ON IDENTIFICATION VIA MULTI–WAY CHANNELS WITH FEEDBACK

Rudolf Ahlswede and Bart Verboven

Rudolf Ahlswede
Fakultät für Mathematik
Universität Bielefeld
Postfach 100131
33501 Bielefeld
Germany

Bart Verboven
Katholieke Universiteit Leuven
Departement Wiskunde
Celestijnenlaan 200B
3030 Heverlee
Belgium

Abstract

"Identification for Multi–way channels" was mentioned by Ahlswede and Dueck [1] as a challenging direction of research.

Here we present in case of complete feedback a rather unified theory of identification. (For the classical transmission problem the dream of such a theory did not get fulfilled for more than twenty years.)

Its guiding principle is the discovery of [2], that communicators (sender and receiver) must set up a common random experiment with maximal entropy and use it as randomization for a suitable (see [2]) identification technique.

Here we show how this can be done in a constructive way.

The proof of optimality (weak converse) is based on a new entropy bound, which can be viewed as a substitute for Fano's Lemma in the present context.

The "single–letter" characterization of (second order) capacity regions rests now on a new "entropy characterisation problem", which often can be solved.

Here this is done for the multiple–access channel with deterministic and for the broadcast channel with randomized encoding strategies.

Keywords and phrases: identification, multi–way channels, feedback, coding scheme, capacity region, randomized strategies.

Footnote

This paper was presented in the Recent Results Session at the 1990 International Symposium on Information Theory, January 14–19, 1990, San Diego, California and at the IEEE Workshop on Information Theory in Veldhoven, The Netherlands, June 10–15, 1990.

I. Introduction and The results.

In [1] Ahlswede and Dueck have introduced a new model for communication, which they call identification (ID), hereby contrasting Shannon's original transmission (TR) problem. Whereas in [2] one–way channels with feedback were analysed, we present here, as promised in [1], contributions to the theory of multi–way channels. The discussion concentrates on cases where complete feedback links are present. We establish as an always valid principle the idea of [2], that the average maximal entropies of common random experiments among communicators determine the optimal (second order) identification rates. The achievability proof follows the method of [2], page 8, to use keys selected by the common random experiment with blocklength n and short, for instance length $\sqrt{n}$, encryptions for the messages to be identified. The wide applicability of this method is due to the fact that this " $\sqrt{n}$ trick" can be applied independently for several users simply by timesharing without an essential loss in rates.

However, the converse proofs of [2] use special properties of one–way channels and don't seem to be adaptable to multi–way channels. We present here a new method (Lemma 1 under E), which yields weak converses for these channels. Its essence is an elementary relation in terms of entropy between the cardinalities of sets and their probabilities in arbitrary discrete probability spaces.

In our second main contribution we show how the encryption method mentioned above can be made constructive (see paragraphs F and G). Roughly speaking it improves a suboptimal encryption scheme of Mehlhorn and Schmidt ([6]) via Ahlswede's idea of an iterative reduction used originally for the TR problem ([7], [8]). Finally we emphasize that the determination of the maximal entropies obtainable with common random experiments can be difficult for some channels (see Examples 1 and 2 under D). This shows that the theory is not trivial. It cannot be expected from a general and not trivial theory that it gives detailed answers to all special questions. We remind the reader that after the foundation of mechanics there was still no explicit answer to the motion of three bodies. This hint may help to judge the state of our theory. Some examples are discussed in detail. We give now the formal statements of our concepts and results.

A. Review of known concepts and results.

We first briefly review concepts concerning identification via one–way channels with feedback. Extensions to multy–way channels then almost suggest themself. Unless stated otherwise we use the notation of [1] and [2], in particular, script capitals $\mathcal{X}, \mathcal{Y}, \dots$ denote finite sets. $|\mathcal{A}|$ stands for the cardinality of set $\mathcal{A}$. The letters P, Q always denote probability distributions (PD's) on finite (or countable) sets. $X, Y, \dots$ are random variables (RV's) with PD's $P_X, P_Y, \dots$.

$\mathcal{P}(\mathcal{A})$ is the set of all PD's on $\mathcal{A}$. For a stochastic $|\mathcal{X}| \times |\mathcal{Y}|$–matrix W we denote by W^n the transmission probabilities for n–length words of a DMC. Other notions such as entropies and information quantities are either standard or those from [1] and [2]. The functions "log" and "exp" are understood to be to the base 2.

Let us now turn to the identification problems of [2]. There are two concepts, deterministic and randomized feedback strategies, with corresponding code concepts. The vector–valued function

$$f^n = [f_1, \dots, f_n] \tag{1.1}$$

is a deterministic encoding strategy of blocklength n, if $f_1 \in \mathcal{X}$ and $f_t : \mathcal{Y}^{t-1} \to \mathcal{X}$ for $t > 1$. It is understood that after the received elements $Y_1, \dots, Y_{t-1}$ have been made known to the sender by the feedback channel, the sender transmits $f_t(Y_1, \dots, Y_{t-1})$. At $t = 1$ the sender transmits f_1.

The distribution of the random variables $Y_t (t = 1, 2, \dots)$ is determined by f and W. We denote the probability of receiving $y^n = (y_1, \dots, y_n) \in \mathcal{Y}^n$ by $W^n(y^n|f) = W(y_1|f_1) \cdot W\big(y_2|f_2(y_1)\big) \dots W\big(y_n|f_n(y_1, \dots, y_{n-1})\big)$.

Let $\mathcal{F}_n^d$ be the set of all possible encoding functions as defined in (1.1). A (deterministic) (n, N, λ) IDF code for W is a system $\{(f_i^n, \mathcal{D}_i)|i = 1, 2, \dots, N\}$ with $f_i^n \in \mathcal{F}_n^d, \mathcal{D}_i \subset \mathcal{Y}^n$ for $i \in \{1, 2, \dots, N\}$ and

$$W^n(\mathcal{D}_i^c|f_i^n) \le \lambda \qquad\qquad W^n(\mathcal{D}_j|f_i^n) \le \lambda \tag{1.2}$$

for all $i, j \in \{1, 2, \dots, N\}$ with $i \neq j$.

A randomized (n, N, λ) IDF code for W is a system $\{(Q_F(\cdot|i), \mathcal{D}_i)|i = 1, 2, \dots, N\}$ with $Q_F(\cdot|i) \in \mathcal{P}(\mathcal{F}_n^d), \mathcal{D}_i \subset \mathcal{Y}^n$, with

$$\sum_{g \in \mathcal{F}_n} Q_F(g|i) W^n(\mathcal{D}_i^c|g) \le \lambda \tag{1.3}$$

$$\sum_{g \in \mathcal{F}_n} Q_F(g|j) W^n(\mathcal{D}_i|g) \le \lambda \tag{1.4}$$

for all $i, j \in \{1, 2, \dots, N\}$ with $i \neq j$.

Let $N_d(n, \lambda)$ (resp. $N_r(n, \lambda)$) be the maximal N for which a deterministic (resp. randomized) (n, N, λ) IDF code exists. We summarize the results of [2] as follows.

Theorem AD. *(Coding Theorems and strong converses): If the transmission capacity C of W is positive, then, for all $\lambda \in (0, \frac{1}{2})$, we have:*

(i) $\lim_{n \to \infty} \frac{1}{n} \log \log N_d(n, \lambda) = \max_{x \in \mathcal{X}} H\big(W(\cdot|x)\big)$

(ii) $\lim_{n \to \infty} \frac{1}{n} \log \log N_r(n, \lambda) = \max_{P \in \mathcal{P}(\mathcal{X})} H(P \cdot W)$.

