Semi-noisy Deterministic Multiple Access Channels: Coding Theorems for List Codes and Codes with Feedback

Rudolf Ahlswede and Ning Cai

Abstract

Whereas the average error capacity region $\mathcal{R}_a$ for the MAC $W : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ is known for a long time (Ahlswede 71), very little is known about the capacity region $\mathcal{R}_m$ for the maximal error concept (as predicted also by Ahlswede in 71).

Inspite of great efforts during the past three decades even for some special examples of deterministic MAC, for which the maximal error concept coincides with the concept of unique decodability, the progress has been slow.

It is known that the permission of list codes can be of great help, even if list sizes are of negligible rates (c.f. AVC and especially Shannon's zero–error capacity problem for the one–way channels).

Therefore it is theoretically appealing to look at their regions $\mathcal{R}_{m,\ell}$ for the MAC. For a nice class of deterministic MAC, which we call "semi–noisy", we completely characterized $\mathcal{R}_{m,\ell}$. For these channels the $\mathcal{Y}$–input is determined uniquely by the output. Dueck's example with $\mathcal{R}_a \neq \mathcal{R}_m$ and Vanroose's "Noiseless binary switching MAC" with $\mathcal{R}_a = \mathcal{R}_m$ fall into this class.

Finally, for this class the capacity region $\mathcal{R}_{m,f}$, which concerns complete feedback, equals $\mathcal{R}_{m,\ell}$.

Keywords: Multiple access channel, listcodes, feedback, average, maximal and zero error probabilities

1 Introduction

After Shannon laid the foundations of a theory of transmission in the presence of noise for one-way channels in [14] he made subsequently significant extensions of his model in [15] and [16]. Actually, in [15] we find two additional issues: the performance of channels under the *zero-error* criterion and in the presence of *feedback*. In [16] Shannon introduces models of channels with several senders and receivers. Quite remarkably the so called two-way channel, which is discussed there as the only channel in greater detail, envolves also feedback. It must be emphasized that both papers lead to hard mathematical problems, on which until today the progress has been relatively small. Incidentally, the first author came on independent paths to the MAC, as a channel without feedback, and the DMC with complete feedback and obtained two striking results: a complete characterisation of the capacity region of the MAC [1] (and thus the first result of this kind for a channel with several senders and receivers) and a constructive coding scheme for the DMC with complete feedback achieving capacity [2].

The first result was largely responsible for a period of stormy developments in multi-user channel and source coding theory concerning existence theorems, but which got stuck for instance already for the interference channel and the general broadcast channel (see [7], [8]). It also gave the start for code constructions for special MAC (like the adder channel etc.) under the maximal error criterion (that is, equivalently, for uniquely decodable or zero-error codes in case of deterministic channels).

Already in [1] it was mentioned that maximal error capacity regions are in general smaller than those for average error and that its determination, if it ever should be achieved, must be very difficult and requires new methods. Example 1 in Section 2 is Dueck's [9] MAC with different regions. On the feedback side the scheme of [2] mentioned also gave fruits in connection with the MAC. Since here the feedback creates "correlated knowledge" for the senders it obviously enlarges the capacity region of the MAC, which is based on independent input assignments and no general capacity formula in case of feedback is known even today. However, a series of partial results have been obtained. N.T. Gaarder and J.K. Wolf [12] set up a simple example of a MAC for which feedback enlarges the capacity region for average probability of errors.

Then T.M. Cover and C.S.K. Leung [6] obtained inner bounds of capacity regions of the discrete memoryless and the white Gaussian noise MAC with feedback and average probability of errors. Actually, in the second case the true bound was determined by L. Ozarow. It is strictly larger than the inner bound in [6]. F.M.J. Willems [18] proved that the inner bound for discrete memoryless MAC in [6] is tight in the case the MAC satisfies the condition

($*$): There is (at least) one input alphabet whose letters are a function of the other input and output letters.

By our knowledge (besides this paper) the only paper concerning the capacity regions of MAC with feedback and maximal or zero probability of error is [10]. There G. Dueck determined the zero-error capacity region with feedback of the class of discrete memoryless MAC satisfying the condition

$(**)$: Each input letter is a function of the other input and output letters.

In another direction the coding scheme of [2] led to the solution of the capacity problem for AVC with complete feedback in [3], under a constraint, and recently to the full solution in [5]. It also led to the idea of binning in [4] and to a capacity formula of AVC for list codes (introduced by Elias [11]) of negligible list rates. In this context the main observations are that

(a) list codes are especially connected to feedback coding problems.

(b) list codes are mathematically a more natural concept than ordinary codes (for instance for zero-error problems and more generally for AVC).

Returning to the foundations of 1970 the aim of this paper is to start a systematic exploration of these observations for the MAC. What are here the connections between ordinary codes, list codes and feedback codes while the criteria change from zero errors over maximal errors to average errors.

The contribution of this paper is the discovery of another subclass of deterministic MAC satisfying $(*)$, which we call semi-noisy deterministic channels, and for which we obtain conclusive results on capacity regions.

2 Basic definitions: channels, codes and capacity regions

Let us consider a memoryless multiple access channel (MAC) W with input alphabets $\mathcal{X}$ and $\mathcal{Y}$ and output alphabet $\mathcal{Z}$ i. e., when $x^n = (x_1, x_2, \dots, x_n) \in \mathcal{X}^n$ and $y^n = (y_1, y_2, \dots, y_n) \in \mathcal{Y}^n$ are input sequences, $z^n = (z_1, z_2, \dots, z_n) \in \mathcal{Z}^n$ is output sequence with probability

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

We say an MAC is deterministic (or a d-MAC) if there is a function $\phi : \mathcal{X} \times \mathcal{Y} \longrightarrow \mathcal{Z}$ such that for all $x \in \mathcal{X}, y \in \mathcal{Y}$, and $z \in \mathcal{Z}$,

$$W(z|x, y) = \begin{cases} 1 & \text{if } z = \phi(x, y) \\ 0 & \text{else .} \end{cases}$$

A d-MAC is semi-noisy, or sd-MAC (for $\mathcal{X}$), if there is a function $\psi : \mathcal{Z} \longrightarrow \mathcal{Y}$ such that $\psi(z) = y$ whenever there is an $x \in \mathcal{X}$ with $W(z|x, y) > 0$. In other words, a d-MAC is an MAC whose output is determined by the inputs uniquely with probability one and an sd-MAC for $\mathcal{X}$ is a d-MAC whose $\mathcal{Y}$-input is determined by the output uniquely.

The class of sd-MAC's includes as members examples 1 and 2 below, which are used by Dueck [9] resp. Vanrose [17] to contribute to an understanding of the relation between the capacity regions of codes with maximal error (or zero error) and average error.

Recall that an (n, M_1, M_2, λ) code with maximal (resp. average) probability of error, or briefly an (n, M_1, M_2, λ) m-code (resp. a-code), is a system $\{u_i, v_j, \mathcal{D}_{i,j} : i = 1, \ldots, M_1 \text{ and } j = 1, \ldots, M_2\}$, such that $u_i \in \mathcal{X}^n$ for $i = 1, \ldots, M_1$, $v_j \in \mathcal{Y}^n$ for $j = 1, \ldots, M_2$, $\mathcal{D}_{i,j} \subset \mathcal{Z}^n$ with $\mathcal{D}_{i,j} \cap \mathcal{D}_{i',j'} = \emptyset$ for $i, i' = 1, \ldots, M_1, j, j' = 1, \ldots, M_2, (i,j) \neq (i'j')$ and $W^n(\mathcal{D}_{i,j}|u_i, v_j) > 1 - \lambda$ for all i, j (resp. $\frac{1}{M_1} \frac{1}{M_2} \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} W^n(\mathcal{D}_{i,j}|u_i, v_j) > 1 - \lambda$).

