Lóránt Farkas; Tamás Kói

On capacity regions of discrete asynchronous multiple access channels

Kybernetika, Vol. 50 (2014), No. 6, 1003–1031

Persistent URL: `http://dml.cz/dmlcz/144120`

KYBERNETIKA — VOLUME *50* (2014), NUMBER 6, PAGES 1003–1031

ON CAPACITY REGIONS OF DISCRETE ASYNCHRONOUS MULTIPLE ACCESS CHANNELS

Lóránt Farkas and Tamás Kói

A general formalization is given for asynchronous multiple access channels which admits different assumptions on delays. This general framework allows the analysis of so far unexplored models leading to new interesting capacity regions. The main result is the single letter characterization of the capacity region in case of 3 senders, 2 synchronous with each other and the third not synchronous with them.

Keywords: partly asynchronous, delay, multiple-access, rate splitting, successive decoding

Classification: 94A24

1. INTRODUCTION

Ahlswede [1] and Liao [12] showed that if two senders communicate synchronously over a discrete memoryless multiple access channel (MAC) which is characterized by a stochastic matrix $W(y|x_1, x_2)$, it is possible to communicate with arbitrary small average probability of error if the rate pair is inside the following pentagon:

$$\begin{aligned} 0 \le R_1 &\le I(X_1 \wedge Y|X_2) \\ 0 \le R_2 &\le I(X_2 \wedge Y|X_1) \\ R_1 + R_2 &\le I(X_1, X_2 \wedge Y) \end{aligned} \tag{1}$$

for some independent input random variables X_1, X_2, where $P(Y = y|X_1 = x_1, X_2 = x_2) = W(y|x_1, x_2)$. Moreover, the convex hull of the union of these pentagons can also be achieved, via time sharing, while no rate pair outside this convex hull is achievable.

The discrete memoryless asynchronous multiple access channel (AMAC) arises when the senders cannot synchronize the starting times of their codewords, rather, there is an unknown delay between these starting times. Cover, McEliece and Posner [3] showed that if the delay is bounded by b_n depending on the codeword length n such that $\frac{b_n}{n} \to 0$ then the convex closure is still achievable by a generalized time sharing method.

Poltyrev [15] and Hui and Humblet [11] addressed models with arbitrary delays known (in [15]) or unknown (in [11]) to the receiver. For such models, the capacity region was shown to be the union of the pentagons above although with some gaps in the proofs,

DOI: 10.14736/kyb-2014-6-1003

see Appendix A. Verdú [18] studied asynchronous channels with memory. His model slightly differs from common models: the time runs over a torus rather than from $-\infty$ to ∞. Later, Grant, Rimoldi, Urbanke and Whiting [8] showed that in the informed receiver case the union can be achieved by rate splitting and successive decoding. The gap in the achievability proof of [11] for the uninformed receiver case has been filled in the book of El Gamal and Kim [7].

This paper is an extended version of the ISIT 2011 contribution [5], originating from the authors' effort to derive the AMAC capacity region without gaps in the proof (the result in [7] was unknown to us at the time, as was, apparently, to the reviewers of [5]). More than doing that, in [5] a general formalization for AMACs was introduced leading to the first (somewhat artificial) example that the capacity region could be strictly between the union and its convex closure.

Here, the general model of [5] is developed in more detail. The capacity region depends on the distribution of the delays typically through the support of that distribution. Even for a given delay distribution, several model versions are analyzed in parallel, and shown not to differ substantially. The general (not single letter) converse of [5] is also treated more deeply.

The main result of this paper is a single letter characterization of the capacity region for 3 senders, two synchronized with each other and the third unsynchronized with them. Its converse part is derived from the general converse. To prove the achievability we had to combine the techniques of rate splitting and successive decoding developed by Grant, Rimoldi, Urbanke, Whiting [8] and Rimoldi [13] with time sharing.

2. MODEL OF CODING FOR THE AMAC

In this paper vectors (finite sequences) will be denoted by boldface symbols. Furthermore, $[K]$ denotes the set $\{1, 2, \dots, K\}$.

A K-senders asynchronous discrete memoryless multiple-access channel (K-AMAC) is defined in terms of K finite input alphabets $\mathcal{X}_m, m \in [K]$, a finite output alphabet $\mathcal{Y}$, and a stochastic matrix $W : \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_K \to \mathcal{Y}$ describing the probability distribution of the output given the inputs.

Definition 2.1. A codebook system of block-length n for a K-AMAC W consists of K codebooks $C_1, C_2, \dots, C_K$, where the codebook C_m of the mth sender has 2^{nR_m} codewords of length n whose symbols are from $\mathcal{X}_m$. The rate vector of this codebook system is $\mathbf{R} = (R_1, R_2 \dots, R_K)$.

The system is symbol synchronized but not frame synchronized. The differences between the timing of the receiver and the timings of the senders are represented by a K-tuple of delays as in Definition 2.3.

The senders have two-way infinite sequences of random messages, and assign codewords to their consecutive messages. The codewords go through the channel. The sequences of the senders' codewords and hence also the output symbol sequence are two-way infinite sequences. Fix the location of the 0th output symbol. The message of sender $m \in [K]$ whose codeword affects the 0'th output is denoted by $M_{m,0}$. The time difference between the 0th output and the first output influenced by the message $M_{m,0}$

is referred to as the delay of sender m, see Figure 1. Formally, we use the following definitions, where n denotes the block-length in Definition 2.1.

Definition 2.2. For each integer $j \in \mathbb{Z}$ and for each $m \in [K]$ let $M_{m,j}$ be a uniformly distributed random variable taking values in the set $\{1, 2, \ldots, 2^{nR_m}\}$. All these random variables are mutually independent. The two-way infinite sequence $\{M_{m,j}, j \in \mathbb{Z}\}$ represents the message flow sent by the mth sender. For each integer $j \in \mathbb{Z}$ and for each $m \in [K]$ let $\mathbf{X}_{m,j}$ denote the $M_{m,j}$th codeword in the codebook of sender m. Let $X_{m,nj+i}$ be the ith symbol of $\mathbf{X}_{m,j}$ where $i \in \{0, 1, \ldots, n-1\}$.

Note, though the codebooks are fixed, the sent codewords are random because the underlying random messages.

Definition 2.3. Let

$$\mathbf{D}(n) = (D_1(n), D_2(n), \ldots, D_K(n)) \tag{2}$$

be a K-tuple of random variables, not necessarily independent of each other but independent of the message random variables $M_{m,j}$ (and hence also of $\mathbf{X}_{m,j}$), taking values in the set $\{0, 1, \ldots, n-1\}$. $D_m(n)$ will represent the delay of sender m relative to the receiver's timing. The joint distribution of delays is known to the senders and the receiver. The realizations of the random variables $D_1(n), D_2(n), \ldots, D_K(n)$ are not known to the senders and, depending on the model, may be known or unknown to the receiver. The sequence $\mathbf{D} = \{\mathbf{D}(1), \mathbf{D}(2), \ldots, \mathbf{D}(n), \ldots\}$ will be called the delay system. When dealing with a fixed block-length n, we also write $\mathbf{D}$ instead of $\mathbf{D}(n)$.

Remark 2.4. Our definition allows arbitrary distributions for the delays for each block-length n. Clearly, in practical models these distributions can not be arbitrary, but have to satisfy consistency conditions. We have chosen this general model since we think that any practical model can be described this way.

Example 2.5. Let $D_m(n)$, $m \in [K]$, be mutually independent random variables with uniform distribution on $\{0,\ 1, \ldots,\ n-1\}$. Following [11] it is called the totally asynchronous case in the paper.

Example 2.6. Let $K=2$, let $D_1(n), D_2(n)$ be independent random variables uniformly distributed on the even numbers in $\{0, 1, \ldots, n-1\}$. It is called the even delays case in the paper.

Example 2.7. Let $K = 3$, let $D_1(n) = D_2(n)$ be random variable uniformly distributed on $\{0, 1, \ldots, n-1\}$ and let $D_3(n)$ be a random variable independent of $D_1(n)$ and uniformly distributed on $\{0, 1, \ldots, n-1\}$. It is called the partly asynchronous three senders case in the paper.

For fixed n, the output sequence is defined as follows:

Definition 2.8. Let Y_{nj+i} be the output random variable of the channel with transition matrix W when the inputs are $X_{1,nj+i+D_1(n)}$, $X_{2,nj+i+D_2(n)}$, $\ldots$, $X_{K,nj+i+D_K(n)}$ where $i \in \{0, 1, \ldots, n-1\}$.

Formally, the conditional distribution of Y_{nj+i} given all messages, delays (hence also all inputs) and other outputs, only depends on the values $x_1, x_2, \ldots, x_n$ of $X_{1,nj+i+D_1(n)}$, $X_{2,nj+i+D_2(n)}, \ldots, X_{K,nj+i+D_k(n)}$, and equals $W(\cdot|x_1, x_2, \ldots, x_K)$.

It is possible to define the decoder in several ways. We will consider two different definitions, which give the strongest version of the converse respectively the direct parts of the coding theorems.

Definition 2.9. An informed infinite decoder is defined as a function which assigns to each two way infinite output sequence $\{y_l, l \in \mathbb{Z}\}$ and each realization of $\mathbf{D}(n) = (D_1(n), D_2(n), \ldots, D_K(n))$, a K-tuple of messages $\{\hat{m}_{m,0}, m \in [K]\}$.

Definition 2.10. An uninformed L-block decoder, $L \in \mathbb{Z}^+$, is defined as a function which assigns to each $(2Ln-1)$-tuple $\{y_l, l \in \{-Ln+1, \ldots, 0, \ldots, Ln-1\}\}$ of possible output realizations a K-tuple of messages $\{\hat{m}_{m,0}, m \in [K]\}$.

The codebooks and the decoder form an n-length coding/decoding system.

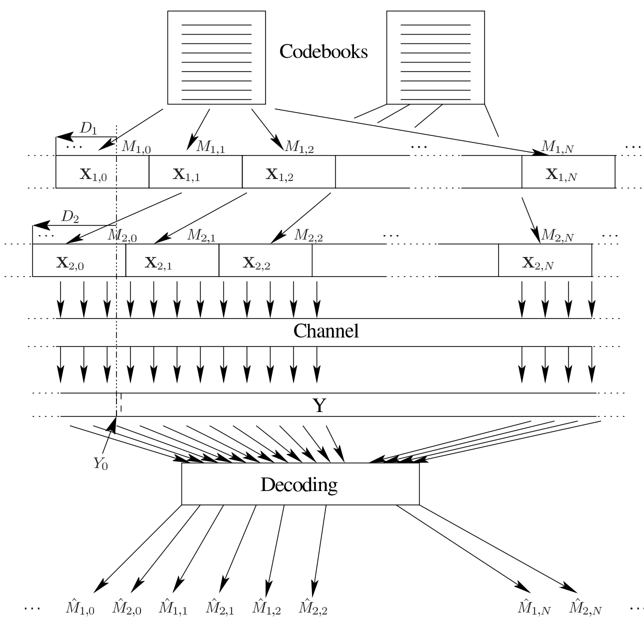

Fig. 1. The setting for two senders.

The definitions above determine the probability structure of the model, for each fixed $n \in \mathbb{Z}^+$. For each $m \in [K]$ the random variable sequence $\{M_{m,j}, j \in \mathbb{Z}\}$ is the two way infinite message flow of the mth sender. The corresponding flow of codewords is $\{\mathbf{X}_{m,j}, j \in \mathbb{Z}\}$. The flows of the senders, the channel transition and the delay system $\mathbf{D}$, define a two way infinite output random variable sequence $\{Y_l, l \in$

$\mathbb{Z}\}$. In case of uninformed L-block decoder the receiver examines the output block $Y_{-Ln+1}, Y_{-Ln+2}, \ldots, Y_0, Y_1, \ldots Y_{Ln-1}$ from which estimates $\{\hat{M}_{m,0}, m \in [K]\}$ are created. In case of an informed infinite decoder the whole output sequence and the realizations of delays can be used to form estimates $\{\hat{M}_{m,0}, m \in [K]\}$. The receiver is assumed to perform the same but shifted decoding procedure at each time instance $\{nk,\, k \in \mathbb{Z}\}$. Hence the estimates $\{\hat{M}_{m,j}, m \in [K]\,, j \in \mathbb{Z}\}$ are also defined. See Fig. 1. for this model, in case $K = 2$.

We will consider two different error definitions. As is standard for multiple-access channels, both errors are averages over messages. However, our first error type also involves averaging over delays, while the second one takes maximum over the possible delays.

