The AVC with noiseless feedback and maximal error probability: A capacity formula with a trichotomy

Rudolf Ahlswede and Ning Cai

Abstract

To use common randomness in coding is a key idea from the theory of identification. Methods and ideas of this theory are shown here to have also an impact on Shannon's theory of transmission. As indicated in the title, we determine the capacity for a classical channel with a novel structure of the capacity formula. This channel models a robust search problem in the presence of noise (see R. Ahlswede and I. Wegner, Search Problems, Wiley 1987).

1 Introduction

Let $\mathcal{X}, \mathcal{Y}$ be the finite input and output alphabets of an AVC defined by the class of $|\mathcal{X}| \times |\mathcal{Y}|$–stochastic matrices $\mathcal{W}$, which we assume to be finite. Eventhough our results hold for every $\mathcal{W}$, we assume here $\mathcal{W}$ to be finite, because already under this restriction the proofs are highly sophisticated and we don't want to burden the reader with additional technical, but known, approximation arguments (like i.e. in [2]).

It was assumed in [1] that $\mathcal{W}$ equals its row–convex hull $\overline{\overline{\mathcal{W}}}^{=}$ and it was shown that in the presence of noiseless feedback under the maximal error probability criterion its capacity $C_F(\overline{\overline{\mathcal{W}}}^{=})$ has the formula

$$C_F(\overline{\overline{\mathcal{W}}}^{=}) = \max_{P \in \mathcal{P}(\mathcal{X})} \min_{W \in \overline{\overline{\mathcal{W}}}^{=}} I(P, W), \quad \text{if the capacity is positive.} \tag{1.1}$$

Here $\mathcal{P}(\mathcal{X})$ is the set of probability distributions (PD) on $\mathcal{X}$ and I is the mutual information.

Actually, this result was shown with an explicit coding strategy. Clearly, the known (in [11]) exact condition for positivity in the absence of feedback, namely,

$$\overline{\mathcal{W}}(x) \cap \overline{\mathcal{W}}(x') = \varnothing \quad \text{for some} \quad x, x' \in \mathcal{X}, \tag{1.2}$$

where $\overline{\mathcal{W}}(x) =$ convex hull $(\mathcal{W}(x))$ and $\mathcal{W}(x) = \{W(\cdot|x) : W \in \mathcal{W}\}$, is also sufficient for positivity in the presence of feedback.

However, it is not necessary for positivity of $C_F(\overline{\overline{\mathcal{W}}})$.

On the other hand (see Lemma 3 of [1]) condition (1.2) is necessary and sufficient for positivity of $C_F(\overline{\overline{\mathcal{W}}})$ (and also of $C_F(\mathcal{W})$), if $\mathcal{W}$ contains only 0–1–matrices.

Furthermore, Example 2 of [1] shows that $C_F(\overline{\overline{\mathcal{W}}})$ and $C_F(\mathcal{W})$ can be different. This construction shows that in cases where (1.2) does not hold (for letters) its extension for feedback strategies can still hold.

In this paper we determine $C_F(\mathcal{W})$ completely. The formula distinguishes three cases and therefore we speak of a trichotomy. It is an absolute novelty for capacity formulas in Information Theory.

A *dichotomy occurred* — quite surprisingly at its time — for AVC without feedback under the average error criterion ([2]): $C_{av}(\mathcal{W})$ is zero or else equals the random code capacity $C_R(\mathcal{W}) = \max_P \min_{W \in \overline{\mathcal{W}}} I(P, W)$, where $\overline{\mathcal{W}}$ is the convex hull of $\mathcal{W}$.

We settle now the positivity problem for $C_F(\mathcal{W})$ and we prove the Trichotomy Theorem. The Positivity Theorem and the easy direction of its proof are presented in Section 2. The much harder direction is given in Section 6. It uses a Balanced Coloring Lemma, which we establish in Section 3.

The Trichotomy Theorem is stated in Section 4. It incorporates the Positivity Theorem and the Capacity Theorem for 0–1–matrices of [1], which also readily leads to the Converse of the Trichotomy Theorem. Its direct part, however, is far more complex. The main ingredients are the List Reduction Lemma of [1], the Elimination Technique of [2], and the Balanced Coloring Lemma (see [2], [7]) in the version of Section 3.

Finally we mention that the coding problem for the AVC with feedback has another appealing interpretation. One of the simplest search problems is to find an unknown element $x \in \mathcal{X}$ by sequentially "Yes–No" questions like "Is $x \in A$?" where A is any subset of $\mathcal{X}$. It is easy to see that the minimal number of such questions which specify x is in the worst case $\lceil \log |\mathcal{X}| \rceil$. Now, if the answers are false with probability ε, allowing an error probability λ, then this problem is equivalent to the coding problem for the BSC $W = \left(\begin{smallmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{smallmatrix}\right)$ with complete feedback. A proof can be found in the book mentioned in the abstract.

More generally there is the same connection for a–ary questions with b–ary answers with noise, that is, the BSC can be replaced by a general DMC. In a robust noise model this DMC is to be replaced by an AVC.

Needless to say that channels with feedback links are of practical interest (see [13]) in error control coding (ARQ, FEC systems etc.). Here we settle the capacity problem for the robust channel model AVC.

2 Positivity of the capacity $C_F(\mathcal{W})$

We are given the set of transmission matrices

$$\mathcal{W} = \big\{W(\cdot|\cdot, s) : s \in \mathcal{S}\big\}, |\mathcal{S}| < \infty. \tag{2.1}$$

For state sequence $s^n \in \mathcal{S}^n$ the n–length feedback transmission matrix $W_F^n(\cdot|\cdot, s^n)$ is an $|\mathcal{X}|^{\sum_{t=0}^{n-1} |\mathcal{Y}^t|} \times |\mathcal{Y}^n|$–stochastic matrix with entries $W(y_1|f_1, s_1) \prod_{t=2}^{n} W\big(y_t|f_t(y^{t-1}), s_t\big)$, where the feedback strategy $f^n = (f_1, \ldots, f_n)$ is defined by $f_1 \in \mathcal{X}$ and $f_t : \mathcal{Y}^{t-1} \to \mathcal{X}$ for $t = 2, \ldots, n$.

We denote the set of those strategies by $\mathcal{F}^n$ and then write $W_F^n(\cdot|\cdot, s^n) = (W^n(\cdot|f^n, s^n))_{f^n \in \mathcal{F}^n}$ and

$$\mathcal{W}_F^n = \big\{W_F^n(\cdot|\cdot, s^n) : s^n \in \mathcal{S}^n\big\} \tag{2.2}$$

and draw an immediate consequence of (1.2).

Lemma 1. *$C_F(\mathcal{W}) > 0$ iff for some n there are two n–length strategies $f^n, f'^n \in \mathcal{F}^n$ with disjoint corresponding convex hulls, that is, convex hull $\big(\big\{W^n(\cdot|f^n, s^n) : s^n \in \mathcal{S}^n\big\}\big) \cap$ convex hull $\big(\big\{W^n(\cdot|f'^n, s^n) : s^n \in \mathcal{S}^n\big\}\big) = \varnothing$.*

Next we need for our analysis two concepts, namely, for $x \in \mathcal{X}$

$$\mathcal{S}_x = \big\{s \in \mathcal{S} : \text{ for some } \ y \quad W(y|x, s) = 1\big\} \tag{2.3}$$

and

$$\mathcal{Y}_x = \big\{y \in \mathcal{Y} : \text{ for some } \ s \quad W(y|x, s) = 1\big\}. \tag{2.4}$$

Notice that both, $\mathcal{S}_x$ and $\mathcal{Y}_x$, can be empty and that $\mathcal{S}_x = \varnothing$ iff $\mathcal{Y}_x = \varnothing$.

Lemma 2. *If $C_F(\mathcal{W}) > 0$, then necessarily*

(i) $C_R(\mathcal{W}) > 0$ (2.5)

and

(ii) $\mathcal{Y}_x \cap \mathcal{Y}_{x'} = \varnothing$ *for some* $x \neq x'$. (2.6)

Proof: If (i) does not hold, then there is a distribution P on $\mathcal{S}$ such that the matrix $\sum_s P(s)W(\cdot|\cdot, s)$ has identical rows. Therefore for all n and $P^n(s^n) = \prod_{t=1}^{n} P(s_t)$ also $\sum_{s^n} P^n(s^n)W_F^n(\cdot|\cdot, s^n)$ has identical rows and (as a special case of Lemma 1) $C_F(\mathcal{W}) = 0$.

If (ii) does not hold, then for all $x, x'(x \neq x')$ there are $y(x, x') \in \mathcal{Y}$ and $s(x, x')$, $s'(x, x') \in \mathcal{S}$ with the property $W\big(y(x, x')|x, s(x, x')\big) = W\big(y(x, x')|x', s'(x, x')\big) = 1$.

This implies that for all n and any two rows of W_F^n corresponding to the feedback strategies $f^n = (f_1, f_1, \ldots, f_n)$ and $f'^n = (f'_1, f'_1, \ldots, f'_n)$ we can choose $y_1 = y(f_1, f'_1)$, $s_1 = s(f_1, f'_1)$,

$s'_1 = s'(f_1, f'_1)$ and; for $t = 2, 3, \ldots, n$; $y_t = y\big(f_t(y^{t-1}), f'_t(y^{t-1})\big)$, $s_t = s\big(f_t(y^{t-1}), f'_t(y^{t-1})\big)$, and $s'_t = s\big(f_t(y^{t-1}), f'_t(y^{t-1})\big)$ such that

$W(y^n|f^n, s^n) = W(y^n|f'^n, s'^n) = 1$ and thus $C_F(\mathcal{W}) = 0$.

Quite remarkably also the converse of Lemma 2 holds. This is a much deeper result.

Positivity Theorem. *$C_F(\mathcal{W}) > 0$ iff (i) and (ii) in Lemma 2 hold.*

The rather sophisticated proof is based on the Coloring Lemma of Section 3, which is closely related to its predecessors in [3] and [7]. We give it in the last section so that readers, who are interested only in our coding scheme of Section 4 can skip it.