In identification the receiver does not necessarily want to know the message $i \in \mathcal{N} = \{1, 2, \dots, N\}$ given to the sender, he only wants to know the answer to the question "Is it $\hat{\imath}$?". Here $\hat{\imath}$ could be any member of $\mathcal{N}$. The quantities

$e_1^n = 1 - \min_{i \in \mathcal{N}} W^n(\mathcal{D}_i|f_i^n)$ and $e_2^n = \max_{i \neq \hat{\imath}} W^n(\mathcal{D}_{\hat{\imath}}|f_i^n)$

are called first kind and second kind error probabilities. (n,N,λ) IDF codes guarantee these quantities not to exceed λ . The quantity $\frac{1}{n}\log\log N$ is called (second order) rate of the code.

Clearly, analogous definitions can be given for multi–way channels. The receivers want to identify with small error probabilities of both kinds. Senders may or may not be allowed to randomize. Achievable rates are replaced by (second order) capacity regions.

Insofar we have spoken about multi–way channels without very specific definitions. We describe now a sufficiently general class, introduce then mystery numbers and use them to characterize capacity regions.

B. A general model for communication systems.

To describe a communication system in general, we introduce the following parameters:

- Ω , the set of terminals: at each terminal $\omega\in\Omega$ information can be sent and/or received;
- Γ , the set of messengers: for each $\gamma\in\Gamma$, there will be a message set $\mathcal{N}_\gamma$.

The situation at each terminal $\omega\in\Omega$ is further described by

- $\mathcal{A}_\omega\subset\Gamma$, the set of messengers reporting to ω ;
- $\mathcal{B}_\omega\subset\Gamma$, where $\gamma\in\mathcal{B}_\omega$ indicates that the messages of $\mathcal{N}_\gamma$ should be decoded at $\mathcal{B}_\omega$;
- $\Phi_\omega\subset\Gamma$, the set of feedback signals linked back to ω , i.e., $\omega'\in\Phi_\omega$ indicates that all symbols received at ω' are also available at ω .

Finally, the communication between terminals is governed by a discrete memoryless channel matrix, i.e., a stochastic matrix

$$W:\prod_{\omega\in\Omega}\mathcal{X}_\omega\longrightarrow\prod_{\omega\in\Omega}\mathcal{Y}_\omega,$$

with input alphabets $\{\mathcal{X}_\omega\}_{\omega\in\Omega}$ and output alphabets $\{\mathcal{Y}_\omega\}_{\omega\in\Omega}$. Notice that we assume an input and output alphabet at each terminal. However, allowing $|\mathcal{X}_\omega|=1$ or $|\mathcal{Y}_\omega|=1$, resp., we can effectively model also situations where ω only receives or sends signals, respectively.

The reader can convince himself that the following axioms provide plausible assumptions:

$A_1:\quad \mathcal{A}_\omega\cap\mathcal{B}_\omega=\phi$ and $\bigcup_{\omega\in\Omega}\mathcal{A}_\omega=\bigcup_{\omega\in\Omega}\mathcal{B}_\omega=\Gamma$;

$A_2:\quad \max\{|\mathcal{X}_\omega|,|\mathcal{Y}_\omega|\}\geq 2$;

$A_3:$ if $|\mathcal{X}_\omega| = 1$, then $\mathcal{A}_\omega = \phi$; if $|\mathcal{Y}_\omega| = 1$, then $\mathcal{B}_\omega = \phi$;

$A_4:$ if $\mathcal{A}_\omega = \phi$ and $|\mathcal{X}_\omega| \geq 2$ then $|\mathcal{Y}_\omega| \geq 2$; if $\mathcal{B}_\omega = \phi$ and $|\mathcal{Y}_\omega| \geq 2$ then $|\mathcal{X}_\omega| \geq 2$;

and as a convention to simplify notation further on, we also assume

$A_5:$ $\omega \in \Phi_\omega$.

These definitions and axioms define <u>a general discrete memoryless communication system.</u>

We will restrict our attention to the class of systems with <u>supervisory feedback</u>, i.e., where for all $\omega, \omega' \in \Omega$ it holds that

$A_6:$ if $\mathcal{A}_\omega \cap \mathcal{B}_{\omega'} \neq \phi$, then $\Phi_{\omega'} \subset \Phi_\omega$.

This assures each terminal encoding γ's messages $(\gamma \in \Gamma)$ of at least all the output signals that are known at the terminals decoding these messages.

The set of decoders Δ is defined by

$$\Delta \stackrel{\text{def}}{=} \{\omega \in \Omega \,|\mathcal{B}_\omega \neq \phi\}\,. \tag{1.5}$$

We mainly consider the case of <u>passive decoders</u>, i.e., where

$A_7:$ for all $\omega \in \Delta$, $|\mathcal{X}_\omega| = 1$,

to avoid decoders to influence the communication.

To illustrate our model, we state the following explicit communication systems (CS):

CS 1 one–way channel: $\Omega = \{1,2\}$, $|\Gamma| \geq 1$, $\mathcal{A}_1 = \Gamma$, $\mathcal{A}_2 = \phi$, $\mathcal{B}_1 = \phi$, $\mathcal{B}_2 = \Gamma$, $\Phi_1 = \{1,2\}$, $\Phi_2 = \{2\}$, $\mathcal{X}_2 = \mathcal{Y}_1 = \{0\}$, and $W : \mathcal{X}_1 \times \{0\} \to \{0\} \times \mathcal{Y}_2$.

CS 2 multiple–access channel (MAC): $\Omega = \{1,2,3\}$, $\Gamma = \{a,b\}$, $\mathcal{A}_1 = \{a\}$, $\mathcal{A}_2 = \{b\}$, $\mathcal{A}_3 = \phi$, $\mathcal{B}_1 = \phi = \mathcal{B}_2$, $\mathcal{B}_3 = \Gamma$, $\Phi_1 = \{1,3\}$, $\Phi_2 = \{2,3\}$, $\Phi_3 = \{3\}$, $\mathcal{X}_3 = \mathcal{Y}_1 = \mathcal{Y}_2 = \{0\}$, and $W : \mathcal{X}_1 \times \mathcal{X}_2 \times \{0\} \to \{0\} \times \{0\} \times \mathcal{Y}_3$.

CS 3 the broadcast channel (BC): $\Omega = \{1,2,3\}$, $\Gamma = \{a,b\}$, $\mathcal{A}_1 = \Gamma$, $\mathcal{A}_2 = \mathcal{A}_3 = \phi$, $\mathcal{B}_2 = \phi$, $\mathcal{B}_2 = \{a\}$, $\mathcal{B}_3 = \{b\}$, $\Phi_1 = \{1,2,3\}$, $\Phi_2 = \{2\}$, $\Phi_3 = \{3\}$, $\mathcal{X}_2 = \mathcal{X}_3 = \mathcal{Y}_1 = \{0\}$, and $W : \mathcal{X}_1 \times \{0\} \times \{0\} \to \{0\} \times \mathcal{Y}_2 \times \mathcal{Y}_3$.