Definition 2.11. The average error is:

$$P_e^n = \Pr\left\{\bigcup_{m=1}^{K}\left\{M_{m,0} \neq \hat{M}_{m,0}\right\}\right\}. \tag{3}$$

Definition 2.12. The maximal error is:

$$P_e^n(*) = \max_{\mathbf{d}(n):\Pr\{\mathbf{D}(n)=\mathbf{d}(n)\}>0} \Pr\left\{\bigcup_{m=1}^{K}\left\{M_{m,0} \neq \hat{M}_{m,0}\right\}|\mathbf{D}(n) = \mathbf{d}(n)\right\}. \tag{4}$$

Remark 2.13. The average error depends on the joint distribution of delays $(D_1(n), D_2(n), \ldots, D_K(n))$, while the maximal error depends on the joint distribution of the delays only through its support.

Remark 2.14. The two kinds of error are related very closely. If $P_e^n(*) \to 0$ then $P_e^n \to 0$. On the other hand, if $P_e^n \to 0$ exponentially as $n \to \infty$ and if $\min_{\mathbf{d}(n):\Pr\{\mathbf{D}(n)=\mathbf{d}(n)\}>0} \Pr\{\mathbf{D}(n) = \mathbf{d}(n)\}$ tends to 0 slower than exponentially then also $P_e^n(*) \to 0$ exponentially.

We have defined several types of models according to the various definitions of decoder and of error. For the sake of brevity, the following definition is meant to define a capacity region simultaneously for all cases. Here, in case of L-block decoder, a proper choice of L is understood (not depending on n). In particular cases, a suitable L will be specified, not entering the question whether a smaller L would also do.

Definition 2.15. Corresponding to the delay system $\mathbf{D}$, a vector $(R_1, R_2, \ldots, R_K)$ is an achievable rate vector if for every $\varepsilon > 0$, $\delta > 0$ for all $N_0 \in \mathbb{Z}^+$ there exists a coding/decoding system with blocklength $n > N_0$ with rates coordinate-wise exceeding $(R_1 - \delta, R_2 - \delta, \ldots, R_K - \delta)$ and with error less than ε. The set of achievable rate vectors is the capacity region of the K-AMAC.

Remark 2.16. In the definition above we used the 'optimistic' definition of capacity region, rather than the more usual 'pessimistic one', see [4][1]. The reason is that in the even delays case there are differences in the performance of coding/decoding systems of even and odd blocklength (see Theorem 5.1).

[1]In short, in the 'optimistic' definition it is enough to show that there is a "good" coding/decoding system for a sequence of blocklength $n_k \to \infty$.

Remark 2.17. If for some region achievability is proved in case of uninformed L-block decoder with maximal error, and the converse is proved in case of informed infinite decoder with average error, then for any combination of the model assumptions above the capacity region is equal to this region.

Lemma 2.18. For either type of AMAC model, if $\mathbf{D}$ and $\mathbf{D}'$ are two delay systems such that for some $0 < \alpha \le 1$ for all $n \in \mathbb{Z}^+$ and $\mathbf{d}(n) \in \{0, 1, \dots, n-1\}^K$

$$\Pr\{\mathbf{D}'(n) = \mathbf{d}(n)\} \ge \alpha \Pr\{\mathbf{D}(n) = \mathbf{d}(n)\}, \tag{5}$$

then the capacity region under delay system $\mathbf{D}'$ is contained (perhaps strictly) in the capacity region under delay system $\mathbf{D}$.

Proof. Consider an arbitrary n length coding/decoding system. Then $P^n_{e,\mathbf{D}'(n)} \ge \alpha P^n_{e,\mathbf{D}(n)}$ and $P^n_{e,\mathbf{D}'(n)}(*) \ge P^n_{e,\mathbf{D}(n)}(*)$ hold, where the lower indices indicate the underlying delay system. This proves the lemma. □

Remark 2.19. For either type of decoder, if two delay systems $\mathbf{D}$ and $\mathbf{D}'$ have the same support set for each n, then the capacity regions corresponding to delay systems $\mathbf{D}$ and $\mathbf{D}'$ coincide in case of maximal error. Furthermore, if the equation (5) is fulfilled by $\mathbf{D}$ and $\mathbf{D}'$ and it is also fulfilled when the roles of $\mathbf{D}$ and $\mathbf{D}'$ are reversed, then by Lemma 2.18 the capacity regions also coincide in case of average error.

3. A GENERAL CONVERSE

In this section a general converse theorem is proved, which depends on the delay system. In the following sections, this general converse is used to derive the capacity region of special cases.

For all subset S of $[K]$ write

$$\mathbf{X}_S = (X_i)_{i\in S},\ S^c = [K] \setminus S, \tag{6}$$

and for all $\mathbf{R} = (R_1, R_2, \dots, R_K)$ write

$$R(S) = \sum_{i\in S} R_i. \tag{7}$$

Let $\mathbf{D} = \mathbf{D}(n)$ denote the delay vector. Let $\mathbf{X}_{B,i+D_B}$ denote the random vector with components $X_{l,i+D_l}$, $l \in B$ where $B \subset [K]$ and $X_{m,j}$ is defined as in definition 2.2; similar notation is used where $+$ is replaced by $\oplus$ which means addition modulo n.

Theorem 3.1. For any n length coding/decoding system with rate vector $\mathbf{R} = (R_1, R_2, \dots, R_k)$ for a K senders AMAC W with informed infinite decoder, the following bounds hold for all $S \subset [K]$:

$$R(S) \le \mathrm{I}(\mathbf{X}_{S,Q\oplus D_S} \wedge \tilde{Y}_Q | \mathbf{X}_{S^c,Q\oplus D_{S^c}}, Q, \mathbf{D}) + \varepsilon_n. \tag{8}$$

Here $\varepsilon_n = (R([K]))P^n_e + \frac{1}{n}$, the random variable Q is uniformly distributed on $\{0, 1, \dots, n-1\}$ and is independent of $\mathbf{D}$ and the message flows of the senders. Further, the conditional distribution of $\tilde{Y}_Q$ on the condition that the values of Q, $\mathbf{D}$ and the messages $M_{m,j}$ (hence also of $X_{1,Q\oplus D_1}, X_{2,Q\oplus D_2}, \ldots, X_{K,Q\oplus D_K}$) are given, depends only on the values $x_1, x_2, \ldots, x_K$ of the latter random variables and is equal to $W(\cdot|x_1, x_2, \ldots, x_K)$.

Remark 3.2. Theorem 3.1 will be used for sequences of coding/decoding systems with $P_e^n \to 0$. In this case ε_n also tends to 0.

P r o o f. For the sake of clarity just the two senders special case is addressed here, the full proof of Theorem 3.1 can be found in Appendix B. In case of two senders the bounds (8) are:

$$R_1 \le \mathrm{I}(X_{1,Q\oplus D_1} \wedge \tilde{Y}_Q | X_{2,Q\oplus D_2}, Q, D_1, D_2) + \varepsilon_n \tag{9}$$

$$R_2 \le \mathrm{I}(X_{2,Q\oplus D_2} \wedge \tilde{Y}_Q | X_{1,Q\oplus D_1}, Q, D_1, D_2) + \varepsilon_n \tag{10}$$

$$R_1 + R_2 \le \mathrm{I}(X_{1,Q\oplus D_1}, X_{2,Q\oplus D_2} \wedge \tilde{Y}_Q | Q, D_1, D_2) + \varepsilon_n. \tag{11}$$

Note that $n\varepsilon_n = n(R_1 + R_2)P_e^{(n)} + 1$. Hence

$$n\varepsilon_n \ge \mathrm{H}(M_{1,0}, M_{2,0} | \hat{M}_{1,0}, \hat{M}_{2,0}) \tag{12}$$

by Fano's inequality.

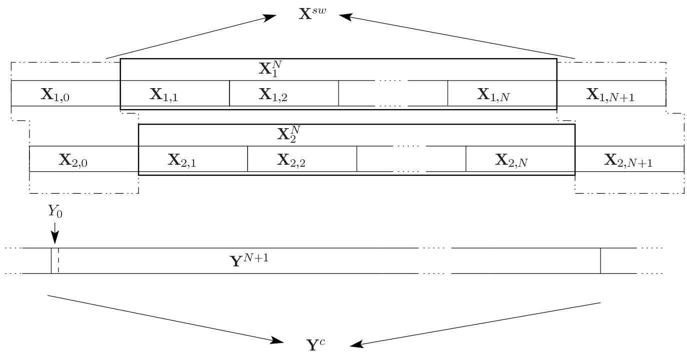

Fig. 2. The random variables that play role in the bound on the sum $R_1 + R_2$.

We just bound $R_1 + R_2$. The bounds for R_1 and R_2 can be derived similarly (See also Appendix B).

Take a window of the receiver consisting of $N+1$ n-length blocks $\mathbf{Y}^{N+1} = \{Y_0, Y_1, \ldots, Y_{n(N+1)-1}\}$. This window fully covers the code-blocks $\mathbf{X}_{1,1}, \mathbf{X}_{1,2}, \ldots, \mathbf{X}_{1,N}$ of sender 1 and $\mathbf{X}_{2,1}, \mathbf{X}_{2,2}, \ldots, \mathbf{X}_{2,N}$ of sender 2, denoted by $\mathbf{X}_1^N$ and $\mathbf{X}_2^N$ respectively. The code-words at the sides of the output window are $\mathbf{X}_{1,0}, \mathbf{X}_{2,0}, \mathbf{X}_{1,N+1}, \mathbf{X}_{2,N+1}$, denote this quadruple by $\mathbf{X}^{sw}$ (where sw stands for "side of the window"). Then we have

$$Nn(R_1 + R_2) = \mathrm{H}(\mathbf{M}_1^N, \mathbf{M}_2^N) \tag{13}$$

$$= \mathrm{I}(\mathbf{M}_1^N, \mathbf{M}_2^N \wedge \hat{\mathbf{M}}_1^N, \hat{\mathbf{M}}_2^N) + \mathrm{H}(\mathbf{M}_1^N, \mathbf{M}_2^N | \hat{\mathbf{M}}_1^N, \hat{\mathbf{M}}_2^N) \tag{14}$$

$$\leq \mathrm{I}(\mathbf{M}_1^N, \mathbf{M}_2^N \wedge \hat{\mathbf{M}}_1^N, \hat{\mathbf{M}}_2^N) + \sum_{i=1}^{N} \mathrm{H}(M_{1,i}, M_{2,i} | \hat{M}_{1,i}, \hat{M}_{2,i}) \tag{15}$$

$$\leq \mathrm{I}(\mathbf{M}_1^N, \mathbf{M}_2^N \wedge \hat{\mathbf{M}}_1^N, \hat{\mathbf{M}}_2^N) + Nn\varepsilon_n \tag{16}$$

$$\leq \mathrm{I}(\mathbf{X}_1^N, \mathbf{X}_2^N \wedge \mathbf{Y}^{N+1}, \mathbf{X}^{sw}, D_1, D_2) + Nn\varepsilon_n \tag{17}$$

where (16) comes from (12) and (17) comes from the Markov relation

$$(\mathbf{M}_1^N, \mathbf{M}_2^N) \ominus (\mathbf{X}_1^N, \mathbf{X}_2^N) \ominus (\mathbf{Y}^{N+1}, \mathbf{X}^{sw}, D_1, D_2) \ominus \\ \ominus (\mathbf{Y}^{N+1}, \mathbf{Y}^c, D_1, D_2) \ominus (\hat{\mathbf{M}}_1^N, \hat{\mathbf{M}}_2^N). \tag{18}$$

Here $\mathbf{Y}^c$ denote the whole output sequence except $\mathbf{Y}^{N+1}$. Note that in [5] the Markov relation $(\mathbf{X}_1^N, \mathbf{X}_2^N) \ominus (\mathbf{Y}^{N+1}, D_1, D_2) \ominus (\mathbf{Y}^{N+1}, \mathbf{Y}^c)$ was assumed, which need not hold in general[2]. It seems that Poltyrev [15] also made this error.