3 Balanced coloring

Lemma 3. *Let $\mathcal{Q} \subset \mathcal{P}(\mathcal{V})$ be a finite set of PD's on $\mathcal{V}$ and let there be associated with every $P \in \mathcal{Q}$ a family $\mathcal{E}(P)$ of subsets of $\mathcal{V}$ such that*

$$\alpha(P) \triangleq \max\left\{P(v) : v \in \bigcup_{E \in \mathcal{E}(P)} E\right\} < 1. \tag{3.1}$$

Now, if there are positive numbers $\eta(P)$ for all $P \in \mathcal{Q}$ such that for $k \geq 2$, $\delta \in (0, 1)$ and all $E \in \mathcal{E}(P)$

$$\left(\frac{1}{\alpha(P)}\right)^{1-\delta}\left[\eta(P) - \frac{e}{2k}\alpha(P)^{\delta}P(E)\right] > \ell n\left\{2k\sum_{P\in\mathcal{Q}}|\mathcal{E}(P)|\right\}, \tag{3.2}$$

then there is a function $g : \mathcal{V} \to \{1, 2, \ldots, k\}$ which satisfies for all $P \in \mathcal{Q}, E \in \mathcal{E}(P)$, and $i \in \{1, 2, \ldots, k\}$

$$\left|P\big(g^{-1}(i) \cap E\big) - \frac{1}{k}P(E)\right| < \eta(P). \tag{3.3}$$

Furthermore, for $\delta = \frac{1}{4}$, $\eta(P) = 2\alpha(P)^{\frac{1}{4}}$, and $\alpha \triangleq \max\limits_{P\in\mathcal{Q}} \alpha(P)$

$$\alpha^{-\frac{1}{2}} > \ell n\left[2k\sum_{P\in\mathcal{Q}}|\mathcal{E}(P)|\right] \tag{3.4}$$

implies (3.2) and thus (3.3) holds.

Proof: The idea behind the following probabilistic existence proof is to use a union bound argument to show that the probability of a randomly chosen coloring to be "bad" is less than 1. We color all $v \in \mathcal{V}$ at random independently and uniformly with k colors.

Next we introduce the RV's

$$\Psi_i(v) = \begin{cases} 1, & \text{if } \ v \ \text{ gets color } \ i \\ 0 & \text{otherwise} \end{cases}$$

and

$$Z_i^P(E) = \sum_{v \in E} P(v)\Psi_i(v) \quad \text{for} \quad P \in \mathcal{Q}.$$

With Bernstein's version of Chebyshev's inequality

$$\begin{aligned}
&\Pr\big(Z_i^P(E) > \frac{1}{k}P(E) + \eta(P)\big\} \\
&\leq \exp_e\left\{-\alpha(P)^{-(1-\delta)}\left[\frac{1}{k}P(E) + \eta(P)\right]\right\} \cdot \mathbb{E}\exp_e\left\{\alpha(P)^{-(1-\delta)}\sum_{v\in E} P(v)\Psi_i(v)\right\} \\
&= \exp_e\left\{-\alpha(P)^{-(1-\delta)}\left[\frac{1}{k}P(E) + \eta(P)\right]\right\} \cdot \prod_{v\in E} \mathbb{E}\exp_e\big\{\alpha(P)^{-(1-\delta)}P(v)\Psi_i(v)\big\} \\
&= \exp_e\left\{-\alpha(P)^{-(1-\delta)}\left[\frac{1}{k}P(E) + \eta(P)\right]\right\} \cdot \prod_{v\in E}\left(\frac{k-1}{k} + \frac{1}{k}\exp_e\big\{\alpha(P)^{-(1-\delta)}P(v)\big\}\right).
\end{aligned}$$

Using Lagrange's remainder formula for the Taylor series of the exponential function we continue with the upper bound

$$\exp_e\left\{-\alpha(P)^{-(1-\delta)}\left[\frac{1}{k}P(E) + \eta(P)\right]\right\} \cdot \prod_{v\in E}\left\{1 + \frac{1}{k}\left[\alpha(P)^{-(1-\delta)}P(v) + \frac{[\alpha(P)^{-(1-\delta)}P(v)]^2 \cdot e}{2}\right]\right\}$$

and since $\ell n(1+x) < x$ for $x > 0$ with the upper bound

$$\begin{aligned}
&\exp_e\left\{-\alpha(P)^{-(1-\delta)}\left[\frac{1}{k}P(E) + \eta(P) - \frac{1}{k}\sum_{v\in E}P(v) - \frac{e}{2k}\alpha(p)^{-(1-\delta)}\sum_{v\in E}P^2(v)\right]\right\} \\
=&\exp_e\left\{-\alpha(P)^{-(1-\delta)}\left[\eta(P) - \frac{e}{2k}\alpha(P)^{-(1-\delta)}\sum_{v\in E}P^2(v)\right]\right\} \\
\leq&\exp_e\left\{-\alpha(P)^{-(1-\delta)}\left[\eta(P) - \frac{e}{2k}\alpha(P)^{-(1-\delta)}\cdot\sum_{v\in E}\alpha(P)P(v)\right]\right\},
\end{aligned}$$

because $P(v) \leq \alpha(P)$ for $v \in E$.

The last upper bound equals

$$\exp_e\left\{-\alpha(P)^{-(1-\delta)}\left[\eta(P) - \frac{e}{2k}\alpha(P)^{\delta}P(E)\right]\right\}.$$

Analogously,
$\Pr\left\{Z_i^P(E) < \frac{1}{k}P(E) - \eta(P)\right\} \le \exp_e\left\{-\alpha(P)^{-(1-\delta)}\left[\eta(P) - \frac{e}{2k}\alpha(P)^{\delta}P(E)\right]\right\}$ for all $P \in \mathcal{Q}$, $E \in \mathcal{E}(P)$ and $i \in \{1, 2, \ldots, k\}$. This together with (3.2) implies (3.3).

Finally, since $\left(\frac{1}{\alpha(P)}\right)^{\frac{3}{4}}\left[2\alpha(P)^{\frac{1}{4}} - \frac{e}{2k}\alpha(P)^{\frac{1}{4}}P(E)\right] > \left(\frac{1}{\alpha(P)}\right)^{\frac{1}{2}} \ge \alpha^{-\frac{1}{2}}$
(3.4) implies (3.2).

4 The Trichotomy Theorem

For the formulation of our main result we need a concept from [1].

With our set of matrices $\mathcal{W}$ we associate the set of stochastic $|\mathcal{X}| \times |\mathcal{Y}| - (0-1)$ matrices

$$\hat{\mathcal{W}} = \left\{\hat{W} : \hat{W}(\cdot|x) \in \mathcal{W}(x) \text{ for all } x \in \mathcal{X} \text{ and } \hat{W}(y|x) \in \{0,1\} \text{ for all } y \in \mathcal{Y}\right\}. \quad (4.1)$$

Let this set be indexed by the set $\hat{\mathcal{S}}$. Then we have that for all $\hat{s} \in \hat{\mathcal{S}}$ and $x \in \mathcal{X}$ there is an $s \in \mathcal{S}_x$ with

$$\hat{W}(\cdot|x, \hat{s}) = W(\cdot|x, s). \quad (4.2)$$

Of course, $\hat{\mathcal{W}}$ (and thus also $\hat{\mathcal{S}}$) can be empty. This happens exactly, if for some x $\mathcal{S}_x = \varnothing$ or (equivalently) $\mathcal{Y}_x = \varnothing$. These sets are defined in (2.3) and (2.4).

Shannon determined in [12] the zero–error feedback capacity $C_{0,F}(W)$ of a DMC W.

An alternate formula — called for by Shannon — was given in [1]. For

$$\hat{V}(\cdot|\cdot) = |\hat{\mathcal{S}}|^{-1}\sum_{\hat{s}\in\hat{\mathcal{S}}}\hat{W}(\cdot|\cdot, \hat{s})$$

this formula asserts

$$C_{0,F}(\hat{V}) = C_F(\hat{\mathcal{W}}) = \begin{cases} \max\limits_{P}\min\limits_{W\in\overline{\hat{\mathcal{W}}}} I(P, W), & \text{if } \mathcal{Y}_x \cap \mathcal{Y}_{x\prime} = \varnothing \text{ for some } x, x' \\ 0 & \text{otherwise.} \end{cases} \quad (4.3)$$

Moreover, we have an inequality for this quantity.

Lemma 4. *$C_F(\mathcal{W}) \le C_F(\hat{\mathcal{W}})$, if $\hat{\mathcal{W}} \neq \varnothing$.*

Proof: It suffices to show that every feedback code with maximal error probability $\varepsilon < 1$ for $\mathcal{W}$ is a code for $\hat{\mathcal{W}}$. Indeed, otherwise there exists a feedback code for $\mathcal{W}$ with two encoding functions $f^n = (f_1, \ldots, f_n)$ and $f'^n = (f'_1, \ldots, f'_n)$ such that for some $y^n \in \mathcal{Y}^n$ and $\hat{s}^n, \hat{s}'^n \in \hat{\mathcal{S}}^n$

$$\hat{W}^n(y^n|f^n, \hat{s}^n) = \hat{W}^n(y^n|f'^n, \hat{s}'^n) = 1.$$

But then, if we choose s_t, s'_t corresponding to $\big(f_t(y^{t-1}), \hat{s}_t\big)$ and $\big(f'_t(y^{t-1}), \hat{s}'_t\big)$, respectively, according to (4.2), we get

$$W^n(y^n|f^n, s^n) = W^n(y^n|f'^n, s'^n\big) = 1,$$

a contradiction.

Clearly by averaging we see that an ε–code with feedback for the AVC $\mathcal{W}$ is an ε–code for the AVC with feedback and therefore $C_F(\mathcal{W}) = C_F(\overline{\mathcal{W}}$. Furthermore, since feedback does not increase the capacity of an individual DMC $W \in \overline{\mathcal{W}}$ we have that

Lemma 5. $C_F(\mathcal{W}) = C_F(\overline{\mathcal{W}}) \leq C_R(\mathcal{W})$.

We are now ready to state our main result.

Trichotomy Theorem.

$$C_F(\mathcal{W}) = \begin{cases} 0, & \textit{iff } C_R(\mathcal{W}) = 0 \textit{ or } \mathcal{Y}_x \cap \mathcal{Y}_{x'} \neq \varnothing \textit{ for all } x, x' \in \mathcal{X} \text{(i)} \\ C_R(\mathcal{W}), & \textit{if } C_F(\mathcal{W}) > 0 \textit{ and } \mathcal{Y}_x = \varnothing \textit{ for some } x \quad \text{(ii)} \\ \min\big\{C_R(\mathcal{W}), C_F(\hat{\mathcal{W}})\big\}, & \textit{if } C_F(\mathcal{W}) > 0 \textit{ and } \mathcal{Y}_x \neq \varnothing \textit{ for all } x. \quad \text{(iii)} \end{cases}$$

Remark 1: There is almost no connection between the values of $C_R(\mathcal{W})$ and $C_F(\hat{\mathcal{W}})$.

Example 1:

Choose $\mathcal{X} = \mathcal{S} = \{1, 2, \ldots, a\}$, $\mathcal{Y} = \{1, 2, \ldots, a, b\}$, and $\mathcal{W}$ as set of matrices W with

$$W(y|x, s) = 1, \quad \text{if} \quad x \neq s \quad \text{and} \quad y = x \quad \textbf{or} \quad x = s, y = b.$$

Then $C_F(\hat{\mathcal{W}}) = 0$, but with P as uniform distribution on $\mathcal{X}$,

$$C_R(\mathcal{W}) \geq \min_{W \in \overline{\mathcal{W}}} I(P, W) = \left(1 - \frac{1}{a}\right) \log a$$

and this goes to infinity with a going to infinity.