CS 4 the interference channel (IC): $\Omega = \{1,2,3,4\}$, $\Gamma = \{a,b\}$, $\mathcal{A}_1 = \{a\}$, $\mathcal{A}_2 = \{b\}$, $\mathcal{A}_3 = \mathcal{A}_4 = \phi$, $\mathcal{B}_1 = \mathcal{B}_2 = \phi$, $\mathcal{B}_3 = \{a\}$, $\mathcal{B}_4 = \{b\}$, $\Phi_1 = \{1,3\}$, $\Phi_2 = \{2,4\}$, $\Phi_3 = \{3\}$, $\Phi_4 = \{4\}$, $\mathcal{X}_3 = \mathcal{X}_4 = \{0\} = \mathcal{Y}_1 = \mathcal{Y}_2$, and $W : \mathcal{X}_1 \times \mathcal{X}_2 \times \{0\} \times \{0\} \to \{0\} \times \{0\} \times \mathcal{Y}_3 \times \mathcal{Y}_4$.

CS 5 the relay channel (RC): $\Omega = \{1,2,3\}$, $|\Gamma| \geq 1$, $\mathcal{A}_1 = \Gamma$, $\mathcal{A}_2 = \mathcal{A}_3 = \phi$, $\mathcal{B}_1 = \mathcal{B}_2 = \phi$, $\mathcal{B}_3 = \Gamma$, $\Phi_1 = \{1,3\}$, $\Phi_2 = \{2\}$, $\Phi_3 = \{3\}$, $\mathcal{X}_3 = \{0\} = \mathcal{Y}_1$, and $W : \mathcal{X}_1 \times \mathcal{X}_2 \times \{0\} \to \{0\} \times \mathcal{Y}_2 \times \mathcal{Y}_3$.

CS 6 the two–way channel (TWC): $\Omega = \{1,2\}$, $\Gamma = \{a,b\}$, $\mathcal{A}_1 = \{a\}$, $\mathcal{A}_2 = \{b\}$, $\mathcal{B}_1 = \{b\}$, $\mathcal{B}_2 = \{a\}$, $\Phi_1 = \Phi_2 = \{1,2\}$, and $W : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}_1 \times \mathcal{Y}_2$.

C. Classes of feedback strategies, common random experiments and their mystery numbers.

In dealing with different kinds of feedback strategies it is convenient to have the following concept. Let $\mathcal{F}_n (n = 1,2,\dots)$ be a subset of the set of all randomized feedback strategies $\mathcal{F}_n^r$ of a DMC W with blocklength n and let it contain the set $\mathcal{F}_n^d$ of all deterministic strategies.

We call $(\mathcal{F}_n)_{n=1}^\infty$ a smooth class of strategies if for all $n_1, n_2 \in \mathbb{N}$ and $n = n_1 + n_2$

$$\mathcal{F}_n \supset \mathcal{F}_{n_1} \times \mathcal{F}_{n_2} \tag{1.6}$$

and the product means concatenation of strategies.

Now for $f^n \in \mathcal{F}_n$ the channel induces an output sequence $Y^n(f^n)$. For any smooth class we define numbers

$$\mu(\mathcal{F}_n) = \max_{f^n \in \mathcal{F}_n} H(Y^n(f^n)) \ . \tag{1.7}$$

By (1.6) and the memoryless character of the channel

$$\mu(\mathcal{F}_n) \geq \mu(\mathcal{F}_{n_1}) + \mu(\mathcal{F}_{n_2}) \ , \tag{1.8}$$

and therefore $\mu = \mu\left((\mathcal{F}_n)_{n=1}^\infty\right) = \lim_{n\to\infty} \frac{1}{n}\mu(\mathcal{F}_n)$ exists. It is called mystery number in order to attract attention.

We call $\overline{\mathcal{F}}^r = (\mathcal{F}_n^r)_{n=1}^\infty$ also the complete class of strategies. We mentioned already the class of deterministic strategies $\overline{\mathcal{F}}^d = \left(\mathcal{F}_n^d\right)_{n=1}^\infty$. Both classes are smooth. Between those classes there is a natural smooth class $\overline{\mathcal{F}}^s = (\mathcal{F}_n^s)_{n=1}^\infty$ of what may be termed stochastic strategies. For every member $F^n = \left(F_1, \dots, F_n\right) \in \mathcal{F}_n^s$ F_1 is a RV on $\mathcal{X}$ and $F_t : \mathcal{Y}^{t-1} \to \mathcal{X}$ for $t \geq 2$ are stochastic functions, that is, for each y^{t-1} , $F_t(y^{t-1})$ is a RV with values in $\mathcal{X}$. Stochastic functions are like channels, stochastic strategies are "stochastic versions" of deterministic strategies. One readily verifies that for a DMC

$$\mu\left(\overline{\mathcal{F}}^s\right) = \mu\left(\overline{\mathcal{F}}^r\right) , \tag{1.9}$$

however for multi–way channels there are differences (see Example 1 below). For these channels each sender has his class of feedback strategies. If they are all smooth, then a region $\mathcal{V}$ of achievable mystery tuples is well–defined. Also, by concatenation all common random experiments are of the i.i.d. type and the AEP holds and the " $\sqrt{n}$–trick " can be applied. It yields the direct coding parts in the Main Theorem and Theorem 2 in Paragraph D.

Stochastic strategies for multi–way channels shall have the property that for $t \geq 2$ and given outputs $y_1^{t-1}, y_2^{t-1}, \dots$ at all receivers the RV's $F_{1t}(y_1^{t-1}, y_2^{t-1}, \dots)$, $F_{2t}(y_1^{t-1}, y_2^{t-1}, \dots), \dots$ in the strategies of all senders are independent. This condition seems reasonable, if the senders share only the knowledge of all outputs at each step.

Remark 1: Of course the complete class $\overline{\mathcal{F}}^r$ gives the largest rates. However, ratewise inferior classes often have other advantages such as smaller coding efforts. They therefore also should be studied.

Finally we give the formal definitions for the general communication system of Paragraph B. We assume that each terminal ω uses feedback strategies from a smooth set $\mathcal{F}_{n,\omega}$ for encoding. We will denote $\mathcal{G}_n$ for the smooth class of composite strategies,

$$\mathcal{G}_n \stackrel{\text{def}}{=} \left\{ g^n = \left(f_\omega^n \right)_{\omega \in \Omega} | f_\omega^n \in \mathcal{F}_{n,\omega} \right\} \tag{1.10}$$

As before, we denote $\{\mathcal{G}_n\}_{n=1}^\infty$ by $\overline{\mathcal{G}}$. The channel outputs produced via the composite encoding strategy g^n can then be denoted as $Y_\omega^n(g^n)$.

For every decoder $\omega \in \Delta$ (cf. (1.5)), we introduce

$$Z_\omega^n(g^n) = (Y_{\omega'}^n(g^n))_{\omega' \in \Phi_\omega} \tag{1.11}$$

The set of mystery vectors for the system is then defined as

$$\mathcal{V}_\Delta(\overline{\mathcal{G}}) \stackrel{\text{def}}{=} \lim_{n \to \infty} \left\{ (v_\omega)_{\omega \in \Delta} \middle| \exists g^n \in \mathcal{G}_n : \forall \omega \in \Delta : 0 \leq v_\omega \leq \frac{H\big(Z_\omega^n(g^n)\big)}{n} \right\}, \tag{1.12}$$

where the convergence of sets is understood in the Hausdorff metric and follows here by the memoryless character of our channel and the smoothness assumptions for the classes of strategies.