Introduce the notation $|\mathcal{X}| = \max(|\mathcal{X}_1|, |\mathcal{X}_2)|$. Continuing the bounds (13) – (17),

$$Nn(R_1 + R_2) \leq \mathrm{I}(\mathbf{X}_1^N, \mathbf{X}_2^N \wedge \mathbf{Y}^{N+1}, \mathbf{X}^{sw}, D_1, D_2) + Nn\varepsilon_n \tag{19}$$

$$= \mathrm{H}(\mathbf{X}_1^N, \mathbf{X}_2^N) + Nn\varepsilon_n - \mathrm{H}(\mathbf{X}_1^N, \mathbf{X}_2^N | \mathbf{Y}^{N+1}, \mathbf{X}^{sw}, D_1, D_2) \tag{20}$$

$$= \mathrm{H}(\mathbf{X}_1^N, \mathbf{X}_2^N | D_1, D_2) + Nn\varepsilon_n - \mathrm{H}(\mathbf{X}_1^N, \mathbf{X}_2^N | \mathbf{Y}^{N+1}, D_1, D_2) \\ + \mathrm{H}(\mathbf{X}_1^N, \mathbf{X}_2^N | \mathbf{Y}^{N+1}, D_1, D_2) - \mathrm{H}(\mathbf{X}_1^N, \mathbf{X}_2^N | \mathbf{Y}^{N+1}, \mathbf{X}^{sw}, D_1, D_2) \tag{21}$$

$$= \mathrm{I}(\mathbf{X}_1^N, \mathbf{X}_2^N \wedge \mathbf{Y}^{N+1} | D_1, D_2) + Nn\varepsilon_n + \mathrm{I}(\mathbf{X}^{sw} \wedge \mathbf{X}_1^N, \mathbf{X}_2^N | \mathbf{Y}^{N+1}, D_1, D_2) \tag{22}$$

$$\leq \mathrm{H}(\mathbf{Y}^{N+1} | D_1, D_2) - \mathrm{H}(\mathbf{Y}^{N+1} | \mathbf{X}_1^N, \mathbf{X}_2^N, D_1, D_2) + 4n \log |\mathcal{X}| + Nn\varepsilon_n \tag{23}$$

$$= \mathrm{H}(\mathbf{Y}^{N+1} | D_1, D_2) + 4n \log |\mathcal{X}| + Nn\varepsilon_n \\ - \sum_{j=0}^{N} \sum_{i=0}^{n-1} \mathrm{H}(Y_{nj+i} | \mathbf{Y}_0^{nj+i-1} \mathbf{X}_1^N, \mathbf{X}_2^N, D_1, D_2) \tag{24}$$

$$\leq \sum_{j=0}^{(N+1)n-1} \mathrm{H}(Y_j | D_1, D_2) + 4n \log |\mathcal{X}| + Nn\varepsilon_n \\ - \sum_{j=1}^{N-1} \sum_{i=0}^{n-1} \mathrm{H}(Y_{nj+i} | X_{1,nj+i+D_1}, X_{2,nj+i+D_2}, D_1, D_2). \tag{25}$$

In (25) we dropped some negative terms (notice that j runs from 1 to $N-1$). Introduce the random variable $\tilde{Y}_i$ such that its conditional distribution given D_1, D_2 and the messages $M_{m,j}$, depends only on the values x_1, x_2 of $X_{1,i\oplus D_1}$ and $X_{2,i\oplus D_2}$, and is equal to $W(\cdot|x_1, x_2)$. For all j the joint distribution of $(D_1, D_2, X_{1,nj+i+D_1}, X_{2,nj+i+D_2}, Y_{nj+i})$ is the same as the joint distribution of $(D_1, D_2, X_{1,i\oplus D_1}, X_{2,i\oplus D_2}, \tilde{Y}_i)$. Using this substi-

[2]For a simple counterexample, let W be the 2-user binary adder channel with $\mathcal{X}_1 = \mathcal{X}_2 = \{0,1\}$ and $\mathcal{Y} = \{0,1,2\}$, and let $n = 2$, $N = 1$. Both senders have the codewords 00 and 11. Elementary calculations show that the conditional probability that Y_{-1} is equal to 2 given that $Y_0 = Y_1 = Y_2 = Y_3 = 1, D_1 = 1, D_2 = 0$ is $\frac{1}{4}$, while given also that $M_{1,1} = 1, M_{2,1} = 2$ (and hence $\mathbf{X}_{1,1} = 00, \mathbf{X}_{2,1} = 11$) this probability becomes 0.

tution, (25) can be further bounded from above by:

$$\begin{aligned}
\leq &(N-1)\sum_{i=0}^{n-1} \mathrm{H}(\tilde{Y}_i|D_1, D_2) + 2n\log|\mathcal{Y}| + 4n\log|\mathcal{X}| \\
&- (N-1)\sum_{i=0}^{n-1} \mathrm{H}(\tilde{Y}_i|X_{1,i\oplus D_1}X_{2,i\oplus D_2}, D_1, D_2) + Nn\varepsilon_n \qquad (26)\\
= &(N-1)\sum_{i=0}^{n-1} \mathrm{I}(X_{1,i\oplus D_1}, X_{2,i\oplus D_2} \wedge \tilde{Y}_i|D_1, D_2) + Nn\varepsilon_n + 4n\log|\mathcal{X}| + 2n\log|\mathcal{Y}|. \qquad (27)
\end{aligned}$$

Divide by nN and introduce the random variable Q uniformly distributed on $\{0, 1, \ldots, n-1\}$, independent of D_1, D_2 and the messages $M_{m,j}$ and introduce $\tilde{Y}_Q$ such that its conditional distribution given Q, D_1, D_2 and the messages $M_{m,j}$, depends only on the values x_1, x_2 of $X_{1,Q\oplus D_1}$ and $X_{2,Q\oplus D_2}$, and is equal to $W(\cdot|x_1, x_2)$. Then

$$\begin{aligned}
&R_1 + R_2 \\
&\leq \frac{N-1}{Nn}\sum_{i=1}^{n} \mathrm{I}(X_{1,i\oplus D_1}, X_{2,i\oplus D_2} \wedge \tilde{Y}_i|D_1, D_2) + \varepsilon_n + \frac{2\log|\mathcal{Y}|}{N} + \frac{4\log|\mathcal{X}|}{N} \qquad (28)\\
&\leq \frac{N-1}{N}\mathrm{I}(X_{1,Q\oplus D_1}, X_{2,Q\oplus D_2} \wedge \tilde{Y}_Q|Q, D_1, D_2) + \varepsilon_n + \frac{2\log|\mathcal{Y}|}{N} + \frac{4\log|\mathcal{X}|}{N}. \qquad (29)
\end{aligned}$$

If $N \to \infty$ then

$$R_1 + R_2 \leq \mathrm{I}(X_{1,Q\oplus D_1}, X_{2,Q\oplus D_2} \wedge \tilde{Y}_Q|Q, D_1, D_2) + \varepsilon_n. \qquad (30)$$

□

Corollary 3.3. Under the assumptions of Theorem 3.1 the following bounds hold in the 2-senders case:

$$R_1 \leq \mathrm{I}(X_{1,Q} \wedge \hat{Y}_Q|X_{2,Q\ominus D}, Q, D) + \varepsilon_n \qquad (31)$$

$$R_2 \leq \mathrm{I}(X_{2,Q\ominus D} \wedge \hat{Y}_Q|X_{1,Q}, Q, D) + \varepsilon_n \qquad (32)$$

$$R_1 + R_2 \leq \mathrm{I}(X_{1,Q}, X_{2,Q\ominus D} \wedge \hat{Y}_Q|Q, D) + \varepsilon_n. \qquad (33)$$

Here Q is uniformly distributed on $\{0, 1, \ldots, n-1\}$ and independent of D_1, D_2 and the message flows of the senders, $\ominus$ denotes the subtraction modulo n, $D = D_1 \ominus D_2$ is the relative delay between the two senders. Further, the conditional distribution of $\hat{Y}_Q$ on the condition that the values of Q, D_1, D_2 and the messages $M_{m,j}$ (hence also of D, $X_{1,Q}, X_{2,Q\ominus D}$) are given, depends only on the values x_1, x_2 of the last two random variables and is equal to $W(\cdot|x_1, x_2)$.

P r o o f. Expand the right sides of the equations (9),(10),(11) as sums for the possible values of Q, D_1, D_2, e. g.

$$\begin{aligned}
&\mathrm{I}(X_{1,Q\oplus D_1}, X_{2,Q\oplus D_2} \wedge \tilde{Y}_Q|Q, D_1, D_2) \\
&= \sum_q \sum_{d_1} \sum_{d_2} \frac{1}{n} \cdot \Pr(D_1 = d_1)\Pr(D_2 = d_2)\,\mathrm{I}(X_{1,q\oplus d_1}, X_{2,q\oplus d_2} \wedge \tilde{Y}_{q,d_1,d_2}). \qquad (34)
\end{aligned}$$

Substituting $q' = q \oplus d_1$ and $d = d_1 \ominus d_2$, and renaming q' to q, the Corollary is proved. □

4. KNOWN CAPACITY REGIONS WITH A NEW INSIGHT

4.1. The asynchronous one-sender model

In this section the asynchronous model from section 2 is analyzed where there is just one sender ($K = 1$). This will provide the basics for the decoding method of the K-AMAC in general.

In case of $K = 1$, $W : \mathcal{X} \to \mathcal{Y}$ denotes a classical DMC. For the sake of clarity, we omit from the notations of Section 2 the index corresponding to the unique sender. Let $\{\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(\mathcal{M})\}$ denote the codewords of the codebook of the sender, where $\mathcal{M} = 2^{nR}$ is the number of codewords in the codebook of the sender. The coordinates of $\mathbf{x}(i)$ are denoted by $(x_0(i), x_1(i), \dots, x_{n-1}(i))$.

The difference between this model and the classical one is that the task of the receiver is not just decoding the codewords but also to find the beginning of the codewords. Note that related problems have been considered in the literature, for example in [16, 17]. The known results, however, do not directly apply for our purposes.

Theorem 4.1. For each version of the model, the capacity region of the one sender asynchronous model is $[0, \max_p(I(p, W))]$, the same as that of the classical model, in case of arbitrary delay system.

Remark 4.2. It has crucial importance in the proofs of Theorems 4.10 and 6.1 that in the achievability proof below, beyond decoding the codewords, the receiver also finds the delay of the sender.

Proof. The converse part follows from Theorem 3.1.

In order to prove the direct part it is enough to restrict attention to uninformed L-block decoder and to maximal error; actually $L = 1$ suffices. Moreover, it is enough to show that $\max_p(I(p, W))$ is achievable rate for delay system uniformly distributed on the set $\{0, 1, \dots, n-1\}$ (see also Remark 2.13).

Standard random coding argument is used, the main difficulty is summarized in Remark 4.3. Let p be an arbitrary distribution over the input alphabet $\mathcal{X}_1$. Choose the symbols of codewords in the codebook of rate $0 < R = I(p, W) - 2\delta$ independently according to p. Let $P^n(\mathbf{x}^n, \mathbf{y}^n)$ be the joint distribution on $\mathcal{X}^n \times \mathcal{Y}^n$ induced by the nth power of p and by the memoryless channel W. Let q^n be the marginal of P^n on $\mathcal{Y}^n$. We define the decoder as follows. In order to estimate the 0th sent message $M_{1,0}$, the receiver first examines the n-tuple of outputs $(Y_{-n+1}, Y_{-n+2}, \dots, Y_0)$, then it examines the next n-tuple $(Y_{-n+2}, Y_{-n+3}, \dots, Y_1)$, etc. until the n-tuple $(Y_0, Y_1, \dots Y_{n-1})$. The estimate will be $\hat{M}_{1,0} = s$ if among the examined n-tuples there is a unique one denoted by Y^n, for which $((X_0(s), X_1(s), \dots, X_{n-1}(s)), Y^n)$ belongs to the typical set

$$S_n^\delta := \left\{ (\mathbf{x}^n, \mathbf{y}^n) : P^n(\mathbf{x}^n, \mathbf{y}^n) > 0 \text{ and } \left| \frac{1}{n} log \frac{P^n(\mathbf{x}^n, \mathbf{y}^n)}{p^n(\mathbf{x}^n) q^n(\mathbf{y}^n)} - I(p, W) \right| \leq \delta \right\}, \quad (35)$$

and also this s is unique. The following argument shows that, for this random code with the above decoder,

$$P_e^n(d) = \Pr\left\{M_{1,0} \neq \hat{M}_{1,0} | D_1 = d\right\} \tag{36}$$

is exponentially small for each delay d.

By the symmetry of the random code it can be assumed that $M_{1,-1} = 2$, $M_{1,0} = 1$, and $M_{1,1} = 3$ (the probability that they are not different is exponentially small). It is clear from the classical channel coding theorem that if the decoder examines an n-tuple Y^n which is the output of the whole codeword $\mathbf{X}(1)$, then the decoder will find $\mathbf{X}(1)$ but no other codewords jointly typical with Y^n, with probability exponentially close to 1.

