. \tag{AI.4}$$

It is also clear that row supports

$$\mathcal{Z}_{V,r}(z) = \{z'\colon V(z'|z) > 0\} \tag{AI.5}$$

and columns supports

$$\mathcal{Z}_{V,c}(z) = \{z'\colon V(z|z') > 0\} \tag{AI.6}$$

being equal simultaneously means that the corresponding letters can be contracted for zero-error detection codes for V^n. This leads to the notion of an irreducible V. Moreover, one readily verifies the next fact.

Lemma AI.2: For an irreducible W the corresponding V is also irreducible and conversely.

Next we characterize those matrices V which can be obtained via (AI.1), (AI.2) from some irreducible W and so for the maximal code sizes

$$M^n_{m\,de}(W) = M^n_{de}(V), \qquad n \geq 1. \tag{AI.7}$$

Using this characterisation (Theorem AI.1 below) we then show by Example AI.1, that matching zero-error detection coding is indeed more special than zero-error detection coding.

A few more definitions are needed. Set

$$A_x = \{x\} \times \mathcal{Y}_W(x) = \{(x,y)\colon y \in \mathcal{Y}, (x,y) \in \mathcal{Z}\} \tag{AI.8}$$

$$B_y = \mathcal{X}_W(y) \times \{y\} = \{(x,y)\colon x \in \mathcal{X}, (x,y) \in \mathcal{Z}\}. \tag{AI.9}$$

Clearly, $(A_x)_{x\in\mathcal{X}}$ and $(B_y)_{y\in\mathcal{Y}}$ are both partitions of $\mathcal{Z}$ and

$$a, a' \in A_x \Rightarrow \mathcal{Z}_{V,r}(a) = \mathcal{Z}_{V,r}(a') \tag{AI.10}$$

$$b, b' \in B_y \Rightarrow \mathcal{Z}_{V,c}(b) = \mathcal{Z}_{V,c}(b'). \tag{AI.11}$$

On the other hand, every $V\colon \mathcal{Z} \to \mathcal{Z}$ (not necessary generated by (AI.1) and (AI.2)), whose row partition $(A_x)_{x\in\mathcal{X}}$ and column partition $(B_y)_{y\in\mathcal{Y}}$ satisfies (AI.10), and (AI.11) (where $\mathcal{X}$ and $\mathcal{Y}$ serve only as index sets) has the following properties.

It partitions for all $x \in \mathcal{X}$ the set $\mathcal{Y}$ into $\mathcal{Y}^*(x)$ and $\mathcal{Y}^{*c}(x) = \mathcal{Y} \smallsetminus \mathcal{Y}^*(x)$ such that for all $a \in A_x$

$$\mathcal{Z}_{V,r}(a) = \bigcup_{y\in\mathcal{Y}^*(x)} B_y$$
$$\mathcal{Z}_{V,r}(a) \cap B_y = \varnothing, \qquad \text{if } y \in \mathcal{Y}^{*c}(x). \tag{AI.12}$$

Similarly, it partitions, for all $y \in \mathcal{Y}, \mathcal{X}$ into $\mathcal{X}^*(y)$ and $\mathcal{X}^{*c}(y) = \mathcal{X} \smallsetminus \mathcal{X}^*(y)$ such that for all $b \in B_y$

$$\mathcal{Z}_{V,c}(b) = \bigcup_{x\subset\mathcal{X}^*(y)} A_x$$
$$\mathcal{Z}_{V,c}(b) \cap A_x = \varnothing, \qquad \text{if } x \in \mathcal{X}^{*c}(y). \tag{AI.13}$$