D. Main Theorem and Consequences.

Using the notation of a general (Ω, Γ) communication system in Paragraph B, we define an $\big(n, \{N_\gamma\}_{\gamma \in \Gamma}, \lambda\big)$ IDF–code for a general (Ω, Γ) communication system and a smooth class of feedback strategies $\overline{\mathcal{G}}$ as a system

$$\Big\{\big(g_m^n, \{\mathcal{D}_{m_\omega}^{(\omega)}\}_{\omega \in \Delta}\big)\Big\}, \tag{1.13}$$

with encoding strategies $g_m^n \in \mathcal{G}_n$, message vectors

$$m = (i_\gamma)_{\gamma \in \Gamma},\ m_\omega = (i_\gamma)_{\gamma \in \mathcal{B}_\omega},\ i_\gamma \in \mathcal{N}_\gamma \stackrel{\text{def}}{=} \{1, \dots, N_\gamma\}, \tag{1.14}$$

and decoding sets

$$\mathcal{D}_{m_\omega}^{(\omega)} \subseteq \mathcal{X}_\omega \times \prod_{\omega' \in \Phi_\omega} \mathcal{Y}_{\omega'}^n, \tag{1.15}$$

that satisfies the upperbound λ on both kinds of error probability (which can be defined similarly to (1.2)–(1.4)). Achievable ID–rates $(R_\gamma)_{\gamma \in \Gamma}$ are defined as usual, and the region $\mathcal{C}(\overline{\mathcal{G}})$ of all these rates is then the (second order) ID–capacity region.

Main Theorem.

Consider an (Ω, Γ) communication system with passive decoders (i.e. A_7 holds) and supervisory feedback, and a smooth class of feedback strategies, $\overline{\mathcal{G}}$.

a) If all messengers $\gamma \in \Gamma$ can transmit at positive rate, then $(R_\gamma)_{\gamma \in \Gamma} \in \mathcal{C}(\overline{\mathcal{G}})$ if and only if there exists some $(\nu_\omega)_{\omega \in \Delta} \in \mathcal{V}_\Delta(\overline{\mathcal{G}})$ such that

$$0 \le R_\gamma \le \nu_\omega, \quad \text{for all } \omega \in \Delta \text{ and } \gamma \in \mathcal{B}_\omega. \tag{1.16}$$

b) If Γ_o is the set of messengers which can have only transmission rate O, then $\mathcal{C}(\overline{\mathcal{G}})$ is obtained as a projection of the region described in (1.16) into the intersection of the hyperplanes $\mathcal{R}_\gamma = O$ $(\gamma \in \Gamma_o)$.

The proof of this theorem will be given in Section III.

Remark 2: The Main Theorem of course presents a non–single–letter characterization of $\mathcal{C}(\overline{\mathcal{G}})$ in the usual language of Information Theory. Still, we want to state its merits:

– first of all, the machinery for deriving such a characterization had to be devello-ped for the ID–situation (substitutes for Shannon's Random Coding Argument and Fano's Lemma);

– secondly, the characterization involves only optimization over single strategies, rat-her than over codebooks of strategies;

– as entropy quantities, mystery numbers are easier to determine than quantities involving mutual information; this is largely responsible for the fact that we can derive a single–letter characterization from the limiting characterization in the Main Theorem directly (see Corollaries 1,2,3,4 below); we remind the reader that in the present literature on transmission there is no such direct derivation of the single–letter capacity region for the MAC from its non–single–letter characterization, and that for none of the situations studied in the corollaries below, a complete transmission result is known.

Our methods of proof for the Main Theorem also apply for communication systems not satisfying A_7 , if all strategies permitted are deterministic.

Theorem 2.

For a general communication system with supervisory feedback, and for the set of deterministic strategies $\overline{\mathcal{G}}^d$ *, the characterization of* $\mathcal{C}\big(\overline{\mathcal{G}}^d\big)$ *is also given by the parts a) and b) of the Main Theorem.*

We now state some applications of the Main Theorem. We restrict the discussion to the genuine part a) since the situation in part b) is always obvious from there.

Corollary 1.

For the MAC as described by communication system CS2 in Paragraph B, and the set of deterministic feedback strategies $\overline{\mathcal{G}}^d$ *, it holds under the condition of part a) of the Main Theorem that*

$$\mathcal{C}_{MA}\big(\overline{\mathcal{G}}^d\big) = \big[0, \mu_{MA}^d\big] \times \big[0, \mu_{MA}^d\big] \,,$$

where

$$\mu_{MA}^d = \max_{x_1,x_2} H\left(W(\cdot|x_1,x_2)\right).$$

Notice that in the explicit entropy expressions we will discard the degenerate in– and outputs of the channel W .

Proof of Corollary 1: Since $\Delta = \{3\}$, $\mathcal{V}_\Delta\big(\overline{\mathcal{G}}^d\big)$ is a one–dimensional region.

Now, for $g^n = \big(f_1^n, f_2^n\big) \in \mathcal{G}_n^d$ and $Y^n = Y_3^n(g^n)$, since $H(Y^n) = \sum_{t=1}^n H\big(Y_t|Y^{t-1}\big)$ and $H\big(Y_t|Y^{t-1}=y^{t-1}\big) = H\left(W\big(\cdot|f_{1t}\big(y^{t-1}\big), f_{2t}\big(y^{t-1}\big)\big)\right) \le \max_{x_1,x_2} H\big(W\big(\cdot|x_1,x_2\big)\big)$, it obviously follows that $\mathcal{V}_\Delta\big(\overline{\mathcal{G}}^d\big) = \big[0, \mu_{MA}^d\big]$ and from the Main Theorem (or Theorem 2) hence also the corollary.

□

Corollary 2.

For the MAC as described in CS2 and for the set of stochastic strategies $\overline{\mathcal{G}}^s$ *, it holds under the condition of part a) of the Main Theorem that*

$$\mathcal{C}_{MA}\big(\overline{\mathcal{G}}^s\big) = [0, \mu^s_{MA}] \times [0, \mu^s_{MA}],$$

where

$$\mu^s_{MA} = \max_{P_1 \in \mathcal{P}(\mathcal{X}_1)} \max_{P_2 \in \mathcal{P}(\mathcal{X}_2)} H(Y),$$

and

$$P_Y(y) = \sum_{x_1} \sum_{x_2} P_1(x_1) P_2(x_2) W\big(y|x_1, x_2\big).$$

Proof: Let $G^n = \big(F_1^n, F_2^n\big) \in \mathcal{G}^s_n$, and denote $Y^n = Y_3^n(G^n)$. Since, given $Y^{t-1} = y^{t-1}$, the RV's $F_{1t}\big(y^{t-1}\big)$ and $F_{2t}\big(y^{t-1}\big)$ are independent, we have

$$H(Y^n) \leq n\ \mu^s_{MA},$$

which proves the corollary by the Main Theorem.

$\square$

Remark 3: Using the converse methods in [2] and regarding the MAC as a one–way channel, one can obtain a strong converse to the above characterizations.