We speak of codes with feedback when there are noiseless channels to connect the outputs and the encoders so that both encoders (senders) are able to choose the next input letters according to the previous outputs. Thus an (n, M_1, M_2, λ) code with (noiseless) feedback and the criterion of maximal probability of error , or an (m,f)-code, is a system $\{f_i^n, g_j^n, \mathcal{D}_{i,j} : i = 1, \ldots, M_1, \text{ and } j = 1, \ldots, M_2\}$ such that f_i^n's are functions from $\mathcal{Z}^n$ to $\mathcal{X}^n$ such that for all $i = 1, \ldots, M_1$ and $z^n \in \mathcal{Z}^n$,

$$f_i^n(z^n) = (f_i^{(1)}, f_i^{(2)}(z_1), \ldots, f_i^{(t)}(z^{t-1}), \ldots, f_i^{(n)}(z^{n-1})), \tag{1}$$

and g_j^n's are functions from $\mathcal{Z}^n$ to $\mathcal{Y}^n$ such that all $j = 1, \ldots, M_2$ and $z^n \in \mathcal{Z}^n$,

$$g_j^n(z^n) = (g_j^{(1)}, g_j^{(2)}(z_1), \ldots g_j^{(t)}(z^{t-1}), \ldots g_j^{(n)}(z^{n-1})), \tag{2}$$

where $f_i^{(1)}, f_i^{(2)}(z_1), \ldots, f_i^{(n)}(z^{n-1}) \in \mathcal{X}, g_j^{(1)}, g_j^{(2)}(z_1), \ldots, g_j^{(n)}(z^{n-1}) \in \mathcal{Y}$, and $z^{(t-1)} = (z_1, \ldots, z_{t-1})$, and for all$i, j, W^n(\mathcal{D}_{i,j}|f_i^n, g_j^n) > 1 - \lambda$.

Obviously, for any $\lambda < 1$ an (n, M_1, M_2, λ) m-code for a d-MAC is an $(n, M_1, M_2, 0)$ m-code for the same channel, which is also called a zero-error, error-free or uniquely decodable code. The capacity regions for m-codes, a-codes, and (m, f)-codes are defined in the standard way and we denote them by $\mathcal{R}_m, \mathcal{R}_a$ and $\mathcal{R}_{m,f}$ respectively. $\mathcal{R}_a$ was determined for all MAC by R. Ahlswede [1].

Finally , we define an $(n, M_1, M_2, \lambda, L)$ list code with maximal probability of error (an (m, l)-code) as a system $\{u_i, v_j, \mathcal{D}_{i,j} : i = 1, \ldots, M_1 \text{ and } j = 1, \ldots, M_2\}$ such that $u_i \in \mathcal{X}^n, v_j \in \mathcal{Y}^n, \mathcal{D}_{i,j} \subset \mathcal{Z}^n, W^n(\mathcal{D}_{i,j}|u_i, v_j) > 1-\lambda$ and for all $z^n \in \mathcal{Z}^n$, for the list size $|\mathcal{L}(z^n)| := |\{(i,j) : z^n \in \mathcal{D}_{i,j}\}| \leq L$. We speak of (m, l)-codes and the capacity region $\mathcal{R}_{m,l}$ for (m, l)-codes is the set of real pairs (r_1, r_2) such that for all $\epsilon, \delta, \lambda > 0$ there is an $(n, M_1, M_2, \lambda, L)$ (m, l)-code with $\frac{1}{n} \log M_k > r_k - \epsilon$ for $k = 1, 2$ and $\frac{1}{n} \log L < \delta$. It is also clear that for any $\lambda < 1$ and any d-MAC, an $(n, M_1, M_2, \lambda, L)$ (m,l) code is an error-free list code.

Analogously, one can also define (a, f)- and (a, l)- codes if the average probability of error criterion is used, but we shall not discuss them here.

By their definitions, we immediately have that for all MAC $\mathcal{R}_m \subset \mathcal{R}_a$ and $\mathcal{R}_m \subset \mathcal{R}_{m,l}$. Moreover we observe that the proof to the converse coding theorem for a-codes of an MAC [1] can be modified to yield a converse result for list codes and therefore $\mathcal{R}_{m,l} \subset \mathcal{R}_a$. It is easy to see that $\mathcal{R}_{m,l} \subset \mathcal{R}_{m,f}$, if $\mathcal{R}_{m,f}$ has a pair of positive rates by using the following scheme. The two encoders first employ an (m, l)-code with sufficiently small list size (for all outputs) to send their messages. Due to the feedback, both of them learn the output z^n and therefore $\mathcal{L}(z^n)$. Thus they can inform the receiver which pair of messages on the list is correct by an (m, f)-code with sufficiently small rate (depending on the list size of the previous code). Summarizing the above facts, we have the chains of containments

$$\mathcal{R}_m \subset \mathcal{R}_{m,l} \subset \mathcal{R}_a \tag{3}$$

and if $\mathcal{R}_{m,f}$ contains a pair of positive rates (c.f. Lemma 6.1)

$$\mathcal{R}_m \subset \mathcal{R}_{m,l} \subset \mathcal{R}_{m,f}. \tag{4}$$

In general, there is no containment relation for MAC between $\mathcal{R}_{m,f}$ and $\mathcal{R}_a$.

Example 1 (G. Dueck [9]): Let $\mathcal{X} = \{A, B, a, b\}, \mathcal{Y} = \{0, 1\}$, and $\mathcal{Z} = \{A, B, C, a, b, c\}$. The channel W is defined such that for $y = 0, W(z|x, y) = 1$ iff $x = z = A$ or B, or $x = a$ or b, and $z = c$; and for $y = 1$, $W(z|x, y) = 1$ iff $x = A$, or B, and $z = C$ or $x = z = a$ or b. Notice that A, B, or c is output iff $y = 0 \in \mathcal{Y}$ is input and so it is an sd-MAC. Dueck proved that for this channel $\mathcal{R}_m, \mathcal{R}_{m,l}$ and $\mathcal{R}_{m,f}$ are strictly smaller than $\mathcal{R}_a$.

Example 2 The "Noiseless Binary Switching MAC" (P. Vanroose [17]): Let $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ and $\mathcal{Z} = \{0, 1, \infty\}$. Define the channel W such that for $y = 0$ and all $x \in \mathcal{X}$, $W(\infty|x, y) = 1$ and for $y = 1$, $W(z|x, y) = 1$ iff $z = y$. Obviously this is an sd-MAC. Vanroose showed that $\mathcal{R}_m = \mathcal{R}_a$ and this gave an example of a nontrivial channel with this property.

It is interesting that these two extremal examples of MAC's fall into our class of sd-MAC.

3 Incomplete (one sided) feedback is sufficient

Usually in the MAC coding theory feedback plays two roles.

-*Reducing list size*: The idea was started by Shannon in his pioneering work [15] on zero-error capacities of memoryless channels and led to the List Reduction Lemma for maximal error probabilities in [2] and its extension in [3]. According to the initial part $(z_1, \ldots, z_{t-1})$ of the output sequence, which he has received at time $t-1$, the receiver lists all possible messages (corresponding to $(z_1, \ldots, z_{t-1})$). The encoders learn

$(z_1, \ldots, z_{t-1})$ from the feedback at the same time and therefore the list. So they are able to cooperate to work on the list instead of the whole sets of the messages. Thus the encoders choose the next input symbols according to the previous outputs to reduce the list sizes step by step until the size of the list reaches one so that the receiver learns the messages sent by the encoders.

-*Shifting the private messages to the common messages*: The idea was used to show that feedback enlarges the capacity region of MAC in [12] and [6] and also for determining the zero-error capacity regions for a fairly large class of MAC in [10]. A difficulty for the encoders of an MAC (without feedback) to cooperate is that their (private) messages are independent. When feedback is present, an encoder can obtain certain information about the message sent by the other encoder via his own inputs and the previous outputs which he got via the feedback. Thus he can shift the information from the private messages (of the other encoder) to "common messages" and cooperate with the other encoder to send them. Notice that their "information" must be "unknown" to the receiver because otherwise it is not necessary to send it. On the other hand it is possible for the encoder to learn more than the receiver because he knows more, namely his own inputs.

Let us turn to sd-MAC for $\mathcal{X}$ and take a look at the roles of feedback at the $\mathcal{Y}$-encoder. The channel is "noiseless" for the inputs $y^n \in \mathcal{Y}^n$. So the receiver knows the input y^n with probability one and therefore after the transmission the list size on the "component" of the message from the second encoder is automatically one even when the feedback is absent. Thus reducing the list size seems to be not necessary for the second, the $\mathcal{Y}$-encoder. On the other hand the $\mathcal{Y}$-encoder does not know more than the receiver even if he knows the output. So the above second role seems not to be well played by the feedback at $\mathcal{Y}$. We expect that the feedback at $\mathcal{Y}$ is not necessary if the feedback is present at $\mathcal{X}$. So we define the code with incomplete feedback (for maximal probability of error), the (m,f-)-code, as a system $\{f_i^n, v_j, \mathcal{D}_{i,j} : i = 1, \ldots, M_1 \text{ and } j = 1, \ldots, M_2\}$. Here we replace the encoding functions in (2) by the codewords, the sequences v_j's in $\mathcal{Y}^n$. Since a code for a d-MAC with maximal probability of error λ, for any $\lambda < 1$, is a zero-error code for this channel, the following theorem only concerns the zero-error codes. It confirms our intuition.