Theorem AI.1: For $W\colon \mathcal{X} \to \mathcal{Y}$ and corresponding $\mathcal{Z}, V$ (defined by (AI.1) and (AI.2))

i) the identity (AI.7) holds;

ii) (AI.4) and the identities

$$|A_x| + |\mathcal{Y}^{*c}(x)| = |\mathcal{Y}|, \qquad \text{for all } x \in \mathcal{X} \tag{AI.14}$$

$$|B_y| + |\mathcal{X}^{*c}(y)| = |\mathcal{X}|, \qquad \text{for all } y \in \mathcal{Y} \tag{AI.15}$$

hold, if $A_x, B_y, \mathcal{Y}^*(x), \mathcal{Y}^{*c}(x), \mathcal{X}^*(y)$, and $\mathcal{X}^{*c}(y)$ are defined as in (AI.8), (AI.9), (AI.12), and (AI.13)

iii) Conversely, any irreducible matrix $V\colon \mathcal{Z} \to \mathcal{Z}$ satisfying (AI.4) corresponds to an irreducible matrix W so that (AI.7) holds, if there are partitions $(A_x)_{x\subset\mathcal{X}}$ and $(B_y)_{y\subset\mathcal{Y}}$ for some index sets $\mathcal{X}$ and $\mathcal{Y}$, such that (AI.10), (AI.11), and (AI.14) hold for $\mathcal{Y}^{*c}(x)$ defined by (AI.12). Symmetrically, here (AI.14) can be replaced by (AI.15)

Proof:

i) Here Lemma AI.1 is just restated.

ii) By (AI.12) and (AI.13)

$$\mathcal{Y}^{*c}(x) = \{y\colon W(y|x) = 0\}, \qquad \text{for all } x \in \mathcal{X}$$

and

$$\mathcal{X}^{*c}(y) = \{x\colon W(y|x) = 0\}, \qquad \text{for all } y \in \mathcal{Y}.$$

Thus (AI.14) and (AI.15) follow from (AI.8) and (AI.9), respectively.

iii) Define $W\colon \mathcal{X} \to \mathcal{Y}$ as follows:

$$\text{For all } x \in \mathcal{X}, y \in \mathcal{Y}\colon W(y|x) > 0, \qquad \text{iff } y \in \mathcal{Y}^*(x). \tag{AI.16}$$

In order to see that $\mathcal{Z}$ is generated by W via (AI.1), we have to show that there is a bijection from

$$\{(x,y)\colon w(y|x) > 0\} = \bigcup_{x\in\mathcal{X}} \{x\} \times \mathcal{Y}^*(x)$$

to $\mathcal{Z}$. To achieve this goal, we first characterize $\mathcal{Y}^*(x)$ and $\mathcal{Y}^{*c}(x)$. Indeed, we shall show that for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$

$$y \in \mathcal{Y}^{*c}(x), \qquad \text{iff } A_x \cap B_y = \varnothing \tag{AI.17}$$

$$y \in \mathcal{Y}^*(x), \qquad \text{iff } |A_x \cap B_y| = 1. \tag{AI.18}$$

We begin with the assumption $y \in \mathcal{Y}^{*c}(x)$. If now $A_x \cap B_y \neq \varnothing$ and $z \in A_x \cap B_y$, then by (AI.4), (AI.10), (AI.11), and the definition of $\mathcal{Z}_{V,r}$, we have for all $a \in A_x$ $B_y \subset \mathcal{Z}_{V,r}(a)$ and therefore by (AI.12) $y \in \mathcal{Y}^*(x)$, a contradiction.

Moreover, since V is irreducible, we also know that

$$|A_x \cap B_y| \leq 1, \qquad \text{for } x \in \mathcal{X}, y \in \mathcal{Y}. \qquad \text{(AI.19)}$$

So we know that

$$|A_x| \leq |\mathcal{Y}^*(x)| = |\mathcal{Y}| - |\mathcal{Y}^{*c}(x)|.$$

The reverse implications in (AI.17) and (AI.18) follow immediately from our assumption (AI.14), which gives also the direct implication in (AI.17).

