LOCALIZED RANDOM AND ARBITRARY ERRORS IN THE LIGHT OF AV CHANNEL THEORY

R. Ahlswede [1], L.A. Bassalygo [2], and M.S. Pinsker [2]

1.
Universität Bielefeld
Fakultät für Mathematik
Postfach 100131
33501 Bielefeld
Germany

2.
Institute for Problems
of Information Transmission
Ermolovoistr. 19
Moscow 101447
Russia

Abstract

We introduce probabilistic communication models with localized errors and determine the optimal rates of codes, if a priori error patterns or actual errors or both occur at random according to uniform distributions. There are strong connections to the theory of AV channels.

Key words: Localized errors, arbitrary and random errors, capacities, compression lemmas, elemination and robustification techniques, arbitrary varying channels with and without side information at sender.

1. Introduction and Results

A central problem in coding theory consists in finding bounds for the maximal size, say $N(n, 2t+1, q)$, of a t–error correcting code over a q–ary alphabet with blocklength n .

This code concept is suited for communication over a q–ary channel with input and output alphabet $\mathcal{X} = \{0, 1, \dots, q-1\}$, when a word of length n sent by the encoder is changed by the channel in atmost t letters. Here neither the encoder nor the decoder knows in advance where the errors, that is changes of letters, occur.

Bassalygo, Gelfand, and Pinsker introduced in [1] the concept of *localized* errors. They assume that the encoder, who wants to encode message m , knows the t–element set $E \subset [n] = \{1, 2, \dots, n\}$ of positions, in which errors may occur. The encoder can make the codeword, representing m , dependent on $E \in \mathcal{E}_t$, the family of t–elements subsets of $[n]$.

We call them *a priori error pattern.* The set of associated (a posteriori) errors is

$$V(E) = \{e^n = (e_1, \dots, e_n) \in \mathcal{X}^n : e_t = 0 \text{ for } t \notin E\}.$$

We endow $\mathcal{X}^n$ with a group structure by adding componentwise modulo q . For a set $\mathcal{M} = \{1, 2, \dots, M\}$ of messages a family $\{u(m, E) : m \in \mathcal{M}, E \in \mathcal{E}_t\}$ of words in $\mathcal{X}^n$ is an (M, n, t, q) code, if for all $E, E' \in \mathcal{E}_t$

$$\big(u(m, E) + V(E)\big) \cap \big(u(m', E') + V(E')\big) = \varnothing \text{ for all } m \neq m'.$$

A quantity of basic interest is $M(n, t, q)$ the maximal M for which an (M, n, t, q) code exists. Rather sharp estimates for this quantity were obtained for $q = 2$ in [1] and for general q , but constant t , in [6].

Notice that there both, the a priori pattern and the a posteriori errors, occur *arbitrarily.* We refer to this model as (A, A) . Here three new models are introduced and analysed.

<u>Model (A, R) :</u> The a priori pattern $E \in \mathcal{E}_t$ occur arbitrarily and the errors occur at random according to the uniform distribution on $V(E)$.

<u>Model (R, A) :</u> The a priori pattern occur at random according to the uniform distribution on $\mathcal{E}_t$ and the errors occur arbitrarily.

<u>Model (R, R) :</u> Both events occur at random according to the previous distributions and independently.

In these probabilistic models for a message set $\mathcal{M}$ a code is specified by codewords and *decoding sets*, that is, a family of pairs

$$\Big\{\Big(\big(u(m, E)\big)_{E \in \mathcal{E}_t}, D_m\Big) : m \in \mathcal{M}\Big\} \tag{1.1}$$

where $D_m \subset \mathcal{X}^n (m \in \mathcal{M})$ and $D_m \cap D_{m'} = \varnothing (m \neq m')$. With such a code and every model we associate two kinds of error probabilities, which we call maximal and average error.

They can be described in terms of the error function $\lambda : \mathcal{M} \times \mathcal{E}_t \times \mathcal{X}^n \to \{0,1\}$, defined by

$$\lambda(m,E,e) = \begin{cases} 1 & \text{if } u(m,E) + e \notin D_m \\ 0 & \text{otherwise,} \end{cases} \tag{1.2}$$

as follows:

$$\lambda_{AR} = \max_{m \in \mathcal{M}} \max_{E \in \mathcal{E}_t} q^{-t} \sum_{e \in V(E)} \lambda(m,E,e) \tag{1.3}$$

$$\overline{\lambda}_{AR} = M^{-1} \sum_{m \in \mathcal{M}} \max_{E \in \mathcal{E}_t} q^{-t} \sum_{e \in V(E)} \lambda(m,E,e) \tag{1.4}$$

$$\lambda_{RA} = \max_{m \in \mathcal{M}} \binom{n}{t}^{-1} \sum_{E \in \mathcal{E}_t} \max_{e \in V(E)} \lambda(m,E,e) \tag{1.5}$$

$$\overline{\lambda}_{RA} = M^{-1} \sum_{m \in \mathcal{M}} \binom{n}{t}^{-1} \sum_{E \in \mathcal{E}_t} \max_{e \in V(E)} \lambda(m,E,e) \tag{1.6}$$

$$\lambda_{RR} = \max_{m \in \mathcal{M}} \binom{n}{t}^{-1} \sum_{E \in \mathcal{E}_t} q^{-t} \sum_{e \in V(E)} \lambda(m,E,e) \tag{1.7}$$

$$\overline{\lambda}_{RR} = M^{-1} \sum_{m \in \mathcal{M}} \binom{n}{t}^{-1} \sum_{E \in \mathcal{E}_t} q^{-t} \sum_{e \in V(E)} \lambda(m,E,e). \tag{1.8}$$

We denote the corresponding maximal code sizes by $M_{AR}(n,t,\varepsilon), \overline{M}_{AR}(n,t,\varepsilon)$ etc., if the respective error probabilities don't exceed ε.

In this paper we always assume that

$$t = \lfloor \tau \cdot n \rfloor \text{ for } n = 1,2,\dots, \tag{1.9}$$

that is, t is proportional to the blocklength n. Under this assumption we characterize capacities such as

$$C_{AR}(\tau) = \inf_{\varepsilon > 0} \underline{\lim}_{n \to \infty} \frac{1}{n} \log M_{AR}(n, \lfloor \tau \cdot n \rfloor, \varepsilon).$$

Since $\overline{M}_{AR} \geq M_{AR}$, $\overline{M}_{RA} \geq M_{RA}, \overline{M}_{RR} \geq M_{RR}$ and since by Chebyshev's inequality for any $\gamma \in (0,1)$

$$M_{AR}(n,t,\varepsilon) \geq (1-\gamma)\overline{M}_{AR}(n,t,\gamma\varepsilon) \tag{1.10}$$

$$M_{RA}(n,t,\varepsilon) \geq (1-\gamma)\overline{M}_{RA}(n,t,\gamma\varepsilon) \tag{1.11}$$

$$M_{RR}(n,t,\varepsilon) \geq (1-\gamma)\overline{M}_{RR}(n,t,\gamma\varepsilon), \tag{1.12}$$

we see that

$$C_{AR} = \overline{C}_{AR},\ C_{RA} = \overline{C}_{RA} \text{ and } C_{RR} = \overline{C}_{RR}. \tag{1.13}$$

Thus only three quantities have to be determined. Notice also that

$$C_{AR} \leq C_{RR}. \tag{1.14}$$

We state our results.

Theorem 1.

$$C_{RA}(\tau) = \overline{C}_{RA}(\tau) = \begin{cases} \log q - \tau \log(q-1) - h(\tau) & \textit{for } \tau < \frac{1}{2} \\ 0 & \textit{for } \tau \geq \frac{1}{2}. \end{cases}$$

In particular, for $q = 2$, $C_{RA}(\tau) = 1 - h(\tau)$ *for* $\tau \leq \frac{1}{2}$.

Remark 1. For $q \geq 3$ $C_{RA}(\tau)$ has a jump at $\tau = \frac{1}{2}$. A discontinuity in capacity formulas occurs also for AV channels (see [2]).

Theorem 2. *For all* $\tau \in [0,1]$

$$C_{AR}(\tau) = \overline{C}_{AR}(\tau) = C_{RR}(\tau) = \overline{C}_{RR}(\tau) = \log q - h\big(\tau \frac{q-1}{q}\big) - \tau \frac{q-1}{q} \log(q-1).$$

In particular, for $q = 2$, $C_{AR}(\tau) = 1 - h\big(\frac{\tau}{2}\big)$.

The paper is organized as follows:

We first address the upper bounds on C_{RA} and $\overline{C}_{RR}$ and then we establish the lower bounds for C_{RA} and C_{AR} . The derivation of the upper bounds in Sections 3 and 4 simulates the approach of [1] in conjunction with some elementary approximations. The key tools are Compression Lemmas 1 and 2 in Section 2. Whereas the first Lemma presents a familiar inequality in terms of cardinalities, the second one provides a novel entropy inequality.