Hence, we only have to discuss the case when the decoder examines output symbols in a window of length n when the input symbols were $(X_{n-l}(2), \ldots, X_{n-1}(2), X_0(1), \ldots, X_{n-l-1}(1))$ for some $n > l > 0$ (the opposite cases, i. e., when the first part of the input symbols come from $\mathbf{X}(1)$ and the second part from $\mathbf{X}(3)$, can be analyzed similarly). We will show that the probability of incorrectly recognizing typicality in this window is exponentially small. The probability, conditioned on the previously presented structure of the examined window, that codewords $\mathbf{X}(s)$ will be typical with this examined output n tuple can be written as:

$$\begin{aligned}
&\Pr_{cond}\left\{(X_0(s), \ldots, X_{n-1}(s), Y^n) \in S_n^\delta\right\} \\
&= \sum_{(\mathbf{x}^n(s), \mathbf{y}^n) \in S_n^\delta} p^n(\mathbf{x}^n(s)) \cdot \Pr_{cond}\left\{Y^n = \mathbf{y}^n | (X_0(s), \ldots, X_{n-1}(s)) = \mathbf{x}^n(s)\right\} &(37)\\
&= \sum_{(\mathbf{x}^n(s), \mathbf{y}^n) \in S_n^\delta} p^n(\mathbf{x}^n(s)) \frac{q^n(\mathbf{y}^n)}{q^n(\mathbf{y}^n)} \cdot \Pr_{cond}\left\{Y^n = \mathbf{y}^n | (X_0(s), \ldots, X_{n-1}(s)) = \mathbf{x}^n(s)\right\} &(38)\\
&\leq \sum_{(\mathbf{x}^n(s), \mathbf{y}^n) \in S_n^\delta} \frac{2^{-n(I(p,W)-\delta)} P^n(\mathbf{x}^n(s), \mathbf{y}^n)}{q^n(\mathbf{y}^n)} \\
&\qquad \cdot \Pr_{cond}\left\{Y^n = \mathbf{y}^n | (X_0(s), \ldots, X_{n-1}(s)) = \mathbf{x}^n(s)\right\} &(39)\\
&= 2^{-n(I(p,W)-\delta)} \sum_{(\mathbf{x}^n(s), \mathbf{y}^n) \in S_n^\delta} P^n(\mathbf{x}^n(s) | \mathbf{y}^n) \\
&\qquad \cdot \Pr_{cond}\left\{Y^n = \mathbf{y}^n | (X_0(s), \ldots, X_{n-1}(s)) = \mathbf{x}^n(s)\right\}. &(40)
\end{aligned}$$

In the above derivation the definition of the set S_n^δ is used, and in the last equation $P^n(\mathbf{x}^n(s)|\mathbf{y}^n)$ denotes the conditional probability induced by the joint distribution P^n. We will show that the sum in (40) is ≤ 1. We have to use the structure of the examined window. Recall the assumption that in this window the second part of $\mathbf{X}(2)$ and the first part of $\mathbf{X}(1)$ were sent. We should distinguish three cases: $\{s \neq 2, s \neq 1\}$, $\{s = 1\}$, $\{s = 2\}$. In the first case $\Pr_{cond}\{Y^n = \mathbf{y}^n | (X_0(s), \ldots, X_{n-1}(s)) = \mathbf{x}^n(s)\}$ is equal to $\Pr_{cond}\{Y^n = \mathbf{y}^n\}$. This proves that in this case the sum in (40) is indeed ≤ 1. The remaining two cases can be treated very similarly. For the sake of brevity, concentrate on the case $\{s = 1\}$. Then the sum in (40) is bounded from above by

$$\sum_{\mathbf{x}^n(2)\in\mathcal{X}^n} p^n(\mathbf{x}^n(2)) \sum_{(\mathbf{x}^n(1),\mathbf{y}^n)\in\mathcal{X}^n\times\mathcal{Y}^n} \left[\prod_{h=0}^{n-1} P(x_h(1)|y_h)\right]\left[\prod_{h=0}^{l-1} W(y_h|x_{n-l+h}(2))\right]\left[\prod_{h=l}^{n-1} W(y_h|x_{h-l}(1))\right]. \quad (41)$$

Sum in the following order: $x_{n-1}(1)$, y_{n-1}, $x_{n-2}(1)$, y_{n-2}, $\ldots$, $x_l(1)$, y_l. We get that (41) is equal to

$$\sum_{\mathbf{x}^n(2)\in\mathcal{X}^n} p^n(\mathbf{x}^n(2)) \sum_{(\mathbf{x}^l(1),y^l)\in\mathcal{X}^l\times\mathcal{Y}^l} \left[\prod_{h=0}^{l-1} P(x_h(1)|y_h)\right]\left[\prod_{h=0}^{l-1} W(y_h|x_{n-l+r}(2))\right]. \quad (42)$$

This is equal to 1, because the terms of the inner sum may be regarded as a joint distribution for a memoryless channel with inputs y_i and outputs $x_i(1)$.

In the above derivation we demonstrated that the probability that the receiver finds the sth codeword typical with an output n tuple whose input symbols consist of two different codewords, can be bounded from above by $2^{-n(I(p,W)-\delta)}$. Using the union bound over all codewords and over all the n tuples examined by the decoder, gives that the probability of recognizing one of the codewords in a window where the inputs are from two different codewords is less then $n2^{-n\delta}$.

The above paragraphs show that, for the random code, $P_e^n(d)$ in (36) and hence also the average error over delays is exponentially small. We can conclude that there exists a sequence of deterministic coding-decoding systems with exponentially small average error over delays. Optimizing the distribution p and taking into account Remark 2.14 we can see that $\max_p I(p,W)$ is achievable in case of uninformed $L = 1$-block decoder and maximal error. □

Remark 4.3. Cases $\{s = 1\}$, $\{s = 2\}$ are the main difficulties in this proof. The tricky summation in equation (41) which solves these difficulties is adopted from Gray [9].

Remark 4.4. Note that using [9] a stronger result can be proved: any sequence of deterministic codes which work well for the classical channel coding model, can be modified to work well for the asynchronous model. Namely, if the same random synchronization sequence of length $k \approx \log^2(n)$ is appended to each of the original codewords, then with probability tending to 1 it is possible to detect the synchronization sequence and decode the original codewords. However, the authors believe that the proof without using synchronization sequence is somewhat simpler.

4.2. The totally asynchronous case

From this point on, the paper strongly relies on the results of [8] and [13]. Though the reader is assumed familiar with the concepts of successive decoding and rate splitting, the basics will be summarized below.

Let W be a K-AMAC.

Definition 4.5. The convex polytope $\mathcal{R}\left[W; p(x_1, x_2, \ldots, x_K)\right]$ is the set of K tuples $\mathbf{R} \in (\mathbb{R}^+)^K$ such that

$$R(S) \leq I(\mathbf{X}_S \wedge Y | \mathbf{X}_{S^c}) \ , S \subseteq [K] \, , \tag{43}$$

where the joint distribution of $X_1, X_2 \ldots, X_K$ is $p(x_1, x_2 \ldots, x_K)$ and Y is connected to $X_1, X_2, \ldots, X_K$ by the channel W.

Definition 4.6. Let C denote the following set:

$$C := \bigcup_{p_{X_1} \times p_{X_2} \times \cdots \times p_{X_K}} \mathcal{R}\left[W; p_{X_1} \times p_{X_2} \times \cdots \times p_{X_K}\right] \tag{44}$$

where the union is over all product distributions.

Definition 4.7. The dominant face of $\mathcal{R}\left[W; p_{X_1} \times p_{X_2} \times \cdots \times p_{X_K}\right]$ is its subset consisting of all vectors $(R_1, R_2, \ldots, R_K)$ with $R([K]) = \mathrm{I}(\mathbf{X}_{[K]} \wedge Y)$. It is denoted by $D(\mathcal{R}\left[W; p_{X_1} \times p_{X_2} \times \cdots \times p_{X_K}\right])$.

Definition 4.8. We say that $(R_1, R_2, \ldots, R_K)$ is dominated by $(\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_K)$ if $R_1 \leq \tilde{R}_1$, $R_2 \leq \tilde{R}_2$,..., $R_K \leq \tilde{R}_K$.

It can be seen[3] that the points of $D(\mathcal{R}\left[W; p_{X_1} \times p_{X_2} \times \cdots \times p_{X_K}\right])$ cannot be dominated by other points of $\mathcal{R}\left[W; p_{X_1} \times p_{X_2} \times \cdots \times p_{X_K}\right]$, but any point from $\mathcal{R}\left[W; p_{X_1} \times p_{X_2} \times \cdots \times p_{X_K}\right]$ can be dominated by a point from the dominant face.

Remark 4.9. According to Definition 2.15, if $(R_1, R_2, \ldots, R_K)$ is dominated by an achievable rate vector then the rate vector $(R_1, R_2, \ldots, R_K)$ is also achievable.

According to [8] the vertices of $D(\mathcal{R}\left[W; p_{X_1} \times p_{X_2} \times \cdots \times p_{X_K}\right])$ can be described in the following way. Let $\pi = (\pi_1, \pi_2, \ldots, \pi_K)$ be an ordering of $[K]$. For all $i \in [K]$ let $R^{\pi}_{\pi_i}$ be equal to $I(X_{\pi_i} \wedge Y | \mathbf{X}_{\{\pi_1, \ldots, \pi_{i-1}\}})$. For example if $K = 3$, and $\pi = (2, 3, 1)$, then $R^{\pi}_2 = I(X_2 \wedge Y)$, $R^{\pi}_3 = I(X_3 \wedge Y | X_2)$, $R^{\pi}_1 = I(X_1 \wedge Y | X_2, X_3)$. Then the rate vector $\mathbf{R}^{\pi} = (R^{\pi}_1, R^{\pi}_2, \ldots, R^{\pi}_K)$ is a vertex, and all vertices of $D(\mathcal{R}\left[W; p_{X_1} \times p_{X_2} \times \cdots \times p_{X_K}\right])$ can be written in this way with appropriate π. Note that the vertices $\mathbf{R}^{\pi}$ need not be all distinct.

In the Appendix of [8] it is proved for informed $L = K$ block decoder that in the totally asynchronous case $\mathbf{R}^{\pi} \in D(\mathcal{R}\left[W; p_{X_1} \times p_{X_2} \times \cdots \times p_{X_K}\right])$ can be achieved by successive decoding with ordering π. We summarize the proof for $\mathbf{R}^{\{1,2,\ldots,K\}}$. The coding/decoding system is randomly constructed the following way. The symbols of the codebooks of the senders are chosen independently according to the appropriate input distributions. The receiver first decodes by joint typicality the codewords of the first sender, considering the random codewords of the other senders as noise. This means that the receiver behaves as if there were only one sender and the channel was the following:

$$W^1(y|x_1) = \sum_{x_2 \in \mathcal{X}_2} \sum_{x_3 \in \mathcal{X}_3} \cdots \sum_{x_k \in \mathcal{X}_K} p_{X_2}(x_2) p_{X_3}(x_3) \cdots p_{X_K}(x_K) W(y | x_1, x_2, \ldots, x_K). \tag{45}$$

[3] According to [8] it is a consequence of the fact that $\mathcal{R}\left[W; p_{X_1} \times p_{X_2} \times \cdots \times p_{X_K}\right]$ is a polymatroid, which was observed in [10, 14].

Next the receiver decodes the codewords of the second sender by joint typicality using the already decoded codewords of the first sender, considering the other senders as noise. This means that the receiver behaves as it would in a one sender model when the channel was the following:

$$W^2(y, x_1|x_2) = \sum_{x_3\in\mathcal{X}_3}\sum_{x_4\in\mathcal{X}_3}\cdots\sum_{x_k\in\mathcal{X}_K} p_{X_1}(x_1)p_{X_3}(x_3)\cdots p_{X_K}(x_K)W(y|x_1, x_2, \ldots, x_K). \tag{46}$$

The codewords of the other senders are decoded similarly. In the final decoding step the receiver decodes the codewords of the K'th sender by joint typicality using the already decoded codewords of all the other senders. This means that the receiver behaves as it would in a one sender model when the channel was the following:

$$W^K(y, x_1, x_2, \ldots, x_{K-1}|x_K) = p_{X_1}(x_1)p_{X_2}(x_2)\cdots p_{X_{K-1}}(x_{K-1})W(y|x_1, x_2, \ldots, x_K). \tag{47}$$

More detail can be found in the Appendix of [8].

Now recall the notion of individual split from [8] with splitting function $f(x_a, x_b) = \max(x_a, x_b)$. A split of sender i with input distribution p_{X_i} on $\mathcal{X}_i = \{0, 1, \ldots, |\mathcal{X}_i| - 1\}$ results in two virtual senders ia, ib with distributions $p_{X_{ia}}$ and $p_{X_{ib}}$, also on $\mathcal{X}_i$, explicitly determined by p_{X_i} and a splitting parameter, such that the splitting function $f(x_a, x_b) = \max(x_a, x_b)$ maps $p_{X_{ia}} \times p_{X_{ib}}$ into p_{X_i}.

Section 2 of [8] shows in the totally asynchronous case that each $\mathbf{R} \in D(\mathcal{R}[W; p_{X_1} \times p_{X_2} \times \cdots \times p_{X_K}])$ can be achieved with Rate Splitting via at most $K - 1$ splits[4]. This means that a good code for W with rate vector $\mathbf{R}$ can be obtained from a code with successive decoding for an auxiliary channel $W'_{\mathbf{R}}$ with $2K-1$ virtual senders constructed by splitting the original senders, perhaps some of them split repeatedly and others not at all; the rate vector of this code equals the vertex $\mathbf{R}'^{\pi}$ of the dominant face of $\mathcal{R}\left[W'_{\mathbf{R}}; p_{X'_1} \times p_{X'_2} \times \cdots \times p_{X'_{2K-1}}\right]$ for some ordering π and distributions $p_{X'_1} \times p_{X'_2} \times \cdots \times p_{X'_{2K-1}}$. In particular, the i'th coordinate of $\mathbf{R}$ is the sum of those coordinates of $\mathbf{R}'^{\pi}$ that correspond to the virtual senders into which the i'th sender has been split, $i = 1, 2, \ldots, K$.