Now, (AI.18) gives rise to a mapping defined on $\cup_{x \in \mathcal{X}} \{x\} \times \mathcal{Y}^*(x)$ and sending (x,y) to the unique element in $A_x \cap B_y$. Further, by (AI.17), $\{A_x \cap B_y \colon y \in \mathcal{Y}^*(x)\}$ is a partition of $\mathcal{Z}$ and this guarantees the mapping to be bijective. Thus we can rename the elements of $\mathcal{Z}$ by

$$\{(x,y) \colon W(y|x) > 0\} = \cup_{x \subset \mathcal{X}} \{x\} \times \mathcal{Y}^*(x)$$

and rewrite $z \in \mathcal{Z}$ as (x,y), if $z \in A_x \cap B_y$.

So only (AI.2) remains to be verified. For the new names $(x,y), (x',y') \in \mathcal{Z}$

$$\begin{aligned} V((x',y')|(x,y)) > 0 &\overset{1)}{\Longleftrightarrow} A_{x'} \cap B_{y'} \subset \mathcal{Z}_{V,r}((x,y)) \\ &\overset{2)}{\Longleftrightarrow} y' \in \mathcal{Y}^*(x) \\ &\overset{3)}{\Longleftrightarrow} W(y'|x) > 0, \end{aligned}$$

that is, (AI.2).

Here the equivalences are justified as follows.

1) Use definitions of (x',y') and $\mathcal{Z}_{V,r}((x,y))$.
2) Use (AI.12) and that by definition of (x,y) necessarily $(x,y) \in A_x$.
3) Use (AI.16).

Example AI.1: Consider $\mathcal{Z} = \{0,1,2\}, V = W_1$, defined in Example 3.

The only partitions satisfying (AI.10) and (AI.11) are

$$\{A_0, A_1, A_2\} = \{\{0\},\{1\},\{2\}\}$$

and

$$\{B_0, B_1, B_2\} = \{\{0\},\{1\},\{2\}\}.$$

However, since $0 \in A_0$ and $\mathcal{Z}_{V,r}(0) = \{0,1\} = B_0 \cup B_1, \mathcal{Y}^{*c}(0) = \{2\}$ we have $|A_0| + |\mathcal{Y}_0^{*c}| = 2 \neq 3 = |\mathcal{Y}|$. Thus by Theorem AI.1 no W can generate V.

Example AI.2: $W = W_1$, defined in Example 3, generates by (A.1) and (A.2),

$$\mathcal{Z} = \{(0,0),(0,1),(1,1),(1,2),(2,0),(2,2)\}$$

$$V = \begin{array}{c} \\ (0,0) \\ (0,1) \\ (1,1) \\ (1,2) \\ (2,0) \\ (2,2) \end{array} \begin{array}{c} \begin{array}{cccccc} (0,0) & (0,1) & (1,1) & (1,2) & (2,0) & (2,2) \end{array} \\ \begin{pmatrix} + & + & + & 0 & + & 0 \\ + & + & + & 0 & + & 0 \\ 0 & + & + & + & 0 & + \\ 0 & + & + & + & 0 & + \\ + & 0 & 0 & + & + & + \\ + & 0 & 0 & + & + & + \end{pmatrix} \end{array}$$

and

$$\begin{aligned} A_0 &= \{(0,0),(0,1)\} \\ A_1 &= \{(1,1),(1,2)\} \\ A_2 &= \{(2,0),(2,2)\} \\ B_0 &= \{(0,0),(2,0)\} \\ B_1 &= \{(1,1),(1,2)\} \end{aligned}$$

and

$$B_2 = \{(1,2),(2,2)\}.$$

Appendix II

At this moment our only motivation for the code concepts below is mathematical interest in searching for new combinatorial structures and their connections to the zero-error detection and matching zero-error detection codes. We use the abbreviations

$$\begin{aligned} \mathcal{Y}(x) &= \mathcal{Y}_W(x) & \mathcal{X}(y) &= \mathcal{X}_W(y) \\ \mathcal{Y}^n(x^n) &= \prod_{t=1}^{n} \mathcal{Y}(x_t) & \mathcal{X}^n(y^n) &= \prod_{t=1}^{n} \mathcal{X}(y_t). \end{aligned} \qquad \text{(AII.1)}$$