Next we turn to the lower bounds. In Section 5 we present a coding procedure for the model (R, A) . It is based on the idea to devide the information given to the decoder into protocol information about the a priori pattern and into useful information about the message.

In Section 6 we establish the lower bound for the model (A,R). This completes the proofs of the capacity formulas (Theorems 1, 2) for the models with localized errors (R,A), (A,R), and (R,R).

Finally, in Sections 7 and 8 we consider the model (A,R) in the light of the theory of arbitrarily varying channels, which we shortly call AV channel theory. The model associated with the model (A,R) is denoted as (A,R,V). Here pattern $E \in \mathcal{E}_t$ takes the role of a *state sequence known to the encoder* and the communicators, the encoder and the decoder, are interested in codes with error probabilities (maximal or average), which are small for all state sequences. The appropriate error concepts are

$$\lambda_{ARV} = \max_{E \in \mathcal{E}_t} \max_{m \in \mathcal{M}} q^{-t} \sum_{e \in V(E)} \lambda(m,E,e) \tag{1.15}$$

and

$$\overline{\lambda}_{ARV} = \max_{E \in \mathcal{E}_t} M^{-1} \sum_{m \in \mathcal{M}} q^{-t} \sum_{e \in V(E)} \lambda(m,E,e). \tag{1.16}$$

Certainly, as for capacities

$$C_{AR} = C_{ARV} \leq \overline{C}_{ARV} \leq \overline{C}_{RR}, \tag{1.17}$$

Theorem 2 gives us the formulas for C_{ARV} and $\overline{C}_{ARV}$.

Corollary 1. *For all* $\tau \in [0,1]$

$$C_{ARV}(\tau) = \overline{C}_{ARV}(\tau) = \log q - h\left(\tau \frac{q-1}{q}\right) - \tau \frac{q-1}{q} \log(q-1).$$

However, we establish stronger results.

In addition to the model (A,R,V) we consider the AV channel model $(A,R,V)^*$, which is characterized by the assumption that there is *no side information about the state sequences at the encoder*. We determine its capacity $\overline{C}^*_{ARV}$ for average error.

Theorem 3. *For all* $\tau \in [0,1]$

$$\overline{C}^*_{ARV}(\tau) = \log q - h\left(\tau \frac{q-1}{q}\right) - \tau \frac{q-1}{q} \log(q-1).$$

We give two proofs of this result. The first one, presented in Section 7, relies upon special symmetry properties (such as additivity) of the channel. The second one, presented in Section 8, proceeds via general AV channel theory.

Technically, we choose to use a canonical approximation by a q–ary symmetric channel $W : \mathcal{X}^n \rightsquigarrow \mathcal{Y}^n$ with transmission matrix

$$W_q = \begin{pmatrix} 1-\frac{q-1}{q}\tau, \frac{\tau}{q}, \dots, \frac{\tau}{q} \\ \frac{\tau}{q}, 1-\frac{q-1}{q}\tau, \dots, \frac{\tau}{q} \\ \ddots \\ \frac{\tau}{q}, \dots, \frac{\tau}{q}, 1-\frac{q-1}{q}\tau \end{pmatrix}.$$

Notice that its capacity equals

$\log q - H\left(1-\frac{q-1}{q}\tau, \frac{\tau}{q}, \dots, \frac{\tau}{q}\right) = \log q - h\left(\tau\frac{q-1}{q}\right) - \tau\frac{q-1}{q}\log(q-1)$

(by the grouping axiom), our familiar quantity.

Recall that in our model (R, R) a member $E \in \mathcal{E}_t$ is choosen according to a uniform distribution p on $\mathcal{E}_t$ and then a member $e \in V(E)$ is choosen independently according to a uniform distribution q_E on $V(E)$. This generates a channel $\underline{W} : \mathcal{X}^n \rightsquigarrow \mathcal{Y}^n$. It is not a memoryless channel, but its transmission probabilities are close to those of W. Their capacities are equal also for the model $(R, R)^*$, where the *encoder has no knowledge about the a priori pattern* E.

The details are presented in an Appendix, because we can't give a reference.

Finally we derive also in Section 8 by a standard argument of [2] an important consequence of Theorem 3.

Corollary 2. *For all* $\tau \in [0, 1]$

$$C_{ARV}(\tau) = \overline{C}^*_{ARV}(\tau).$$

Remark 2. Using (1.17) we see that Theorem 3 and Corollary 2 imply Theorem 2. This result thus can also be derived with AV channel theory.

Remark 3. We draw attention to the fact that generally $C^*_{ARV}(\tau) \neq \overline{C}^*_{ARV}(\tau)$, whereas in case of side information $C_{ARV}(\tau) = \overline{C}_{ARV}(\tau)$. This phenomenon was observed in [4].

Actually we have here the same phenomenon, because

$$C^*_{ARV}(\tau) = C^*_{AR}(\tau) = \overline{\lim}_{n\to\infty}\frac{1}{n}\log M(n, \tau n), \tag{1.18}$$

where $M(n, \tau n)$ is the maximal size of an error correcting code with pairwise Hamming distances at least τn. In particular,

$$C^*_{AR}(\tau) = 0 \quad \text{for} \quad \tau \geq \frac{1}{2}.$$

2. Compression Lemmas for proving converses

It is instructive to start with the results of [1] for the model (A, A) . They can be summarized as follows:

Theorem BGP.

For the binary alphabet, that is $q = 2$,

(a) $M_{AA}(n,t) \le \dfrac{2^n}{\sum_{i=0}^t \binom{n}{i}}$

(b) $M_{AA}(n,t) \ge \frac{1}{2n} \dfrac{2^n}{\sum_{i=0}^t \binom{n}{i}}$

(c) $C_{AA}(\tau) = 1 - h\big(\min(\tau, \frac{1}{2})\big)$ *for* $0 \le \tau \le 1$.

The upper bound in (a) is Hamming's bound. It is remarkable that here it is asymptotically tight. For $q > 2$ Hamming's bound has the form (see [6])

$$M_{AA}(n,t) \le \frac{q^n}{\sum_{i=1}^t \binom{n}{i}(q-1)^i}. \tag{2.1}$$

This is a consequence of a generalisation of the Lemma in [1], which we now state and prove.

Compression Lemma 1. For any *distinct* nonempty subsets $E(i), i \in I$, of $[n]$ and any elements $u(i), i \in I$, of $\mathcal{X}^n$ we have

$$\left| \bigcup_{i\in I} u(i) + V\big(E(i)\big) \right| \ge \left| \bigcup_{i\in I} V\big(E(i)\big) \right|.$$

Proof. We partition I into $J \dot\cup K$, where $J = \big\{ i \in I : 1 \in E(i) \big\}$ and $K = I \smallsetminus J$, and define the associated sets

$$A = \bigcup_{i\in J} u(i) + V\big(E(i)\big) \text{ and } B = \bigcup_{i\in K} u(i) + V\big(E(i)\big).$$

We want to lower bound $|A \cup B|$. For this we first replace B by

$$B' = \{0\} \times \left(\bigcup_{i\in K} \big(u(i)_2, \dots, u(i)_n\big) + V'\big(E(i)\big) \right), \tag{2.2}$$

where $V'\big(E(i)\big) = \big\{ (e_2, \dots, e_n) : (0, e_2, \dots, e_n) \in V\big(E(i)\big) \big\}$. It is true that

$$|A \cup B| \ge |A \cup B'|, \tag{2.3}$$

because $|B \smallsetminus A| \geq |B' \smallsetminus A|$.