Theorem 3.1 *For given n, M_1, M_2, there exists an $(n, M_1, M_2, 0)$ (m, f-)-code for an sd-MAC if there exists an $(n, M_1, M_2, 0)$ (m, f)-code for the same channel.*

Proof: We employ induction on the length n of codes. For $n = 1$ the statement is trivially true since in this case the feedback has no room to play its role. Assume that the statement is true for n-1 and we are given an $(n, M_1, M_2, 0)$ (m, f)-code for an sd-MAC W, say $\{f_i^n, g_j^n, \mathcal{D}_{i,j}\}$, where f_i^n's and g_j^n's have the forms in (1) and (2), respectively. Denote by $\mathcal{M}_k = \{1, 2, \ldots, M_k\}$, $k = 1, 2$, the sets of the messages for the two encoders. Let $\mathcal{M}_1(x) = \{i \in \mathcal{M}_1 : f_i^{(1)} = x\}$ for $x \in \mathcal{X}$, $\mathcal{M}_2(y) = \{j \in \mathcal{M}_2 : g_j^{(1)} = y\}$ for $y \in \mathcal{Y}$, and $\mathcal{M}_1(A) = \bigcup_{x \in A} \mathcal{M}_1(x)$ for $A \subset \mathcal{X}$, where $f_i^{(1)}$ is the first component of f_i^n in (1) and $g_j^{(1)}$ is the first component of g_j^n in (2).

By the definition of an sd-MAC there are functions $\phi : \mathcal{X} \times \mathcal{Y} \longrightarrow \mathcal{Z}$ and $\psi : \mathcal{Z} \longrightarrow \mathcal{Y}$ such that $W(z|x, y) = 1$ if $z = \phi(x, y)$ and $\psi(z) = y$ if there is an $x \in \mathcal{X}$ with $z = \phi(x, y)$. Denote by $\mathcal{Z}(y) = \psi^{-1}(y) = \{z \in \mathcal{Z} : \psi(z) = y\}$ for all $y \in \mathcal{Y}$ and $\mathcal{X}_y(z) = \{x \in \mathcal{X} : \phi(x, y) = z\}$ for $y \in \mathcal{Y}$ and $z \in \mathcal{Z}(y)$. Then $\{\mathcal{Z}(y) : y \in \mathcal{Y}\}$ is a partition of $\mathcal{Z}$ and for a given $y \in \mathcal{Y}$, $\{\mathcal{X}_y(z) : z \in \mathcal{Z}(y)\}$ is a partition of $\mathcal{X}$.

For given $z \in \mathcal{Z},\ y = \psi(z),\ i \in \mathcal{M}_1(\mathcal{X}_y(z)),\ j \in \mathcal{M}_2(y)$, we define $f^{n-1}_{(y,z),i}$ as

$$f^{n-1}_{(y,z),i}(z_2, \dots, z_n) = (f_i^{(2)}(z), f_i^{(3)}(z, z_2), \dots, f_i^{(n)}(z, z_2, \dots, z_n)), \tag{5}$$

and $g^{n-1}_{(y,z),j}$ as

$$g^{n-1}_{(y,z),j}(z_2, \dots, z_n) = (g_j^{(2)}(z), g_j^{(3)}(z, z_2), \dots, g_j^{(n)}(z, z_2, \dots, z_n)), \tag{6}$$

for all $(z_2, \dots, z_n) \in \mathcal{Z}^{n-1}$, and

$$\mathcal{D}_{i,j}(y, z) = \{(z_2, \dots, z_n) : (z, z_2, \dots, z_n) \in \mathcal{D}_{i,j}\}. \tag{7}$$

Namely, for all $z \in \mathcal{Z}$, $i \in \mathcal{M}_1(\mathcal{X}_{\psi(z)}(z))$, and $j \in \mathcal{M}_2(\psi(z))$, the communicators, the two encoders and the receiver, cooperate to simulate the encoding-decoding procedure of code $\{f_i^n, g_j^n, \mathcal{D}_{i,j} : i = 1, \dots, M_1 \text{ and } j = 1, \dots, M_2\}$ under the assumption that z has been received by them. Thus $\{f^{n-1}_{(y,z),i}, g^{n-1}_{(y,z),j}, \mathcal{D}_{i,j}(y, z) : i \in \mathcal{M}_1(\mathcal{X}_y(z)),\ j \in \mathcal{M}_2(y)\}$ for $y = \psi(z)$ is an (error-free) (m, f)-code for the sd-MAC W since $\{f_i^n, g_j^n, \mathcal{D}_{i,j} : i = 1, \dots, M_1 \text{ and } j = 1, \dots, M_2\}$ is an (error-free) (m, f)-code for the same channel. Moreover for $y \in \mathcal{Y}$ let $z(y)$ be the letter in $\mathcal{Z}(y)$ achieving $\max_{z \in \mathcal{Z}(y)} |\mathcal{M}_1(\mathcal{X}_y(z))|$ and α_z be any injection from $\mathcal{M}_1(\mathcal{X}_y(z)), z \in \mathcal{Z}$, to $\mathcal{M}_1(\mathcal{X}_y(z(y))$ if $z \neq z(y)$. α_z is the identity mapping if $z = z(y)$. α_z is well defined because $y = \psi(z)$ and therefore $z(y)(= z(\psi(z))$ are uniquely determined by z. Denote by $\alpha_z^{-1}(i)$ the unique i' with $\alpha(i') = i$ if it exists. Since we have an (m, f)-code of length $n-1$ with the sets $\mathcal{M}_1(\mathcal{X}_\dagger(\ddagger(\dagger)))$, $\mathcal{M}_\in(\dagger)$ of messages for all $y \in \mathcal{Y}$, by the induction hypothesis there is an (m, f-)-code of length $n-1$ with the (same) sets $\mathcal{M}_1(\mathcal{X}_\dagger(\ddagger(\dagger)))$, $\mathcal{M}_\in(\dagger)$ of messages for each $y \in \mathcal{Y}$, say $\{f^{*n-1}_{y,i}, u^*_{y,j}, \mathcal{D}^*_{i,j}(y) : i \in \mathcal{M}_1(\mathcal{X}_\dagger(\ddagger(\dagger)))$ and $| \in \mathcal{M}_\in(\dagger)\}$.

Next to complete our proof we define an (m f-)-code of length n with the sets $\mathcal{M}_1$, $\mathcal{M}_2$ of messages based on the above (m, f-)-codes of length $n-1$. For an $i \in \mathcal{M}_1$ we let $f_i^{*(1)} = x$ if $i \in \mathcal{M}_1(x)$ (or in other words $f_i^{(1)} = x$). For $t \geq 2$ and $z^{t-1} = (z_1, \dots, z_{t-1}) \in \mathcal{Z}^{t-1}$ we let

$$f_i^{*(t)}(z^{t-1}) = \begin{cases} f^{*(t-1)}_{\psi(z_1),\alpha_{z_1}(i)}(z_2, \dots, z_{t-1}) & \text{if } i \in \mathcal{M}_1(\mathcal{X}_{\psi(z_1)}(z_1)) \\ \text{any fixed letter } x \in \mathcal{X} & \text{else,} \end{cases}$$

where $f^{*(t-1)}_{y,i'}$ is the $(t-1)$th components of $f^{*n-1}_{y,i'}$ and

$$f_i^{*n}(z^n) = (f_i^{*(1)}, f_i^{*(2)}(z_1), \dots, f_i^{*(n)}(z^{n-1})).$$

For $j \in \mathcal{M}_2$ we let $u_j^*(1) = y$ and $u_j^* = (u_j^*(1), u_{y,j}^*)$ if $j \in \mathcal{M}_2(y)$. Moreover for $i \in \mathcal{M}_1(x) \subset \mathcal{M}_1, j \in \mathcal{M}_2(y) \subset \mathcal{M}_2$, we define $\mathcal{D}_{i,j}^* = \{z^n : z_1 = \phi(x,y), (z_2, \ldots, z_n) \in \mathcal{D}^*_{\alpha_{z_1}(i),j}(\psi(z_1))\}$. That is, the communicators simulate the encoding-decoding procedure of $\{\{f_{y,i}^{*n-1}, u_{y,j}^*, \mathcal{D}_{i,j}^*(y) : i \in \mathcal{M}_1(z(y))$ and $j \in \mathcal{M}_2(y)\}, y \in \mathcal{Y}\}$ as follows.

Let us assume that the first encoder wants to send $i \in \mathcal{M}_1$ and the second encoder wants to send $j \in \mathcal{M}_2$. Abbreviate the code $\{f_{y,i}^{*,n-1}, u_{y,j}^*, \mathcal{D}_{i,j}^*(y) : i \in \mathcal{M}_1(\mathcal{X}_1(z(y)))$ and $j \in \mathcal{M}_2(y)\}$ as $C(y)$.