Example 1: In [10] it was shown that the rate point $(R_1, R_2) = (0.76, 0.76)$ is achievable for transmission via the binary erasure MAC defined by

$$W\big(y|x_1, x_2\big) = 1 \ \text{ iff } \ y = x_1 + x_2,$$

and

$$\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\},\ \mathcal{Y}_3 = \{0, 1, 2\}.$$

If both senders each choose a key at random, transmit it at this rate 0.76 and decode each other's key from the feedback signal, they can each apply the " $\sqrt{n}$ trick " to the pair of keys, and achieve the ID–rate pair $(R_1 + R_2, R_1 + R_2) = (1.52, 1.52)$. As one can easily calculate that $\mu^s_{MA} = 1.5$ for this MAC, this clearly shows that the randomized ID–capacity region $\mathcal{C}_{MA}\big(\overline{\mathcal{G}}^r\big)$ exceeds $\mathcal{C}_{MA}\big(\overline{\mathcal{G}}^s\big)$.

Corollary 3.

For the BC (see CS3), it holds under condition a) of the Main Theorem that

$$\mathcal{C}_{BC}\left(\overline{\mathcal{G}}^r\right) = \mathcal{C}_{BC}\left(\overline{\mathcal{G}}^s\right) = \mathcal{R}^*_{BC}$$

where

$$\mathcal{R}^*_{BC} = \Big\{(\nu_1, \nu_2)|\exists\ P \in \mathcal{P}(\mathcal{X}_1) : 0 \le \nu_1 \le H(P \cdot W_2),\ 0 \le \nu_2 \le H(P \cdot W_3),$$
$$\textit{where } W_2 \textit{ and } W_3 \textit{ are the marginal channels}\Big\}.$$

Proof: Let $Q \in \mathcal{G}^r_n = \mathcal{P}\big(\mathcal{G}^d_n\big)$ and let Y_2^n and Y_3^n denote the corresponding channel outputs at terminals 2 and 3, respectively. Now,

$$H\big(Y_2^n\big) \le \sum_{t=1}^n H\big(Y_{2t}\big), H\big(Y_3^n\big) \le \sum_{t=1}^n H\big(Y_{3t}\big),$$

where $Pr\big[Y_{2t} = y_2,\ Y_{3t} = y_3\big] = \sum_{x \in \mathcal{X}_1} P_t(x) W\big(y_2, y_3|x\big)$ for some $P_t \in \mathcal{P}(\mathcal{X}_1)$. Therefore

$$\mathcal{V}_\Delta = \mathcal{R}^*_{BC}.$$

Since this region is also achievable with the stochastic strategies of $\overline{\mathcal{G}}^s$, this proves the corollary.

□

Example 2: For $\overline{\mathcal{G}} = \overline{\mathcal{G}}^d$, a natural candidate for the single–letterization of $\mathcal{V}_\Delta\big(\overline{\mathcal{G}}^d\big)$ would be the region

$$\mathcal{R} = \Big\{(R_1, R_2)|\exists\ x \in \mathcal{X}_1 : 0 \le R_1 \le H\big(W_2(\cdot|x)\big), 0 \le R_2 \le H\big(W_3(\cdot|x)\big)\Big\}.$$

However, let us consider the BC with $\mathcal{Y}_3 = \mathcal{X}_1$ and $W(y_2, y_3|x) = W_2(y_2|x)\ Q^*(y_3)$, where Q^* satisfies $H(Q^* \cdot W_2) = \max_{P \in \mathcal{P}(\mathcal{X}_1)} H(P \cdot W_2)$. If now the sender uses the deterministic strategy g, defined by

$$g_t\big(y_2^{t-1}, y_3^{t-1}\big) = y_{3,t-1},$$

he generates in this way the maximal entropy $n \cdot H(Q^* \cdot W_2)$ at terminal 2. This proves the ID–achievability of $\big(H(Q^* \cdot W_2), 0\big)$, which clearly in general is not contained in the region $\mathcal{R}$.

Remark 4: In [3] the deterministic BC was treated. In that case feedback is implicitly present and therefore its non–feedback capacity region equals the feedback capacity region, in particular for randomized encoding. In [3] the direct part is proven by applying the " $\sqrt{n}$ trick " twice to $H(Y_2^n)$ and $H(Y_3^n)$ resp., and an application of the converse from [1] gives an upperbound which coincides with this inner bound in the case of a deterministic channel.

Theorem 2 has the following consequence.

Corollary 4.

For the TWC (see CS6) it holds under the condition of part a) of the Main Theorem that

$$\mathcal{C}_{TW}\left(\overline{\mathcal{G}}^d\right) = \left[0, \nu_{TW}^d\right] \times \left[0, \nu_{TW}^d\right]$$

where

$$\nu_{TW}^d = \max_{x_1, x_2} H\big(\omega(\cdot, \cdot | x_1, x_2)\big).$$

E. A method for proving converses in case of feedback.

For one–way channels the approach of [2] gives sharp upper bounds (strong converse). However, it does not seem to generalize to multi–way channels with complicated interactions of feedback strategies. Settling for a weaker bound (weak converse but stronger than soft converse of [1]) we found a method (Lemma 2 below) which always works, that is, it relates rates to the number $\mu(\mathcal{F}_n)$, that is the maximal entropy of a common random experiment of blocklength n which can be produced under the restrictions present. For example for the DMC with restriction to deterministic strategies this number equals $n \max_x H\big(W(\cdot|x)\big)$. The common random experiment with this entropy uses one coding strategy and not a whole codebook! In case of randomized strategies the number is $n \max_{P \in \mathcal{P}(\mathcal{X})} H(P \cdot W)$.

For an (n, N, λ) IDF code the encoding strategy $f_i^n \in \mathcal{F}_n$ generates a RV Y_i^n with distribution

$$Pr\left[Y_i^n = y^n\right] = W^n(y^n | f_i^n). \tag{1.17}$$

Of course for $i = 1, 2, \ldots, N$

$$H(Y_i^n) \leq \mu_n \stackrel{\text{def}}{=} \mu(\mathcal{F}_n). \tag{1.18}$$

The basis of our method is now a very general entropy–setsize relation, which we prove in Section II.

Lemma 1. *For* $P = (P_1, P_2, \dots) \in \mathcal{P}(\mathbb{N})$ *, the set of PD's on the positive integers, define*

$$\varepsilon(d, P) = \max \left\{ \sum_{j \in J} P_j : J \subset \mathbb{N}, |J| = \lceil 2^{H(P)d} \rceil + 1 \right\}$$

and $\varepsilon(d) = \min_{P \in \mathcal{P}(\mathbb{N})} \varepsilon(d, P)$ *. Then*

$$\varepsilon(d) = 1 - \frac{1}{d} \ \textit{for all} \ \ d \geq 1.$$

Remark 5: For discrete memoryless sources $X^n = (X_1, \dots, X_n)$ Shannon proved that

$$\varepsilon(d, P_X^n) = \max \left\{ \sum_{x^n \in J} P_X^n(x^n) : J \subset \mathcal{X}^n, |J| = \lceil 2^{dH(X^n)} \rceil \right\}$$

satisfies

$$\lim_{n \to \infty} \varepsilon(d, P_X^n) = 1, \text{ if } d > 1.$$

Lemma 1 shows what can be done for arbitrary discrete sources.