Example 2:

Choose $\mathcal{X}' = \{0, 1, \ldots, a\}$, $\mathcal{S}' = \{1, 2, \ldots, a\}$, $\mathcal{Y}' = \{0, 1, \ldots, a, b\}$ and define $\mathcal{W}'$ as set of matrices with $W(y|x, s) = 1$, if $x = y = 0$ (for every s) **or** $x \neq 0$, $x \neq s$ and $y = x$ **or** $x = s$, $y = b$, $x \neq 0$.

Then $C_F(\hat{\mathcal{W}}') = \log 2 > 0$, however for $\mathcal{W}$ in Example 1 $C_R(\mathcal{W}') > C_R(\mathcal{W})$. So $C_R(\mathcal{W}')$ can be arbitrary large and much larger than a positive $C_F(\hat{\mathcal{W}})$.

Example 3:

Choose $\mathcal{X} = \mathcal{Y} = \mathcal{S} = \{0, 1\}$, $W(\cdot|\cdot, 0) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$, $W(\cdot|\cdot, 1) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Then $C_R(\mathcal{W}) = 0$ and $C_F(\hat{\mathcal{W}}) = 1$.

Finally, we formulate the Trichotomy Theorem in a more elegant, but less informative way. For this we define

$$C_F^{\infty}(\hat{\mathcal{W}}) = \begin{cases} C_F(\hat{\mathcal{W}}), & \text{if } \hat{\mathcal{W}} \neq \varnothing \\ \infty, & \text{if } \hat{\mathcal{W}} = \varnothing. \end{cases} \tag{4.4}$$

Then Lemma 4 says that always

$$C_F(\mathcal{W}) \leq C_F^{\infty}(\hat{\mathcal{W}})$$

and with Lemma 5 we conclude that

$$C_F(\mathcal{W}) \leq \min\big(C_R(\mathcal{W}), C_F^{\infty}(\hat{\mathcal{W}})\big). \tag{4.5}$$

Furthermore, now (ii) and (iii) say that there is equality in (4.5), if $C_F(\mathcal{W}) > 0$. Finally, if $C_F(\mathcal{W}) = 0$, then by (i) and (4.3) either $C_R(\mathcal{W}) = 0$ **or** $C_F(\hat{\mathcal{W}}) = 0$.

We summarize our findings.

Capacity Theorem. $C_F(\mathcal{W}) = \min\big\{C_R(\mathcal{W}), C_F^{\infty}(\hat{\mathcal{W}})\big\}$.

5 Proof of the Trichotomy Theorem

It remains to be seen that for $C_F(\mathcal{W}) > 0$

(ii) $C_F(\mathcal{W}) \geq C_R(\mathcal{W})$, if $\mathcal{S}_x = \varnothing$ for some x,

and

(iii) $C_F(\mathcal{W}) \geq \min\big\{C_R(\mathcal{W}), C_F(\hat{\mathcal{W}})\big\}$ otherwise.

For the convenience of the reader we mention first that in the case, where $\mathcal{W}$ contains only 0–1–matrices, we are in the case (iii) and (4.3) gives the desired result.

In the other extreme case (ii) we have $\hat{\mathcal{W}} = \varnothing$ and can use Lemma 3 (to establish a common random experiment) in conjunction with the elimination technique of [2]. (This approach of [7] works here even for maximal errors, because the "edges E" are big enough, if 0–1–distributions are excluded. In contrast to the previous work now the sender cannot randomize!)

To be specific, for any $\gamma > 0$ choose $\ell \sim \frac{n}{2}\ \gamma\ C_R^{-1}(\mathcal{W})$, an $x_0 \in \mathcal{X}$ with $\mathcal{S}_{x_0} = \varnothing$, and the encoding

$$f_t(y^{t-1}) = x_0 \quad \text{for} \quad 1 \leq t \leq \ell. \tag{5.1}$$

Next, clearly for $x_0^{\ell} = (x_0, \ldots, x_0)$ and all y^{ℓ}, s^{ℓ}

$$W^{\ell}(y^{\ell} | x_0^{\ell}, s^{\ell}) \leq \omega^{*\ell} < 1, \tag{5.2}$$

where

$$\omega^* = \max\big\{W(y|x,s) : W(y|x,s) \neq 1, x \in \mathcal{X}, s \in \mathcal{S}, \text{ and } y \in \mathcal{Y}\big\}. \tag{5.3}$$

By applying Lemma 3 to $\mathcal{Q} = \big\{W^\ell(\cdot|x_0^\ell, s^\ell) : s^\ell \in \mathcal{S}^\ell\big\}$, $k = (n-\ell)^2$, $\mathcal{E}(P) = \{\mathcal{Y}^\ell\}$ for all P, $\alpha = {w^*}^\ell$ in (3.4) then when ℓ is sufficiently large, so that $w^{*-\frac{1}{2}\ell} > \ell n(n-\ell)^2|\mathcal{S}|^\ell$, i.e. (3.4) holds, there is a coloring or equivalently a partition $\{A_i\}_{i=1}^{(n-\ell)^2}$ of $\mathcal{Y}^\ell$ such that for all $s^\ell \in \mathcal{S}^\ell$ and $i = 1, 2, \ldots, (n-\ell)^2$

$$\left|W^\ell(A_i|x_0^\ell, s^\ell) - \frac{1}{(n-\ell)^2}\right| < 2^{-\ell\tau} \tag{5.4}$$

for a positive $\tau \left(= -\frac{1}{8}\log w^*\right)$, which is independent of ℓ.

For this we have used ℓ letters and for the remaining $n-\ell$ letters we use a random code with rate $C_R(\mathcal{W}) - \frac{\gamma}{2}$, maximum error probability $\frac{\lambda}{2}$, and with ensemble size $(n-\ell)^2$. Its existence is guaranteed by the elimination technique of [2].

Now, after having sent x_0^ℓ and received $y^\ell \in A_i$, which is also known to the sender, because of the feedback, for any message m the m–th codeword in the i–th code of the ensemble is send next.

This n–length feedback code achieves a rate

$$\frac{1}{n}\left(n - \frac{n}{2}\,\gamma\, C_R^{-1}(\mathcal{W})\right)\left(C_R(\mathcal{W}) - \frac{\gamma}{2}\right) \geq C_R(\mathcal{W}) - \delta$$

and a maximum error probability less than $(n-\ell)^2 2^{-\ell\tau} + \frac{\lambda}{2} < \lambda$, when ℓ is large enough.

The main issue is really to prove the direct part for the mixed case:

$$\hat{\mathcal{W}} \neq \varnothing \quad \text{and} \quad \mathcal{W} \setminus \hat{\mathcal{W}} \neq \varnothing,\ C_F(\mathcal{W}) > 0.$$

We design a strategy by compounding *four types* of codes. There germ is the iterative list reduction code of [1].

However, now we must achieve a higher rate by incorporating also codes based on common randomness. The detailed structure will become clear at the end of our description.

We begin with the codes announced.

1. List reducing or coloring code (LROCC)

As in [1] we start with $\mathcal{T}_P^\ell$, the set of P–typical sequences in $\mathcal{X}^\ell$, where $P \in \mathcal{P}_\ell(\mathcal{X}) = \big\{P \in \mathcal{P}(\mathcal{X}) : \mathcal{T}_P^\ell \neq \varnothing\big\}$.

However, right in the beginning we gain a certain freedom by deviating from [1] by choosing parameters such that $|\mathcal{T}_P^\ell|$ is much smaller than the size of the set of messages $\mathcal{M}$. An (ℓ, ξ, c) LROCC (where the role of parameter ξ becomes clear in (5.6) and (5.7)) is defined by a triple (g, L, K) of functions, which we now explain.

Function $g : \mathcal{L} \to \mathcal{T}_P^\ell$ (called *balanced partition function*) is chosen such that

$$\big||g^{-1}(x^\ell)| - |g^{-1}(x'^\ell)|\big| \leq 1 \quad \text{for all} \quad x^\ell, x'^\ell \in \mathcal{T}_P^\ell. \tag{5.5}$$

Function $L : \mathcal{Y}^\ell \to 2^{\mathcal{L}}$

This function, which we call *list function*, assigns to every $y^\ell \in \mathcal{Y}^\ell$ a sublist of $\mathcal{L}$ as follows. Define first for $x^\ell \in \mathcal{X}^\ell$, $y^\ell \in \mathcal{Y}^\ell$, and $\mathcal{Y}_x$

$$\tilde{\delta}(x^\ell, y^\ell) = |\{t : y_t \notin \mathcal{Y}_{x_t}\}|, \tag{5.6}$$

the *discriminator*.

Then set

$$L(y^\ell) = \big\{v \in \mathcal{L} : \tilde{\delta}\big(g(v), y^\ell\big) < \xi\big\} \quad \text{for} \quad y^\ell \in \mathcal{Y}^\ell. \tag{5.7}$$

We need later interpretations for the relation $v \in L(y^\ell)$. Since by our assumptions $\mathcal{Y}_x \neq \varnothing$ for all x, $\tilde{\delta}(x^\ell, y^\ell) < \xi$ implies that a $y'^\ell \in \mathcal{Y}^\ell$ can be found so that (in the Hamming distance)

$$d_H(y^\ell, y'^\ell) < \xi \tag{5.8}$$

and

$$y'_t \in \mathcal{Y}_{x_t} \quad \text{for all} \quad t = 1, 2, \ldots, \ell. \tag{5.9}$$

Equivalently, we can say that there is a

$$\overline{\hat{W}} \in \overline{\hat{\mathcal{W}}} \quad \text{with} \quad y'^\ell \in \mathcal{T}^\ell_{\overline{\hat{W}}}(x^\ell).$$

Also, by (5.7) – (5.9) for all $y^\ell \in \mathcal{Y}^\ell$

$$\frac{1}{\ell} \log |L(y^\ell)| < \frac{1}{\ell} \log |\mathcal{L}| - \min_{\overline{\hat{W}} \in \overline{\hat{\mathcal{W}}}} I(P, \overline{\hat{W}}) + u(\ell, \xi), \tag{5.10}$$

where u is a function with

$$u(\ell, \xi) \to 0 \quad \text{as} \quad \frac{\xi}{\ell} \to 0 \quad \text{and} \quad \ell \to \infty. \tag{5.11}$$

(Notice: when $\xi = 1$, then L is a list reduction via $\hat{\mathcal{W}}$ as in [1].)

Function $K : \mathcal{Y}^\ell \to \{1, 2, \dots, c\}$

In this *coloring function* we choose c of *polynomial growth in* ℓ. Let $\mathcal{Q} = \big\{W^\ell(\cdot|x^\ell, s^\ell) : x^\ell \in \mathcal{X}^\ell, s^\ell \in \mathcal{S}^\ell\big\}$, $\mathcal{E}\big(W^\ell(\cdot|x^\ell, s^\ell)\big) = \big\{\{y^\ell : \tilde{\delta}(x^\ell, y^\ell) \geq \xi\}\big\}$ and $k = c$ in Lemma 3.