Theorem 4.10. In the totally asynchronous case (Example 2.5), for each model version the capacity region is C.

P r o o f. In the converse part it is enough to treat the case of an informed infinite decoder and average error. The right side of eq. (8) can be bounded from above as follows.

$$\mathrm{I}(\mathbf{X}_{S,Q\oplus D_S} \wedge \tilde{Y}_Q|\mathbf{X}_{S^c,Q\oplus D_{S^c}}, Q, \mathbf{D}) \tag{48}$$
$$= \mathrm{H}(\tilde{Y}_Q|\mathbf{X}_{S^c,Q\oplus D_{S^c}}, Q, \mathbf{D}) - \mathrm{H}(\tilde{Y}_Q|\mathbf{X}_{[K],Q\oplus D_{[K]}}, Q, \mathbf{D}) \tag{49}$$
$$= \mathrm{H}(\tilde{Y}_Q|\mathbf{X}_{S^c,Q\oplus D_{S^c}}, Q, \mathbf{D}) - \mathrm{H}(\tilde{Y}_Q|\mathbf{X}_{[K],Q\oplus D_{[K]}}) \tag{50}$$
$$\le \mathrm{H}(\tilde{Y}_Q|\mathbf{X}_{S^c,Q\oplus D_{S^c}}) - \mathrm{H}(\tilde{Y}_Q|\mathbf{X}_{[K],Q\oplus D_{[K]}}) \tag{51}$$
$$= \mathrm{I}(\mathbf{X}_{S,Q\oplus D_S} \wedge \tilde{Y}_Q|\mathbf{X}_{S^c,Q\oplus D_{S^c}}) \tag{52}$$

[4]The stronger result of Section 3 of [8] is not needed in this paper.

where (50) comes from the fact that the output depends only on the input variables.

From the fact that the delays are independent and uniform it follows that the random variables $\{Q \oplus D_i, i \in [K]\}$ are independent, hence the random variables $\{X_{i,Q\oplus D_i}, i \in [K]\}$ are also independent. On account of this, the converse statement follows from Theorem 3.1.

The achievability part needs one modification of the proof in the Appendix of [8] of the assertion that C is achievable with informed $L = 2K - 1$-block decoder, considering maximal error. In order to get rid of the assumption that the delays are known to the receiver, it is enough to use the synchronization method from Subsection 4.1 in the successive steps of achievability of vertices. Note that it is important that in the successive steps the decoder finds the exact delay of the actual sender (see Remark 4.2). □

Remark 4.11. In case of two senders Corollary 3.3 leads to a stronger result. If the relative delay $D = D_1 \ominus D_2$ is uniformly distributed on the set $\{0, 1, \dots, n-1\}$ then, for each model version, the capacity region is C.

5. EVEN DELAYS

In [7] an artificial but interesting (from theoretical point of view) model is mentioned as open problem: the possible delays are in the set $\{0, 1, \dots, \alpha n\}$ for some $\alpha \in (0, 1)$. In this section, though this problem is not solved, a similar artificial model is analyzed which also has theoretical interest.

Theorem 5.1. In the even delays case (Example 2.6), for each version of the model the capacity region consists of those rate pairs that either belong to C or are linear combinations with weights $\frac{1}{2}, \frac{1}{2}$ of points in C. Moreover, using coding/decoding systems of odd block-length, only C can be achieved.

Proof. In order to prove the direct part it is enough to restrict attention to uninformed L-block decoder and to maximal error. Let n be even. Then the senders can do time sharing with weights $\frac{1}{2}, \frac{1}{2}$ using separately the even and the odd symbols and using the coding/decoding method of Theorem 4.10. Figure 3 demonstrates this fact. Note that in this case L can be chosen as 3.

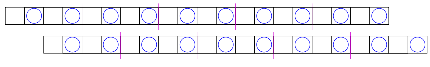

Fig. 3. Time sharing when the relative delay is uniform on even numbers.

In the converse part it is enough to treat the case of an informed infinite decoder and average error. The proof uses Corollary 3.3.

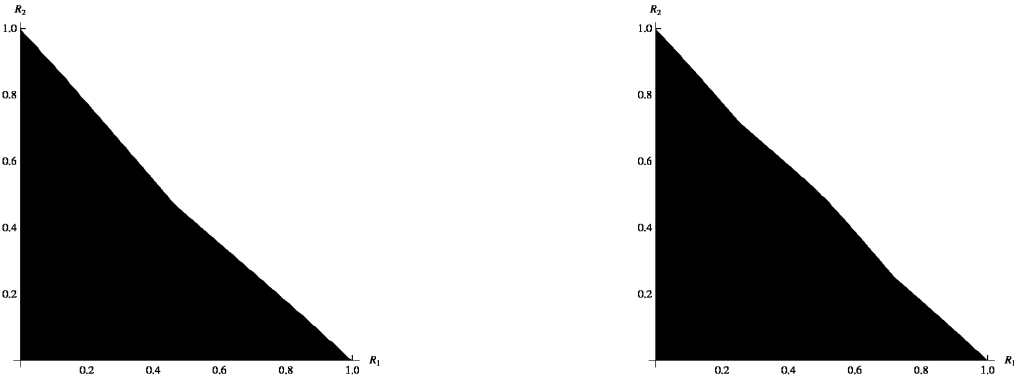

Fig. 4. Capacity region in the totally asynchronous and in the even delays case.

In case of coding/decoding systems of even length the relative delay is uniformly distributed on the even numbers in $\{0, 1, \ldots, n-1\}$. Write the upper bounds in Corollary 3.3 as a sum for the possible values of Q and define two random variables Q_1, Q_2 as uniform on even/odd numbers and independent of each other and everything else. Then the following can be written:

$$R_1 \leq \mathrm{I}(X_{1,Q} \wedge \hat{Y}_Q | X_{2,Q\ominus D}, Q, D) + \varepsilon_n \tag{53}$$

$$= \frac{1}{n} \sum_{i=0}^{n-1} \mathrm{I}(X_{1,i} \wedge \hat{Y}_i | X_{2,i\ominus D}, D) + \varepsilon_n \tag{54}$$

$$\leq \frac{1}{2}\frac{2}{n} \sum_{i\in\mathrm{odd}} \mathrm{I}(X_{1,i} \wedge \hat{Y}_i | X_{2,i\ominus D}) + \frac{1}{2}\frac{2}{n} \sum_{i\in\mathrm{even}} \mathrm{I}(X_{1,i} \wedge \hat{Y}_i | X_{2,i\ominus D}) + \varepsilon_n \tag{55}$$

$$\leq \frac{1}{2} \mathrm{I}(X_{1,Q_1} \wedge \hat{Y}_{Q_1} | X_{2,Q_1\ominus D}) + \frac{1}{2} \mathrm{I}(X_{1,Q_2} \wedge \hat{Y}_{Q_2} | X_{2,Q_2\ominus D}) + \varepsilon_n. \tag{56}$$

Similarly we get

$$R_2 \leq \frac{1}{2} \mathrm{I}(X_{2,Q_1\ominus D} \wedge \hat{Y}_{Q_1} | X_{1,Q_1}) + \frac{1}{2} \mathrm{I}(X_{2,Q_2\ominus D} \wedge \hat{Y}_{Q_2} | X_{1,Q_2}) + \varepsilon_n \tag{57}$$

$$R_1 + R_2 \leq \frac{1}{2} \mathrm{I}(X_{1,Q_1}, X_{2,Q_1\ominus D} \wedge \hat{Y}_{Q_1}) + \frac{1}{2} \mathrm{I}(X_{1,Q_2}, X_{2,Q_2\ominus D} \wedge \hat{Y}_{Q_2}) + \varepsilon_n \tag{58}$$

where X_{1,Q_1},$X_{2,Q_1\ominus D}$ and X_{1,Q_2},$X_{2,Q_2\ominus D}$ are independent. This proves the converse result for even blocklength (see [4] Lemma 14.4+, or its generalization, Lemma 6.3 in Section 6 of this paper).

In the subsequent part of this proof the symbol n denotes odd integer. Now we prove that with coding/decoding systems of odd length, just the union of the pentagons can be achieved. Given such sequence of coding/decoding systems, where $P_e^n \to 0$, let c_n be a sequence with $c_n \to 0$ and $\frac{P_e^n}{c_n} \to 0$ such that $c_n n$ is integer.

Recall that the delays $D_1(n)$ and $D_2(n)$ are independent and uniformly distributed random variables on the set $\{0, 2, \ldots, n-1\}$. For all $i \in \{0, 1, \ldots, n-1\}$ let $K(i)$ be the number of those pairs $d_1, d_2 \in \{0, 2, \ldots, n-1\}$ for which the relative delay $d = d_1 \ominus d_2$ is equal to i, then:

$$K(i) = \begin{cases} \frac{n-i+1}{2} & \text{if } i \text{ is even} \\ \frac{i+1}{2} & \text{if } i \text{ is odd.} \end{cases}$$

Let $D_1^{'}(n)$ and $D_2'(n)$ be two random variables taking values in the set $\{0, 2, \ldots, n-1\}$ with the following joint distribution. For each $d_1, d_2 \in \{0, 2, \ldots, n-1\}$, if $d_1 \ominus d_2 \in \{c_n n, c_n n+1, \ldots, n-1-c_n n\}$ let $\Pr\{D_1'(n) = d_1, D_2'(n) = d_2\}$ be equal to $\frac{1}{n(1-2c_n)K(d_1 \ominus d_2)}$, otherwise 0. Then for each $d_1, d_2 \in \{0, 2, \ldots, n-1\}$ the following bound holds if n is large enough:

$$\frac{4}{(n+1)^2} = \Pr\{D_1(n) = d_1, D_2(n) = d_2\} \geq c_n \Pr\{D_1'(n) = d_1, D_2'(n) = d_2\}. \tag{59}$$

Using the same idea as in the proof of Lemma 2.18 and the fact that $\frac{P_e^n}{c_n} \to 0$ we can conclude that the given sequence of coding/decoding systems has average error also tending to 0 under the delay system $\mathbf{D}'$ described by the random variables $D_1'(n)$ and $D_2'(n)$. Hence if we show that under delay system $\mathbf{D}'$ only C can be achieved, the assertion is proved.

Under delay system $\mathbf{D}'$ the relative delay $D'(n) = D_1'(n) \ominus D_2'(n)$ is uniformly distributed on the set $\{c_n n, c_n n+1, \ldots, n-1-c_n n\}$. By Corollary 3.3 the following bounds hold for the rates:

$$\begin{aligned} R_1 &\leq \mathrm{I}(X_{1,Q} \wedge \hat{Y}_Q | X_{2,Q\ominus D'}, Q, D') + \varepsilon_n \\ R_2 &\leq \mathrm{I}(X_{2,Q\ominus D'} \wedge \hat{Y}_Q | X_{1,Q}, Q, D') + \varepsilon_n \\ R_1 + R_2 &\leq \mathrm{I}(X_{1,Q}, X_{2,Q\ominus D'} \wedge \hat{Y}_Q | Q, D') + \varepsilon_n. \end{aligned} \tag{60}$$

Let $\bar{D}(n)$ be a random variable uniformly distributed on the set $\{0, 1, \ldots, n-1\}$. As the variation distance between the product joint distributions of (Q, D') and $(Q, \bar{D})$ tends to 0, the following differences also tend to 0 as $n \to \infty$:

$$\begin{aligned} &\mathrm{I}(X_{1,Q}, X_{2,Q\ominus D'} \wedge \hat{Y}_Q | Q, D') - \mathrm{I}(X_{1,Q}, X_{2,Q\ominus \bar{D}} \wedge \hat{Y}_Q | Q, \bar{D}), \\ &\mathrm{I}(X_{1,Q} \wedge \hat{Y}_Q | X_{2,Q\ominus D'}, Q, D') - \mathrm{I}(X_{1,Q} \wedge \hat{Y}_Q | X_{2,Q\ominus \bar{D}}, Q, \bar{D}), \\ &\mathrm{I}(X_{2,Q\ominus D'} \wedge \hat{Y}_Q | X_{1,Q}, Q, D') - \mathrm{I}(X_{2,Q\ominus \bar{D}} \wedge \hat{Y}_Q | X_{1,Q}, Q, \bar{D}). \end{aligned} \tag{61}$$

Taking into account that $X_{1,Q}$ and $X_{2,Q\ominus \bar{D}}$ are independent, the assertion is proved (See also Remark 4.11). □

Example 5.2. There are two well-known examples ([2, 4]) which show that the convex hull operation can be useful. Here we use [4]. Let the channel be defined by $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y} = \{0, 1\}$, $W(0|0, 0) = 1$, $W(1|1, 0) = W(1|0, 1) = 1$ and $W(1|1, 1) = W(0|1, 1) = \frac{1}{2}$. The capacity regions in the totally asynchronous and in the even delays case are shown on Figure 4. In the even delays case a hill appears in the middle of the picture.