Definition AII.1: A *pairwise* zero-error detection code for W^n is a set $\{u_i \colon 1 \leq i \leq M\} \subset \mathcal{X}^n$ such that for all pairs (i,j) there is a $v_{ij} \in \mathcal{Y}^n$ with

$$W^n(v_{ij}|u_j) = 0 \quad \text{and} \quad W^n(v_{ij}|u_i) > 0. \qquad \text{(AII.2)}$$

For two sets A, B we write

$$A \supset\!\!\subset B, \qquad \text{iff } A \subset B \text{ or } B \subset A$$

and we write

$$A \not\supset\!\!\subset B, \qquad \text{iff } A \not\subset B \text{ and } B \not\subset A. \qquad \text{(AII.3)}$$

In this terminology (AII.2) is equivalent to

$$\mathcal{Y}^n(u_i) \not\supset\!\!\subset \mathcal{Y}^n(u_j), \qquad \text{for all } i \neq j. \qquad \text{(AII.4)}$$

Hence, if we define $U \colon \mathcal{X} \to \mathcal{X}$ by

$$U(x|x') > 0, \qquad \text{iff } \mathcal{Y}(x) \subset \mathcal{Y}(x') \qquad \text{(AII.5)}$$

we get the following characterization.

Lemma AII.1: $\{u_i \colon 1 \leq i \leq M\} \subset \mathcal{X}^n$ is a pairwise zero-error detection code for W^n iff it is a zero-error detection code for U^n.

Of course, every zero-error detection code is also a pairwise zero-error detection code for the same channel.

Since by (AII.5) every U generated by a W must have a positive diagonal, we ask whether arbitrary U with positive diagonal can be generated this way. This is not the case.

Example AII.1: Choose again $U = W_1$ in Example 3. Every W generating U must satisfy

$$\mathcal{Y}(0) \supset \mathcal{Y}(1), \mathcal{Y}(1) \supset \mathcal{Y}(2) \quad \text{and} \quad \mathcal{Y}(2) \supset \mathcal{Y}(0)$$

and hence $\mathcal{Y}(0) = \mathcal{Y}(1) = \mathcal{Y}(2)$. However, such a W generates

$$\begin{pmatrix} + & + & + \\ + & + & + \\ + & + & + \end{pmatrix}$$

and not U.

Definition AII.2: A component-pairwise zero-error detection code for W^n is a set $\{u_i\colon 1 \le i \le M\} \subset \mathcal{X}^n$ such that for all pairs $(u_i, u_j) = (u_{i1} \cdots u_{in}, u_{j1} \cdots u_{jn})$ $(i \ne j)$ there is a component $t = t(i,j)$ and there are letters y, y' with

$$W(y|u_{it}) > 0, W(y'|u_{jt}) > 0 \quad \text{and}$$
$$W(y'|u_{it}) = W(y|u_{jt}) = 0$$

(or equivalently $\mathcal{Y}(u_{it}) \supset|\subset \mathcal{Y}(u_{jt})$).

For a suitable set $\mathcal{Y}'$ we define now $T\colon \mathcal{X} \to \mathcal{Y}'$ by requiring that for all $x, x' \in \mathcal{X}$

$$\mathcal{Y}(x) \supset\!\!\!|\!\!\!\subset \mathcal{Y}(x'), \qquad \text{iff for all } y \in \mathcal{Y}' \; T(y|x)T(y|x') = 0. \tag{AII.6}$$

One notices that $\{u_i\colon 1 \le i \le M\}$ is a component-pairwise zero-error detection code for W^n iff it is a zero-error code (in Shannon's sense) for T^n.