For $i \in K$ replace now $u(i) = \big(u(i)_1, \dots, u(i)_n\big)$ by $\big(0, u(i)_2, \dots, u(i)_n\big)$ and make no changes for $i \in J$. Denote the resulting words by $u'(i)(i \in I)$. The $E(i)$'s and the sets $V\big(E(i)\big)(i \in I)$ remain unchanged. Reiterate this transformation for all components $t = 1, 2, \dots, n$ until we arrive at $u''(i) \quad (i \in I)$ with $u''_t(i) = 0$, if $t \notin E(i)$, and

$$\left| \bigcup_{i \in I} u(i) + V\big(E(i)\big) \right| \geq \left| \bigcup_{i \in I} u''(i) + V\big(E(i)\big) \right| = \left| \bigcup_{i \in I} V\big(E(i)\big) \right| .$$

Suppose now that

$$\eta(\ell) = \big|\{i : i \in I, |E(i)| = \ell\}\big| \tag{2.4}$$

and that for any $F \subset [n]$

$$V^+(F) = \big\{e \in V(F) : e_t \neq 0 \text{ for all } t \in F\big\}. \tag{2.5}$$

Then by the Lemma

$$\left| \bigcup_{i \in I} \big(u(i)\big) + V\big(E(i)\big) \right| \geq \left| \bigcup_{i \in I} V^+\big(E(i)\big) \right| \geq \sum_{\ell=0}^{n} \eta(\ell)(q-1)^\ell. \tag{2.6}$$

We derive now (2.1). First notice that for any code $\big\{u(m, E) : m \in \mathcal{M}, E \in \mathcal{E}_t\big\}$ by (2.6) for every message $m \in \mathcal{M}$

$$\left| \bigcup_{E \in \mathcal{E}_t} u(m, E) + V(E) \right| \geq \binom{n}{t} (q-1)^t.$$

This implies

$$M_{AA}(n, t) \leq \frac{q^n}{\binom{n}{t}(q-1)^t}, \tag{2.7}$$

which implies already

$$C_{AA}(\tau) \leq \log_2 q - h(\tau) - \tau \log_2 (q-1). \tag{2.8}$$

However, there is a more efficient way to use (2.6). Choose any map $f : \bigcup_{i=0}^{t} \mathcal{E}_i \to \mathcal{E}_t$ with the property $E \subset f(E)$. Then we can write

$$\bigcup_{E \in \mathcal{E}_t} u(m, E) + V(E) = \bigcup_{E \in \mathcal{E}_t} \bigcup_{F : f(F) = E} u(m, E) + V(F) \tag{2.9}$$

and by (2.6) the cardinality of the set to the right is at least

$$\sum_{i=0}^{t} |\mathcal{E}_i|(q-1)^i = \sum_{i=0}^{t} \binom{n}{i}(q-1)^i.$$

This gives (2.1).

Next we address the model (R,R) . Let S be a RV with values equally distributed in $\mathcal{E}_t$ and let T_E be a RV with values equally distributed in $V(E)$ $(E \in \mathcal{E}_t)$. For any function $u : \mathcal{E}_t \to \mathcal{X}^n$ we consider the entropy $H\big(u(S)+T_S\big)$, where T_S is a RV with $Pr(T_S = e^n) = \sum_{E\in\mathcal{E}_t} Pr(S=E)\cdot Pr(T_E = e^n)$.

Compression Lemma 2.

$$H\big(u(S)+T_S\big) \geq H(T_S).$$

Proof. $u(S)+T_S$ takes its values in $\bigcup\limits_{E\in\mathcal{E}_t} u(E)+V(E)$.

Denote the distribution by P . Now we follow literally the transformations applied to u in the previous proof. In particular recall the definitions of A, B, B' and u' . Let P' denote the distribution of $u'(S)+T_S$.

Notice the following facts:

$Pr\big(u(S)+T_S = xx_2\dots x_n|1\in S\big)$ is independent of x for fixed $x_2\dots x_n$.

$Pr\big(u'(S)+T_S = xx_2\dots x_n|1\in S\big) = Pr\big(u(S)+T_S = xx_2\dots x_n|1\in S\big)$ for all $x\in\mathcal{X}$ and

$Pr\big(u'(S)+T_S = 0x_2\dots x_n|1\notin S\big) = \sum\limits_{x\in\mathcal{X}} Pr\big(u(S)+T_S = xx_2\dots x_n|1\notin S\big)$.

Therefore for every fixed $x_2\dots x_n \in \prod_2^n \mathcal{X}$

$Pr\big(u(S)+T_S \in \mathcal{X}\times\{x_2\dots x_n\}\big) = Pr\big(u'(S)+T_S \in \mathcal{X}\times\{x_2\dots x_n\}\big)$.

However on the subexperiment $\mathcal{X}\times\{x_2\dots x_n\}$ we have $\frac{P'}{P(\mathcal{X}\times\{x_2\dots x_n\})}$ is Schur dominated by $\frac{P}{P(\mathcal{X}\times\{x_2\dots x_n\})}$. By the grouping axiom and the Schur convexity of entropy $H(P)\geq H(P')$. (Schur's theory can be found in [8].)

3. Upper bound for $M_{RA}(n,t,\varepsilon)$

Observe first that (1.5) implies that for every $m \in \mathcal{M}$ there is a set $\mathcal{E}_t(m) \subset \mathcal{E}_t$ with the properties

$$|\mathcal{E}_t(m)| \geq (1-\lambda_{RA})|\mathcal{E}_t|, \tag{3.1}$$

$$\max_{e\in V(E)} \lambda(m,E,e) = 0 \text{ for } E \in \mathcal{E}_t(m) \tag{3.2}$$

or (equivalently)

$$\bigcup_{E\in\mathcal{E}_t(m)} u(m,E)+V(E) \subset D_m. \tag{3.3}$$

Moreover, the sets $\mathcal{E}_i(m) = \big\{F \in \mathcal{E}_i : F \subset E \text{ for some } E \in \mathcal{E}_t(m)\big\}$ satisfy

$$|\mathcal{E}_i(m)| \geq (1-\lambda_{RA})\binom{n}{i}, \tag{3.4}$$

because by counting containments in two ways we see that $|\mathcal{E}_{t-1}(m)| \geq |\mathcal{E}_t(m)|\frac{t}{n-t+1} \geq (1-\lambda_{RA})|\mathcal{E}_t|\frac{t}{n-t+1} = (1-\lambda_{RA})\binom{n}{t-1}$ and so on.

Choosing a map $f : \bigcup\limits_{i=0}^{t} \mathcal{E}_i(m) \to \mathcal{E}_t(m)$ with the property $F \subset f(F)$ and keeping in mind that now $\mathcal{E}_t(m)$ takes the role of $\mathcal{E}_t$ in the derivation in Section 2 we conclude from (3.3), (2.9), (2.6), and (3.4) that

$$|D_m| \geq \left| \bigcup_{E\in\mathcal{E}_t(m)} u(m,E)+V(E) \right| \geq (1-\lambda_{RA})\sum_{i=0}^{t}\binom{n}{i}(q-1)^i.$$

We summarize our findings.

Proposition 1

$$M_{RA}(n,t,\varepsilon) \leq \frac{q^n}{(1-\varepsilon)\sum_{i=0}^{t}\binom{n}{i}(q-1)^i}.$$

Next we analyse the case $\tau \geq \frac{1}{2}$.

Proposition 2 If $\tau \geq \frac{1}{2}$ and $\varepsilon < \frac{1}{2}$, then $M_{RA}(n,\tau n,\varepsilon) = 1$.

Proof. Consider a code $\{((u(m,E))_{E\in\mathcal{E}_t}, D_m) : m=1,2\}$ with two code words and maximal error probability $\lambda_{RA} < \varepsilon$. Then there are two subsets $\mathcal{E}_t(1)$ and $\mathcal{E}_t(2)$ of $\mathcal{E}_t$, which both have cardinality at least $(1-\varepsilon)\binom{n}{t}$, and such that message m is correctly decoded, if $E \in \mathcal{E}_t(m) (m=1,2)$.

Let us define
$\text{shadow}_{n-t}(\mathcal{E}_t(2)) = \{F \in \mathcal{E}_{n-t} : \exists E \in \mathcal{E}_t(2) \text{ with } F \subset E\}$. By counting containment relations we get

$$|\text{shadow}_{n-t}(\mathcal{E}_t(2))| \geq |\mathcal{E}_t(2)| \binom{t}{n-t} \binom{t}{t-(n-t)}^{-1} \tag{3.5}$$

and hence $|\text{shadow}_{n-t}(\mathcal{E}_t(2))| \geq |\mathcal{E}_2|$.

Since $|\mathcal{E}_t(1)| > \frac{1}{2}\binom{n}{t}$ and $|\text{shadow}_{n-t}(\mathcal{E}_t(2))| > \frac{1}{2}\binom{n}{t}$ there must be a pair (E,F) with $E \in \mathcal{E}_t(1)$, $F \in \text{shadow}_{n-t}(\mathcal{E}_t(2))$, and $F = E^c$.

Let $E' \in \mathcal{E}_t(2)$ contain E^c . Then $u(1,E)+V(E) \cap u(2,E')+V(E') \neq \varnothing$ in contradiction to our definitions.

Remark 4. Perhaps the following problem is unsolved. Let $\mathcal{A}, \mathcal{B} \subset \mathcal{E}_t$, $|\mathcal{A}| = |\mathcal{B}|$, satisfy $A \cup B \neq [n]$ for all $A \in \mathcal{A}, B \in \mathcal{B}$.

What is $\max|\mathcal{A}|$? The answer can be used to improve Proposition 2 to ranges $\varepsilon \geq \frac{1}{2}$.

4. Upper bound for $\overline{M}_{RR}(n,t,\varepsilon)$

Consider the code $\{(u(m,E),D_m) : m\in\mathcal{M}, E\in\mathcal{E}_t\}$ with average error $\overline{\lambda}_{RR}$.