Remark 5.3. Similar results can be achieved if the distribution is uniform on numbers which are divisible by 3. In this case time sharing with weights $\frac{1}{3}, \frac{2}{3}$ becomes possible.

Remark 5.4. This also means that if the senders of the totally asynchronous AMAC want to time share with weights $(\frac{1}{2}, \frac{1}{2})$, they can do that if a one-shot 1-bit side-information about the delays is available to the senders.

6. PARTLY ASYNCHRONOUS THREE-SENDERS CASE

In this section we will prove coding theorem in case of $K = 3$, when $D_1(n) = D_2(n)$ and $D_3(n)$ are independent and uniformly distributed on the set $\{0, 1, \ldots, n-1\}$.

Theorem 6.1. In the partly asynchronous three senders case (Example 2.7), for each version of the model the capacity region is

$$\bigcup_{p_{X_3}} Conv \left(\bigcup_{p_{X_1} \times p_{X_2}} \mathcal{R}\left[W; p_{X_1} \times p_{X_2} \times p_{X_3}\right] \right). \tag{62}$$

In words, it consists of the convex combination of rate triples from C whose corresponding convex polytopes are defined by the same third distribution.

Remark 6.2. Using the Carathéodory-Frenchel Theorem (e. g. Chapter 15 of [4]) in Theorem 6.1, it suffices to take convex combinations involving at most three rate triples.

Proof. [Converse part of Theorem 6.1]

It is enough to treat the case of an informed infinite decoder and average error. Theorem 3.1 can be used as follows.

If $S \subset \{1, 2, 3\}$ then the following bound holds:

$$R(S) \leq \sum_{i=1}^{n} \frac{1}{n} \mathrm{I}(\mathbf{X}_{S, i \oplus D_S} \wedge \tilde{Y}_i | \mathbf{X}_{S^c, i \oplus D_{S^c}}, \mathbf{D}) + \varepsilon_n. \tag{63}$$

Summing over the possible values of $D_1 = D_2$ we get the following bounds:

$$\begin{aligned}
R_1 &\leq \sum_{i=0}^{n-1} \sum_{d=0}^{n-1} \frac{1}{n} \frac{1}{n} \mathrm{I}(X_{1, i\oplus d} \wedge \tilde{Y}_i | X_{2, i\oplus d}, X_{3, i\oplus D_3}, D_3) + \varepsilon_n \\
R_2 &\leq \sum_{i=0}^{n-1} \sum_{d=0}^{n-1} \frac{1}{n} \frac{1}{n} \mathrm{I}(X_{2, i\oplus d} \wedge \tilde{Y}_i | X_{1, i\oplus d}, X_{3, i\oplus D_3}, D_3) + \varepsilon_n \\
R_3 &\leq \sum_{i=0}^{n-1} \sum_{d=0}^{n-1} \frac{1}{n} \frac{1}{n} \mathrm{I}(X_{3, i\oplus D_3} \wedge \tilde{Y}_i | X_{1, i\oplus d}, X_{2, i\oplus d}, D_3) + \varepsilon_n \\
R_1 + R_2 &\leq \sum_{i=0}^{n-1} \sum_{d=0}^{n-1} \frac{1}{n} \frac{1}{n} \mathrm{I}(X_{1, i\oplus d}, X_{2, i\oplus d} \wedge \tilde{Y}_i | X_{3, i\oplus D_3}, D_3) + \varepsilon_n \\
R_2 + R_3 &\leq \sum_{i=0}^{n-1} \sum_{d=0}^{n-1} \frac{1}{n} \frac{1}{n} \mathrm{I}(X_{2, i\oplus d}, X_{3, i\oplus D_3} \wedge \tilde{Y}_i | X_{1, i\oplus d}, D_3) + \varepsilon_n
\end{aligned}$$

$$R_1 + R_3 \le \sum_{i=0}^{n-1}\sum_{d=0}^{n-1} \frac{1}{n}\frac{1}{n}\,\mathrm{I}(X_{1,i\oplus d}, X_{3,i\oplus D_3} \wedge \tilde{Y}_i | X_{2,i\oplus d}, D_3) + \varepsilon_n$$

$$R_1 + R_2 + R_3 \le \sum_{i=0}^{n-1}\sum_{d=0}^{n-1} \frac{1}{n}\frac{1}{n}\,\mathrm{I}(X_{1,i\oplus d}, X_{2,i\oplus d}, X_{3,i\oplus D_3} \wedge \tilde{Y}_i | D_3) + \varepsilon_n. \tag{64}$$

Note that, $X_{3,i\oplus D_3}$ is independent of $X_{1,i\oplus d}$ and $X_{2,i\oplus d}$, and has the same distribution for all i. Note also that the above inequalities can be overestimated by dropping D_3 from the condition (same argument as in Theorem 4.10). Hence the converse part follows from Lemma 6.3 below. □

The achievability part in Theorem 6.1 is proved later in this section.

Lemma 6.3. Given k sets $\mathcal{R}\left[W; p_{X_1^i} \times p_{X_2^i} \times p_{X_3^i}\right]$, $i \in [k]$, a vector (R_1, R_2, R_3) equals a convex combination with weights α_i of k vectors from these sets if and only if they are contained in $\mathcal{R}(\alpha_1, \alpha_2, \dots, \alpha_k)$ which is defined by the following inequalities:

$$0 \le R_1 \le \sum_{i=1}^{k} \alpha_i\,\mathrm{I}(X_1^i \wedge Y^i | X_2^i, X_3^i)$$

$$0 \le R_2 \le \sum_{i=1}^{k} \alpha_i\,\mathrm{I}(X_2^i \wedge Y^i | X_1^i, X_3^i)$$

$$0 \le R_3 \le \sum_{i=1}^{k} \alpha_i\,\mathrm{I}(X_3^i \wedge Y^i | X_1^i, X_2^i)$$

$$R_1 + R_2 \le \sum_{i=1}^{k} \alpha_i\,\mathrm{I}(X_1^i, X_2^i \wedge Y^i | X_3^i)$$

$$R_1 + R_3 \le \sum_{i=1}^{k} \alpha_i\,\mathrm{I}(X_1^i, X_3^i \wedge Y^i | X_2^i)$$

$$R_2 + R_3 \le \sum_{i=1}^{k} \alpha_i\,\mathrm{I}(X_2^i, X_3^i \wedge Y^i | X_1^i)$$

$$R_1 + R_2 + R_3 \le \sum_{i=1}^{k} \alpha_i\,\mathrm{I}(X_1^i, X_2^i, X_3^i \wedge Y^i). \tag{65}$$

Proof. This proof follows the proof of Lemma 14.4+ in [4]. The sets $\mathcal{R}\left[W; p_{X_1^i} \times p_{X_2^i} \times p_{X_3^i}\right]$, $i \in [k]$, and the set $\mathcal{R}(\alpha_1, \alpha_2, \dots, \alpha_k)$ are convex polytopes with 16 vertices. Using the fact that the mutual and the conditional mutual information are always non-negative, it can be easily derived that there are no redundant inequalities between the defining equations of the sets $\mathcal{R}\left[W; p_{X_1^i} \times p_{X_2^i} \times p_{X_3^i}\right]$, $i \in [k]$, and $\mathcal{R}(\alpha_1, \alpha_2, \dots, \alpha_k)$. This means for example that it is not possible that the sum of the bounds for $R_1 + R_2$

and R_3 is strictly less then the bound for $R_1 + R_2 + R_3$. Using this fact the vertices of $\mathcal{R}\left[W; p_{X_1^i} \times p_{X_2^i} \times p_{X_3^i}\right]$, $i \in [k]$, can be written in the following way. First $\mathbf{v}_i^0 = (0, 0, 0)$ is a vertex. The remaining 15 vertices can be divided into three groups of equal size. The first group consists of those vertices (R_1, R_2, R_3) for which R_1 is equal to its own bound, i.e, of the vertices:

$$\begin{aligned}
\mathbf{v}_i^1 &= (\mathrm{I}(X_1^i \wedge Y^i | X_2^i, X_3^i), 0, 0) \\
\mathbf{v}_i^2 &= (\mathrm{I}(X_1^i \wedge Y^i | X_2^i, X_3^i), \mathrm{I}(X_2^i \wedge Y^i | X_3^i), 0) \\
\mathbf{v}_i^3 &= (\mathrm{I}(X_1^i \wedge Y^i | X_2^i, X_3^i), \mathrm{I}(X_2^i \wedge Y^i | X_3^i), \mathrm{I}(X_3^i \wedge Y^i)) \\
\mathbf{v}_i^4 &= (\mathrm{I}(X_1^i \wedge Y^i | X_2^i, X_3^i), 0, \mathrm{I}(X_3^i \wedge Y^i | X_2^i)) \\
\mathbf{v}_i^5 &= (\mathrm{I}(X_1^i \wedge Y^i | X_2^i, X_3^i), \mathrm{I}(X_2^i \wedge Y^i), \mathrm{I}(X_3^i \wedge Y^i | X_2^i))
\end{aligned} \tag{66}$$

The two other groups $(\mathbf{v}_i^6, \mathbf{v}_i^7, \ldots, \mathbf{v}_i^{10})$ and $(\mathbf{v}_i^{11}, \mathbf{v}_i^{12}, \ldots, \mathbf{v}_i^{15})$ are obtained similarly.

Note that $\mathcal{R}\left[W; p_{X_1^i} \times p_{X_2^i} \times p_{X_3^i}\right]$ can be degenerate in the sense that these sixteen vertices need not be all distinct. The vertices of $\mathcal{R}(\alpha_1, \alpha_2, \ldots, \alpha_k)$ are the points $\sum_{i=1}^k \alpha_i \mathbf{v}_i^j$, $0 \le j \le 15$. As these vertices are contained in the (convex) set of convex combinations with weights α_i of vectors in the sets $\mathcal{R}\left[W; p_{X_1^i} \times p_{X_2^i} \times p_{X_3^i}\right]$, $i \in [k]$, therefore whole $\mathcal{R}(\alpha_1, \alpha_2, \ldots, \alpha_k)$ is contained. The reverse inclusion is obvious. □

By Definition 4.7, the points (R_1, R_2, R_3) of $D(\mathcal{R}[W; p_{X_1} \times p_{X_2} \times p_{X_3}])$ satisfy the inequalities in (43), with $R_1 + R_2 + R_3 = I(X_1, X_2, X_3 \wedge Y)$. An edge of this dominant face is characterized by another inequality in (43) fulfilled with equality. The set S corresponding to that inequality will be called the type of this edge.

The following lemma states that rate triples lying on edges with a fixed type S behave similarly in context of rate splitting and successive decoding. It can be considered as a remark to the general theory of [8, 13] in the special case $K = 3$.

Lemma 6.4. For every fixed nonempty $S \subsetneq \{1, 2, 3\}$ there exists a 4-senders channel W' derived from W by splitting the first or the second sender, and an ordering $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)$ of the 4 senders with the following property. If W' is derived from W by splitting the first sender, then to any input distributions $p_{X_1}, p_{X_2}, p_{X_3}$ of W and for all $(R_1, R_2, R_3) \in D(\mathcal{R}[W; p_{X_1} \times p_{X_2} \times p_{X_3}])$ lying on the edge of type S, there exist input distributions $p_{X_{1a}}, p_{X_{1b}}$ and non-negative numbers R_{1a}, R_{1b} with $R_{1a} + R_{1b} = R_1$ such that $(R_{1a}, R_{1b}, R_2, R_3)$ is the vertex of $D(\mathcal{R}[W'; p_{X_{1a}} \times p_{X_{1b}} \times p_{X_2} \times p_{X_3}])$ described by ordering π. If W' is derived from W by splitting the second sender, then to any input distributions $p_{X_1}, p_{X_2}, p_{X_3}$ of W and for all $(R_1, R_2, R_3) \in D(\mathcal{R}[W; p_{X_1} \times p_{X_2} \times p_{X_3}])$ lying on the edge of dominant face with type S, there exist input distributions $p_{X_{2a}}, p_{X_{2b}}$ and non-negative numbers R_{2a}, R_{2b} with $R_{2a} + R_{2b} = R_2$ such that $(R_1, R_{2a}, R_{2b}, R_3)$ is the vertex of $D(\mathcal{R}[W'; p_{X_1} \times p_{X_{2a}} \times p_{X_{2b}} \times p_{X_3}])$ described by ordering π.