Example AII.2: That

$$T = \begin{array}{c} a_0 \\ a_1 \\ a_2 \\ b_0 \\ b_1 \\ b_2 \end{array} \begin{pmatrix} + & 0 & + & 0 & 0 \\ + & 0 & 0 & + & 0 \\ + & 0 & 0 & 0 & + \\ 0 & + & + & 0 & 0 \\ 0 & + & 0 & + & 0 \\ 0 & + & 0 & 0 & + \end{pmatrix}$$

cannot be generated by any W via (AII.6) can be seen as follows. Such a W would have to satisfy $\mathcal{Y}(a_0) \supset\!\subset \mathcal{Y}(b_0)$, $\mathcal{Y}(a_0) \supset\!\subset \mathcal{Y}(a_1)$, and $\mathcal{Y}(a_0) \supset\!\subset \mathcal{Y}(a_2)$. Since $\mathcal{Y}(b_0) \supset|\subset \mathcal{Y}(a_1)$ and $\mathcal{Y}(b_0) \supset|\subset \mathcal{Y}(a_2)$, we get $\mathcal{Y}(a_0) \supset \mathcal{Y}(b_0)$, $\mathcal{Y}(a_0) \supset \mathcal{Y}(a_1)$, and $\mathcal{Y}(a_0) \supset \mathcal{Y}(a_2)$ or

$$\mathcal{Y}(a_0) \subset \mathcal{Y}(b_0), \mathcal{Y}(a_0) \subset \mathcal{Y}(a_1) \quad \text{and} \quad \mathcal{Y}(a_0) \subset \mathcal{Y}(a_2). \tag{AII.7}$$

The corresponding relations for a_1 and a_2 are

$$\mathcal{Y}(a_1) \supset \mathcal{Y}(b_1), \mathcal{Y}(a_1) \supset \mathcal{Y}(a_2) \quad \text{and} \quad \mathcal{Y}(a_1) \supset \mathcal{Y}(a_2)$$

or

$$\mathcal{Y}(a_1) \subset \mathcal{Y}(b_1), \mathcal{Y}(a_1) \subset \mathcal{Y}(a_2) \quad \text{and} \quad \mathcal{Y}(a_1) \subset \mathcal{Y}(a_0) \tag{AII.8}$$

and

$$\mathcal{Y}(a_2) \supset \mathcal{Y}(b_2), \mathcal{Y}(a_2) \supset \mathcal{Y}(a_0) \quad \text{and} \quad \mathcal{Y}(a_2) \supset \mathcal{Y}(a_1)$$

or

$$\mathcal{Y}(a_2) \subset \mathcal{Y}(b_2), \mathcal{Y}(a_2) \subset \mathcal{Y}(b_2) \quad \text{and} \quad \mathcal{Y}(a_2) \subset \mathcal{Y}(a_1) \tag{AII.9}$$

respectively.

However, for $i \ne j$ $\mathcal{Y}(a_i) \ne \mathcal{Y}(a_j)$ because, for example, for $i = 0, j = 1$, $\mathcal{Y}(a_0) \supset\!\!\!|\!\!\!\subset \mathcal{Y}(b_1)$ and $\mathcal{Y}(a_1) \supset\!\subset \mathcal{Y}(b_1)$.

So there is no way to satisfy (AII.7)–(AII.9) simultaneously (with $\mathcal{Y}(a_i) \ne \mathcal{Y}(a_j)$ for $i \ne j$).

Definition AII.3: $\{(u_i, v_i)\colon 1 \le i \le M\} \subset \mathcal{X}^n \times \mathcal{Y}^n$ is a pseudomatching zero-error detection code, if

$$W^n(v_j|u_i) = 0, \qquad \text{for } i \ne j.$$

(We have given up (9.3)) in the definition of an MDC.

At first notice that for MDC necessarily $u_i \ne u_j$ and $v_i \ne v_j$ for $i \ne j$. For the new codes (without (9.3)) this no longer is true. We can study the following four conditions (of increasing strength)

$$(u_i, v_i) \ne (u_j, v_j), \qquad \text{for } i \ne j \tag{AII.10}$$
$$u_i \ne u_j, \qquad \text{for } i \ne j \tag{AII.11}$$
$$v_i \ne v_j, \qquad \text{for } i \ne j \tag{AII.12}$$
$$u_i \ne u_j \text{ and } v_i \ne v_j, \qquad \text{for } i \ne j. \tag{AII.13}$$

We denote the corresponding maximal code sizes for W^n by $M_-^n(W)$, $M_{+-}^n(W) M_{-+}^n(W)$,and $M_{++}^n(W)$.