Let Z be equidistributed with values in $\mathcal{M}$, let S_m be equidistributed with values in $\mathcal{E}_t$, and let $T_{E,m}$ be equidistributed with values in $V(E)(m\in\mathcal{M}, E\in\mathcal{E}_t)$. The RV's are independent.

Then $u(Z,S_Z)+T_{S_Z,Z}=Y^n$ describes the received sequence. By Fano's inequality

$$\log M \le \frac{H(Y^n)-H(Y^n|Z)+1}{1-\overline{\lambda}_{RR}} \tag{4.1}$$

and by the second Compression Lemma $H(Y^n|Z=m)\ge H(T_{S_m,m})$. Therefore

$$\log M \le \left[n\left(\log q - \frac{1}{n}H(T_S)\right)+1\right]\frac{1}{1-\lambda_{RR}}. \tag{4.2}$$

We calculate $H(T_S)$ first for $q=2$. For this we look at the structure of $V(E)$.

Define $V_r(E)=\{e=(e_1,\dots,e_n)\in V(E):\sum_{i=1}^n e_i=r\}$ and notice that by Chebyshev's inequality the set $V^\alpha(E)=\bigcup_{r=\frac{t}{2}-\sqrt{\frac{t}{4\alpha}}}^{\frac{t}{2}+\sqrt{\frac{t}{4\alpha}}} V_r(E)$ satisfies

$$|V^\alpha(E)|\ge(1-\alpha)\cdot 2^t, \alpha>0. \tag{4.3}$$

Now $H(T_S)\ge(1-\alpha)H\left(T_S|T_S\in\bigcup_E V^\alpha(E)\right)$ and since elements in $\bigcup_{E\in\mathcal{E}_t}V_r(E)$ are equiprobable

$$
\begin{aligned}
&H\left(T_S|T_S\in\bigcup_E V^\alpha(E)\right)\\
&\ge \sum_{r=\frac{t}{2}-\sqrt{\frac{t}{4\alpha}}}^{\frac{t}{2}+\sqrt{\frac{t}{4\alpha}}} \mathrm{Prob}\left(T_S\in\bigcup_E V_r(E)\right) H\left(T_S|T_S\in\bigcup_E V_r(E)\right)\cdot \mathrm{Prob}\left(T_S\in\bigcup_E V^\alpha(E)\right)^{-1}\\
&\ge \min_{\frac{t}{2}-\sqrt{\frac{t}{4\alpha}}\le r\le\frac{t}{2}+\sqrt{\frac{t}{4\alpha}}} H\left(T_S|T_S\in\bigcup_E V_r(E)\right)=\log\binom{n}{\frac{t}{2}-\sqrt{\frac{t}{4\alpha}}}=n\big(h(\tau/2)-g(\tau,\alpha)\big)
\end{aligned}
$$

for $n\ge n_0(\tau,\alpha)$ and a function g with $\lim_{\alpha\to 0}g(\tau,\alpha)=0$.

We give now our result for general q .

Proposition 3
$\overline{C}_{RR}(\tau) \leq \log q - h\big(\tau \cdot \frac{q-1}{q}\big) - \tau\frac{q-1}{q}\log(q-1)$ for all $\tau \in [0,1]$.

Proof.
Set $V_{r_0,\dots,r_{q-1}}(E) = \big\{e = (e_1,\dots,e_n) \in V(E) : x$ occurs r_x times in $e, x \in \mathcal{X}\big\}$ and set

$$V^{\alpha}(E) = \bigcup_{\frac{t}{q}-\sqrt{\frac{t}{\alpha}}\leq r_i \leq \frac{t}{q}+\sqrt{\frac{t}{\alpha}}} V_{r_0,\dots,r_{q-1}}(E).$$

Then $|V^{\alpha}(E)| \geq (1-\alpha)q^t$ follows by applying Chebychev's inequality q times. The previous arguments extend to this general case.

5. A coding scheme for the model (R, A)

We need an auxiliary result.

Covering Lemma ([3]).

For a hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ there is a covering $\mathcal{C}, \mathcal{C} \subset \mathcal{E}$, of the vertex set $\mathcal{V}$ with

$$|\mathcal{C}| \leq \lceil |\mathcal{E}| d^{-1} \log |\mathcal{V}| \rceil,$$

where $d = \min_{v \in \mathcal{V}} |\{E \in \mathcal{E} : v \in E\}|$.

Corollary. Let $t < \ell < n$ be positive integers. For the hypergraph $\mathcal{H} = \left(\binom{[n]}{t}, \binom{[n]}{\ell} \right)$ there is a covering $\mathcal{C}_\ell \subset \binom{[n]}{\ell}$ with

$$|\mathcal{C}_\ell| \leq \binom{n}{t} \binom{\ell}{t}^{-1} \cdot n.$$

Proof.

Since $|\mathcal{E}| d^{-1} \log |\mathcal{V}| = \binom{n}{\ell} \binom{n-t}{\ell-t}^{-1} \log \binom{n}{t} \leq \binom{n}{t} \binom{\ell}{t}^{-1} n$, the result follows from the Covering Lemma.

The guiding idea in deriving a lower bound on $M_{RA}(n,t,\varepsilon)$ is based on the following calculation for the "useful information". Choose a function $g : \mathcal{E}_t \to \mathcal{C}_\ell$ with the property $g(E) \supset E$. The encoder, knowing E , also knows $g(E)$. Now, if the decoder would also know $g(E)$, then the communicators could transmit $M = q^{n-\ell}$ messages. However, since $g(E)$ is not known to the decoder, $|\mathcal{C}_\ell|$ of these messages must be reserved for the "protocol" and there are only $M|\mathcal{C}_\ell|^{-1} \geq q^{n-\ell} \binom{\ell}{t} \binom{n}{t}^{-1} n^{-1}$ (see corollary) "useful messages". An elementary calculation shows that this expression attains its maximum for $\ell = \frac{q}{q-1} t$. Since $q^{\frac{q}{q-1}t} \approx (q-1)^t \binom{\frac{q}{q-1}t}{t}$, its value is

$$q^{n-\ell} \binom{\ell}{t} \binom{n}{t}^{-1} n^{-1} = \frac{q^n \binom{\frac{q}{q-1}t}{t}}{n \, q^{\frac{q}{q-1}t} \binom{n}{t}} \approx \frac{q^n}{(q-1)^t \binom{n}{t}} \tag{5.1}$$

and (in rate) corresponds to the Hamming bound.

How can the information be coded?

1.) Write the blocklength n in the form $n = m_0 + m = m_0 + m_1 + \cdots + m_r$, where

$$m_i = \left\lfloor \frac{m}{r} \right\rfloor \text{ or } \left\lceil \frac{m}{r} \right\rceil \quad \text{for } i = 1, \ldots, r$$

and m_0 and r are specified later.

Furthermore define $B_0 = [1, m_0]$ and for $i \geq 1$ $B_i = \left[\sum_{j=0}^{i-1} m_j + 1, \sum_{j=0}^{i} m_j\right]$. Set

$$E_i = B_i \cap E.$$

The encoder, knowing E , knows also the sets E_i and he orders the intervals $B_i (i = 1, \dots, r)$ as $B_{i_1}, \dots, B_{i_r}$ according to increasing cardinalities $t_i = |E_i|$ and, in cases of ties, according to increasing i's .

2.) For any $\gamma > 0$ with $\tau + \gamma < \frac{1}{2}$ the randomly chosen E has with a probability at least

$$1 - \exp\{-c(\gamma) m_0\},$$

where $c(\gamma) > 0$, the property

$$|E_0| \leq (\tau + \gamma) m_0. \tag{5.2}$$

If (5.2) is violated, an error is declared. With increasing m_0 this error probability tends to zero.

3.) By Theorem BGP there is a code over the interval B_0 , which uses only the letters 0 and 1 and has size $M_{AA}\big(m_0, (\tau + \gamma) m_0\big) \geq \dfrac{1}{2\ m_0} \dfrac{2^{m_0}}{\sum_{i=0}^{(\tau+\gamma)m_0} \binom{m_0}{i}}$.

This code is used to inform the decoder about the order defined above and about the values $t_{i_1}, \dots, t_{i_r}$. This requires at most $r! t^r$ messages. Furthermore, this code is used to inform the decoder about $E_{i_1} \in \binom{B_{i_1}}{t_{i_1}}$. Clearly, a total of $M_1 = r! t^r \cdot 2^{\lceil \frac{m}{r} \rceil}$ messages suffices for all three purposes.