Then by Lemma 3 we can also assume that for all $x^\ell \in \mathcal{T}_P^\ell$, $s^\ell \in \mathcal{S}^\ell$, and $j \in \{1, 2, \dots, c\}$

$$\big|W^\ell\big(K^{-1}(j) \cap \big\{y^\ell : \tilde{\delta}(x^\ell, y^\ell) \geq \xi\big\}|x^\ell, s^\ell\big) - c^{-1}W^\ell\big(\big\{y^\ell : \tilde{\delta}(x^\ell, y^\ell) \geq \xi\big\}|x^\ell, s^\ell\big)\big| < 2\omega^{*\frac{1}{4}\xi}, \tag{5.12}$$

because $\tilde{\delta}(x^\ell, y^\ell) \geq \xi$ implies $W^\ell(y^\ell|x^\ell, s^\ell) \leq \omega^{*\xi}$ for all s^ℓ (ω^* was defined in (5.3)) and consequently, $w^{*-\frac{1}{2}\xi} > \log[2c|\mathcal{X}^|\ell|\mathcal{S}|^\ell]$, i.e. (3.4) holds for sufficiently large ξ satisfying (5.11).

2. Index Code (IC)

This code has two codewords of length j and error probability μ. The codewords stand for messages L, K. They are used by the sender (based on the discriminator) to inform the receiver whether next he uses reducing the list, by sending L, or coloring on the output, by sending K.

3. Eliminated correlated code (ECC)

An m–length and (maximal) μ–error probability ECC is a *family*

$$\big\{\{(u_i^q, D_i^q) : 1 \leq i \leq M\} : 1 \leq q \leq m^2\big\}$$

of m^2 codes with the properties

$$m^{-2}\sum_{q=1}^{m^2} W^m(D_i^q|u_i^q, s^n) > 1 - \mu \ \text{ for all } \ s^n \in \mathcal{S}^n \ \text{ and all } \ i = 1, \dots, M \tag{5.13}$$

and

$$m^{-1}\log M > C_R(\mathcal{W}) - \delta'. \tag{5.14}$$

Their existence was proved in [2].

4. $(k, 2^{\gamma k}, \mu)$–Code

This is just an ordinary feedback code for $\mathcal{W}$ of length k, rate γ, and maximal error probability μ. Its existence is provided by $C_F(\mathcal{W}) > 0$.

Choice of parameters:

Before we present our coding algorithm we adjust the parameters. It is convenient to have the abbreviation

$$C = \min\big(C_R(\mathcal{W}), C_F(\hat{\mathcal{W}})\big). \tag{5.15}$$

a.) Let P attain the maximum in $\max_{P' \in \mathcal{P}_\ell(\mathcal{X})} \min_{\overline{\hat{W}} \in \overline{\hat{\mathcal{W}}}} I(P', \overline{\hat{W}})$.

b.) Fix now any $\delta \in (0, C)$ and $\lambda \in (0, 1)$.

c.) By our assumption $C_F(\mathcal{W}) > 0$ there is a positive number γ so that for large enough k and $\log M \leq k \cdot \gamma$ (k, M, μ)–codes exist.

d.) Define

$$r_o = \left\lceil \frac{2}{\delta} \left(\frac{\log |\mathcal{X}|}{\gamma} + 2 \right) C^2 \left(C - \frac{\delta}{2} \right)^{-1} \right\rceil \tag{5.16}$$

and let j be a fixed integer such that a j–length IC with error probability $\frac{\lambda}{4r_o}$ exists.

e.) Let ξ increase with ℓ, but keep for sufficiently large ℓ $\frac{\xi}{\ell}$ so small that for the u in (5.10)

$$u(\ell, \xi) < \frac{\delta}{4}. \tag{5.17}$$

f.) Insure

$$\ell > r_o j \tag{5.18}$$

and for the message set $\mathcal{M}$ set

$$n_0 = \log |\mathcal{M}| = \left\lceil \frac{2}{\delta} \left(\frac{\log |\mathcal{X}|}{\gamma} + 2 \right) C^2 \ell \right\rceil. \tag{5.19}$$

g.) Require ℓ and also ξ to be so large that the coloring function K for the LROCC can be obtained with Lemma 3 and still

$$n_0^2 \omega^{*\xi/4} < \frac{\lambda^2}{64 r_o}. \tag{5.20}$$

h.) Finally we make ℓ so large that all codes in the following algorithm exist.

Encoding Algorithm

Begin:
Input: $v \in \mathcal{M}$

1. Set $i := 0$ and let $\mathcal{L}_i := \mathcal{M}$, go to 2.
2. If $|\mathcal{L}_i| \geq |\mathcal{T}_P^\ell|$, then let $m_i := \left\lfloor \frac{\log |\mathcal{L}_i|}{C_R(\mathcal{M}) - \frac{\delta}{2}} \right\rfloor$, encode $\mathcal{L}_i$ to an (ℓ, ξ, m_i^2) LROCC (g, L, K) over $\mathcal{T}_P^\ell$, send $g(v) := x^\ell$ to the receiver, go to 3.

   Otherwise, go to 5.
3. Receive the output y^ℓ and encode a j–length IC with $\frac{\lambda}{4r_o}$–error probability.

   If $\tilde{\delta}(x^\ell, y^\ell) < \xi$, send the word "$L$" of the IC to the receiver. Let $i := i + 1$, $\mathcal{L}_i := L(y^\ell)$ and go to 2.

   Otherwise send the word "K" of the IC to the receiver, let $q = K(y^\ell)$, go to 4.
4. Encode $\mathcal{L}_i$ to an m_i–length ECC with $\frac{\lambda}{4}$–error probability and send the codeword u_v^q to the receiver, go to 6.
5. Encode $\mathcal{L}_i$ to a $\left(k, |\mathcal{L}_i|, \frac{\lambda}{4}\right)$–code with rate γ and send the codeword standing for v to the receiver, go to 6.
6. Stop.

End.

Decoding Algorithm

Begin:

1. Set $i := 0$ and let $\mathcal{L}_i = \mathcal{M}$, go to 2.
2. If $|\mathcal{L}_i| \geq |\mathcal{T}_P^\ell|$, go to 3.

   Otherwise go to 5.
3. Receive (y^ℓ, y^j) and decode y^j for the j–length IC.

   If the decoding result is "L", let $i := i + 1$, $\mathcal{L}_i = L(y^\ell)$, go to 2.

   Otherwise let $q = K(y^\ell)$ and go to 4.
4. Let $m_i := \left\lfloor \frac{\log |\mathcal{L}_i|}{C_R(\mathcal{M}) - \frac{\delta}{2}} \right\rfloor$, receive y^{m_i} and decode y^{m_i} for the q–th value–code of the m_i–length ECC, go to 6.
5. Receive y^k and decode it for the $\left(k, |\mathcal{L}_i|, \frac{\lambda}{4}\right)$ code with rate γ and length k, go to 6.
6. Stop

End.

Analysis

According to the choice of our P, by (5.10) and (5.17), for sufficiently large ℓ we have

$$\frac{1}{\ell}\log|\mathcal{L}_{i+1}| < \frac{1}{\ell}\log|\mathcal{L}_i| - C_F(\hat{\mathcal{W}}) + \frac{\delta}{2}, \tag{5.21}$$

or in other words

$$\begin{aligned}\log|\mathcal{L}_i| &< \log|\mathcal{M}| - i\ell C_F(\hat{\mathcal{W}}) + i\ell\frac{\delta}{2}\\ &\le \log|\mathcal{M}| - i\ell C + i\ell\frac{\delta}{2}.\end{aligned} \tag{5.22}$$

Thus, according to our encoding program, by (5.16), (5.19), and (5.22), at most r_o LROCC–IC–pairs may be encoded, and at most one "K". If it exists, it must be in the last IC. Therefore we can define the RV U as

$$U = \begin{cases} r, & \text{if } r \text{ LROCC–IC–pairs are sent and}\\ & \text{the last sent word of IC is "}K\text{"}\\ r_o+1, & \text{if no "}K\text{" is sent,}\end{cases} \tag{5.23}$$

or in other words,

$$\begin{aligned} U = r \le r_o \Leftrightarrow\ & \text{After the message set is reduced } r-1 \text{ times, the "}r\text{–th output"}\\ & \text{is "colored" and then the message is sent by the value "with this}\\ & \text{color" in an ECC.}\\ U = r_o+1 \Leftrightarrow\ & \text{After the size of the message set is reduced to less than } |\mathcal{T}_P^\ell|,\\ & \text{the message is sent by the ordinary (feedback) code with rate}\\ & \gamma. \end{aligned} \tag{5.24}$$

The rate:

Although the encoding algorithm may produce sequences with different lengths, by obvious reasons, we only need their common bound, say b.

Moreover, we only have to show that

$$b \le \left(C-\frac{\delta}{2}\right)^{-1}\log|\mathcal{M}| + \left(\frac{\log|\mathcal{X}|}{\gamma}+2\right)\ell. \tag{5.25}$$

This is so, because by an elementary calculation, for any positive $a, aC^2 \le \frac{\delta}{2}\log|\mathcal{M}|$ implies $\left(C-\frac{\delta}{2}\right)^{-1}\log|\mathcal{M}| + a \le (C-\delta)^{-1}\log|\mathcal{M}|$ and then (5.19) and (5.25) imply that the lengths of the encoding sequences are bounded by $(C-\delta)^{-1}\log|\mathcal{M}|$.

Case $U = r \leq r_o$:

By (5.24), after having been reduced $r-1$ times, the "message list" with size at most $\log|\mathcal{M}| - (r-1)\ell\left(C - \frac{\delta}{2}\right)$ (by (5.22)), is encoded by an

$\left\lfloor \left(C_R(\mathcal{M}) - \frac{\delta}{2}\right)^{-1} \left(\log|\mathcal{M}| - (r-1)\ell\left(C - \frac{\delta}{2}\right)\right)\right\rfloor$ –length ECC.

Therefore the total length of the encoding sequences is not exceeding

$r(\ell + j) + \left(C - \frac{\delta}{2}\right)^{-1}\left(\log|\mathcal{M}| - (r-1)\ell\left(C - \frac{\delta}{2}\right)\right) \leq \left(C - \frac{\delta}{2}\right)^{-1}\log|\mathcal{M}| + r_o j + \ell$

$\leq \left(C - \frac{\delta}{2}\right)^{-1}\log|\mathcal{M}| + 2\ell$ (by (5.18))

Case $U = r_o + 1$:

By (5.16), (5.18), (5.19), (5.24) and the wellknown fact that $|\mathcal{T}_P^\ell| \leq 2^{\ell\log|\mathcal{X}|}$, the total lengths of encoding sequences are bounded by

$$\begin{aligned} r_o(\ell + j) + \frac{\log|\mathcal{X}|}{\gamma}\ell &\leq \left[\left(\ell\left(C - \frac{\delta}{2}\right)\right)^{-1}\log|\mathcal{M}| + 1\right]\ell + r_o j + \frac{\log|\mathcal{X}|}{\gamma}\ell \\ &\leq \left(C - \frac{\delta}{2}\right)^{-1}\log|\mathcal{M}| + \left(2 + \frac{\log|\mathcal{X}|}{\gamma}\right)\ell, \end{aligned}$$

i.e. (5.25).