Application of Lemma 1 to the distribution of Y_i^n gives a set $\mathcal{E}_i \subset \mathcal{Y}^n$ with

$$Pr\left[Y_i^n \in \mathcal{E}_i\right] \geq 1 - \frac{1}{d}, \tag{1.19}$$

$$|\mathcal{E}_i| \leq \lceil 2^{dH(Y_i^n)} \rceil + 1 \leq 2^{d\mu_n} + 2. \tag{1.20}$$

Define now $\mathcal{D}_i^* = \mathcal{D}_i \cap \mathcal{E}_i$. By (1.3) and (1.19) $Pr\left[Y_i^n \in \mathcal{D}_i^*\right] \geq 1 - \lambda - \frac{1}{d}$. Under the assumption $\lambda < 1 - \lambda - \frac{1}{d}$ by (1.4) these $\mathcal{D}_i^*$ are necessarily <u>distinct</u>. With the abbreviation $K = 2^{d\mu_n} + 2$ we get therefore

$$N \leq \sum_{k=1}^{K} \binom{|\mathcal{Y}^n|}{k} \leq K 2^{nK \log |\mathcal{Y}|}$$

and $\log \log N \leq d\mu_n + o(n)$.

We summarize this result.

Lemma 2. *For any* (n, N, λ) *IDF code with coding strategies* $\mathcal{F}_n^*$ *and corresponding* $\mu_n^* = \mu(\mathcal{F}_n^*)$

$$\log\log N \leq d\mu_n^* + o(n),$$

provided that $d > 1$ *and* $\lambda < \frac{1}{2}\left(1 - \frac{1}{d}\right)$.

Remark 6: In case of the DMC for deterministic strategies $\mu_n^* = n \max_{x \in \mathcal{X}} H\big(W(\cdot|x)\big)$ and for randomized strategies $\mu_n^* = n \max_{P \in \mathcal{P}(\mathcal{X})} H(P \cdot W)$. For λ tending to 0 in Lemma 2 we can let d tend to 1 and thus obtain weak converses.

F. A 3–step ID scheme for the noiseless BSC.

We begin right–away with the definition of the scheme. Some heuristic understanding is provided subsequently. The proof of asymptotic optimality is given in Section IV.

We are given a set of messages $\mathcal{M} = \{1, \dots, M\}$. For any constant $\alpha > 1$ define

$$K = \lceil (\log M)^\alpha \rceil \tag{1.21}$$

and $\pi_1 < \pi_2 < \dots \pi_K$ as the K smallest prime numbers.

For $k \in \mathcal{K} = \{1, \dots, K\}$ define a key $\varphi_k : \mathcal{M} \to \{1, \dots, \pi_k\}$ by

$$\varphi_k(m) - 1 \equiv m \mod \pi_k. \tag{1.22}$$

Let $\{\varphi_k : k \in \mathcal{K}\}$ be a cipher and $\mathcal{M}' = \{\varphi_k(m) : m \in \mathcal{M}, k \in \mathcal{K}\} = \{1, 2, \dots, \pi_K\}$ the set of all possible encipherings serving as "message set" for a further cipher $\{\varphi'_\ell : \ell \in \mathcal{K}'\}$, where $\mathcal{K}' = \{1, \dots, K'\}$ with

$$K' = \lceil (\log \pi_K)^\alpha \rceil \tag{1.23}$$

and $\varphi'_\ell : \mathcal{M}' \to \{1, \dots, \pi_\ell\}$ satisfies

$$\varphi'_\ell(m') - 1 \equiv m' \mod \pi_\ell. \tag{1.24}$$

Step 1 The sender chooses $k \in \mathcal{K}$ randomly according to the uniform distribution on $\mathcal{K}$ and transmits it (and therefore also φ_k) over the channel. This requires $\lceil \log K \rceil$ bits.

Step 2 Similarly the sender chooses an $\ell \in \mathcal{K}'$ at random and transmits it (and therefore also the key φ'_ℓ) over the channel. This requires $\lceil \log K' \rceil$ bits.

Step 3 $m \in \mathcal{M}$ being given to the sender he calculates $\varphi'_\ell\big(\varphi_k(m)\big)$ and sends it to the receiver. Knowing both, k and ℓ , the receiver calculates $\varphi'_\ell\big(\varphi_k(\hat{m})\big)$ and compares it with the transmitted encryption $\varphi'_\ell\big(\varphi_k(m)\big)$. In case of equality he decides $m = \hat{m}$ and otherwise he decides $m \neq \hat{m}$.

Theorem 3. *(Optimality of the 3–step scheme)*

The scheme satisfies

$$\lim_{M \to \infty} \frac{\log \log M}{n} = \frac{1}{\alpha}$$

and thus achieves any rate below the capacity 1.

G. Extension of the 3–step ID scheme to the DMC with and without feedback.

Outcomes in random experiments below must be labelled by consecutive integers $1, 2, 3 \dots$ to make the number theoretic setting of the previous scheme possible. Otherwise the only changes on the scheme are the following:

The uniform random experiments for the choice of the two keys, known to sender and receiver, are formed now under the given circumstances. We discuss three cases.

1.) Deterministic feedback strategies.

The sender sends b times a letter $x^* \in \mathcal{X}$ with $H\big(W(\cdot|x^*)\big) = \max_{x \in \mathcal{X}} H\big(W(\cdot|x)\big)$.

As in [2] the generated sequences $\mathcal{D}^* = \bigcup_{V: \|V-W\| \le \varepsilon} \mathcal{T}_V^b(x^{*b})$ and an erasure e for the sequences in $\mathcal{Y}^b \smallsetminus \mathcal{D}^*$ give an essentially uniform random experiment $\big(\mathcal{D}^* \cup \{e\}, W^b(\cdot|x^{*b})\big)$ used for the key selections in the first two steps with appropiate b's . A factor $\big(1 - 2\exp\{-E(\varepsilon)b'\}\big)$ enters the changes in error probabilities of the second kind. The erasure option and also a small error probability in performing step 3 add a small error probability to both kinds of errors. Since $|\mathcal{D}^*| = \exp\{b\ H\big(W(\cdot|x^*)\big) + o(b)\}$ the scheme achieves rates below $H\big(W(\cdot|x^*)\big)$, provided that W has positive transmission capacity C .

2.) Complete randomized feedback strategies

Replace $\mathcal{D}^*$ by $\bigcup_{Q: \|Q-Q^*\| \le \varepsilon} \mathcal{T}_{Q^*}^b$ with $Q^* = P^* \cdot W$ and $H(Q^*) = \max_{P \in \mathcal{P}(\mathcal{X})} H(PW)$. Now rates below $H(Q^*)$ are achievable, if $C > 0$.

3.) Randomized encoding without feedback

As in [2] use now standard transmission codes with uniform distribution on the set of codewords. Here sender and receiver know the outcome of the random experiments in step 1 and 2 with a small error probability only, but this can be digested. Notice that the resulting scheme is totally constructive if the transmission codes used are constructed. Here the ID capacity is C .

II. Proof of Lemma 1.

We can of course assume that $P_1 \ge P_2 \dots$. We show first $\varepsilon(d, P) \ge 1 - \frac{1}{d}$. Set $L = \lceil 2^{H(P)d} \rceil + 1$ and $c = P_L$.