- The first encoder first sends $x_1 = x$ if $i \in \mathcal{M}_1(x)$. Having received z_1, the first encoder figures out the first symbols $y = y_1$ by calculating $\Psi(z_1) = y$. Then he simulates the encoding procedure of $C(y)$ under-the assumption that he wants to send $\alpha_{z_1}(i)$.
- If $j \in \mathcal{M}_2(y)$, the second encoder first sends $y_1 = y$ and then simulates the encoding procedure of $C(y)$ under the assumption that he wants to send the same j.
- After receiving the first symbol z_1 of the output, the receiver obtains $y_1 = y$ via $y = \psi(z_1)$ and then simulates the decoding procedure of $C(y)$ to obtain an $i' = \alpha_{z_1}\ i$ and j. Finally he recovers $i = \alpha_{z_1}^{-1}(i')$.

Since both, the first encoder and the receiver, are able to learn y_1 (with probability one), the communicators work in the same $C(y)$ after z_1 is output. Thus the code which, we construct, is an error-free (m f-)-code.

4 An outer bound to the capacity region $\mathcal{R}_{m,f}$

In this section we present an outer bound to the capacity region $\mathcal{R}_{m,f}$ for the (m. f)-codes. In the next section we shall show that it is also the inner bound to the capacity region $\mathcal{R}_{m,l}$. Thus by (4) we have that it actually is the capacity region for both codes. Intuitively the results are not as obvious as the result in the last section. In fact they are surprising.

Let us fix an sd-MAC W with input alphabets $\mathcal{X}$ and $\mathcal{Y}$ and output alphabet $\mathcal{Z}$. Let $\mathcal{P}(\mathcal{X})$ and $\mathcal{P}(\mathcal{Y})$ be the sets of the probability distributions over $\mathcal{X}$ and $\mathcal{Y}$, respectively. For a triple (X, Y, Z) of random variables, we denote its joint distribution by P_{XYZ} and analogously their marginal and conditional distributions by $P_X, P_Y, P_{Y|X}, P_{X|YZ}$ and so on. For $P \in \mathcal{P}(\mathcal{X})$ and $Q \in \mathcal{P}(\mathcal{Y})$, we define $\mathcal{Q}(P,Q)$ as the set of triples (X, Y, Z) of random variables satisfying the following conditions: For all $x \in \mathcal{X}, y \in \mathcal{Y}$, and $z \in \mathcal{Z}$,

$$P_{XYZ}(x,y,z) = P_{XY}(x,y)W(z|x,y), \tag{8}$$

$$P_X(x) = \sum_{y\in\mathcal{Y}} P_{XY}(x,y) = P(x), \text{ and } P_Y(y) = \sum_{x\in\mathcal{X}} P_{XY}(x,y) = Q(y). \tag{9}$$

In other words (X, Y) is the pair of input random variables with marginal distributions P and Q and Z is the output random variable of the sd-MAC W when (X, Y) is input. With this notation, we define a subset $\mathcal{R}(P, Q)$ of the real plane for $P \in \mathcal{P}(\mathcal{X}), Q \in \mathcal{P}(\mathcal{Y})$ as

$$\mathcal{R}(P,Q) = \{(R_1, R_2) : 0 \leq R_1 \leq H(P) - \max_{(X,Y,Z)\in\mathcal{Q}(P,Q)} H(X|YZ) \text{ and } 0 \leq R_2 \leq H(Q)\}. \tag{10}$$

Notice that for an sd-MAC we always have

$$H(X|YZ) = H(X|Z) \tag{11}$$

in (10), since by the definition of an sd-MAC Y is a function of Z with probability one. Finally let

$$\mathcal{R}^* = conv\left(\bigcup_{P\in\mathcal{P}(\mathcal{X})\ Q\in\mathcal{P}(\mathcal{Y})} \mathcal{R}(P,Q)\right), \tag{12}$$

where $conv(A)$ is the closed convex hull of the set A. Recalling that for our sd-MAC a code with maximal probability of error $\lambda < 1$ is an error-free code, it is sufficient to bound the capacity region for $\lambda = 0$.

Theorem 4.1 *For an $(n, M_1, M_2, 0)$ (m, f)-code for the sd-MAC*

$$(\frac{1}{n}\log M_1, \frac{1}{n}\log M_2) \in \mathcal{R}^*. \tag{13}$$

Therefore

$$\mathcal{R}_{m,f} \subset \mathcal{R}^*. \tag{14}$$

Proof: We prove the theorem by induction on the length n of the code. For $n = 1$ the statement trivially holds and there is actually no room in the case that the code length is one for the feedback to play its role.

Assume that the statement is true for $n - 1$ and we are given an sd-MAC W and an $(n, M_1, M_2, 0)$ (m, f)-code $\{f_i^n, g_j^n, \mathcal{D}_{i,j} : i = 1, \ldots, M_1, \text{ and } j = 1, \ldots, M_2\}$ for it.

Let us use the same notation as in the proof of Theorem 3.1. That is, denote by $\mathcal{M}_k = \{1, \ldots, M_k\}$ for $k = 1, 2$, $\mathcal{M}_1(x) = \{i : f_i^{(1)} = x\}$ for $x \in \mathcal{X}$, $\mathcal{M}_2(y) = \{j : g_j^{(1)} = y\}$ for $y \in \mathcal{Y}$, and $\mathcal{M}_1(A) = \bigcup_{x \in A} \mathcal{M}_1(x)$ for $A \subset \mathcal{X}$. Moreover recalling from the proof of Theorem 3.1 that for the sd-MAC the function ψ partitions $\mathcal{Z}$ into the subsets $\mathcal{Z}(y) = \{z \in \mathcal{Z} : \psi(z) = y\}$, $y \in \mathcal{Y}$, and the function ϕ partitions $\mathcal{X}$ into subsets $\mathcal{X}_y(z) = \{x \in \mathcal{X} : \phi(x, y) = z\}$, $z \in \mathcal{Z}(y)$, for all $y \in \mathcal{Y}$. Let $r_1(z) = \frac{1}{n-1} \log |\mathcal{M}_1(\mathcal{X}_y(z))|$ for $y = \psi(z)$, if $\mathcal{M}_1(\mathcal{X}_y(z)) \neq \emptyset$, and let $r_2(y) = \frac{1}{n-1} \log |\mathcal{M}_2(y)|$, if $\mathcal{M}_2(y) \neq \emptyset$. Since the code $\{f^{n-1}_{(y,z),i}, g^{n-1}_{(y,z),j}, \mathcal{D}_{i,j}(y, z) : i \in \mathcal{M}_1(\mathcal{X}_y(z)),\ \text{and}\ j \in \mathcal{M}_2(y)\}$ for $y = \psi(z)$ defined by (5), (6), and (7) is an $(n-1), |\mathcal{M}_1(\mathcal{X}_y(z))|, |\mathcal{M}_2(y)|, 0)$ (m, f)-code, by the induction hypothesis, we have that, when $\mathcal{M}_1(\mathcal{X}_y(z))$ and $\mathcal{M}_2(y)$ are not empty,

$$(r_1(z), r_2(y)) \in \mathcal{R}^* \text{ for all } y = \psi(z). \tag{15}$$

Let for $x \in \mathcal{X}$

$$P(x) = \frac{|\mathcal{M}_1(x)|}{M_1} \tag{16}$$

and for $y \in \mathcal{Y}$

$$Q(y) = \frac{|\mathcal{M}_2(y)|}{M_2}. \tag{17}$$

Then obviously P is a probability distribution on $\mathcal{X}$ and Q is a probability distribution on $\mathcal{Y}$. Let $R_1 = \frac{1}{n} \log M_1$ and $R_2 = \frac{1}{n} \log M_2$. It follows from the definition of $r_2(y)$ and (17) that

$$R_2 = \frac{1}{n} \log M_2 = \frac{1}{n} [\log |\mathcal{M}_2(y)| - \log Q(y)] = \frac{n-1}{n} r_2(y) - \frac{1}{n} \log Q(y), \tag{18}$$

if $Q(y) > 0$. Similarly, by the definition of $r_1(z)$ with $y = \psi(z)$

$$\begin{aligned} R_1 &= \frac{1}{n} \log M_1 \\ &= \frac{1}{n} [\log |\mathcal{M}_1(\mathcal{X}_y(z))| - \log \frac{|\mathcal{M}_1(\mathcal{X}_y(z))|}{M_1} \\ &= \frac{n-1}{n} r_1(z) - \frac{1}{n} \log \frac{|\mathcal{M}_1(\mathcal{X}_y(z))|}{M_1}, \end{aligned} \tag{19}$$

if $\mathcal{M}_1(\mathcal{X}_y(z)) \neq \emptyset$. Moreover, since by (16) and the definition of $\mathcal{M}_1(A)$

$$\frac{|\mathcal{M}_1(\mathcal{X}_y(z))|}{M_1} = \sum_{x \in \mathcal{X}_y(z)} \frac{\mathcal{M}_1(x)}{M_1} = \sum_{x \in \mathcal{X}_y(z)} P(x) \tag{20}$$

(19) can be re-written as

$$R_1 = \frac{n-1}{n} r_1(z) - \frac{1}{n} \log \sum_{x \in \mathcal{X}_y(z)} P(x) \text{ for all } z \in \mathcal{Z} \text{ and } y = \psi(z) \ , \tag{21}$$

if $\mathcal{M}_1(\mathcal{X}_y(z)) \neq \emptyset$ or equivalently $\sum_{x \in \mathcal{X}_y(z)} P(x) > 0$.