Proof. Assume for example that a rate triple $(R_1, R_2, R_3) \in D(\mathcal{R}[W; p_{X_1} \times p_{X_2} \times p_{X_3}])$ lies on the edge of type $S = \{1, 3\}$, hence $R_1 + R_2 + R_3 = I(X_1, X_2, X_3 \wedge Y)$, $R_1 + R_3 = I(X_1, X_3 \wedge Y | X_2)$. The other cases are similar. Then $R_2 = I(X_2 \wedge Y)$ and

(R_1, R_3) lies on $D(\mathcal{R}\left[\hat{W}; p_{X_1} \times p_{X_3}\right])$, where $\hat{W}(y, x_2|x_1, x_3) = p_{X_2}(x_2)W(y|x_1, x_2, x_3)$ (see [13], beginning of section 3c). Denote by $\hat{W}'$ the three senders channel derived from $\hat{W}$ by splitting the first sender. We could split the third sender instead of the first sender, but we want to leave the third sender unsplit. Using the basic rate splitting result of [8] for two senders channels, there exist input distributions $p_{X_{1a}}, p_{X_{1b}}$ and non-negative numbers R_{1a}, R_{1b} with $R_{1a} + R_{1b} = R_1$ such that (R_{1a}, R_{1b}, R_3) is the vertex of $D(\mathcal{R}\left[\hat{W}'; p_{X_{1a}} \times p_{X_{1b}} \times p_{X_3}\right])$ described by the ordering $(1a, 3, 1b)$. Hence, $(R_{1a}, R_{1b}, R_2, R_3)$ is the vertex of $D(\mathcal{R}\left[W'; p_{X_{1a}} \times p_{X_{1b}} \times p_{X_2} \times p_{X_3}\right])$ described by the ordering $(2, 1a, 3, 1b)$, where W' is the 4-senders channel derived from W by splitting the first sender. This argument shows that for $S = \{1, 3\}$, the channel W' derived from W by splitting the first sender, and the ordering $(2, 1a, 3, 1b)$ on the senders of W' fulfill the requirements of this lemma. □

The next lemma shows that each $\mathbf{r}$ which is not in C but can be written as the convex combination of rate triples from C, can be dominated by a convex combination of rate triples from C which lie on edges of same type.

Lemma 6.5. Given k sets $\mathcal{R}\left[W; p_{X_1^i} \times p_{X_2^i} \times p_{X_3^i}\right]$, $i \in [k]$, if a vector $\mathbf{r}$ is not in C, but can be written as $\mathbf{r} = \sum_{i=1}^k \alpha_i \mathbf{r}_i$, where $\mathbf{r}_i \in \mathcal{R}\left[W; p_{X_1^i} \times p_{X_2^i} \times p_{X_3^i}\right]$, $0 \le \alpha_i < 1$, $i \in [k]$, $\sum_{i=1}^k \alpha_i = 1$, then $\mathbf{r}$ can be dominated by an $\mathbf{r}'$ which can be written as $\sum_{i=1}^k \alpha_i' \mathbf{r}_i'$, where $\mathbf{r}_i' \in D(\mathcal{R}\left[W; p_{X_1^i} \times p_{X_2^i} \times p_{X_3^i}\right])$, $0 \le \alpha_i' < 1, i \in [k]$, $\sum_{i=1}^k \alpha_i' = 1$ and the vectors $\mathbf{r}_i', i \in [k]$, lie on edges of same type.

P r o o f. It can be assumed that $\alpha_i > 0$ for all i. If $\mathbf{r}_i$ is not on $D(\mathcal{R}\left[W; p_{X_1^i} \times p_{X_2^i} \times p_{X_3^i}\right])$ then we can take a dominating $\tilde{\mathbf{r}}_i$ from the dominant face, for all i. Then $\tilde{\mathbf{r}} = \sum_{i=1}^k \alpha_i \tilde{\mathbf{r}}_i$ dominates $\mathbf{r}$. So it can be assumed that the rate triple $\mathbf{r}_i$ is on $D(\mathcal{R}\left[W; p_{X_1^i} \times p_{X_2^i} \times p_{X_3^i}\right])$ for all i.

The dominant face of a set $\mathcal{R}\left[W; p_{X_1^i} \times p_{X_2^i} \times p_{X_3^i}\right]$ is a hexagon[5] on a plane with normal vector $(1, 1, 1)$. We say that the height of the plane with normal vector $(1, 1, 1)$ is a if its equation is $x + y + z = a$. The height of a dominant face is the height of its plane. As in Lemma 6.3 let us consider the set $\mathcal{R}(\alpha_1, \alpha_2, \ldots, \alpha_k)$. This is the set of convex combinations with weights $\alpha_i, 1 \le i \le k$ of the sets $\mathcal{R}\left[W; p_{X_1^i} \times p_{X_2^i} \times p_{X_3^i}\right], 1 \le i \le k$. The dominant face $\mathcal{D}(\alpha_1, \alpha_2, \ldots, \alpha_k)$ of $\mathcal{R}(\alpha_1, \alpha_2, \ldots, \alpha_k)$ consists of those points (R_1, R_2, R_3) for which $R_1 + R_2 + R_3 = \sum_{i=1}^k \alpha_i I(X_1^i, X_2^i, X_3^i \wedge Y^i)$. Note that $\mathbf{r} \in \mathcal{D}(\alpha_1, \alpha_2, \ldots, \alpha_k)$ because the points $\mathbf{r}_i$ are on the dominant face of $\mathcal{R}\left[W; p_{X_1^i} \times p_{X_2^i} \times p_{X_3^i}\right]$ respectively. Any edge of $\mathcal{D}(\alpha_1, \alpha_2, \ldots, \alpha_k)$ consists of those points for which one of the inequalities (65) is fulfilled with equality. Hence the edges of $\mathcal{D}(\alpha_1, \alpha_2, \ldots, \alpha_k)$ consist of points which are convex combinations with weights $\alpha_1, \alpha_2, \ldots, \alpha_k$ of points lying on edges of same type. If $\mathbf{r}$ is on an edge of $\mathcal{D}(\alpha_1, \alpha_2, \ldots, \alpha_k)$ then we proved the

[5] The hexagon can be degenerated since some vertices can be identical

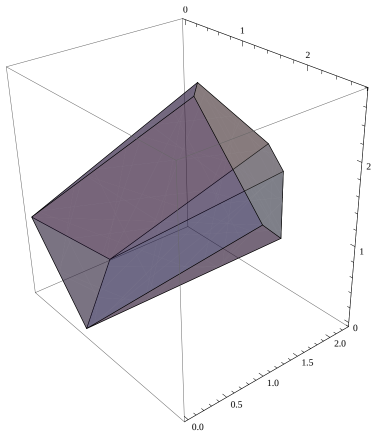

Fig. 5. The set of convex combination of two dominant faces. One of them is degenerate (triangle).

assertion. Hence it can be assumed that $\mathbf{r}$ is an inner point of $\mathcal{D}(\alpha_1, \alpha_2, \ldots, \alpha_k)$. Suppose first that there exists m, l such that $I(X_1^m, X_2^m, X_3^m \wedge Y^m) > I(X_1^l, X_2^l, X_3^l \wedge Y^l)$. Let us define new weights: If $i \neq m, i \neq l$ then let $\alpha_i^{'} = \alpha_i$, and let $\alpha_m^{'} = \alpha_m + \varepsilon$, $\alpha_l^{'} = \alpha_l - \varepsilon$. Then the height of $\mathcal{D}(\alpha_1^{'}, \alpha_2^{'}, \ldots, \alpha_k^{'})$ is larger than the height of $\mathcal{D}(\alpha_1, \alpha_2, \ldots, \alpha_k)$. If ε is small then one of the points of $\mathcal{D}(\alpha_1^{'}, \alpha_2^{'}, \ldots, \alpha_k^{'})$ will dominate $\mathbf{r}$. We increase ε until this property holds or until $\alpha_l^{'}$ becomes 0. Then, using continuity, an edge point of $\mathcal{D}(\alpha_1^{'}, \alpha_2^{'}, \ldots, \alpha_k^{'})$ will dominate $\mathbf{r}$ or $\alpha_l^{'} = 0$ holds. This argument shows that it is enough to restrict attention to the case when $I(X_1^m, X_2^m, X_3^m \wedge Y^m) = I(X_1^l, X_2^l, X_3^l \wedge Y^l)$ for all m, l. This means that the dominant faces of sets $\mathcal{R}\left[W; p_{X_1^i} \times p_{X_2^i} \times p_{X_3^i}\right]$ are in the same plane. Using again the continuous change of $\mathcal{D}(\alpha_1, \alpha_2, \ldots, \alpha_k)$: if $\alpha_i \to 1$, and $\alpha_j \to 0$ if $j \neq i$, then $\mathcal{D}(\alpha_1, \alpha_2, \ldots, \alpha_k)$ tends to $D(\mathcal{R}\left[W; p_{X_1^i} \times p_{X_2^i} \times p_{X_3^i}\right])$. As $\mathbf{r}$ is not in C, it is not in $D(\mathcal{R}\left[W; p_{X_1^i} \times p_{X_2^i} \times p_{X_3^i}\right])$ for either $i \in [k]$, hence there are weights $\alpha_1^*, \alpha_2^*, \ldots, \alpha_k^*$ for which $\mathbf{r}$ is on an edge of $\mathcal{D}(\alpha_1^*, \alpha_2^*, \ldots, \alpha_k^*)$. So it is a convex combination of points lying on edges of same type. □

Proof. [Achievability of Theorem 6.1]

In order to prove the direct part it is enough to restrict attention to uninformed L-block decoder and to maximal error.

Theorem 4.10 shows that the rate triples of C can be achieved in the totally asynchronous case considering uninformed $L = 5$-block decoder with maximal error. It follows that in the party asynchronous three senders case, C is also achievable considering uninformed $L = 5$-block decoder with maximal error with the same coding/decoding method[6].

[6] Actually $L = 3$-block decoder suffices in this partly asynchronous case because if a sender is split, then the delays of the two virtual senders are equal to the delay of the original sender and the third sender can remain unsplit.

Hence, using Remark 6.2, it is enough to consider points which are not in C but can be written as the convex combination of two or three rate triples from C whose corresponding sets $\mathcal{R}[W; p_{X_1} \times p_{X_2} \times p_{X_3}]$ have the same third distribution. Note that the following part of this proof shows that $L = 3$-block decoder suffices. For the sake of clarity we deal only with points which can be written as the convex combination of two rate triples whose corresponding sets $\mathcal{R}[W; p_{X_1} \times p_{X_2} \times p_{X_3}]$ have the same third distribution. The case of convex combination of three rate triples can be derived similarly.

Let (R_1, R_2, R_3) be in $\mathcal{R}[W; p_{X_1} \times p_{X_2} \times p_{X_3}]$ and $(\tilde{R}_1, \tilde{R}_2, \tilde{R}_3)$ in $\mathcal{R}\left[W; p_{\tilde{X}_1} \times p_{\tilde{X}_2} \times p_{X_3}\right]$. Note that the third input distribution is the same in case of both convex polytopes. We want to show that $\alpha(R_1, R_2, R_3) + (1-\alpha)(\tilde{R}_1, \tilde{R}_2, \tilde{R}_3)$ can be achieved in the partly asynchronous three senders case, $\alpha \in (0,1)$. Using Lemma 6.5 it can be assumed that (R_1, R_2, R_3) and $(\tilde{R}_1, \tilde{R}_2, \tilde{R}_3)$ lie on edges of same type of dominant faces $D(\mathcal{R}[W; p_{X_1} \times p_{X_2} \times p_{X_3}])$ and $D(\mathcal{R}\left[W; p_{\tilde{X}_1} \times p_{\tilde{X}_2} \times p_{X_3}\right])$ respectively. Without loss of generality it can be assumed that this common type is $S = \{1, 3\}$.

Let W' and π be the 4-senders channel and the ordering in Lemma 6.4 for $S = \{1, 3\}$. From the proof of Lemma 6.4 it can be seen that W' is the first sender split version of W and $\pi = (2, 1a, 3, 1b)$. As a consequence of Lemma 6.4 there exist R_{1a}, R_{2a} with $R_{1a} + R_{2a} = R_1$ and $\tilde{R}_{1a}, \tilde{R}_{2a}$ with $\tilde{R}_{1a} + \tilde{R}_{2a} = \tilde{R}_1$ and input distributions $p_{X_{1a}}, p_{X_{1b}}, p_{\tilde{X}_{1a}}, p_{\tilde{X}_{1b}}$ such that $(R_{1a}, R_{1b}, R_2, R_3)$ and $(\tilde{R}_{1a}, \tilde{R}_{1b}, \tilde{R}_2, \tilde{R}_3)$ are those vertices of $D(\mathcal{R}[W'; p_{X_{1a}} \times p_{X_{1b}} \times p_{X_2} \times p_{X_3}])$ and $D(\mathcal{R}\left[W'; p_{\tilde{X}_{1a}} \times p_{\tilde{X}_{1b}} \times p_{\tilde{X}_2} \times p_{X_3}\right])$ respectively, which can be described by ordering π.