Our main, but simple, observations areas follows.

I) All these quantities are generally different from the corresponding quantities for zero-error, zero-error detection, and matching zero-error detection codes for W^n.

II) All these quantities can be estimated at least asymptoticalle rather accurately, so that the capacities are known.

III) The bounds on the code sizes are almost trivial. However, an exact determination of the code sizes is perhaps challenging for a combinatorialist.

For the formulation of our results we need a few definitions. $A \times B$ with $A \subset \mathcal{X}$, $B \subset \mathcal{Y}$ is a zero-rectangle for W if

$$W(y|x) = 0, \qquad \text{for all } x \in A,\ y \in B. \tag{AII.14}$$

Let $\mathcal{R}$ be the set of such rectangles, let $A^* \times B^*$ be a maximal rectangle in the sense

$$|A^*||B^*| = \max_{A \times B \in \mathcal{R}} |A||B| \tag{AII.15}$$

and let $A^{(n)} \times B^{(n)}$ be a maximal rectangle in the sense

$$\max_{A \times B \in \mathcal{R}} (\min[|\mathcal{X}|^{n-1}|A|, |\mathcal{Y}|^{n-1}|B|])$$
$$= \min[|\mathcal{X}|^{n-1}A^{(n)}, |\mathcal{Y}|^{n-1}B^{(n)}]. \tag{AII.16}$$

Theorem AII.1: For every memoryless channel W^n with $A^* \times B^* \ne \varnothing$

i) $|\mathcal{X}|^{n-1}|\mathcal{Y}|^{n-1}|A^*||B^*| \le M_-^n(W) \le |\mathcal{X}|^n|\mathcal{Y}|^n$;

ii) $|\mathcal{X}|^n - (\min_{y \in \mathcal{Y}} |\mathcal{X}(y)|)^n \le M_{+-}^n(W) \le |\mathcal{X}|^n - (\min_{y \in \mathcal{Y}} |\mathcal{X}(y)|)^n + 1$;

iii) $|\mathcal{Y}|^n - (\min_{x \in \mathcal{X}} |\mathcal{Y}(x)|)^n \le M_{-+}^n(W) \le |\mathcal{Y}|^n - (\min_{x \in \mathcal{X}} |\mathcal{Y}(x)|)^n + 1$;

iv) $\min[|\mathcal{X}|^{n-1}A^{(n)}, |\mathcal{Y}|^{n-1}B^{(n)}] \le M_{++}^n(W) \le \min(|\mathcal{X}|^n, |\mathcal{Y}|^n)$.

Obviously, all quantities equal 1, if $A^* \times B^* = \varnothing$.

Proof: The upper bounds in i) and iv) are trivial. The lower bounds in i) is achieved by the code

$$\{(x^{n-1}x_n, y^{n-1}y_n): x_n \in A^*, y_n \in B^*\}.$$

The lower bound in iv) is achieved by a maximal matching of the complete bipartite graph with vertex sets

$$\mathcal{X}^{n-1} \times A^{(n)}, \quad \mathcal{Y}^{n-1} \times B^{(n)}.$$

Finally, we have to prove only ii), because iii) is symmetrically the same. Suppose then that $\{(u_i, v_i) : 1 \leq i \leq M\}$ is pseudomatching and satisfies (AII.11). By Definition AII.3 $u_i \notin \mathcal{X}^n(v_1)$ for $i > 1$ and consequently we get the upper bound.

In conclusion we define $y_0 \in \mathcal{Y}$ by

$$|\mathcal{X}(y_0)| = \min_{y \subset \mathcal{Y}} |\mathcal{X}(y)|$$

and observe that the lower bound in ii) is achieved by the code

$$\{(x^n, y_0^n): x^n \notin \mathcal{X}^n(y_0^n)\}$$

where $y_0^n = y_0 \cdots y_0$.