We shall show that there is an $R_1' \geq R_1$ such that (R_1', R_2) can be expressed as a linear combination of the points in $\mathcal{R}^*$ and therefore by the convexity of $\mathcal{R}^*, (R_1', R_2) \in \mathcal{R}^*$. This implies $(R_1, R_2) \in \mathcal{R}^*$, because $\mathcal{R}^*$ is also closed under projections, i.e., (13). To this end, we need the positivity of conditional informational divergence (see for example, [8]). Let $\overline{\mathcal{X}}$ and $\overline{\mathcal{Y}}$, be finite set, $\overline{P}$ be a probability distribution on $\overline{\mathcal{X}}$, and let $\overline{W}_1$ and $\overline{W}_2$ be two stochastic matrices from $\overline{\mathcal{X}}$ to $\overline{\mathcal{Y}}$. Then the conditional informational divergence is

$$D(\overline{W}_1 \parallel \overline{W}_2 | \overline{P}) = \sum_{\overline{x} \in \overline{\mathcal{X}}} \overline{P}(\overline{x}) \sum_{\overline{y} \in \overline{\mathcal{Y}}} \overline{W}_1(\overline{y}|\overline{x}) \log \frac{\overline{W}_1(\overline{y}|\overline{x})}{\overline{W}_2(\overline{y}|\overline{x})}$$

and

$$D(\overline{W}_1 \parallel \overline{W}_2 | \overline{P}) \geq 0. \tag{22}$$

Let the triple $(X, Y, Z) \in \mathcal{Q}(P, Q)$ of random variables achieve $\max_{(X,Y,Z) \in \mathcal{Q}(P,Q)} H(X|YZ)$. Then for all $x \in \mathcal{X}, \ y \in \mathcal{Y}$, and $z \in \mathcal{Z}$ (8) holds,

$$P_X = P \text{ and } P_Y = Q, \tag{23}$$

$H(X|YZ) = \max_{(X,Y,Z) \in \mathcal{Q}(P,Q)} H(X|YZ)$, and by the definition of $\mathcal{R}(P, Q)$ in (10),

$$(H(P) - H(X|YZ), \ H(Q))) \in \mathcal{R}(P, Q) \subset \mathcal{R}^* \tag{24}$$

Since by the definitions of the function ψ and the set $\mathcal{X}_y(z)$, $W(z|x, y) > 0$ only if $y = \psi(z)$ and $x \in \mathcal{X}_y(z)$, for $y \in \mathcal{Y}$ and $z \in \mathcal{Z}, \ P_{YZ}(y, z) = \sum_{x \in \mathcal{X}} P_{XY}(x, y) W(z|x, y) > 0$ yields $y = \psi(z)$.

Moreover, by the same reason, for $y = \psi(z)$, $P_{YZ}(y, z) = \sum_{x \in \mathcal{X}_y(z)} P_{XY}(x, y) W(z|x, y) \leq \sum_{x \in \mathcal{X}_y(z)} P_{XY}(x, y) \leq \sum_{x \in \mathcal{X}_y(z)} P_X(x) = \sum_{x \in \mathcal{X}_y(z)} P(x)$. Thus,

$$P_{YZ}(y, z) > 0 \text{ implies } y = \psi(z) \text{ and } \sum_{x \in \mathcal{X}_y(z)} P(x) > 0. \tag{25}$$

Let

$$V(x|y,z)=\begin{cases}\frac{P(x)}{\sum_{x'\in\mathcal{X}_y(z)}P(x')} & \text{if } P_{YZ}(y,z)>0\ x\in\mathcal{X}_y(z)\\ 0 & \text{if } P_{YZ}(y,z)>0 \text{ and } x\notin\mathcal{X}_y(z)\\ \overline{V}(x) \text{ for any fixed } \overline{V}\in\mathcal{P}(\mathcal{X}) & \text{else.}\end{cases} \tag{26}$$

Notice that (25) guarantees that the stochastic matrix $V:\mathcal{Y}\times\mathcal{Z}\longrightarrow\mathcal{X}$ is well defined. By combining (21) and (26), we obtain that

$$R_1=\frac{n-1}{n}r_1(z)+\frac{1}{n}[\log V(x|y,z)-\log P(x)] \text{ for all } (y,z) \text{ with } P_{YZ}(y,z)>0 \text{ and } x\in\mathcal{X}_y(z) \tag{27}$$

Notice that by (8) and the definition of $\mathcal{X}_y(z)$ $P_{XYZ}(x,y,z)>0$ only if $x\in\mathcal{X}_y(z)$. We multiply both sides of (27) by $P_{XYZ}(x,y,z)$ for $(x,y,z)\in\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$ with $P_{XYZ}(x,y,z)>0$, sum up the resulting formulae, and obtain

$$\begin{aligned}
R_1 &= \sum_{(y,z)\in\mathcal{Y}\times\mathcal{Z}}\frac{n-1}{n}P_{YZ}(y,z)r_1(z)\\
&+ \frac{1}{n}[\sum_{(y,z)\in\mathcal{Y}\times\mathcal{Z}}P_{YZ}(y,z)\sum_{x\in\mathcal{X}}P_{X|YZ}(x|y,z)\log V(x|y,z)-\sum_{x\in\mathcal{X}}P_X(x)\log P(x)]\\
&= \sum_{(y,z)\in\mathcal{Y}\times\mathcal{Z}}\frac{n-1}{n}P_{YZ}(y,z)r_1(z)\\
&+ \frac{1}{n}[\sum_{(y,z)\in\mathcal{Y}\times\mathcal{Z}}P_{YZ}(y,z)\sum_{x\in\mathcal{X}}P_{X|YZ}(x|y,z)\log V(x|y,z)+H(P)]\\
&= \sum_{(y,z):\ y=\psi(z)}\frac{n-1}{n}P_{YZ}(y,z)r_1(z)\\
&+ \frac{1}{n}[\sum_{(y,z)\in\mathcal{Y}\times\mathcal{Z}}P_{YZ}(y,z)\sum_{x\in\mathcal{X}}P_{X|YZ}(x|y,z)\log V(x|y,z)+H(P)]\\
&= \sum_{(y,z):\ y=\psi(z)}\frac{n-1}{n}P_{YZ}(y,z)r_1(z)\\
&- \frac{1}{n}\sum_{(y,z)\in\mathcal{Y}\times\mathcal{Z}}P_{YZ}(y,z)\sum_{x\in\mathcal{X}}P_{X|YZ}(x|y,z)\log\frac{P_{X|YZ}(x|y,z)}{V(x|y,z)}\\
&+ \frac{1}{n}\sum_{(y,z)\in\mathcal{Y}\times\mathcal{Z}}P_{YZ}(y,z)\sum_{x\in\mathcal{X}}P_{X|YZ}(x|y,z)\log P_{X|YZ}(x|y,z)+\frac{1}{n}H(P)\\
&= \sum_{(y,z):\ y=\psi(z)}\frac{n-1}{n}P_{YZ}(y,z)r_1(z)\\
&- \frac{1}{n}D(P_{X|YZ}\parallel V|P_{YZ})-\frac{1}{n}H(X|YZ)+\frac{1}{n}H(P),
\end{aligned} \tag{28}$$

where the second equality follows from the first formula in (23) and the third equality follows from (25). Let

$$R_1' = \sum_{(y,z):\ y=\psi(z)} \frac{n-1}{n} P_{YZ}(y,z) r_1(z) + \frac{1}{n}[H(P) - H(X|YZ)]. \tag{29}$$