By Schur–concavity of the entropy function the distribution P_i' defined by

$$P_i' = \begin{cases} c & \text{for} \quad i = 2, \dots, L \\ a = \sum_{i=1}^{L} P_i - (L-1)c & \text{for} \quad i = 1 \\ P_i & \text{for} \quad i \ge L+1 \end{cases} \tag{2.1}$$

satisfies $H(P') \le H(P)$ and by monotonicity $\varepsilon(d, P') \le \varepsilon(d, P)$. Furthermore, again by Schur–concavity for P'' , defined for a suitable T by

$$P''_i = \begin{cases} P'_i & \text{for} \quad i = 1, \dots, L \\ c & \text{for} \quad L \le i \le T-1 \\ b \le c & \text{for} \quad i = T \\ 0 & \text{for} \quad i > T, \end{cases} \tag{2.2}$$

$H(P'') \le H(P')$ and $\varepsilon(d, P'') \le \varepsilon(d, P')$.

Therefore it suffices to prove $\varepsilon(d, P) \ge 1 - \frac{1}{d}$ only for PD of the form

$$P = (a, c, \dots, c, b, 0, \dots).$$

Since $\lceil 2^{H(P)d} \rceil + 1 \ge 2$, obviously

$$\varepsilon(P, d) \ge a + b. \tag{2.3}$$

If $a + b \ge 1 - \frac{1}{d}$, we are done, and otherwise we have

$$d(1 - a - b) > 1 \tag{2.4}$$

and a fortiori by the grouping axiom $\lceil 2^{H(P)d} \rceil + 1 > \lceil 2^{d(1-a-b)\log(T-2)} \rceil + 1 \ge T - 1 + 1 = T$. This implies $\varepsilon(P, d) = 1 \ge 1 - \frac{1}{d}$ in this case.

Conversely, fix d and consider only distributions $Q = (a, c, \dots, c, 0, \dots)$, that is, with $b = c$ in the notation above and also with

$$a > 1 - \frac{1}{d}. \tag{2.5}$$

Then $\varepsilon(Q, d) \le a + \lceil 2^{H(Q)d} \rceil \cdot c = a + \lceil 2^{[h(a) + (1-a)\log(T-1)]d} \rceil \cdot c \le a + (T-1)^{(1-a)d} \cdot 2^{h(a)d} \cdot c + c$, and since $c = \frac{1-a}{T-1}$ we conclude that

$$\varepsilon(Q, d) \le a + (T-1)^{(1-a)d-1} \cdot 2^{h(a)d}(1-a) + \frac{1-a}{T-1}.$$

Now by (2.5) $(1-a)d - 1 < 0$ and by letting $T \to \infty$ we see that

$$\varepsilon(d) \le \inf_Q \varepsilon(Q, d) \le a.$$

Since this is true for all a obeying (2.5) the result follows.

Remark 7: In our applications only $\varepsilon(d) \ge 1 - \frac{1}{d}$ is used.

III. Proof of Main Theorem and Theorem 2

Since we consider supervisory feedback, the direct part of the Main Theorem follows from smoothness and memorylessness, as discussed in C of Section I. We therefore concentrate on the converse part of the Main Theorem.

Let $\varepsilon > 0$ be arbitrary and fixed. Since $\mathcal{V}_\Delta\big(\overline{\mathcal{G}}\big)$ is compact, there exists a number of vectors $\big(\nu_\omega^{(\ell)}\big)_{\omega\in\Delta}$, $\ell = 1, \dots, L = L(\varepsilon)$ such that

$$\mathcal{V}_\Delta\big(\overline{\mathcal{G}}\big) \subseteq \bigcup_{\ell=1}^{L} \Big\{ (v_\omega)_{\omega\in\Delta} \,\Big|\, \forall\omega\in\Delta : 0 \le v_\omega < \nu_\omega^{(\ell)} + \varepsilon \Big\}. \tag{3.1}$$

Let $n \ge n_0(\varepsilon)$ be sufficiently large such that for all $g^n = \big(f_\omega^n\big)_{\omega\in\Omega}$ there exists a $(v_\omega)_{\omega\in\Delta} \in \mathcal{V}_\Delta\big(\overline{\mathcal{G}}\big)$ such that for all $\omega\in\Delta$

$$\frac{H\big(Z_\omega^n(g^n)\big)}{n} < v_\omega + \varepsilon. \tag{3.2}$$

Finally let an $\big(n, \{N_\gamma\}_{\gamma\in\Gamma}, \lambda\big)$ IDF–code be given as described in (1.13)–(1.15), to which we also refer for notation.

By the passive decoding Axiom A_7 and (1.15)

$$\mathcal{D}_{m_\omega}^{(\omega)} \subset \mathcal{Z}_\omega^n \tag{3.3}$$

where we have denoted

$$\mathcal{Z}_\omega^n = \prod_{\omega'\in\Phi_\omega} \mathcal{Y}_{\omega'}^n. \tag{3.4}$$

For all $\ell = 1, \dots, L$, define

$$\mathcal{M}(\ell) = \left\{ m \,\middle|\, \forall\omega\in\Delta : \frac{H\big(Z_\omega^n(g_m^n)\big)}{n} < \nu_\omega^{(\ell)} + 2\varepsilon \right\}. \tag{3.5}$$

By the definition of $\big(\nu_\omega^{(\ell)}\big)_{\omega\in\Delta}$ in (3.1) and the choice of n in (3.2), $\bigcup_{\ell=1}^{L}\mathcal{M}(\ell)$ covers all messages m .

Therefore we can choose ℓ^* such that

$$|\mathcal{M}(\ell^*)| \ge \frac{\prod_{\gamma\in\Gamma} N_\gamma}{L}. \tag{3.6}$$

We will now consider the marginal channels

$$W_\omega : \prod_{\omega' \in \Omega} \mathcal{X}_{\omega'} \longrightarrow \mathcal{Z}_\omega^n,$$

for all $\omega \in \Delta$, as one–way channels, and derive a marginal IDF–code for each W_ω from the above IDF–code for W .

To this end, denote for $\omega \in \Delta$

$$\mathcal{N}_\omega^* = \left\{ m_\omega \,\middle|\, \exists \ m' \in \mathcal{M}(\ell^*) : \forall \gamma \in \mathcal{B}_\omega : i_\gamma = i'_\gamma \right\} \ . \tag{3.7}$$

Then it follows from (3.6) that for all $\gamma \in \mathcal{B}_\omega$

$$|\mathcal{N}_\omega^*| \geq \frac{N_\gamma}{L} \ .$$

We can assume w.l.o.g. that $L = L(\varepsilon) \geq 4$, and $N_\gamma \geq 4L$ for all $\gamma \in \Gamma$. Since $\log\left(\frac{a}{b}\right) \geq \frac{\log a}{\log b}$ for $b \geq 4$ and $a \geq 4b$, it then holds for all $\gamma \in \mathcal{B}_\omega$ that

$$\log \log |\mathcal{N}_\omega^*| \geq \log \log N_\gamma - \log \log L. \tag{3.8}$$

Let us now fix some injective mapping $\sigma : \mathcal{N}_\omega^* \to \mathcal{M}(\ell^*)$ such that for all $m_\omega = (i_\gamma)_{\gamma \in \mathcal{B}_\omega} \in \mathcal{N}_\omega^*$ it holds that

$$\sigma\big(m_\omega\big) = \big(i'_\gamma\big)_{\gamma \in \Gamma} \ \text{ iff } \ i_\gamma = i'_\gamma, \ \text{ for all } \ \gamma \in \mathcal{B}_\omega. \tag{3.9}$$

Let us now define the encoding strategies $\left\{ h^n_{m_\omega} \,\middle|\, m_\omega \in \mathcal{N}_\omega^* \right\}$ for the one–way channel

$$W_\omega : \mathcal{X} \times \{0\} \longrightarrow \{0\} \times \mathcal{Z}_\omega^n, \tag{3.10}$$

with $\mathcal{X} = \prod_{\omega' \in \Omega} \mathcal{X}_{\omega'}$, by

$$h^n_{m_\omega} = g^n_{\sigma(m_\omega)}. \tag{3.11}$$

Then $\left\{ \left(h^n_{m_\omega}, \mathcal{D}^{(\omega)}_{m_\omega} \right) : m_\omega \in \mathcal{N}_\omega^* \right\}$ forms an $\big(n, |\mathcal{N}_\omega^*|, \lambda\big)$ IDF–code for W_ω in (3.10).