The error probability:

Denote by E, E_I, and E_γ, the events that errors occur *at any step*, at decoding an IC, and at the decoding of the ordinary code with rate γ, respectively, and by $\Pr(\cdot|v, s^n)$, $v \in \mathcal{M}$, $s^n \in \mathcal{S}^n$, the corresponding output probability, when v is sent and the channel is governed by s^n. Notice that E_I, $E_\gamma \subset E$. We have to upperbound $\Pr(E|v, s^n)$. For this we first notice that

$$\Pr(E_I|v, s^n) < \sum_{r=1}^{r_o} \Pr(U = r|v, s^n) \cdot r\frac{\lambda}{4r_o} \leq \frac{\lambda}{4} \tag{5.26}$$

and therefore

$$\Pr(E|v, s^n) < \frac{\lambda}{4} + \Pr(E|E_I^c, v, s^n). \tag{5.27}$$

We are left with upper bounding

$$\Pr(E|E_I^c, v, s^n) = \sum_{r=0}^{r_o+1} \Pr(U = r|E_I^c, v, s^n)\Pr(E|E_I^c, U = r, v, s^n). \tag{5.28}$$

Here the last summand is upper bounded by the error probability $\frac{\lambda}{4}$ in a $\left(k, |\mathcal{L}_r|, \frac{\lambda}{4}\right)$ –code, which is used for $r = r_o + 1$, because

$$\Pr(E|E_I^c, U = r_o + 1, v, s^n) = \Pr(E_\gamma|v, s^n) < \frac{\lambda}{4}, \tag{5.29}$$

Finally, for $r \le r_o$ by our coding rules

$$W^\ell\big(\{y^\ell : \tilde{\delta}(x^\ell, y^\ell) \ge \xi\}|x^\ell, s^\ell(r)\big) \ge \Pr(U = r|E_I^c, v, s^n) \tag{5.30}$$

where $x^\ell \in \mathcal{T}_P^\ell$ is the value of the r–th $g(v)$, $s^\ell(r)$ is the segment of s^n corresponding to the r–th LROCC.

Therefore by (5.12), (5.13), and (5.20) in the case

$$\Pr(U = r|E_I^c, v, s^n) \ge \frac{\lambda}{4r_o}$$

and with the convention that $s^{m_r}(m_r)$ is the last part of s^n

$$\begin{aligned}
&\Pr(E|E_I^c, U = r, v, s^n) = \big(W^\ell\big(\{y^\ell : \tilde{\delta}(x^\ell, y^\ell) \ge \xi\}|x^\ell, s^\ell(r)\big)\big)^{-1} \\
&\times \sum_{q=1}^{m_r^2} W^\ell\big(K^{-1}(q) \cap \{y^\ell : \tilde{\delta}(x^\ell, y^\ell) \ge \xi\}|x^\ell, s^\ell(r)\big) W^{m_r}\big((D_v^q)^c|u_v^q, s^{m_r}(m_r)\big) \\
&\le \sum_{q=1}^{m_r^2} m_r^{-2} W^{m_r}\big((D_v^q)^c|u_v^q, s^{m_r}(m_r)\big) + \left(\frac{\lambda}{4r_o}\right)^{-1} \cdot 2m_r^2 \omega^{*\frac{1}{4}\xi} < \frac{\lambda}{4},
\end{aligned} \tag{5.31}$$

This and (5.27) – (5.29) imply

$$\Pr(E|v, s^n) < \frac{\lambda}{4} + \frac{\lambda}{4} + \left(1 \cdot \frac{\lambda}{4} + \frac{\lambda}{4} \cdot 1\right) \le \lambda.$$

6 Proof of the Positivity Theorem

We shall, in this section, show that the conditions in Lemma 2 are also sufficient for the positivity. To this end we assume a contradiction, (i) and (ii) in Lemma 2 hold, that is,

$$C_R(\mathcal{W}) > 0 \tag{6.1}$$

and w.l.o.g. for $0, 1 \in \mathcal{X}$

$$\mathcal{Y}_0 \cap \mathcal{Y}_1 = \varnothing, \tag{6.2}$$

but that

$$C_F(\mathcal{W}) = 0. \tag{6.3}$$

We establish the desired result by deriving a contradiction. First we rewrite (6.1) in the form

$$\theta \triangleq \min_{\pi \in \mathcal{P}(\mathcal{S})} \max_{x, x', y} \Big| \sum_s \pi(s) W(y|x', s) - \sum_s \pi(s) W(y|x, s) \Big| > 0 \tag{6.4}$$

and with Lemma 1 (6.3) in the following form: for any two encoding functions f_0^n and f_1^n there exist PD's α^n and β^n on $\mathcal{S}^n$ such that for all $y^n \in \mathcal{Y}^n$

$$\sum_{s^n} \alpha^n(s^n) W^n(y^n|f_0^n, s^n) = \sum_{s^n} \beta^n(s^n) W^n(y^n|f_1^n, s^n). \tag{6.5}$$

The proof in this part is much harder than others in the paper and as well as in most papers in this direction, which contain only a few new ideas and techniques. So it may be hard to understand for some readers. Therefore, we first describe the main idea and give an outline of the proof.

For an input, a sequence of states (or a distribution on the sequences of states) governing the channel and a coloring of the output space, a subset in the output is said to be well colored if its members are colored with (nearly) uniform probability. We have seen that if one can find an input such that for all distributions on the sequences of states the output space is well colored (with a large probability), then the positivity follows. In fact, we shall see that by Lemma 1 any well colored subset is sufficient. However it cannot always be done, and actually it is not hard to see that one can never find such an input, if for all $x \in \mathcal{X}$ $\mathcal{S}_x \neq \varnothing$ (unless (6.5) holds). To obtain the well colored subsets we have to construct 2 encoding functions f_0^n and f_1^n and to show that under the assumption (6.5) one is always able to find a well colored subset for both of them. Our functions consist of 3 blocks with lengths m_1, m_2 and 1, here m_1 and m_2 will be chosen carefully.

In the first two blocks and for both encoding functions, only letters "0" and "1" satisfying (6.2) are used. The first blocks of f_0^n and f_1^n are m_1 zeros and ones respectively. At the same time, the output space $\mathcal{Y}^{m_1}$ is colored by 2^{2m_2} colors, say $\big\{(b^{m_2}, b'^{m_2}) : b^{m_2}, b'^{m_2} \in \{0,1\}^{m_2}\big\}$. For the output y^{m_1} colored by (b^{m_2}, b'^{m_2}), the encoding functions f_0^n and f_1^n encode in the second block to b^{m_2} and b'^{m_2}, respectively. We use the Balanced Coloring Lemma 3, and color $\mathcal{Y}^m$ in the following way.

— Let $\delta^*(x^m, s^m) = |\{t : s_t \notin S_{x_t}\}|$. Then for 0^{m_1} and all s^{m_1} with $\delta^*(0^{m_1}, s^{m_1}) \geq \ell_1$ (i.e. the number of t's such that $s_t \in S_{x_t}$ is not "too large") for a properly chosen ℓ_1, $\mathcal{Y}^{m_1}$ is well colored.

— For 1^{m_1} and all $s^{m_1} \in \mathcal{S}^{m_1}$ all subsets in $\mathcal{Y}^{m_1}$ of the form $A^{m_1} = \prod_{t=1}^{m_1} A_t$, $A_t \in \{\mathcal{Y}, \mathcal{Y}_0\}$, and $|\{t : A_t = \mathcal{Y}_0\}| = m_1 - \ell_1 + 1$, are well colored.

We shall show in Lemma 6 below that if for a probability measure μ on $\mathcal{S}^m$ and fixed $x^m \in \mathcal{X}^m$ $\mu\big(s^n : \delta^*(x^m, s^m) < \ell\big)$ is sufficiently small, then (for some coloring for x^m and μ), $\mathcal{Y}^{m_1}$ is well colored.

Thus,

Case 1: If $\alpha^n\big(s^n : \delta^*(0^{m_1}, s^{m_1}) < \ell_1\big)$ is sufficiently small, then for 0^{m_1} and α^m, $\mathcal{Y}^{m_1}$ (and $\mathcal{Y}^{m_1} \times L$ for all $L \subset \mathcal{Y}^{m_2+1}$) is well colored.

Moreover in Lemma 7 below we shall show

Case 2: If the condition in Case 1 does not hold, under condition (6.5) one can always find an A^{m_1} such that for 1^{m_1} and β^n, A^{m_1} (and $A^{m_1} \times L$ for all $L \subset \mathcal{Y}^{m_2+1}$) is well colored. Thus in the first round of coloring at least for one input we can find a well colored subset.

Next we use the Balanced Coloring Lemma 3 again, but this time we color $\mathcal{Y}^{m_2}$ such that for 0^{m_1} and s^{m_1} with $\delta^*(0^{m_2}, s^{m_2}) \geq \ell_2$ (for suitable ℓ_2) and for 1^{m_1} and s^{m_1} with $\delta^*(1^{m_2}, s^{m_2}) \geq \ell_2$, $\mathcal{Y}^{m_2}$ is well colored.

The hard kernel in the proof is Lemma 8, which we call the Crowd Lemma. It means that if the decoding functions (in the second block) take sufficiently many values and those values crowd the input space, one can always find "good pairs".

We shall show there that, because in the first block we can always for at least one encoding function find a well colored subset, we can always find a pair (b^{m_2}, b'^{m_2}) (as values for f_0^n and f_1^n, respectively, in the second block), such that for the probability distribution α^n or its conditional probability under certain conditions (probability distribution β^n or its conditional distribution under certain conditions), the probability of $\alpha^n\big(s^{m_2} : \delta^*(b^{m_2}, s^{m_2}) < \ell_2\big)$, $\big(\beta^n\big(s^{m_2} : \delta(b'^{m_2}, s^{m_2}) < \ell_2\big)\big)$ for suitable ℓ_2 is sufficiently small.

Thus by Lemma 6 again, we show that for both, f_0^n and α^n and f_1^n and β^n, $\mathcal{Y}^{m_2}$ is well colored. This will complete our proof. Now let us start it.