Therefore $\log M_1 \leq r \log r + r \log t + \lceil \frac{m}{r} \rceil$ and with the choice

$$r = \sqrt{\frac{n}{\log n}} \tag{5.3}$$

we obtain

$$\log M_1 \leq 2\ r\ \ \log n + \frac{n}{r} = 3 \sqrt{n\ \log n}.$$

On the other hand $\log M_{AA}\big(m_0, (\tau + \gamma) m_0\big) \geq m_0 \big(1 - h(\tau + \gamma)\big) \cdot \frac{1}{2} \geq 3 \sqrt{n\ \log n}$, if $m_0 = c_1(\tau, \gamma) \sqrt{n\ \log n}$ with a sufficiently large constant $c_1(\tau, \gamma)$.

4.) Apply the Corollary to each interval

$$B_{i_1}, B_{i_2}, \dots, B_{i_r}.$$

In interval B_0 the decoder was informed about $E_{i_1} \subset B_{i_1}$ and thus also about $g_1(E_{i_1})$. In the positions $\big(B_{i_1} \smallsetminus g_1(E_{i_1})\big) \cup \big(B_{i_2} \smallsetminus g_2(E_{i_2})\big) \cup \big(B_{i_3} \smallsetminus g_3(E_{i_3})\big) \cup \dots \cup \big(B_{i_r} \smallsetminus g_r(E_{i_r})\big)$ the decoder will be informed successively about $g_2(E_{i_2}), g_2(E_{i_3}), \dots$. Since the cardinalities $\ell_{i_j} = \frac{q}{q-1} t_{i_j}$ increase, this is possible. The information about $g_j(E_{i_j})$ is given before we start in $B_{i_j} \smallsetminus g_j(E_{i_j})$. After the total protocol information is conveyed the decoder will get the useful information in the remaining free positions.

The attainable number of useful messages exceeds

$$\prod_{j=1}^{r} \frac{q^{m_{i_j}}}{(q-1)^{t_{i_j}} \binom{m_{i_j}}{t_{i_j}}} \geq \frac{q^{n-m_0}}{(q-1)^{t-t_0} \binom{n-m_0}{t-t_0}},$$

because as in (5.1) $q^{m_{i_j}-\ell_{i_j}} \binom{\ell_{i_j}}{t_{i_j}} \binom{m_{i_j}}{t_{i_j}}^{-1} m_{i_j}^{-1} \approx \frac{q^{m_{i_j}}}{(q-1)^{t_{i_j}} \binom{m_{i_j}}{t_{i_j}}}$ and because $\binom{a+b}{c+d} \geq \binom{a}{c} \cdot \binom{b}{d}$.

With our choices of m_0 and t_0 it follows that for any $\varepsilon > 0$

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log M_{RA}(n, \tau n, \varepsilon) \geq \log q - \tau \log(q-1) - h(\tau).$$

This and Propositions 1, 2 imply Theorem 1.

6. A coding scheme for the model (A, R)

We begin with an auxiliary result.

Lemma 3. *For any* $A, B \subset \mathcal{X}^n$ *and* $\alpha > 0$

$$|\{x^n \in \mathcal{X}^n : |(x^n + A) \cap B| \geq \alpha |A|\}| \leq \frac{1}{\alpha}|B|.$$

Proof. $|A|\ |B| = \sum_{x^n \in \mathcal{X}^n} |(x^n + A) \cap B| \geq |\{x^n \in \mathcal{X}^n : |(x^n + A) \cap B| \geq \alpha |A|\}| \alpha |A|$.

We present next the key result.

Lemma 4. *For any* $T, A_1, \ldots, A_I \subset \mathcal{X}^n$, $\alpha \in (0,1)$ *and*

$$M \leq \frac{\alpha\ q^n}{|T| I^{\frac{1}{n}} nq}$$

there exists a sequence of families $\big(\mathcal{C}(1), \ldots, \mathcal{C}(M)\big)$ *with* n *members from* $\mathcal{X}^n$ *each and with the property:*

For all i *and all* $m (1 \leq i \leq I, 1 \leq m \leq M)$

$$\min_{c \in \mathcal{C}(m)} |(c + A_i) \cap \bigcup_{c' \in \cup_{m' \neq m} \mathcal{C}(m')} (c' + T)| \leq \alpha |A_i|. \tag{6.1}$$

Proof. We consider all sequences of families $\big(\widetilde{\mathcal{C}}(1), \ldots, \widetilde{\mathcal{C}}(M)\big)$ with n members from $\mathcal{X}^n$ each. The number of such sequences equals $q^{n^2 \cdot M}$. We call such a sequence bad, if for some i and some m

$$\min_{c \in \widetilde{\mathcal{C}}(m)} |(c + A_i) \cap \bigcup_{c' \in \cup_{m' \neq m} \widetilde{\mathcal{C}}(m')} (c' + T)| > \alpha |A_i|.$$

Using Lemma 3 with choices

$$B = \bigcup_{c' \in \cup_{m' \neq m} \widetilde{\mathcal{C}}} (c' + T) \ \text{ and } \ A = A_i$$

we see that the number of bad sequences does not exceed

$$I \cdot M \cdot q^{n(M-1)n} \left(\frac{|T|(M-1)n}{\alpha} \right)^n = \left(\frac{I^{\frac{1}{n}} M^{\frac{1}{n}} |T|(M-1)n}{\alpha\ q^n} \right)^n \cdot q^{n^2 M}.$$

This is smaller than the total number of sequences $q^{n^2 M}$, because by our assumption on M the first factor is smaller than 1. There exists a good sequence of families.

We describe now our coding scheme.
Write $\mathcal{E}_t = \{E_1, \dots, E_I\}$, $I = \binom{n}{t}$, and define

$$A_i = V(E_i) \cap B\left(n, \frac{q-1}{q}t + \varepsilon t\right), \tag{6.2}$$

where $B(n,r)$ denotes the Hamming ball of radius r in $\mathcal{X}^n$ around the origin.

It is well–known that

$$\left|B\left(n, \frac{q-1}{q}t + \varepsilon t\right)\right| = \sum_{j \le \frac{q-1}{q}t + \varepsilon t} (q-1)^j \binom{n}{j} \sim h\left(\tau \frac{q-1}{q}\right) - \tau \frac{q-1}{q} \log(q-1) \tag{6.3}$$

and that

$$|A_i||V(E_i)|^{-1} \sim 1. \tag{6.4}$$

Apply now Lemma 4 with $T = B\left(n, \frac{q-1}{q}t + \varepsilon t\right)$ and A_i as in (6.2). The bound on M is the desired Hamming bound in rate.

Choose as codeword $u(m, E_i)$ a member from $\mathcal{C}(m)$ for which the minimum in (6.1) is assumed and choose as decoding set for message m

$$D_m = \bigcup_{i=1}^{I} \big(u(m, E_i) + A_i\big) \smallsetminus \bigcup_{c' \in \cup_{m' \neq m} \mathcal{C}(m')} (c' + T).$$

The maximal decoding error probability can be made arbitrarily small, since α in (6.1) can be made arbitrarily small and since (6.4) holds. The disjointness of the decoding sets is guaranteed by our definitions.

7. The first proof of Theorem 3

By exchanging summations we can write the average error probability in the form

$$\overline{\lambda}^*_{ARV} = \max_{E\in\mathcal{E}_t} q^{-t} \sum_{e\in V(E)} \frac{1}{M} \sum_{m=1}^{M} \lambda(m,E,e). \tag{7.1}$$

Translations invariance of the transmission probabilities of our channel allow a simple analysis of the term $\frac{1}{M}\sum\limits_{m=1}^{M} \lambda(m,E,e)$ in a random ensemble of codes.

Lemma 5. *There exists a code* $\mathcal{U} = \{u_1,\dots,u_M\} \subset \mathcal{X}^n$ *with*

(a) $M \geq \left\lfloor \frac{q^n}{n^2|B(n,\ell)|} \right\rfloor$, $1 \leq \ell \leq \frac{q-1}{q} n - q\, n^{\frac{2}{3}}$,

and

(b) $|\{u \in \mathcal{U} : d(u+e, \mathcal{U} - \{u\}) \leq \ell\}| \leq \frac{M}{n}$

for all $e = (e_1,\dots,e_n)$ *of weight less than* ℓ *.*

(Here d *denotes the Hamming distance and* $B(n,\ell)$ *denotes again the Hamming ball of radius* ℓ *and the origin as center.)*

Proof. The number of families $\mathcal{U} = \{u_1,\dots,u_M\}$ with members from $\mathcal{X}^n$ equals q^{nM} . A family is called bad *for a fixed* $e \in B(n,\ell)$, if $|\{u \in \mathcal{U} : d(u+e,\mathcal{U} - \{u\}) \leq \ell\}| > \frac{M}{n}$.