Remarks:

11} The notion of pseudomatching detection codes in the sense of (AII.10) is paralled by the notion of a pseudo-zero-error detection code, where in the definition of a zero-error detection code the condition

$$W^n(u_i|u_i) > 0, 1 \leq i \leq M \qquad \text{(AII.17)}$$

is dropped.

Also, Theorem AI.1 is then paralled by pseudocodes: every pseudomatching detection code for W^n generates a pseudo-zero-error detection code for V^n. Further, a $\mathcal{Z}, V$ is generated this way by some W iff $\mathcal{Z}$ has partitions $\{A_x\}_{x \in \mathcal{X}}, \{B_y\}_{y \in \mathcal{Y}}$ in the sense of (AI.10), (AI.11), and

$$|A_x \cap B_y| = 1, \qquad \text{for all } x \in \mathcal{X}, y \in \mathcal{Y}. \qquad \text{(AII.18)}$$

12) In Theorem AII.1, ii) (similarly iii)) can be improved to

$$M^n_{+-}(W) = \max(L_n, M^n_{m\ de}(W)) \qquad \text{(AII.19)}$$

where L_n is the lower bound in ii). Indeed, let $\{(u_i, v_i): 1 \leq i \leq M\}$ be a code achieving $M^n_{+-}(W)$, then in case $W^n(v_i|u_i) > 0$ for all i we really have a MDC and thus $M \leq M^n_{m\ de}(W)$. Otherwise, we can find an i_0 with $W^n(v_{i_0}|u_i) = 0$ for all i and therefore $M \leq L_n$.

Consequently, the lower bound L_n is tight for "most" channels, for the $\binom{\alpha}{\beta}$-uniform complete hypergraph channels of Section X, but for "some" channels like

$$W = \begin{pmatrix} + & 0 & 0 & + \\ 0 & + & 0 & + \\ 0 & 0 & + & + \end{pmatrix}$$

the upper bound and *not* the lower bound is tight.

13) Pseudomatching zero-error detection codes for W^n can be formulated also in graph-theoretic terminology. We begin with the bipartite graph $\mathcal{G}^{\otimes n} = (\mathcal{X}^n, \mathcal{Y}^n, \mathcal{E}_n)$ associated with W^n as in Section III and finally, define a graph $\mathcal{G}_n = (\mathcal{X}^n \times \mathcal{Y}^n, \tilde{\mathcal{E}}_n)$, where

$$\begin{aligned} \tilde{\mathcal{E}}_n = &\{\{(x^n, y^n), (x'^n, y^n)\}: x^n, x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n\} \\ &\cup \{\{(x^n, y^n), (x^n, y'^n)\}: x^n \in \mathcal{X}^n, y^n, y'^n \in \mathcal{Y}^n\} \\ &\cup \{\{(x^n, y^n), (x'^n, y'^n)\}: (x^n, y^n) \in \mathcal{E}_n\}. \end{aligned}$$

Then a pseudomatching zero-error detection code for W^n corresponds to an independent set of $\tilde{\mathcal{G}}_n$.

REFERENCES

[1] R. Ahlswede, "On set coverings in Cartesian product spaces," Preprint 92-005, SFB 343 "Diskrete Strukturen in der Mathematik," Universität Bielefeld, Bielefeld, Germay.

[2] R. Ahlswede and N. Cai, "Cross disjoint pairs of clouds in the interval lattice," in "Diskrete Sturkturen in der Mathematik," Universität Bielefeld, Bielefeld, Germany, Preprint 93-038, SFB 343.
R. L. Graham and J. Nešetřil, Eds., *The Mathematics of Paul Erdös, Algorithm and Combinatorics 13.* Berlin, Heidelnerg, Germany: Springer, 1997, pp. 155–164.

[3] I. Anderson, *Combinatorics of Finite Sets.* Oxford, U.K.: Claredon, 1987.