Then it follows from (22), (28), and (29) that

$$R_1 \leq R_1'. \tag{30}$$

Next for $(y,z) \in \mathcal{Y} \times \mathcal{Z}$ with $P_{YZ}(y,z) > 0$, we multiply both sides of (18), then sum up the resulting formulae, and finally by the second formula in (23) and (25), obtain that

$$\begin{aligned} R_2 &= \sum_{(y,z)\in\mathcal{Y}\times\mathcal{Z}} \frac{n-1}{n} P_{ZY}(y,z) r_2(y) - \frac{1}{n} \sum_{y\in\mathcal{Y}} P_Y \log Q(y) \\ &= \sum_{(y,z):\ y=\psi(z)} \frac{n-1}{n} P_{ZY}(y,z) r_2(y) + \frac{1}{n} H(Q). \end{aligned} \tag{31}$$

Finally we combine (29) and (31) and obtain that

$$(R_1',\ R_2) = \sum_{(y,z):\ y=\psi(z)} \frac{n-1}{n} P_{YZ}(y,z)(r_1(z),\ r_2(y)) + \frac{1}{n}(H(P) - H(X|YZ),\ H(Q)) \tag{32}$$

Since by (25) $\sum_{(y,z):y=\psi(z)} \frac{n-1}{n} P_{YZ}(y,z) + \frac{1}{n} = \frac{n-1}{n} + \frac{1}{n} = 1$, (15), (24), (32), and the convexity of $\mathcal{R}^*$ yield

$$(R_1',\ R_2) \in \mathcal{R}^*,$$

which with (30) completes our proof i.e., $(R_1, R_2) \in \mathcal{R}^*$ or (13).

5 The outer bound is an inner bound to the capacity region $\mathcal{R}_{m,l}$ of the (m, l)-codes

In this section let us turn to the (m, l)-codes, the list codes with maximal probability of error, for sd-MAC and show that the region $\mathcal{R}^*$ in (12) is also an inner bound for their capacity region $\mathcal{R}_{m,l}$. Consequently $\mathcal{R}^*$ actually equals both, $\mathcal{R}_{m,f}$ and $\mathcal{R}_{m,l}$. The main result in this section is the following coding theorem.

Theorem 5.1 *Let $\mathcal{R}^*$ be defined by (8), (9), (10) and (12). For an sd-MAC, any $(R_1, R_2) \in \mathcal{R}^*$, and any $\delta,\ \epsilon > 0$, there exists an $(n, M_1, M_2, 0, L)$ (m, l)-code with $\frac{1}{n}\log M_k > R_k - \epsilon,\ k = 1, 2$ and $\frac{1}{n}\log L < \delta$ for all sufficiently large n. Consequently,*

$$\mathcal{R}^* \subset \mathcal{R}_{m,l}. \tag{33}$$

Proof : Let us fix an sd-MAC W and an integer n, denote by $\mathcal{P}_n(\mathcal{X})$ and $\mathcal{P}_n(\mathcal{Y})$, the n-types over $\mathcal{X}$ and $\mathcal{Y}$ (empirical distributions for n-samples), and by $\mathcal{T}_P^n$, $\mathcal{T}_{XYZ}^n$, and $\mathcal{T}_{X|YZ}^n(y, z)$, the sets of typical sequences with (n-) type P, joint type P_{XYZ}, and conditional type $P_{X|YZ}$, respectively (see for example [2], [8] or [7]). For fixed n, $P \in \mathcal{P}_n(\mathcal{X})$, $Q \in \mathcal{P}_n(\mathcal{Y})$, and $\mathcal{Q}(P, Q)$ defined by (8) and (9) let $\mathcal{Q}_n(P, Q) = \mathcal{Q}(P, Q) \cap \{(X, Y, Z) : P_{X,Y,Z} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}) \times \mathcal{P}(\mathcal{Z})$. By time sharing and the fact that the n-types for n taking values in the set of positive integers are dense in the respective sets of distributions, it is sufficient for us to show that there is an $(n, M_1, M_2, 0, L)$ (m, l)-code with $\frac{1}{n}\log M_k > R_k - \epsilon\ k = 1, 2$ and $\frac{1}{n}\log L < \delta$ for all $P \in \mathcal{P}_n(\mathcal{X}),\ Q \in \mathcal{P}_n(\mathcal{Y}),\ \delta,\ \epsilon > 0$ and sufficiently large n.

Suppose that we are given $P \in \mathcal{P}_n(\mathcal{X})$ and $Q \in \mathcal{P}_n(\mathcal{Y})$ for a sufficiently large n, which will be specified later. We choose $M_2 = |\mathcal{T}_Q^n|$ and label the sequences in $\mathcal{T}_Q^n$ by $1, 2, \ldots, M_2$. We choose the jth sequence v_j as the codeword for the jth message sent by the second encoder, (or the $\mathcal{Y}$-encoder), for $j = 1, 2, \ldots, M_2$. Thus for arbitrarily small $\epsilon > 0$ and sufficiently large n we have

$$\frac{1}{n}\log M_2 > H(Q) - \epsilon. \tag{34}$$

Next we choose the size M_1 of the codebook for the first encoder, (or the $\mathcal{X}$-encoder) such that

$$H(P) - \max_{(X,Y,Z)\in\mathcal{Q}(P,Q)} H(X|YZ) - \epsilon < \frac{1}{n}\log M_1 < H(P) - \max_{(X,Y,Z)\in\mathcal{Q}(P,Q)} H(X|YZ) - \frac{1}{2}\epsilon. \tag{35}$$

Notice that

$$\max_{(X,Y,Z)\in\mathcal{Q}_n(P,Q)} H(X|YZ) \leq \max_{(X,Y,Z)\in\mathcal{Q}(P,Q)} H(X|YZ) \tag{36}$$

since $\mathcal{Q}_n(P, Q) \subset \mathcal{Q}(P, Q)$. Now we choose the codewords for the first encoder $U_1, U_2, \ldots, U_{M_1}$ randomly independently with uniform distribution over $\mathcal{T}_P^n$. We take an output sequence $z^n \in \mathcal{Z}^n$ in the decoding set $\mathcal{D}_{i,j}$ iff z^n is received with a positive probability by the receiver when the first encoder's ith codeword and the second encoder's jth codewords are sent. Thus the error probability is always zero. So by (34) and (35) we only need to show that with a positive probability one can obtain a deterministic code from the above random code such that for all output sequences the list sizes are smaller than $2^{n\delta}$.

Since for all j, $v_j \in \mathcal{T}_Q^n$, and the channel is semi-noisy, an output sequence z^n is received with positive probability only if a $v_j = y^n = (y_1, \ldots, y_n) = (\psi(z_1), \ldots, \psi(z_n)) = \psi(z^n)$ (say) $\in \mathcal{T}_Q^n$ is sent by the second encoder. Moreover for such z^n and $y^n = \psi(z^n)$, z^n is received with positive probability only if a codeword x^n such that the joint type of (x^n, y^n, z^n) is in $\mathcal{Q}_n(P, Q)$ is sent by the first encoder. For an output sequence $z^n \in \mathcal{Z}^n$ with $\psi(z^n) = y^n = v_j \in \mathcal{T}_Q^n$ we let the random variable $K(z^n)$ be the list size of z^n and define the random variables

$$K(z^n, i) = \begin{cases} 1 & \text{if } W(z^n|U_i, v_j) > 0 \\ 0 & \text{else,} \end{cases} \tag{37}$$

for $i = 1, 2, \ldots, M_1$. Then $K(z^n) = \sum_{i=1}^{M_1} K(z^n, i)$ and for all i the probability

$$\begin{aligned} Pr(K(z^n, i) = 1) &= \sum_{(X,Y,Z) \in \mathcal{Q}_n(P,Q)} \frac{|\mathcal{T}_{X|YZ}^n(\psi(z^n), z^n)|}{|\mathcal{T}_P^n|} \\ &\leq (n+1)^{|\mathcal{X}||\mathcal{Y}||\mathcal{Z}|} \max_{(X,Y,Z) \in \mathcal{Q}_n(P,Q)} \frac{|\mathcal{T}_{X|YZ}^n(\psi(z^n), z^n)|}{|\mathcal{T}_P^n|} \\ &\leq 2^{-n[H(P) - \max_{(X,Y,Z) \in \mathcal{Q}_n(P,Q)} H(X|YZ) - \frac{1}{4}\epsilon]} \end{aligned} \tag{38}$$

when n is sufficiently large. Thus for all output sequences $z^n \in \mathcal{Z}^n$ we have that