Now assume that $\lambda < \frac{1}{2} \frac{\varepsilon}{1+\varepsilon}$ and apply Lemma 2 with $d = 1 + \varepsilon > 1$. Since $\sigma(m_\omega) \in \mathcal{M}(\ell^*)$, it follows from (3.5) that

$$\frac{1}{n} \log \log |\mathcal{N}_\omega^*| < (1+\varepsilon) \left(\nu_\omega^{(\ell^*)} + 2\varepsilon \right) + \varepsilon \ , \tag{3.12}$$

if $n \geq n_1(\varepsilon)$, say.

Combination of (3.8) and (3.12) gives for all $\omega \in \Delta$ and all $\gamma \in \mathcal{B}_\omega$

$$\frac{1}{n}\log\log N_\gamma < (1+\varepsilon)(\nu_\omega^{(\ell^*)} + 2\varepsilon) + \varepsilon + \frac{1}{n}\log\log L(\varepsilon).$$

Letting $n \to \infty$ and $\varepsilon \to 0$, this proves the converse part of the Main Theorem.

$\square$

Proof of Theorem 2: It can be seen from (1.15) that (3.3) does not hold in general for all communication systems (cf. two–way channel), so that $\left|\mathcal{D}^{(\omega)}_{m_\omega}\right|$ can no longer be directly related to the output–entropy $H\big(Z^n_\omega\big)$.

However, if the feedback strategies are deterministic, this functional relationship also enables us to describe $\mathcal{D}^{(\omega)}_{m_\omega}$ in such a way that (3.3) does hold. Thus the previous arguments apply and give the converse part of Theorem 2. The direct part goes by the " $\sqrt{n}$–trick " as usual.

IV. Proof of Theorem 3, Optimality of our coding scheme.

Two elementary facts from number theory are used(see e.g. [9]). The first follows from the Prime Factorisation Theorem and the second from a weak version of the Prime Number Theorem originally due to Chebyshev. Here are the statements.

Lemma 3.

(a) The number of prime divisors of an integer m *does not exceed* $\log m$

(b) The k*–th prime number* π_k *satisfies* $\pi_k = 0(k\log k)$.

It is clear from the definition (1.22) that $\varphi_k(m) = \varphi_k(\hat{m})$ *exactly if* $|m - \hat{m}| \equiv 0$ mod φ_k *. Lemma 3a) then implies a result basic for our analysis.*

Lemma 4. *For any* $m, \hat{m} \in \mathcal{M}, m \neq \hat{m}$

$$K^{-1}\left|\left\{k \in \mathcal{K} : \varphi_k(m) = \varphi_k(\hat{m})\right\}\right| \le K^{-1}\log M.$$

<u>Error performance of scheme</u>

Since the transmission is noiseless, the error probability of the first kind is zero. The total error probability of the second kind equals

$$\begin{aligned}
&Pr\left[\varphi'_\ell\big(\varphi_k(\hat{m})\big) = \varphi'_\ell\big(\varphi_k(m)\big)|\hat{m} \neq m\right]\\
&\le Pr\left[\varphi_k(\hat{m}) = \varphi_k(m)|\hat{m} \neq m\right]\\
&+ Pr\left[\varphi'_\ell\big(\varphi_k(\hat{m})\big) = \varphi'_\ell\big(\varphi_k(m)\big)|\varphi_k(\hat{m}) \neq \varphi_k(m)\right].
\end{aligned}$$

By Lemma 4 we have with $K = \lceil (\log M)^\alpha \rceil$

$$Pr\left[\varphi_k(\hat{m}) = \varphi_k(m) | \hat{m} \neq m\right] \leq \frac{1}{(\log M)^{\alpha-1}}$$

and in exact analogy

$$Pr\left[\varphi'_\ell\big(\varphi_k(\hat{m})\big) = \varphi'_\ell\big(\varphi_k(m)\big) | \varphi_k(\hat{m}) \neq \varphi_k(m)\right] \leq \frac{1}{(\log \pi_K)^{\alpha-1}}.$$

Since with M also π_K tends to infinity and since $\alpha > 1$, the total error probability tends to zero.

<u>Blocklength n of the scheme</u>

Clearly, $n = \lceil \log K \rceil + \lceil \log K' \rceil + \lceil \log \pi_{K'} \rceil$.

From (b) in Lemma 3 we have $\pi_{K'} = 0(K' \log K')$. Also, $\log K' = 0(\log \log K)$.

Therefore

$$n = [1 + o(1)] \log K = \alpha [1 + o(1)] \log \log M$$

and thus the result is proved.

Remark 8:

1.) In the method of [2] there is not our second step. However, there only the existence of appropriate keys is shown.

2.) In [6] there is also not our second step. After transmission of φ_k then $\varphi_k(m)$ is <u>transmitted directly</u> (and <u>not only identified</u>, as in our scheme). Therefore $2n$ bits are used and the rate is only $1/2$.

3.) By further iteration one can reduce blocklength slightly at the price of a larger error probability and, vice versa.

4.) In step 3, transmission could be replaced by a sub–optimal constructive ID–scheme.

References

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

[2] R. Ahlswede and G. Dueck, "Identification in the presence of feedback – a discovery of new capacity formulas", *IEEE Trans. Inform. Theory*, vol. 35, pp. 30–39, Jan. 1989.

[3] B. Verboven and E.C. van der Meulen, "Identification via a deterministic broadcast channel", to appear in the *IEEE Trans. Inform. Theory.*

[4] A.C. Yao, "Some complexity questions related to distributive computing", *Proc. 11th ACM Symposium on the Theory of Computing*, pp. 209–213, 1979.

[5] B. Halstenberg, *Zweiprozessor–Kommunikationskomplexität.* Diplomarbeit Universität Bielefeld, May 1986.

[6] K. Melhorn and E.M. Schmidt, "Las Vegas is better than Determinism in VLSI and ditributed computing", *Proc. ACM 14th Symposium on the Theory of Computing*, pp. 330–337, 1982.

[7] R. Ahlswede, "A constructive proof of the coding theorem for discrete memoryless channels with feedback", *Trans. 6th Prague Conf. on Inform. Theory, Stat. Dec. Functions, Random Proc.*, pp. 39–50, 1971.

[8] R. Ahlswede and I. Wegener, *"Search Problems".* Wiley interscience series in discrete mathematics and optimization, J. Wiley & Sons, 1987.

[9] T.M. Apostol, *Introduction to analytic number theory.* Springer–Verlag, New York, 1976.

[10] N.T. Gaarder and J.K. Wolf, "The capacity region of a multiple–access discrete memoryless channel can increase with feedback". *IEEE Trans. Inform. Theory*, vol. 21, no. 1, pp. 100–102, 1975.