First we define a pair (f_0^n, f_1^n) of encoding functions and then show that for them (6.4) and (6.5) cannot hold simultaneously. *The definition is given in four steps.*

1. Let $m_1 > l_1 > m_2 > l_2$ and $n = m_1 + m_2 + 1$ be (large) integers depending on a (small) real $\varepsilon > 0$, to be specified later, such that

$$\frac{l_2}{m_2}, \frac{m_2}{l_1}, \frac{l_1}{m_1} \sim \varepsilon. \tag{6.6}$$

2. Recall the definition of $\mathcal{S}_0, \mathcal{S}_1$ in (2.3). For $b^m \in \{0,1\}^m, s^m \in \mathcal{S}^m$ we introduce the "distance"

$$\delta^*(b^m, s^m) \triangleq |\{t : s_t \notin S_{b_t}\}| \tag{6.7}$$

and for m_1 the sets of PD's

$$\mathcal{P}_1 \triangleq \{W^{m_1}(\cdot|0^{m_1}, s^{m_1}) : \delta^*(0^{m_1}, s^{m_1}) \geq l_1\}, \tag{6.8}$$

$$\mathcal{P}_2 \triangleq \{W^{m_1}(\cdot|1^{m_1}, s^{m_1}) : s^{m_1} \in \mathcal{S}^{m_1}\}, \tag{6.9}$$

and the set of output sets

$$\mathcal{A} \triangleq \{A^{m_1} = \prod_{t=1}^{m_1} A_t : A_t \in \{\mathcal{Y}, \mathcal{Y}_0\} \text{ and } |\{t : A_t = \mathcal{Y}_0\}| = m_1 - l_1 + 1\}. \tag{6.10}$$

We now apply the (balanced coloring) Lemma 3 for the choices $\mathcal{V} = \mathcal{Y}^{m_1}, \mathcal{Q} = \mathcal{P}_1 \cup \mathcal{P}_2$, and

$$\mathcal{E}(P) = \left\{ \begin{array}{ll} \{\mathcal{Y}^{m_1}\}, & \text{if } P \in \mathcal{P}_1 \\ \mathcal{A}, & \text{if } P \in \mathcal{P}_2. \end{array} \right\}, \tag{6.11}$$

and color $\mathcal{Y}^{m_1}$ with a coloring function $g = (\Phi_1, \Psi_1) : \mathcal{Y}^{m_1} \to \{0,1\}^{m_1} \times \{0,1\}^{m_1}$ with $k = 2^{2m_2}$ colors.

Let

$$w \triangleq \max\{W(y|x,s) : W(y|x,s) \neq 1, x = 0,1, s \in \mathcal{S} \text{ and } y \in \mathcal{Y}\}. \tag{6.12}$$

Denote the inverse image of the coloring function g for (b^{m_2}, b'^{m_2}) by

$$\Omega_1(b^{m_2}, b'^{m_2}) \triangleq g^{-1}(b^{m_2}, b'^{m_2}) = \Phi_1^{-1}(b^{m_2}) \cap \Psi_1^{-1}(b'^{m_2}) \tag{6.13}$$

and the subset of A^m colored by (b^{m_2}, b'^{m_2}) by

$$A^{m_1}(b^{m_2}, b'^{m_2}) \triangleq A^{m_1} \cap \Omega_1(b^{m_2}, b'^{m_2}), \tag{6.14}$$

(where $A^{m_1} \in \mathcal{A}$ is defined in (6.10)).

(This change is wrong, please keep my original parameters.)

To apply Lemma 3, we check (3.4) i.e.

$\alpha^{-\frac{1}{2}} > \ell n \left[2k \sum_{P \in Q} |\mathcal{E}(P)| \right] = \ell n \big[2k(1 + |\mathcal{A}||\mathcal{S}^{m_2}|) \big]$, which is true since $\alpha(P) \leq w^{\ell_1}$ for $P \in \mathcal{P}_1$ and by (6.2) $\alpha(P) \leq w^{m_1 - \ell_1 + 1}$.

Then by Lemma 3 we have that (c.f. the choices in (3.4))

$$|W^{m_1}(\Omega_1(b^{m_2}, b'^{m_2})|0^{m_2}, s^{m_1}) - \frac{1}{2^{2m_2}}| < 2w^{\frac{l_1}{4}} \tag{6.15}$$

for all $b^{m_2}, b'^{m_2} \in \{0,1\}^{m_2}$ and all s^{m_1} with

$$\delta^*(0^{m_1}, s^{m_1}) \geq l_1 \tag{6.16}$$

and

$$|W^{m_1}(A^{m_1}(b^{m_2}, b'^{m_2})|1^{m_1}, s^{m_1}) - \frac{1}{2^{m_2}} W^{m_1}(A^{m_1}|1^{m_1}, s^{m_1})| < 2w^{\frac{1}{4}(m_1 - l_1 + 1)} \tag{6.17}$$

for all $b^{m_2}, b'^{m_2} \in \{0,1\}^{m_2}$, for all $A^{m_1} \in \mathcal{A}$, and for all $s^{m_1} \in \mathcal{S}^{m_1}$.

3. Next apply Lemma 3 for the choices $\mathcal{V} = \mathcal{Y}^{m_2}, \mathcal{Q} = \mathcal{P}' = \{W^{m_2}(\cdot|b^{m_2}, s^{m_2}) : b^{m_2} \in \{0,1\}^{m_2}, s^{m_2} \in \mathcal{S}^{m_2}, \text{ and } \delta^*(b^{m_2}, s^{m_2}) \geq l_2\}, \mathcal{E}(P) = \{\mathcal{Y}^{m_2}\}$ for all $P \in \mathcal{P}'$, $k = |\mathcal{X}|^2$ and $g' = (\Phi_2, \Psi_2) : Y^{m_2} \to \mathcal{X} \times \mathcal{X}$. Similarly as in 2. we have for

$$\Omega_2(x, x') \triangleq g'^{-1}(x, x') = \Phi_2^{-1}(x) \cap \Psi_2^{-1}(x') \tag{6.18}$$

$$|W^{m_2}(\Omega_2(x, x')|b^{m_2}, s^{m_2}) - \frac{1}{|\mathcal{X}|^2}| < 2w^{\frac{l_2}{4}} \tag{6.19}$$

for all $x, x' \in \mathcal{X}, b^{m_2} \in \{0,1\}^{m_2}$, and $s^{m_2} \in \mathcal{S}^{m_2}$ with $\delta^*(b^{m_2}, s^{m_2}) \geq l_2$ since here $\alpha = w^{l_2}$ and the right hand side of (3.4) polynomially increases, i.e. (3.4) holds.

4. Finally define the announced encoding functions

$$f_0^n = (0^{m_1}, \Phi_1, \Phi_2) \text{ and } f_1^n = (1^{m_1}, \Psi_1, \Psi_2) \tag{6.20}$$

which lead to the desired contradiction. If they satisfy (6.5) for some α^n and β^n, then we can express this also by saying that for the pairs of $RV's$ (S^n, Y^n) and (S'^n, Y'^n) with PD's $\alpha^n(\cdot)W^n(\cdot|f_0^n, \cdot)$ and $\beta^n(\cdot)W^n(\cdot|f_1^n, \cdot)$, resp., Y^n and Y'^n have the same (marginal) distributions.

For the analysis of these $RV's$ we need the following simple Lemmas 6 and 7 and finally the crucial Crowd Lemma 8.

In the sequel we write (with some abuse of notation) $S^{m_1}S^{m_2+1}$ or $S^{m_1}S^{m_2}S$ for S^n and $Y^{m_1}Y^{m_2+1}$ or $Y^{m_1}Y^{m_2}Y$ for Y^n.

We notice that Y^{m_1} or Y^{m_2} following into $\Omega_1(b^{m_2}, b'^{m_2})$, i.e. it getting color (b^{m_2}, b'^{m_2}), implies that in the second block f_0^n and f_1^n will take values b^{m_2} and b'^{m_2}. A similar event will happen in the third block, when the output in the second block gets color (x, x'). These facts will repeatedly be used in our proof.

Lemma 6. *(i) Suppose that*

$$Pr\left(\delta^*(0^{m_1}, S^{m_1}) < l_1\right) < w^{l_1}, \tag{6.21}$$

then for all $b^{m_2}, b'^{m_2} \in \{0,1\}^{m_2}$ *and* $L \subset \mathcal{Y}^{m_2+1}$

$$\begin{aligned}&|Pr(Y^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2}), Y^{m_2+1} \in L)\\ &- \frac{1}{2^{2m_2}} \sum_{s^{m_2+1}} Pr(S^{m_2+1} = s^{m_2+1}), Pr(Y^{m_2+1} \in L | S^{m_2+1} = s^{m_2+1}, Y^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2}))|\\ &< 2w^{\frac{l_1}{4}} + w^{l_1}\end{aligned} \tag{6.22}$$

and one can choose l_1, m_1, *and* m_2 *in (6.6) such that*

$$\begin{aligned}&|Pr(Y^{m_2+1} \in L | Y^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2})) -\\ &\sum_{s^{m_2+1}} Pr(S^{m_2+1} = s^{m_2+1}) Pr(Y^{m_2+1} \in L | S^{m_2+1} = s^{m_2+1}, Y^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2})| < w^{\frac{l_1}{8}}\end{aligned} \tag{6.23}$$

(ii) Suppose that for some $b^{m_2} \in \{0,1\}^{m_2}$ *and* $E \subset \mathcal{Y}^{m_1}$

$$Pr(\delta^*(b^{m_2}, S^{m_2}) < l_2 | Y^{m_1} \in E) < w^{l_2}, \tag{6.24}$$

then for all $x, x' \in \mathcal{X}, K \subset \mathcal{Y}$, *and* $b'^{m_2} \in \{0,1\}^{m_2}$

$$\begin{aligned}&|\sum_{s^{m_2+1}} Pr(S^{m_2+1} = s^{m_2+1} | Y^{m_1} \in E) Pr(Y^{m_2} \in \Omega_2(x, x'), Y \in K | S^{m_2+1} = s^{m_2+1}, Y^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2}))\\ &- \frac{1}{|\mathcal{X}|^2} \sum_{s \in S} Pr(S = s | Y^{m_1} \in E) W(K|x,s)| < 2w^{\frac{l_2}{4}} + w^{l_2}.\end{aligned} \tag{6.25}$$

Moreover, one can replace (S^{m_2}, Y^{m_1}) *and* $W(K|x,s)$ *in (6.24) and (6.25) by* (S'^n, Y'^n) *and* $W(K|x', s)$.