[4] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic, 1981.

[5] A. Blokhuis, "On the Sperner capacity of the cyclic triangle," *J. Algeb. Combin.*, vol. 2, pp. 123–124, 1993.

[6] P. Hall, "On representatives of subsets," *J. London Math. Soc.*, vol. 10, pp. 26–30, 1935.

[7] L. Lovasz and M. D. Plummer, *Matching Theory.* Amsterdam, The Netherlands: North-Holland, 1986.

[8] C. E. Shannon, "The zero-error capacity of a noisy channel," *IRE Trans. Inform. Theory*, vol. IT-2, pp. 8–19, 1956.

[9] R. Ahlswede, N. Cai, and Z. Zhang, "Erasure, list, and detection zero-error capacities for low noise and a relation to identification," *IEEE Trans. Inform. Theory*, vol. 42, pp. 55–62, Jan. 1996.

[10] D. König, "Über Graphen und ihre Anwendung auf Determinantentheorie und Mengenlehre," *Math. Annalen*, vol. 77, pp. 453–465, 1916.

[11] L. Gargano, J. Körner, and U. Vaccaro, "Sperner capacities," *Graphs and Combinatorics*, to be published.

[12] R. Ahlswede and G. Dueck, "Identification via channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 15–29, Jan. 1989.

[13] ——, "Identification in the presence of feedback—A discovery of new capacity formulas," *IEEE Trans. Inform. Theory*, vol. 35, pp. 30–39, Jan. 1989.

[14] R. Ahlswede and N. Cai, "On extremal set partitions in Cartesian product spaces," *Comb., Probab. Comput.*, vol. 2, pp. 211–220, 1993.

[15] R. Ahlswede, "Channels with arbitrary varying channel probability functions in the presence of feedback," *Z. Wahrscheinlichkeitstheorie und verw. Geb.*, vol. 25, pp. 239–252, 1973.

[16] R. Ahlswede and G. Dueck, "Good codes can be produced by a few permutations," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 430–443, May 1982.

[17] R. Ahlswede, "Coloring hypergraphs: A new approach to multi-user source coding," Part I, *J. Comb,, Inform., Syst. Sci.*, vol. 4, pp. 76–115, 1979; Part II, vol. 5, pp. 220–268, 1980.

[18] R. Ahlswede and G. Simonyi, "Reusable memories in the light of the old varying and a new outputwise varying channel theory," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1143–1150, July 1991.

[19] M. S. Pinsker, "Capacity region of noiseless broadcast channels" (in Russian), *Probl. Pered. Inform.*, vol. 14, no. 2, pp. 28–32, 1978.

[20] T. M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 2–14, Jan. 1972.

[21] E. C. van der Meulen, "Random coding theorems for the general discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 180–190, 1975.

[22] R. Ahlswede and A. Kaspi, "Optimal coding strategies for certain permuting channels," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 310–314, May 1987.

[23] R. Ahlswede, J. Ye, and Z. Zhang, "Creating order in sequence spaces with simple machines," *Inform. Comput.*, vol. 89, no. 1, pp. 47–94, 1990.

[24] K. Kobayashi, "Combinatorial structure and capacity of the permuting relay channel," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 813–826, Nov. 1987.

[25] R. Ahlswede, "Certain results in coding theory for compound channels," in *Proc. Coll. on Information Thheory* (Debrecen, Hungary), 1967, pp. 35–60.

[26] R. Ahlswede and Z. Zhang, "Coding for write efficient memories," *Inform. Comput.*, vol. 83, no. 1, pp. 80–97, 1989.

[27] R. Ahlswede, "A method of coding and an application to arbitrarily varying channels," *J. Comb., Inform. Syst. Sci.*, vol. 5, no. 1, pp. 10–35, 1980.

[28] ——, "Information and control: Matching channels," Preprint 95-035, SFB 343 "Diskrete Strukturen in der Mathematik, Universität Bielefeld, Bielefeld, Germany.