Clearly,

$$\begin{aligned} |\{u \in \mathcal{U} : d(u+e,\mathcal{U} - \{u\}) \leq \ell\}| &\leq |\{u_m \in \mathcal{U} : d(u_m + e, \{u_1,\dots,u_{m-1}\}) \leq \ell\}| \\ &\quad + |\{u_m \in \mathcal{U} : d(u_m + e, \{u_{m+1},\dots,u_M\}) \leq \ell\}| \end{aligned}$$

and so for a bad $\mathcal{U}$

$|\{u_m : d(u_m + e, \{u_1,\dots,u_{m-1}\}) \leq \ell\} \geq \frac{M}{2n}$ or

$|\{u_m : d(u_m + e, \{u_{m+1},\dots,u_M\}) \leq \ell\}| \geq \frac{M}{2n}$.

For each of these two cases the number of families realizing it does not exceed

$$\binom{M}{\frac{M}{2n}} \big(M \cdot b(n,\ell)\big)^{\frac{M}{2n}} q^{n\left(M - \frac{M}{2n}\right)},$$

if we use the notation $b(n,\ell) = |B(n,\ell)|$, and the total number of $\mathcal{U}$'s , which are bad for *any* e , therefore does not exceed

$$b(n,\ell)2\binom{M}{\frac{M}{2n}} \left(\frac{Mb(n,\ell)}{q^n}\right)^{\frac{M}{2n}} \cdot q^{nM}.$$

We have choosen ℓ and M judicially, so that this quantity is smaller than q^{nM} . A good $\mathcal{U}$ exists.

Now Theorem 3, that is, its direct part, is readily established. The choice $\ell = \frac{q-1}{q}t + \varepsilon t$ insures that (a) gives asymptotically the desired bound on the rate. As decoding rule we use a minimal distance decoding, that is, we define

$$D_m = \big(u_m + B(n,\ell)\big) \smallsetminus \bigcup_{m' \neq m} \big(u_{m'} + B(n,\ell)\big).$$

Then

$$\lambda(m,E,e) = \lambda(m,e) = \begin{cases} 1 & \text{if } u_m + e \notin D_m \\ 0 & \text{otherwise} \end{cases}$$

and for $e \in V(E) \cap B(n,\ell)$ by (b)

$$\frac{1}{M}\sum_{m=1}^{M} \lambda(m,E,e) = \frac{1}{M}\big|\big\{u \in \mathcal{U} : d(u+e, \mathcal{U} - \{u\}) \leq \ell\big\}\big| \leq \frac{1}{n}.$$

Furthermore, $|V(E) \cap B(n,\ell)|\ |V(E)|^{-1} \to 1$ as $n \to \infty$ and we see from (7.1) that $\overline{\lambda}^{*}_{ARV} \to 0$ as $n \to \infty$.

8. The second Proof of Theorem 3 and derivation of Corollary 2

We derive first the inequality $\overline{C}^*_{ARV}(\tau) \geq C^*_{RR}(\tau)$. The reverse inequality is obvious.

The proof uses methods from the theory of AV–channels, namely, a simple version of the robustification technique and a novel version of the elimination technique, which is based on several ensembles of codes.

An AVC is defined here by a sequence $\mathcal{A} = \big(\{w(\cdot|\cdot|s^n) : s^n \in \mathcal{S}^n\}\big)_{n=1}^{\infty}$ of sets of transmission probabilities, where for a finite input alphabet $\mathcal{X}$, a finite output alphabet $\mathcal{Y}$ and a finite set $\big\{w(\cdot|\cdot|s) : s \in \mathcal{S}\big\}$ of stochastic $|\mathcal{X}| \times |\mathcal{Y}|$–matrices

$$w(y^n|x^n|s^n) = \prod_1^n w(y_t|x_t|s_t) \tag{8.1}$$

for all $x^n = (x_1, \dots, x_n) \in \mathcal{X}^n = \prod_1^n \mathcal{X}$, for all $y^n \in \mathcal{Y}^n$, and for all $s^n \in \mathcal{S}^n$.

In case of $(A,R,V)^*$ it is appropriate to choose $\mathcal{S} = \{0,1\}$, $w(\cdot|\cdot|0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $w(\cdot|\cdot|1) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$, and to replace $\mathcal{S}^n$ by $\mathcal{S}^n(\tau) = \left\{ s^n \in \mathcal{S}^n : \sum_{i=1}^n s_i = \tau n \right\}$.

Consider now the symmetric group (the set of all permutations) Σ_n acting on $\{1,2,\dots,n\}$. We then define for $s^n \in \mathcal{S}^n(\tau)$, $A \subset \mathcal{S}^n(\tau)$ and $\pi \in \Sigma_n$

$$\pi s^n = \pi(s_1, \dots, s_n) = \big(s_{\pi(1)}, \dots, s_{\pi(n)}\big), \tag{8.2}$$

$$\pi(A) = \{\pi s^n : s^n \in A\}. \tag{8.3}$$

Robustification Lemma.

If $g : \mathcal{S}^n(\tau) \to \mathbb{R}$ satisfies for a $\beta \in \mathbb{R}$ the inequality

$$\binom{n}{t}^{-1} \sum_{s^n \in \mathcal{S}^n(\tau)} g(s^n) < \beta,$$

then it satisfies also the inequality

$$\frac{1}{n!} \sum_{\pi \in \Sigma_n} g(\pi s^n) < \beta \text{ for all } s^n \in \mathcal{S}^n(\tau).$$

Proof. Since $\pi : \mathcal{S}^n(\tau) \to \mathcal{S}^n(\tau)$ is bijective, the first inequality is equivalent to $\binom{n}{t}^{-1} \sum_{s^n \in \mathcal{S}^n(\tau)} g(\pi s^n) < \beta$ for $\pi \in \Sigma_n$ and thus $\binom{n}{t}^{-1} \frac{1}{n!} \sum_{s^n \in \mathcal{S}^n(\tau)} \sum_{\pi \in \Sigma_n} g(\pi s^n) < \beta$.

Since $\sum_{\pi \in \Sigma_n} g(\pi s^n)$ does not depend on s^n , we conclude that $\frac{1}{n!} \sum_{\pi \in \Sigma_n} g(\pi s^n) < \beta$.

We use now for any finite set $\mathcal{Z}$ the following notions: $\mathcal{P}(\mathcal{Z}) \triangleq$ set of all distributions on $\mathcal{Z}$, $\mathcal{P}(n, \mathcal{Z}) \triangleq \{P \in \mathcal{P}(\mathcal{Z}) : P(z)n$ is integer for all $z \in \mathcal{Z}\}$, $z^n = (z_1, \dots, z_n)$ is said to be (P, δ)–typical , if $\|\{i : z_i = z\}| - P(z)n| \leq \delta n$ for all $z \in \mathcal{Z}$, and $\mathcal{Z}^n(P, \delta)$ is the set of those sequences in $\mathcal{Z}^n$. $(P, 0)$–typical sequences are also said to be of type P . Often, if the reference set is clear, with a hint to typicality $\mathcal{Z}^n(P, \delta)$ is written as $\mathcal{T}^n_{P,\delta}$. For $\delta = 0$ we omitt the δ ,

We turn now to the channel $\underline{W}$. For codes of this channel we shall apply the Robustification Lemma. Codes with desired properties are obtained by approximation with the q–ary symmetric channel W . Both channels are defined in Section 1. It is well–known that for any $P \in \mathcal{P}(n, \mathcal{X})$ and $Q = P \cdot W$ we have codes $\{(u_i(P), D_i(P)) : 1 \leq i \leq M\}$ for the q–ary symmetric channel W of rate $R = I(P|W) - \delta_1$ (where $I(P|W)$ is the mutual information in terms of input distribution P and channel W) with the properties

(i) $u_i(P) \in \mathcal{T}^n_P$ for $i = 1, \dots, M$

(ii) $D_i(P) \subset \mathcal{T}^n_{Q,\delta_2}$

(iii) $W(D^c_i(P)|u_i(P)) \leq \exp\{-f(R, \delta_2)n\}$ for $i = 1, \dots, M$

for arbitrarily small $\delta_1, \delta_2 > 0$ and $f(R, \delta_2) > 0$.

By the Approximation Lemmas 1, 2 in the Appendix (see Remark 5) we have also

(iv) $V(D^c_i(P)|u_i(P)) < \exp\{-\frac{1}{2} f(R, \delta_2)n\}$ for $i = 1, \dots, M$ and $P \in \mathcal{P}(n, \mathcal{X})$ and thus codes for V .

Define now

$$g(s^n) = \frac{1}{M} \sum_{i=1}^{M} w\big(D^c_i(P)|u_i(P)|s^n\big), s^n \in \mathcal{S}^n(\tau),$$

and notice that

$$\binom{n}{t}^{-1} \sum_{s^n \in \mathcal{S}^n(\tau)} g(s^n) = \frac{1}{M} \sum_{i=1}^{M} V\big(D^c_i(P)|u_i(P)\big) < \beta$$

if $\beta \triangleq \exp\{-\frac{1}{2} f(R, \delta_2)n\}$.