Proof: Let $L = \mathcal{Y}^{m_2+1}$ in (6.22). Then the resulting inequality $\left|\Pr\left(Y^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2})\right) - \frac{1}{2^{2m_2}}\right| \leq 2w^{\frac{\ell_2}{4}} + w^{\ell_1}$ and (6.22) imply (6.23) (c. f. (6.6)). We show now (6.22). By the definition of

(S^n, Y^n) $\Pr\big(Y^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2}), Y^{m_2+1} \in L\big) = \sum_{s^{m_1}s^{m_2}} \Pr(S^n = s^{m_1}s^{m_2+1}) W^{m_1}\big(\Omega_1(b^{m_2}, b'^{m_2})|0^{m_1}s^{m_1}\big)$ $\Pr\big(Y^{m_2+1} \in L|S^{m_2+1} = s^{m_2+1}, Y^{m_1+1} \in \Omega_1(b^{m_2}, b'^{m_2})\big)$ and then the LHS of (6.22) does not exceed

$$\sum_{s^{m_1}s^{m_2+1}} Pr(S^n = s^{m_1}s^{m_2+1})|W^{m_1}(\Omega_1(b^{m_1}, b'^{m_1})|0^{m_1}, s^{m_1}) - \frac{1}{2^{2m_2}}|$$
$$\times Pr(Y^{m_2+1} \in L|S^{m_2+1} = s^{m_2+1}, Y^{m_1+1} \in \Omega_1(b^{m_2}, b'^{m_2})),$$

which together with (6.15), (6.16) and (6.21) yields (6.22) (by splitting $\mathcal{S}^n$ to $\big\{s^{m_1+m_2+1} : \delta^*(0^{m_1}, s^{m_1}) \geq \ell_1\big\}$ and $\big\{s^{m_1+m_2+1} : \delta^*(0^{m_1}, s^{m_1}) < \ell_1\big\}$).

Notice that by the definition of (Y^n, S^n) and (6.20) for $s^{m_2+1} = s^{m_2}s$ in (6.25)

$$Pr(Y^{m_2} \in \Omega_2(x, x'), Y \in K|S^{m_2+1} = s^{m_2+1}, Y^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2}))$$
$$= W^{m_2}(\Omega_2(x, x')|b^{m_2}, s^{m_2})W(K|x, s)$$

and hence (ii) can be established exactly like (i).

The importance of (6.22) and (6.23) (resp. (6.25)) is that S^{m_2+1} (resp. S) in the second terms (resp. term) is independent of $\Phi_1(Y^{m_1})$ (resp. $\Phi_2(Y^{m_2})$). Intuitively speaking, the jammer has very little knowledge about the output to come. The same phenomenon can be encountered in the next auxiliary result.

Lemma 7. *For all $A^{m_1} \in \mathcal{A}, b^{m_2}, b'^{m_2} \in \{0,1\}^{m_2}$ and $L \subset \mathcal{Y}^{m_2+1}$*

$$|Pr(Y'^{m_1} \in A^{m_1}(b^{m_2}, b'^{m_2}), Y'^{m_2+1} \in L)$$
$$- \frac{1}{2^{2m_2}} Pr(Y'^{m_1} \in A^{m_1}) \sum_{s^{m_2+1}} Pr(S'^{m_2+1} = s^{m_2+1}|Y'^{m_1} \in A^{m_1})$$
$$\times Pr(Y'^{m_2+1} \in L|S'^{m_2+1} = s^{m_2+1}, \Psi_1(Y'^{m_1}) = b'^{m_2})| < 2w^{\frac{m_1-l_1+1}{4}}. \tag{6.26}$$

Moreover, if (6.21) does not hold, one can always choose the parameters according to (6.6) and find an $A^{m_1} \in \mathcal{A}$ in such a way that

$$|Pr(Y'^{m_2+1} \in L|Y'^{m_1} \in A^{m_1}(b^{m_2}, b'^{m_2})) - \sum_{s^{m_2+1}} Pr(S'^{m_2+1} = s^{m_2+1}|Y'^{m_1} \in A^{m_1})\times$$
$$Pr(Y'^{m_2+1} \in L|S'^{m_2+1} = s^{m_2+1}, \Psi_1(Y'^{m_1}) = b'^{m_2})| < w^{l_1}. \tag{6.27}$$

Proof: (6.26) is proved analogously to (6.22). However, notice that here all $W^{m_1}(\cdot|1^{m_1}, s^{m_1})$ are contained in $\mathcal{P}_2 \subset Q$ (see (6.9)) and therefore no condition analogous to (6.21) is necessary. To obtain (6.27) from (6.26) we let $L = \mathcal{Y}^{m_2+1}$ in (6.26) and get

$$|Pr(Y'^{m_1} \in A^{m_1}(b^{m_2}, b'^{m_2})) - \frac{1}{2^{2m_2}} Pr(Y'^{m_1} \in A^{m_1})| < 2w^{\frac{1}{4}(m_1-l_{1+1})} \tag{6.28}$$

A difficulty now arises. In order to obtain a good bound w^{l_1} at the RHS of (2.27), we have to find an $A^{m_1} \in \mathcal{A}$ such that $Pr(Y'^{m_1} \in A^{m_1})$ is not too small. Assume then that (6.21) does not hold and we now look for our A^{m_1}. Since the set $\{s^m : \delta^*(0^{m_1}, s^{m_1}) < l_1\}$ is covered by the family of sets

$$\mathcal{B} \triangleq \left\{ \prod_{t=1}^{m_1} B_t : B_t \in \{\mathcal{S}_0, \mathcal{S}\} \text{ and } |\{t : B_t = \mathcal{S}_0\}| = m_1 - l_1 + 1 \right\},$$

$\sum_{B^{m_1} \in \mathcal{B}} \Pr(S^{m_1} \in B^{m_1}) \geq \Pr\big(\delta^*(0^{m_1}, S^{m_1}) < \ell_1\big) \geq w^{\ell_1}$ and therefore one member of $\mathcal{B}$, say $B^{m_1} = \mathcal{S}_0^{m_1 - l_1 + 1} \times \mathcal{S}^{l_1 - 1}$, must have the probability

$$Pr(S^{m_1} \in B^{m_1}) \geq \begin{pmatrix} m_1 \\ l_1 - 1 \end{pmatrix}^{-1} w^{l_1}, \tag{6.29}$$

if (6.21) does not hold since $|\mathcal{B}| = \binom{m_1}{\ell_1 - 1}$. We then choose $A^{m_1} = \mathcal{Y}_0^{m_1 - l_1 + 1} \times \mathcal{Y}^{l_1 - 1}$. Notice that for all $s^{m_1} \in B^{m_1}$

$$W^{m_1}(A^{m_1} | 0^{m_1}, s^{m_1}) = 1 \tag{6.30}$$

Recalling Y^n and Y'^n have the same distributions, we conclude from (6.20), (6.29), and (6.30) that

$$Pr(Y'^{m_1} \in A^{m_1}) = Pr(Y^{m_1} \in A^{m_1}) \geq \sum_{s^{m_1} \in B^{m_1}} Pr(S^{m_1} = s^{m_1}) W^{m_1}(A^{m_1} | 0, s^{m_1})$$

$$\geq \begin{pmatrix} m_1 \\ l_1 - 1 \end{pmatrix}^{-1} w^{l_1}.$$

With the above inequality and the relation $2^{2m_2+1}\binom{m_1}{\ell_1 - 1} w^{\frac{m_1 - \ell_1 + 1}{4} - \ell_1} = 0(1)$ (which follows from the assumption in (6.6)) and (6.28), (6.227) can be obtained by dividing (2.26) by $\Pr(Y'^{m_1} \in A^{m_1})$.

Now comes the kernel of the proof.

Crowd Lemma 8. *For suitable parameters in (6.6)*

(i) For all PD σ on $\mathcal{S}^{m_2}$ there exists a $b^{m_2} \in \{0,1\}^{m_2}$ such that

$$\sigma(s^{m_2} : \delta^*(b^{m_2}, s^{m_2}) < l_2) < w^{l_2}. \tag{6.31}$$

(ii) If (6.23) holds, then for all $b^{m_2} \in \{0,1\}^{m_2}$ there exists a $b'^{m_2} \in \{0,1\}^{m_2}$ such that

$$Pr(\delta^*(b'^{m_2}, S'^{m_2}) < l_2 | Y'^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2})) < w^{l_2}. \tag{6.32}$$

(iii) If (6.27) holds, then for all $b'^{m_2} \in \{0,1\}^{m_2}$ there exists a $b^{m_2} \in \{0,1\}^{m_2}$ such that

$$Pr(\delta^*(b^{m_2}, S'^{m_2}) < l_2 | Y^{m_1} \in A^{m_1}(b^{m_2}, b'^{m_2})) < w^{l_2}. \tag{6.33}$$

Proof: Ad(i). Assume to the opposite that for some σ and all b^{m_2}

$$\sigma(s^{m_2} : \delta^*(b^{m_2}, s^{m_2}) < l_2) \geq w^{l_2}.$$

Then we add up these inequalities over all $b^{m_2} \in \{0,1\}^{m_2}$. Since for all $s^{m_2} \in \mathcal{S}^{m_2}$ there are at most $\sum_{j=0}^{l_2-1} \left(\begin{array}{c} m_2 \\ j \end{array} \right) 2^j \;\; b^{m_2'} s$ with $\delta^*(b^{m_2}, s^{m_2}) < l_2$ we obtain that

$$\sum_{j=0}^{l_2-1} \left(\begin{array}{c} m_2 \\ j \end{array} \right) 2^j \geq \sum_{s^{m_2}} \sigma(s^{m_2}) |\{b^{m_2} : \delta^*(b^{m_2}, s^{m_2}) < \ell_2\}| =$$

$$\sum_{b^{m_2} \in \{0,1\}^{m_2}} \sigma(s^{m_2} : \delta^*(b^{m_2}, s^{m_2}) < l_2) \geq 2^{m_2} w^{l_2},$$

which cannot happen for sufficiently small ε and large l_2 in (6.6).