$$\begin{aligned} Pr(K(z^n) > 2^{n\delta}) &= Pr(\sum_{i=1}^{M_1} K(z^n, i) > 2^{n\delta}) \\ &= Pr(e^{\sum_{i=1}^{M_1} K(z^n,i) - 2^{n\delta}} > 1) \\ &\leq e^{-2^{n\delta}} \mathbf{E} e^{\sum_{i=1}^{M_1} K(z^n,i)} \\ &= e^{-2^{n\delta}} \mathbf{E} \prod_{i=1}^{M_1} e^{K(z^n,i)} \\ &= e^{-2^{n\delta}} \prod_{i=1}^{M_1} \mathbf{E} e^{K(z^n,i)} \\ &= e^{-2^{n\delta}} \prod_{i=1}^{M_1} [Pr(K(z^n, i) = 0) + Pr(K(z^n, i) = 1)e] \\ &= e^{-2^{n\delta}} \prod_{i=1}^{M_1} [1 + (e-1) Pr(K(z^n, i) = 1)] \\ &\leq \exp_e[-2^{n\delta} + \sum_{i=1}^{M_1} (e-1) Pr(K(z^n, i) = 1)] \\ &\leq \exp_e\{-2^{n\delta} + eM_1 2^{-n[H(P) - \max_{(X,Y,Z) \in \mathcal{Q}_n(P,Q)} H(X|YZ) - \frac{1}{4}\epsilon]}\} \\ &< \exp_e\{-2^{n\delta} + 2^{-n\frac{1}{4}\epsilon}\} \\ &\leq e^{-2^{n\delta}+1}, \end{aligned} \tag{39}$$

where $\mathbf{E}(\cdot)$ is the expectation of random variables, the fourth equality holds because the U_i's are independent, the second inequality follows from the inequality $1 + t \leq e^t$ for $t \geq 0$, the third inequality follows from (38), and the fourth inequality holds by (35)and (36). Notice that there are totally $|\mathcal{Z}|^n$ output sequences and on the other hand the right hand side of (39) is vanishing double exponentially when n is growing up to infinity. Therefore the probability of the event that there exists a $z^n \in \mathcal{Z}^n$ with list size not smaller than $2^{n\delta}$ is arbitrarily small for sufficiently large n. This completes our proof.

6 Capacity regions for the codes

It immediately follows from (4), Theorem 4.1 and Theorem 5.1 that $\mathcal{R}_{m,l} = \mathcal{R}_{m,f} = \mathcal{R}^*$ for all sd-MAC such that $\mathcal{R}_{m,f}$ has considered a pair of positive rates. What are those capacity regions in the other case? The answer is simple and the proof is straightforward.

Lemma 6.1 *1) There are m-, a-, (m, l)-, (m, f)- (and therefore (m, f-)-) codes with positive rates R_1 for the first encoder iff for some $y \in \mathcal{Y}$ there exist at least two letters, say $x, x' \in \mathcal{X}$ such that $\phi(x, y) \neq \phi\ (x', y)$.*

2) There are m-, a-, (m, l)-, (m, f)- (and therefore (m, f-)-) codes with positive rates R_2 for the second encoder iff $|\mathcal{Y}| \geq 2$.

Combining Theorem 4.1, Theorem 5.1 and Lemma 6.1 , by (4), we have

Theorem 6.2 *For all sd-MAC and $\mathcal{R}^*$ defined by (8), (9), (10), and (12),*

$$\mathcal{R}_{m,l} = \mathcal{R}_{m,f} = \mathcal{R}^*. \tag{40}$$

In the proof of Theorem 3.1 we show that we can always construct an (m, f-)-code from a given (m, f)-code and keep its parameters length and size. We remark here that if we only care for asymptotic results, that is the equalities $\mathcal{R}_{m,f-} = \mathcal{R}_{m,f}$ of rates, then this is an easy consequence of Theorem 6.2, because it is not hard to show that $\mathcal{R}_{m,l} \subset \mathcal{R}_{m,f-}$.

Thus by (3) and Theorem 6.2 we have that for an sd-MAC

$$\mathcal{R}_m \subset \mathcal{R}_{m,l} = \mathcal{R}_{m,f-} = \mathcal{R}_{m,f} = \mathcal{R}^* \subset \mathcal{R}_a. \tag{41}$$

We know very little about the first containment relation in (41). The "Noiseless Binary Switching MAC" ([17]) in Example 2 gives an example of an sd-MAC, for which $\mathcal{R}_m = \mathcal{R}^* = \mathcal{R}_a$. But we do not know whether it is true that $\mathcal{R}_m = \mathcal{R}^*$ for general sd-MAC.

PROBLEM: What is the capacity region $\mathcal{R}_m$ for m-codes (or equivalently for error-free codes) for general sd-MAC? Is it equal to $\mathcal{R}^*$?

On the other hand, the second containment relation in (41) is quite clear. Two extremal examples, for which $\mathcal{R}^*$ is strictly contained by and equal to $\mathcal{R}_a$, respectively, have been given in the Examples 1 and 2 in Section 2. More importantly both regions, $\mathcal{R}^*$ and $\mathcal{R}_a$ in [1], are computable in the sense of single letter characterization. Thus one can at least in principle calculate and therefore compare them. It actually is easy in most cases because of the simplicity of the structure of sd-MAC. To illustrate this, we present the following example.

We say that an sd-MAC is *regular* if for all $y \in \mathcal{Y},\ z, z' \in \mathcal{Z}(y),\ |\mathcal{X}_y(z)| = |\mathcal{X}_y(z')|$, where $\mathcal{Z}(y) = \{z \in \mathcal{Z} : \psi(z) = y\}$ and $\mathcal{X}_y(z) = \{x \in \mathcal{X} : \phi(x, y) = z\}$ for $z \in \mathcal{Z}(y)$ are defined in the proof of Theorem 3.1. In other words, for a regular sd-MAC, for all $y \in \mathcal{Y}$ there are a pair of integers, say $\beta(y)$ and $l(y)$ such that $\beta(y)l(y) = |\mathcal{X}|$, $l(y) = |\mathcal{Z}(y)|$, and for all $z \in \mathcal{Z}(y)$, $|\mathcal{X}_y(z)| = \beta(y)$. $\beta(\cdot)$ and $l(\cdot)$ will be used in the proof in the theorem below. Obviously Vanroose's channel in Example 2 is regular. For a regular sd-MAC, we define

$$\mathcal{R}_r(Q) = \{(R_1, R_2) : 0 \le R_1 \le \sum_{y \in \mathcal{Y}} Q(y) \log l(y),\ 0 \le R_2 \le H(Q)\} \tag{42}$$

for all $Q \in \mathcal{P}(\mathcal{Y})$ and

$$\mathcal{R}_r = conv[\bigcup_{Q \in \mathcal{P}(\mathcal{Y})} \mathcal{R}_r(Q)]. \tag{43}$$

Theorem 6.3 *For a regular sd-MAC,*

$$\mathcal{R}^* = \mathcal{R}_a = \mathcal{R}_r. \tag{44}$$

Proof: Since we know that $\mathcal{R}^* = \mathcal{R}_{m,f} \subset \mathcal{R}_a$, it is sufficient for us to show that

$$\mathcal{R}_a \subset \mathcal{R}_r \subset \mathcal{R}^*.$$

We first show that

$$\mathcal{R}_a \subset \mathcal{R}_r. \tag{45}$$

For $P \in \mathcal{P}(\mathcal{X})$ and $Q \in \mathcal{P}(\mathcal{Y})$, let (X, Y, Z) be the triple of random variables with distribution $P_{XYZ}(x, y, z) = P(x)Q(y)W(z|x, y)$ for all $x \in \mathcal{X}, y \in \mathcal{Y}$ and $z \in \mathcal{Z}$ and

$\mathcal{R}_a(P, Q) = \{(R_1, R_2) : 0 \le R_1 \le I(X \wedge Z|Y),\ 0 \le R_2 \le I(Y \wedge Z|X)$
and $R_1 + R_2 \le I(XY \wedge Z)\}$.

Then the capacity region for a-codes (see [1]) is

$$\mathcal{R}_a = conv \bigcup_{P \in \mathcal{P}(\mathcal{X}), Q \in \mathcal{P}(\mathcal{Y})} \mathcal{R}_a(P, Q).$$

Let $(R_1, R_2) \in \mathcal{R}_a(P, Q)$ and (X, Y, Z) be the triple of random variables with distribution $P_{XYZ}(x, y, z) = P(x)Q(y)W(z|x, y)$.
Since for all $y \in \mathcal{Y}$, $P_{YZ}(y, z) = \sum_{x \in \mathcal{X}} P_{XYZ}(x, y, z) > 0$ only if $z \in \mathcal{Z}(y)$, $H(Z|Y) \leq \sum_{y \in \mathcal{Y}} Q(y) \log |\mathcal{Z}(y)| = \sum_{y \in \mathcal{Y}} Q(y) \log l(y)$. Therefore, we have that $R_1 \leq I(X \wedge Z|Y) \leq H(Z|Y) \leq \sum_{y \in \mathcal{Y}} Q(y) \log l(y)$. Moreover $R_2 \leq I(Y \wedge Z|X) \leq H(Y|X) = H(Y) = H(Q)$. Thus by (42) and (43) $(R_1, R_2) \in \mathcal{R}_r(Q) \subset \mathcal{R}_r$ and consequently (45) holds.