By the Robustification Lemma we know that

$$\frac{1}{n!}\sum_{\pi\in\Sigma_n}\frac{1}{M}\sum_{i=1}^{M} w\big(D_i^c(P)|u_i(P)|\pi s^n\big) < \beta \tag{8.4}$$

and by the permutation invariance of a DMC

$$\frac{1}{n!}\sum_{\pi\in\Sigma_n}\left[\frac{1}{M}\sum_{i=1}^{M} w\big(\pi\big(D_i^c(P)\big)|\pi u_i(P)|s^n\big)\right] < \beta \tag{8.5}$$

for all $s^n\in\mathcal{S}^n(\tau)$.

We have arrived at a random (or correlated) code for $(A,R,V)^*$, namely the collection of deterministic codes $\big\{\big(\pi u_i(P),\pi(D_i(P))\big)_{1\le i\le N} : \pi\in\Sigma_n\big\}$ together with the equidistribution μ on Σ_n . Choose now L codes from this collection at random according to μ and associate with it the equidistribution $\widetilde{\mu}$ on $\{1,2,\dots,L\}$. This results in a new random code $\big(\big\{\big(u_i^\ell(P),D_i^\ell(P)\big)_{1\le i\le M} : \ell\in\{1,2,\dots,L\}\big\},\widetilde{\mu}\big)$.

It was shown in [2] that the probability, that for fixed s^n this random selection fails to lead to a new random code with an average error probability less than λ , is smaller than $e^{-\alpha\lambda L}(1+e^{\alpha}\beta)^L$ (for any $\alpha>0$). Therefore the probability, that it fails for any $s^n\in\mathcal{S}^n(\tau)$, is smaller than

$$|\mathcal{S}^n(\tau)|e^{-\alpha\lambda L}(1+e^{\alpha}\beta)^L.$$

For the choice $\alpha=2$ and $L=n^2$ (as in [2]) the quantity is strictly smaller than 1. This can also be achieved (see [7]) with a constant

$$L > \left(\frac{1}{2}f(R,\delta_2)\lambda\right)^{-1}\log|\mathcal{S}|. \tag{8.6}$$

In [2] a very small blocklength was reserved to transmit the index ℓ of the deterministic code choosen now at random with $\widetilde{\mu}$. For this only positivity of the capacity is necessary.

Presently, however, because of lack of knowledge of s^n (or E) such a time sharing argument is not possible. The new trick is to use L different input distributions $P_1,\dots,P_L$ with $I(P_\ell|W)$ close to C_{RR} , but such that for $Q_\ell=P_\ell\cdot W$ and δ sufficiently small we have

$$\mathcal{T}^n_{Q_\ell,\delta}\cap\mathcal{T}^n_{Q_{\ell'},\delta}=\varnothing(\ell\neq\ell'). \tag{8.7}$$

To each input distribution P_ℓ we use a code with the properties described in (i), ... ,(iv). Then we produce for each P_ℓ a random code as in (8.5). Next we choose at random *one* code *from each* of the L ensembles and form again a random code, say

$$\left(\left\{\big(\widetilde{u}_i(P_\ell),\widetilde{D}_i(P_\ell)\big)_{1\le i\le M} : \ell=1,\dots,L\right\},\widetilde{\mu}\right).$$

The derivation of [2] applies again with a small modification: the choices described by independent, identically distributed RV's are now described by independent but not identically distributed RV's with the same bound β on the expected values. Bernstein's form of Chebyshev's inequality literally also applies in this case.

But now we know that

$$\bigcup_{i=1}^{M} \widetilde{D}_i(P_\ell) \subset \mathcal{T}^n_{Q_\ell,\delta} \tag{8.8}$$

and by (8.5) we can form a code with randomized *encoding* only, namely $(\rho_i, D_i)_{1\leq i\leq M}$, where

$$D_i = \bigcup_{\ell=1}^{L} \widetilde{D}_i(P_\ell), \tag{8.9}$$

and $\rho_i \in \mathcal{P}(\mathcal{X}^n)$ with

$$\rho_i\big(\widetilde{u}_i(P_\ell)\big) = \frac{1}{L} \text{ for } \ell = 1, \ldots, L. \tag{8.10}$$

The error probability is bounded by λ again. By Theorem 3 of [2] there is a deterministic code with average error and essentially the same performance.

Finally the reader easily verifies that distributions $P_1, \ldots, P_L$ with the desired properties can be found in the neighbourhood of $\left(\frac{1}{q}, \ldots, \frac{1}{q}\right)$. This proves Theorem 3.

One way to verify Corollary 2 is to start with the code $(\rho_i, D_i)_{1\leq i\leq M}$ specified in (8.8) to (8.10). Since *now the encoder knows the a priori error patterns* E we can replace via the pigeon–hole principle the randomized encoding ρ_i by deterministic code words $u(i, E)$ $(1 \leq i \leq M; E \in \mathcal{E}_t)$ without increasing the *maximal* error probabilities.

Alternatively, we can pass from a code with average error to a code with maximal error, but randomized encoding, via Theorem 3 of [2]. Thus we derive the Corollary directly from our Theorem 3.

We sketch here observations and ideas. Some of them require for their understanding familiarity with advanced parts of AVC theory or the willingness to move deeply into the subject. Readers without these prerequisites may just skip this section.

a.) **Defects**

Suppose that in the positions of $E \in \mathcal{E}_t$ the transmission is governed by $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and in $[n] \smallsetminus E$ by $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. In case (R, R) (the second R is meaningless and only kept in order to stick to our terminology) the channel can be approximated by a channel with random parameters in the sense of [5]. We get $C_{RR}(\tau) = 1 - \tau$ and by the approach of [4] $C_{AR}(\tau) = 1 - \tau$. This shows that this classical result also is included in AVC theory.

In case of no side information the approximating DMC has transmission matrix $\begin{pmatrix} 1-\tau & \tau \\ 0 & 1 \end{pmatrix}$. An elementary calculation shows that here

$$C^*_{RR}(\tau) < C_{AR}(\tau).$$

b.) Replace the matrix $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ by $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ in the preceding discussion. Obviously $C_{AA}(\tau) = C_{RR}(\tau) = 1$, because the sender can switch letters in the positions of E. However, if the sender does not know the pattern E, then by AVC theory

$$C^*_{RR}(\tau) = \overline{C}^*_{ARV} = 1 - h(\tau).$$

c.) **Two kinds of defects**

Let $w(\cdot|\cdot|0) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, $w(\cdot|\cdot|1) = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 0 \end{smallmatrix}\right)$, $w(\cdot|\cdot|2) = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, $q(0) = 1 - \tau_1 - \tau_2, q(1) = \tau_1$, and $q(2) = \tau_2$ define a channel with randomized parameters. The capacity formula of [5] is

$$C_q = \max_{(U,S,X) \in \mathcal{R}_q} \big[I(U \wedge Y) - I(U \wedge S) \big],$$

where $|\mathcal{U}| \leq |\mathcal{X}| + |\mathcal{S}|$, $\mathcal{R}_q$ is the set of triples of RV's (U, S, X) with

$$P_{USXY}(u, s, x, y) = P_{USX}(u, s, x) w(y|x|s),$$

and $I(U \wedge Y)$ is the mutual information for the RV's U, V etc. Set now $\mathcal{U} = \mathcal{Y}, \mathcal{Y}_s = \big\{ y : w(y|x, s) = 1 \text{ for some } x \in \mathcal{X} \big\}$, and

$$P_{U|S}(u|s) = \begin{cases} |\mathcal{Y}_s|^{-1} & \text{for } u \in \mathcal{Y}_s \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, let $P_{X|SU}$ be deterministic, that is, a function $x : \mathcal{S} \times \mathcal{Y} \to \mathcal{X}$ with the property $w\big(u|x(u, s)|s\big) = 1$ for $u \in \mathcal{Y}_s$ and arbitrary otherwise. We have $U = Y$ with probability 1, $P_U(u) = \sum_s P_{U|S}(u|s) q(s)$, and thus

$$I(U \wedge Y) - I(U \wedge S) = H(U) - H(U) + H(U|S) = \sum_{s \in \mathcal{S}} q(s) \log |\mathcal{Y}_s|.$$

Since $\mathcal{Y}_0 = \mathcal{Y}$, $\mathcal{Y}_1 = \{0\}$, $\mathcal{Y}_2 = \{1\}$ we get

$$I(U \wedge Y) - I(U \wedge S) = q(0) = 1 - \tau_1 - \tau_2 .$$

Again by AVC theory this gives the coding theorem for two types of defects known to the encoder.

d.) If under the conditions of c.) we drop the side information, then for average error (again by AVC theory) the capacity $\overline{C}^*_{ARV}$ equals that of the DMC with transmission matrix

$$\begin{pmatrix} \tau_0 + \tau_1 & \tau_2 \\ \tau_1 & \tau_0 + \tau_2 \end{pmatrix} .$$

e.) If in the previous case only τ_0 is known, then $\overline{C}^*_{ARV}$ equals $1 - h\left(\frac{\tau}{2}\right)$, where $1 - \tau = \tau_0$. The case of side information suggests to consider AV channels with *partial side information*: for the positions in E $\left(\begin{smallmatrix} 1 & 0 \\ 1 & 0 \end{smallmatrix}\right)$ or $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ govern the transmission.

f.) If $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ occurs $n - t$ times and in the remaining t positions $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ or $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ occurs arbitrarily and there is no side information, then this AV channel has average error capacity $1 - h(\tau)$. However, if we replace here the matrices $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ by the matrices $\left(\begin{smallmatrix} 1 & 0 \\ 1 & 0 \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, *that is we describe errors differently*, then the average error capacity equals $1 - h\left(\frac{\tau}{2}\right)$.

g.) The same answer is obtained, if we replace the two matrices by the single matrix $\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$.