Ad (ii) and (iii). We only show that (6.32) holds under (6.23), because (iii) can be proved in the same way, whereas in (i) we dealt with one PD, we deal now with a family of PD's. This makes things harder. Define for all $b'^{m_2} \in \{0,1\}^{m_2}$ and $\tilde{\delta}$ in (5.6).

$$L^*(b'^{m_2}) \triangleq \{y^{m_2} \in \mathcal{Y}^{m_2} : \tilde{\delta}(b'^{m_2}, y^{m_2}) < l_2\}. \tag{6.34}$$

Then for all s^{m_2} with $\delta^*(b'^{m_2}, s^{m_2}) < l_2$ by the definitions of (S'^n, Y'^n) and $\mathcal{S}_x$,

$$Pr(Y'^{m_2} \in L^*(b'^{m_2}) | S'^{m_2} = s^{m_2}, Y'^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2})) = W^{m_2}(L^*(b'^{m_2}) | b'^{m_2}, s^{m_2}) = 1. \tag{6.35}$$

Consequently, if (6.32) is false, i.e. for *some* b^{m_2} and *all* b'^{m_2}.

$$Pr(\delta^*(b'^{m_2}, S'^{m_2}) < l_2 | Y'^{m_2} \in \Omega_1(b^{m_2}, b'^{m_2})) \geq w^{l_2},$$

then for such a b^{m_2} and all b'^{m_2}, by (6.35)

$$Pr(Y'^{m_2} \in L^*(b'^{m_2}) | Y'^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2})) = \sum_{s^{m_2}} Pr(S'^{m_2} = s^{m_2} | Y'^{m_2} \in \Omega_1(b^{m_2}, b'^{m_2}))$$

$$\begin{aligned}
&\times Pr(Y'^{m_2} \in L^*(b'^{m_2}) | S'^{m_2} = s^{m_2}, Y'^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2}) \\
&\geq \sum_{s^{m_2} : \delta^*(b^{m_2}, s^{m_2}) < \ell_2} \Pr\big(S'^{m_2} = s^{m_2} | Y^{m_2} \in \Omega_1(b^{m_2}, b'^{m_2}) \Pr\big(Y'^{m_2} \in L^*(b'^{m_2}) | S'^{m_2} \\
&= s^{m_2}, Y'^{m_2} \in \Omega_1(b^{m_2}, b'^{m_2})\big) \\
&= \Pr\big(\delta^*(b'^{m_2}, S'^{m_2}) < \ell_2 | Y'^{m_2} \in \Omega_1(b^{m_2}, b'^{m_2})\big) > w^{l_2}.
\end{aligned}$$

Therefore, since Y^n and Y'^n have the same distributions,

$$Pr(Y^{m_2} \in L^*(b'^{m_2}) | Y^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2})) \geq w^{l_2}. \tag{6.36}$$

Apply now (6.23) to $L = L^*(b'^{m_2})$ for all b'^{m_2}. Thus

$$\sum_{s^{m_2+1}} Pr(S^{m_2+1} = s^{m_2+1}) Pr(Y^{m_2} \in L^*(b'^{m_2}) | S^{m_2+1} = s^{m_2+1}, Y^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2})) \geq w^{l_2} - w^{\frac{l_1}{8}}. \tag{6.37}$$

Finally, by adding both sides of (6.37) over $\{0,1\}^{m_2}$ and by using the fact that each $y^{m_2} \in Y^{m_2}$ is covered by at most $\sum_{j=0}^{l_2-1} \binom{m_2}{j} 2^j$ sets $L^*(b'^{m_2})$ in (6.34) we arrive at

$$\sum_{j=0}^{l_2-1} \binom{m_2}{j} 2^j \geq$$

$$\sum_{s^{m_2+1}} Pr(S^{m_2+1} = s^{m_2+1}) \sum_{b'^{m_2} \in \{0,1\}^{m_2}} Pr(Y^{m_2} \in L^*(b'^{m_2}) | S^{m_2+1} = s^{m_2+1}, Y^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2}))$$

$$\geq 2^{m_2}(w^{l_2} - w^{\frac{l_1}{8}}), \tag{6.38}$$

which contradicts (6.6).

The idea behind the Crowd Lemma is that an encoding function with enough different values has always "a good" value against the jamming.

Now it's time for the harvest.

Proof of Positivity Theorem: We use Lemmas 6-8 to obtain a contradiction to (6.4). This is done in two cases.

Case 1 (6.21) holds: Then by Lemma 6 also (6.23) holds. We apply Lemma 8 (i) to $\sigma = P_{S^{m_2}}$ and obtain a b^{m_2} such that (6.24) holds with $E = \mathcal{Y}^{m_1}$ (i.e. unconditional distribution). Fix this b^{m_2} and apply Lemma 6 (ii) for $E = \mathcal{Y}^{m_1}$. Thus we obtain (6.25) with $E = \mathcal{Y}^{m_1}$. Choose next $L = \Omega_2(x, x') \times K$ in (6.23) and combine it with (6.25) for $E = \mathcal{Y}^{m_1}$. Thus we get that for the fixed b^{m_2}, all $x, x' \in \mathcal{X}$, all $b'^{m_2} \in \{0,1\}^{m_2}$, and all $K \subset \mathcal{X}$

$$|Pr(Y^{m_2} \in \Omega_2(x, x'), Y \in K | Y^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2}))$$
$$-\frac{1}{|\mathcal{X}|^2} \sum_s Pr(S = s) W(K|x, s)| < w^{\frac{l_1}{8}} + 2w^{\frac{l_2}{4}} + w^{l_2}. \tag{6.39}$$

On the other hand, since (6.23) holds, we can find a b'^{m_2} for the fixed b^{m_2} so that (6.32) holds by (ii) in Lemma 8. That is, after replacing (S^n, Y^n) by (S'^n, Y'^n), (6.24) holds for $E = \Omega_1(b^{m_2}, b'^{m_2})$ and therefore, by Lemma 6 (ii) again, but this time for (S'^n, Y'^n) (instead of (S^n, Y^n)) and $E = \Omega_1(b^{m_2}, b'^{m_2})$ we obtain for the fixed b^{m_2}, b'^{m_2}, all $x, x' \in \mathcal{X}$, and $K \subset \mathcal{Y}$

$$|Pr(Y'^{m_2} \in \Omega_2(x, x'), Y' \in K | Y'^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2}))$$
$$-\frac{1}{|\mathcal{X}|^2} \sum_{s \in \mathcal{S}} Pr(S' = s | Y'^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2})) W(K|x', s)| < 2w^{\frac{l_2}{4}} + w^{l_2}, \tag{6.40}$$

where we use the fact that

$$Pr(Y'^{m_2} \in \Omega_2(x,x'), Y' \in K | Y'^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2}))$$

$$= \sum_{s^{m_2+1}} Pr(S'^{m_2+1} = s^{m_2+1} | Y'^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2}))$$

$$\times Pr(Y'^{m_2} \in \Omega_2(x,x'), Y' \in K | S'^{m_2+1} = s^{m_2+1}, Y'^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2})).$$

Finally, let l_1 and l_2 be sufficiently large, then from (6.39), (6.40), and the fact that Y^n and Y'^n have the same distributions we obtain that for θ in (6.4), all $x, x' \in \mathcal{X}$ and $K \subset \mathcal{Y}$,

$$|\sum_s Pr(S=s)W(K|x,s) - \sum_s Pr(S'=s|Y'^{m_1} \in \Omega_1(b^{m_2}, b'^{m_2})) \cdot W(K|x',s)| \leq \frac{\theta}{3},$$

or, for all $x, x'' \in \mathcal{X}$ and $K \subset \mathcal{Y}$.

$$\left|\sum_s Pr(S=s)W(K|x,s) - \sum_s Pr(S=s)W(K|x'',s)\right| < \frac{2\theta}{3} \tag{6.41}$$

which contradicts (6.4) (for $K = \{y\}$).

Case 2: (6.21) does not hold: Here by Lemma 7 we have (6.27) for an $A^{m_2} \in \mathcal{A}$. Fix this A^{m_2} by applying Lemma 8(i) for $\sigma = \Pr(\cdot|Y'^{m_2} \in A^{m_2})$, we obtain that for a (fixed) b'^{m_2} $Pr(\delta^*(b'^{m_2}, S'^{m_2}) < l_2 | Y'^{m_2} \in A^{m_2}) < w^{l_2}$, i.e. (6.24) in terms of the distribution (S'^n, Y'^n) and with $E = A^{m_2}$. Therefore we have (6.25) in terms of the distribution of (S'^n, Y'^n) with $E = A^{m_2}$ and then an inequality in terms of the distribution of (S'^n, Y'^n), analogous to (6.39), by combining (6.25) and (6.27). Next for the fixed b'^{m_2} (obtained by applying Lemma 8 (i) in this case), we find a b^{m_2} such that (6.33) holds. Now we set $E = A^{m_1}(b^{m_2}, b'^{m_2})$ in Lemma 6 (ii) and obtain an inequality, analogous to (6.40), but in terms of the distribution of (S^n, Y^n). Finally, we get an inequality analogous to (6.41), which contradicts (6.4).

7 References

[1] R. Ahlswede, Channels with arbitrarily varying channel probability functions in the presence of noiseless feedback, Z. Wahrsch. Verw. Gebiete, vol. 25, pp. 239–252, 1973.

[2] R. Ahlswede, Elimination of correlation in random codes for arbitrarily varying channels, Z. Wahrsch. Verw. Gebiete, vol. 44, pp. 159–175, 1978.

[3] R. Ahlswede, Coloring hypergraphs: a new approach to multi–user source coding, J. Combin. Inform. System Sci., Part I, vol. 4, pp. 76–115, 1979 and Part II, vol. 5, pp. 220–268, 1980.

[4] R. Ahlswede and V.B. Balakirsky, Identification under random processes, Preprint 95–098, SFB 343 "Diskrete Strukturen in der Mathematik", Universität Bielefeld, 1995, Problemy peredachii informatsii (special issue devoted to M.S. Pinsker), vol. 32, no. 1, pp. 144–160, Jan.–March 1996.

[5] R. Ahlswede and N. Cai, Two proofs of Pinskers conjecture concerning arbitrarily varying channels, IEEE Trans. Inform. Theory, vol. IT–37, pp. 1647–1649, 1991.

[6] R. Ahlswede and N. Cai, Correlated sources help the transmission over AVC, Preprint 95–106, SFB 343 "Diskrete Strukturen in der Mathematik", Universität Bielefeld, 1995, IEEE Trans. Inf. Theory, Vol. IT–43, No. 1, pp. 37–67, 1997.

[7] R. Ahlswede and I. Csiszár, Common randomness in information theory and cryptography, Part 1: Secret sharing, IEEE Trans. Inform. Theory, vol. IT–39, pp. 1121–1132, 1993 and Part 2: CR capacity, Preprint 95–101, SFB 343 "Diskrete Strukturen in der Mathematik", Universität Bielefeld, 1995, IEEE Trans. Inf. Theory, Vol. 44, No. 1, pp. 55–62, 1998.

[8] R. Ahlswede and G. Dueck, Identification via channels, IEEE Trans. Inform. Theory, vol. IT–35, pp. 15–29, 1989.

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

[10] R. Ahlswede and Z. Zhang, New directions in the theory of identification via channels, IEEE Trans. Inform. Theory, vol. IT–41, pp. 1040–1050,

[11] J. Kiefer and J. Wolfowitz, Channels with arbitrarily varying channel probability functions, Inform. and Control, vol. 5, pp. 44–54, 1962.

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

[13] S. Lin and D.J. Costello, Jr., Error control coding: Fundamentals and Applications, Prentice–Hall, Inc., Englewood Cliffs, N.J. 1983.

[14] J.M. Ooi, A Framework for Low–Complexity Communication Channels with Feedback, Dissertation at MIT, RLE Technical Report No. 617, Nov. 1997.

[15] J.M. Ooi and Gregory W. Wornell, Fast iterative coding for feedback channels, 1997 Proceedings IEEE Int. Symp. on Inf. Theory, page 133, Ulm, Germany, June 29 – July 4, 1997.