Next to complete the proof we show that

$$\mathcal{R}_r \subset \mathcal{R}^*. \tag{46}$$

For all $Q \in \mathcal{P}(\mathcal{Y})$ we let P be the uniform distribution over $\mathcal{X}$. assume $(X, Y, Z) \in \mathcal{Q}(P, Q)$ achieves $max_{(X,Y,Z) \in \mathcal{Q}(P,Q)} H(X|YZ)$. Then $H(P) = \log |\mathcal{X}|$. Moreover, since for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and $z \in \mathcal{Z}$, $P_{XYZ}(x, y, z) > 0$ only if $y = \psi(z)$ and $x \in \mathcal{X}_y(z)$, we have by (25) that $\max_{(X,Y,Z) \in \mathcal{Q}(P,Q)} H(X|YZ) = H(X|YZ) \leq \sum_{(y,z):y=\psi(z)} P_{YZ}(y, z) \log |\mathcal{X}_y(z)| = \sum_{(y,z):y=\psi(z)} P_{YZ}(y, z) \log \beta(y) = \sum_{y \in \mathcal{Y}} P_Y(y) \log \beta(y) = \sum_{y \in \mathcal{Y}} Q(y) \log \beta(y)$.
Now for arbitrary $Q \in \mathcal{P}(\mathcal{Y})$ and $(R_1, R_2) \in \mathcal{R}_r(Q)$, by (42) we have that $R_1 \leq \sum_{y \in \mathcal{Y}} Q(y) \log l(y) = \sum_{y \in \mathcal{Y}} Q(y) \log \frac{|\mathcal{X}|}{\beta(y)} = \log |\mathcal{X}| - \sum_{y \in \mathcal{Y}} P_Y(y) \log \beta(y) \leq H(P) - \max_{(X,Y,Z) \in \mathcal{Q}(P,Q)} H(X|YZ)$ and $R_2 \leq H(Q)$. Thus by (10) $(R_1, R_2) \in \mathcal{R}(P, Q)$, which with (12) and (43) implies (46).

7 Concluding remarks

Let us denote the sets of MAC's satisfying the conditions (*), (**), and sd-MAC by $\mathcal{K}_1, \mathcal{K}_2$ and $\mathcal{K}_{sd}$ respectively. Then obviously

$$\mathcal{K}_2 \subset \mathcal{K}_1 \text{ and } \mathcal{K}_{sd} \subset \mathcal{K}_1. \tag{47}$$

Moreover

$$\mathcal{K}_2 \setminus \mathcal{K}_{sd} \neq \emptyset \tag{48}$$

since the binary adder channel is contained in $\mathcal{K}_2$ but not in $\mathcal{K}_{sd}$ and

$$\mathcal{K}_{sd} \setminus \mathcal{K}_2 \neq \emptyset, \tag{49}$$

since the channels in the Examples 1 and 2 in the Section 2 are contained by $\mathcal{K}_{sd}$ but not by $\mathcal{K}_2$.

However the zero-error and average-error capacity regions may be very different. We have seen an example in [9] (i.e. the MAC in Example 1 in Section 2) for which the zero-error capacity region even with feedback is strictly smaller than the average-error capacity region even without feedback (and therefore strictly smaller than the average-error capacity region with feedback). In Information Theory to determine the zero-error capacities or capacity regions is often much harder than to determine the average-error capacities or capacity regions.

Whereas for deterministic MAC zero error and maximal error capacity regions are equal, for non-deterministic MAC they may be different. By our knowledge, the capacity regions $\mathcal{R}_{m,f}$ with feedback and maximal probability of errors are known only for sd-MAC and the d-MAC in the class $\mathcal{K}_2$.

Although, as we mentioned the Section 2, $\mathcal{R}_{a,l}$ is trivially equal to $\mathcal{R}_a$, we know very little about $\mathcal{R}_{m,l}$ except for sd-MAC. It would be an interesting new direction to study (m,l)-codes for general MAC and this may be easier than to study m-codes.

The readers, who are familiar with list codes, may notice that our definition of rates of (m, l)-codes is not quite "standard". For a code with length n and list size L carrying M messages for a two terminal channel, its rate is traditionally defined as $\frac{1}{n}\log\frac{M}{L}$. Along this line, a list code for MAC "should" have three rates. That is, let $L_{\mathcal{X}} = \max_{z^n} |\{i : z^n \in \mathcal{D}_{i,j}\}|$, $L_{\mathcal{Y}} = \max_{z^n} |\{j : z^n \in \mathcal{D}_{i,j}\}|$ and L be defined as in the Section 2. Then the three "rates" are defined as $R'_{\mathcal{X}} = \frac{1}{n}\log\frac{M_1}{L_{\mathcal{X}}}, R'_{\mathcal{Y}} = \frac{1}{n}\log\frac{M_1}{L_{\mathcal{Y}}}$, and $R' = \frac{1}{n}\log\frac{M_1 M_2}{L}$. For simplicity, we use our definition in Section 2. This makes no difference in the asymptotic sense, as we ask the exponent of the list size asymptotically to go to zero.

References

[1] R. Ahlswede, Multi-way Communication, Proc. 2nd Int. Symp. Inform. Theory, Thakhkadzor, Armenia, 23-25, 1971.

[2] R. Ahlswede, A constructive proof of the coding theorem for discrete memoryless channels in case of complete feedback, Sixth Prague Conf. on Inf. Th., Stat. Dec. Fct's and Rand. Proc., Sept. 1971, Publ. House Chechosl. Academy of Sc., 1–22, 1973.

[3] R. Ahlswede, Channels with arbitrarily varying probability functions in the presence of noiseless feedback, Z. Wahrsch. Verw. Gebiete, V. 25, 239-252, 1973.

[4] R. Ahlswede, Channel capacities for list codes, J. Appl. Probability, lo, 824–836, 1973.

[5] R. Ahlswede and N. Cai, The AVC with noiseless feedback and maximal error probability: a capacity formula with a trichotomy, Numbers, Information, and Complexity, (Festschrift Rudolf Ahlswede), eds: I. Althöfer, N. Cai, G. Dueck, L. Khachatrian, M. Pinsker, A. Sarkozy, I. Wegener, and Z. Zhang, (Kluwer), 2000.

[6] T. M. Cover and C. S. K. Leung, An achievable rate region for the multiple-access channel with feedback, IEEE Trans. Inform. Theory, IT-27, 291-298, 1981.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, (Wiley), 1991

[8] I. Csizár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, (Academic Press), 1981.

[9] G. Dueck, Maximal error capacity regions are smaller than average error capacity regions for multi-user channels, Problems Control & Inform. Theory, Vol. 7, 11-19, 1978.

[10] G. Dueck, The zero error feedback capacity region of a certain class of multiple-access channels, Problems Control & Inform. Theory, Vol. 14, 89-103, 1985.

[11] P. Elias, List decoding for noisy channels, Ire Wescon Convention Record, 2, 94–104, 1957.

[12] N. T. Gaarder and J. K. Wolf, The capacity region of a multiple access discrete memoryless channel can increase with Feedback, IEEE Trans. on Infrom. Theory, IT-21, 100-102, 1975.

[13] L. H. Ozarow, Coding and capacity for additive white gaussian noise multiple-access channel with feedback, Ph. D. Dissertation, M.I.T. Cambridge, 1979.

[14] C.E. Shannon, A mathematical theory of communication, Bell Syst. Techn. J. 27, 339–425, 623–656, 1948.

[15] C. E. Shannon, The zero-error capacity of a noisy channel, IRE Trans. Infrom. Theory, IT-2, 8-19, 1956.

[16] C.E. Shannon, Two–way communication channels, Proc. 4th Berkeley Symp. Math. Statist. and Prob. Univ. of California Press, Berkeley, Vol. 1, 611–644, 1961.

[17] P. Vanroose, Code construction for the noiseless binary switching multiple-access channel, IEEE Trans. Inform. Theory, IT-34, 1100-1106, 1988.

[18] F. M. J. Willems, The feedback capacity region of a class of discrete memoryless multiple access channels, IEEE Trans. Inform. Theory, IT-28, 93-95,1982.