Appendix: Approximation by a discrete memoryless channel

We show now how our channel $\underline{W} : \mathcal{X}^n \rightsquigarrow \mathcal{Y}^n$ relates to the q–ary symmetric channel W , with transmission matrix W_q . Both are defined in Section 1. To simplify matters we treat the case $q = 2$. Since both, $\underline{W}$ and W are translation invariant, that is,

$$\underline{W}(y^n + a^n | x^n + a^n) = \underline{W}(y^n | x^n) \text{ and } W(y^n + a^n | x^n + a^n) = W(y^n | x^n),$$

it suffices to consider the input $x^n = \underline{o} = (o, \dots, o)$.

For $y^n = (y_1, \dots, y_n)$ with $\sum_{i=1}^{n} y_i = s$, $s \leq t$, we have

$$\underline{W}(y^n | \underline{o}) = \binom{n-s}{t-s} \binom{n}{t}^{-1} \cdot \left(\frac{1}{2}\right)^t = \binom{t}{s} \binom{n}{s}^{-1} \left(\frac{1}{2}\right)^t \tag{A.1}$$

$$W(y^n | \underline{o}) = \left(1 - \frac{\tau}{2}\right)^{n-s} \left(\frac{\tau}{2}\right)^s . \tag{A.2}$$

We use the well–known approximation of binomial coefficients (and remind the reader that $\sigma = \frac{s}{n}$, $\tau = \frac{t}{n}$)

$$(n+1)^{-1} \exp\{nh(\sigma)\} \leq \binom{n}{s} \leq \exp\{nh(\sigma)\}, \tag{A.3}$$

$$(n+1)^{-1} \exp\left\{\tau n h\left(\frac{\sigma}{\tau}\right)\right\} \leq \binom{t}{s} \leq \exp\left\{\tau n h\left(\frac{\sigma}{\tau}\right)\right\} \tag{A.4}$$

and conclude that

$$(n+1)^{-1} \exp\left\{n\left(\tau h\left(\frac{\sigma}{\tau}\right) - h(\sigma) - \tau\right)\right\} \leq \underline{W}(y^n | \underline{o}) \leq$$
$$(n+1) \exp\left\{n\left(\tau h\left(\frac{\sigma}{\tau}\right) - h(\sigma) - \tau\right)\right\}. \tag{A.5}$$

We also can write

$$W(y^n | \underline{o}) = \exp\left\{n\left[(1-\sigma) \log\left(1 - \frac{\tau}{2}\right) + \sigma \log \frac{\tau}{2}\right]\right\}. \tag{A.6}$$

We express the difference, say $d\big(\sigma, \frac{\tau}{2}\big)$, of the exponents in (A.5) and (A.6) in terms of Kullback–Leibler divergences. For binary distributions $(\alpha, 1-\alpha)$ and $(\beta, 1-\beta)$ we use for the divergence $D\big((\alpha, 1-\alpha) \| (\beta, 1-\beta)\big)$ the shorthand $D(\alpha \| \beta)$.

Now

$$
\begin{aligned}
d\left(\sigma, \frac{\tau}{2}\right) &= -\left[(1-\sigma)\log\left(1-\frac{\tau}{2}\right) + \sigma\log\frac{\tau}{2}\right] + \left[\tau h\left(\frac{\sigma}{\tau}\right) - h(\sigma) - \tau\right] \\
&= -\left[(1-\sigma)\log\left(1-\frac{\tau}{2}\right) + \sigma\log\frac{\tau}{2} + h(\sigma)\right] - \left[\tau - \tau h\left(\frac{\sigma}{\tau}\right)\right] \\
&= D\left(\sigma \| \frac{\tau}{2}\right) - \tau D\left(\frac{\sigma}{\tau} \| \frac{1}{2}\right).
\end{aligned}
$$

By the continuity of D we have $\lim_{\sigma \to \frac{\tau}{2}} d\big(\sigma, \frac{\tau}{2}\big) = 0$.

We summarize our findings

Approximation Lemma 1.
For $x^n = (x_1, \dots, x_n)$, $y^n = (y_1, \dots, y_n) \in \{0,1\}^n$ *with Hamming distance* $d(x^n, y^n) = s \le t$
$W(y^n|x^n)\, exp\left\{-nd\left(\sigma, \frac{\tau}{2}\right) - \log(n+1)\right\} \le \underline{W}(y^n|x^n)$
$\le W(y^n|x^n)\, exp\left\{nd\left(\sigma, \frac{\tau}{2}\right) + \log(n+1)\right\}$,
where $\lim_{\sigma \to \frac{\tau}{2}} d\left(\sigma, \frac{\tau}{2}\right) = 0$.

We see that the two transmission probabilities are close to each other, if σ is close to $\frac{\tau}{2}$. Next we show that there the probabilities concentrate.

Approximation Lemma 2.
For the set $A_\varepsilon(x^n) = \left\{y^n : \frac{\tau}{2} - \varepsilon \le \frac{1}{n} d(x^n, y^n) \le \frac{\tau}{2} + \varepsilon\right\}$

$$W\big(A_\varepsilon(x^n)|x^n\big) \ge 1 - 2\, exp\left\{-nD\left(\frac{\tau}{2} + \varepsilon \| \frac{\tau}{2}\right)\right\}.$$

Proof. A consequence of Chebyshev's inequality states that for independent, identically distributed RV's $X_1, \dots, X_n$ with values in $[0,1]$ and $\mathbb{E}\, X_i \le \mu < \lambda \le 1$

$$\text{Prob}\left(\sum_{i=1}^{n} X_i > n\lambda\right) \le \exp\{-nD(\lambda\|\mu)\}.$$

This inequality and the inequality $D\left(\frac{\tau}{2} + \varepsilon \| \frac{\tau}{2}\right) \le D\left(\frac{\tau}{2} - \varepsilon \| \frac{\tau}{2}\right)$ give the result.

Remark 5. These Lemmas immediately yield that a code for W with error probability less than $\exp\{-\delta n\}$ is also a code for V with error probability less than $\exp\left\{-\frac{\delta}{2} n\right\}$, if n is sufficiently large.

Remark 6. It is now just an exercise to make the corresponding approximations for general q .

References

[1] L.A. Bassalygo, S.I. Gelfand, and M.S. Pinsker, "Coding for channels with localized errors", Proc Fourth Soviet–Swedish Workshop in Information Theory, Gotland Sweden, pp. 95–99, 1989.

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

[3] R. Ahlswede, "Coloring hypergraphs: A new approach to multi–user source coding", Part I: Journ. of Combinatorics, Information and System Sciences, Vol. 4, No. 1, 76–115, 1979; Part II: Vol. 5, No. 3, 220–268, 1980.

[4] R. Ahlswede, "Arbitrarily varying channels with states sequences known to the sender, invited paper at a Statistical Research Conference dedicated to the memory of Jack Kiefer and Jacob Wolfowitz, held at Cornell University, July 1983, IEEE Trans. Inf. Theory, Vol. IT–32, No. 5, 621–629, 1986.

[5] S.I. Gelfand and M.S. Pinsker, "Coding for a channel with random parameters", Probl. Control and Inf. Theory 9, No. 1, 19–31, 1981.

[6] R. Ahlswede, L.A. Bassalygo, and M.S. Pinsker, "Nonbinary codes correcting localized errors", IEEE Trans. Inf. Theory, Vol. 39, No. 4, 1413–1416, 1993.

[7] R. Ahlswede and N. Cai, "Two proofs of Pinsker's conjecture concerning AV channels", IEEE Trans. Information Theory, Vol. 37, No. 6, 1647–1649, 1991.

[8] A.W. Marshall, I. Olkins, "Inequalities: Theory of Majorization and Its Applications", Mathematics in Science and Engineering, Vol. 14b, Academic Press, 1979.