If a sender is split, then the delays of the two virtual senders are equal to the delay of the original sender. Hence it is enough to prove that $\alpha(R_{1a}, R_{1b}, R_2, R_3) + (1 - \alpha)(\tilde{R}_{1a}, \tilde{R}_{1b}, \tilde{R}_2, \tilde{R}_3)$ can be achieved for channel W' when the delay system is the following: $D_{1a}(n) = D_{1b}(n) = D_2(n)$ and $D_3(n)$ are independent and uniformly distributed on the set $\{0, 1, \ldots, n-1\}$.

Note that the coordinates of the 4-tuple $(R_{1a}, R_{1b}, R_2, R_3)$ can be described as follows: $R_2 = I(X_2 \wedge Y)$, $R_{1a} = I(X_{1a} \wedge Y|X_2)$, $R_3 = I(X_3 \wedge Y|X_{1a}, X_2)$ and $R_{1b} = I(X_{1b} \wedge Y|X_{1a}, X_2, X_3)$, where the joint distribution of $(X_{1a},\ X_{1b},\ X_2,\ X_3,\ Y)$ is determined by the product input distribution $p_{X_{1a}} \times p_{X_{1b}} \times p_{X_2} \times p_{X_3}$ and the channel transition W'. Similarly the coordinates of the 4-tuple $(\tilde{R}_{1a}, \tilde{R}_{1b}, \tilde{R}_2, \tilde{R}_3)$ can be described by the equations: $\tilde{R}_2 = I(\tilde{X}_2 \wedge \tilde{Y})$, $\tilde{R}_{1a} = I(\tilde{X}_{1a} \wedge \tilde{Y}|\tilde{X}_2)$, $\tilde{R}_3 = I(X_3 \wedge \tilde{Y}|\tilde{X}_{1a}, \tilde{X}_2)$ and $\tilde{R}_{1b} = I(\tilde{X}_{1b} \wedge \tilde{Y}|\tilde{X}_{1a}, \tilde{X}_2, X_3)$, where the joint distribution of $(\tilde{X}_{1a}, \tilde{X}_{1b}, \tilde{X}_2, X_3, \tilde{Y})$ is determined by the product input distribution $p_{\tilde{X}_{1a}} \times p_{\tilde{X}_{1b}} \times p_{\tilde{X}_2} \times p_{X_3}$ and the channel transition W'.

Random coding argument is used, assuming without any loss of generality that αn and $(1-\alpha)n$ are integers. The symbols of the random codebooks are independent but not identically distributed random variables. The codewords of the virtual senders $1a$, $1b$, and sender 2 consist of two parts. The first αn symbols have distributions $p_{X_{1a}}$, $p_{X_{1b}}$ and p_{X_2} respectively, while the last $(1-\alpha)n$ symbols have distributions $p_{\tilde{X}_{1a}}$, $p_{\tilde{X}_{1b}}$ and $p_{\tilde{X}_2}$ respectively. The symbols of codewords of sender 3 are identically distributed according to the distribution p_{X_3}. We show that with this codebook structure it is possible to achieve the rate tuple $\alpha(R_{1a}, R_{1b}, R_2, R_3) + (1-\alpha)(\tilde{R}_{1a}, \tilde{R}_{1b}, \tilde{R}_2, \tilde{R}_3)$, by successive decoding with ordering $(2, 1a, 3, 1b)$ for channel W' if senders $1a$, $1b$, 2 are synchronized but sender 3 is not synchronized with them.

Note that we do not assume that the receiver knows the delays.

First the receiver decodes the codewords of sender 2. The situation is now more complicated than in case of identically distributed symbols. From the receiver's point of view the codewords of the second sender go through two different channels according to the different symbols of the codewords of the other senders. From the fact that the senders $1a$, $1b$, 2 are synchronized the receiver knows that the first αn consecutive symbols of codewords go through the channel

$$W^2(y|x_2) = \sum_{x_{1a}\in\mathcal{X}_1} \sum_{x_{1b}\in\mathcal{X}_1} \sum_{x_3\in\mathcal{X}_3} p_{X_{1a}}(x_{1a})p_{X_{1b}}(x_{1b})p_{X_3}(x_3)W'(y|x_{1a}, x_{1b}, x_2, x_3), \quad (67)$$

and the last $(1-\alpha)n$ consecutive symbols of codewords go through the channel

$$\tilde{W}^2(y|x_2) = \sum_{x_{1a}\in\mathcal{X}_1} \sum_{x_{1b}\in\mathcal{X}_1} \sum_{x_3\in\mathcal{X}_3} p_{\tilde{X}_{1a}}(x_{1a})p_{\tilde{X}_{1b}}(x_{1b})p_{X_3}(x_3)W'(y|x_{1a}, x_{1b}, x_2, x_3). \quad (68)$$

The decoder does the following. As in Theorem 4.1 the n tuples $(Y_{-n+1}, \dots Y_0)$, $\dots$, $(Y_0, \dots Y_{n-1})$ are examined. The receiver decodes the sth codeword as the 0th message of sender 2 if there exists an n tuple of examined output $(Y_{-n+i}, \dots Y_{i-1})$ such that the first αn symbols of the sth codewords are jointly typical with the first αn symbols of $(Y_{-n+i}, \dots Y_{i-1})$ and the same is true for the last $(1-\alpha)n$ symbols according to channels W^2 and $\tilde{W}^2$ respectively, and there are no other codewords with this property. With the shifted versions of this decoding technique the receiver also decodes the $-2, -1, 1, 2$th messages of sender 2 to ensure the decoding of the 0'th message of sender $1b$ in the last successive step. Note also that implicitly the receiver learns the delay of sender 2 (See Remark 4.2).

In the following successive step the receiver decodes the $-2, -1, 0, 1, 2$th codewords of sender $1a$ considering typicality according to the channels

$$W^{1a}(y, x_2|x_{1a}) = \sum_{x_{1b}\in\mathcal{X}_1} \sum_{x_3\in\mathcal{X}_3} p_{X_{1b}}(x_{1b})p_{X_2}(x_2)p_{X_3}(x_3)W'(y|x_{1a}, x_{1b}, x_2, x_3) \quad (69)$$

and

$$\tilde{W}^{1a}(y, x_2|x_{1a}) = \sum_{x_{1b}\in\mathcal{X}_1} \sum_{x_3\in\mathcal{X}_3} p_{\tilde{X}_{1b}}(x_{1b})p_{\tilde{X}_2}(x_2)p_{X_3}(x_3)W'(y|x_{1a}, x_{1b}, x_2, x_3). \quad (70)$$

In the third successive step the decoder deals with sender 3. Note that sender 3 is not synchronized with senders $1a$, $1b$, 2, hence using the -2,-1,0,1,2'th codewords of the senders 2 and $1a$ the receiver can decode (surely) just the $-1, 0, 1$th codewords of sender 3. It is also true in this case that the symbols of the codewords of sender 3 go through two different channels:

$$W^3(y, x_{1a}, x_2|x_3) = \sum_{x_{1b}\in\mathcal{X}_1} p_{X_{1a}}(x_{1a})p_{X_{1b}}(x_{1b})p_{X_2}(x_2)W'(y|x_{1a}, x_{1b}, x_2, x_3) \quad (71)$$

and

$$\tilde{W}^3(y, x_{1a}, x_2|x_3) = \sum_{x_{1b}\in\mathcal{X}_1} p_{\tilde{X}_{1a}}(x_{1a})p_{\tilde{X}_{1b}}(x_{1b})p_{\tilde{X}_2}(x_2)W'(y|x_{1a}, x_{1b}, x_2, x_3). \quad (72)$$

But there is an essential difference: due to the assumption on the delays it is not always true that the first part of the codewords goes through the channel W^3, it can be any αn consecutive symbols of the codewords. Here the word consecutive is understood modulo n. When the receiver is looking for typicality, n^2 joint typicality examinations are performed[7] according to the n possible positions of the separating line of the two possible channels and the possible codeword positions.

In the final successive step the receiver decodes the 0th codeword of sender $1b$ considering typicality according to the channels

$$W^{1b}(y, x_{1a}, x_2, x_3|x_{1b}) = p_{X_{1a}}(x_{1a})p_{X_2}(x_2)p_{X_3}(x_3)W'(y|x_{1a}, x_{1b}, x_2, x_3) \tag{73}$$

and

$$\tilde{W}^{1b}(y, x_{1a}, x_2, x_3|x_{1b}) = p_{\tilde{X}_{1a}}(x_{1a})p_{\tilde{X}_2}(x_2)p_{X_3}(x_3)W'(y|x_{1a}, x_{1b}, x_2, x_3). \tag{74}$$

Following the calculation method of Theorem 4.1 and [8, Appendix C], it can be seen that the rate tuple $\alpha(R_{1a}, R_{1b}, R_2, R_3) + (1-\alpha)(\tilde{R}_{1a}, \tilde{R}_{1b}, \tilde{R}_2, \tilde{R}_3)$ is achievable with this method. One part of the complete calculation can be found in Appendix C below. A full formal proof would be rather long, but for readers familiar with [8] it does not appear necessary. In particular, a genie added version of the model and the analysis of a larger error event are necessary. □

7. SUMMARY

This paper provides a general framework for asynchronous multiple access channels, in which the delays are random variables. Several model versions are treated, in the considered cases they yield the same capacity region; moreover, this region depends on the distribution of the delays only through its support (see Remark 2.14). A general (not single-letter) converse is established, which leads to single-letter converses in the considered special cases. This, together with achievability proofs relying on a combination of rate splitting and successive decoding with time sharing, leads to new capacity regions between the familiar ones for the synchronous and totally asynchronous cases. In particular, as the main result, a single letter characterization of the capacity region is obtained for channels with two synchronous senders and a not synchronous one. For further results we refer to [6].

8. APPENDIX A

The coding theorem for totally asynchronous MAC (with two senders) was first stated in [15], for the case when receiver knows the delays. The theorem was stated for maximal error but the converse actually proved even for average error. While the paper [15] is hard to read, with the help of reviewers of a previous version of this paper we have checked that the converse proof is correct, up to a minor gap pointed out after eq. (18) which is first filled here. The achievability proof in [15] is not addressed here since more accessible proofs have been published since then. In [11] the same capacity region

[7] Actually in the considered case ($S = \{1, 3\}$) n examinations are enough, but n^2 are needed in case of $S = \{1, 2\}$.

was claimed to be achievable also when the receiver was uninformed. However, in the delay-detection part of the proof in Appendix 1 of [11] there is a gap, in eq. (12b) an independence is assumed that need not hold when the examined n-block of channel input symbols consists of two parts, a codeword part and a sync sequence part. The other part of the proof addresses decoding when the delay is already known. This part is correct, giving rise to a valid achievability proof in the case of informed receiver. Another such proof was given in [8] via rate splitting and successive decoding, for any number of senders. Achievability in the uninformed receiver case has not been revisited until recently. The mentioned error in [11] was corrected in [7]. Our approach to delay detection differs from that of [11] and [7] not so much in not using a sync sequence but rather in our relying on a technique from [9] to bound the probability of delay detection error.

9. APPENDIX B – PROOF OF THEOREM 3.1.

Let $S \subset [K]$. We will derive a bound for $R(S)$. As in Section 3, take a window of the receiver consisting of $N+1$ n-length blocks $\mathbf{Y^{N+1}} = \{Y_0, Y_1, \dots, Y_{n(N+1)-1}\}$ and the codewords having index between 1 and N from all senders (they are fully covered by this window). Recall that $\mathbf{D}$ denotes the delay vector and $\mathbf{X}_{B,i+D_B}$ denotes the random vector with components $X_{l,i+D_l}$, $l \in B$ where $B \subset [K]$. Denote by $\mathbf{X}^{sw}$ the $2K$ input codewords which overlap with the beginning and end of $\mathbf{Y}^{N+1}$. Then

$$
\begin{aligned}
& NnR(S) && (75)\\
&= \mathrm{H}(\mathbf{M}_S^N) && (76)\\
&= \mathrm{I}(\mathbf{M}_S^N \wedge \hat{\mathbf{M}}_S^N) + \mathrm{H}(\mathbf{M}_S^N|\hat{\mathbf{M}}_S^N) && (77)\\
&\le \mathrm{I}(\mathbf{M}_S^N \wedge \hat{\mathbf{M}}_S^N) + Nn\varepsilon_n && (78)\\
&\le \mathrm{I}(\mathbf{X}_S^N \wedge \mathbf{Y}^{N+1}, \mathbf{X}^{sw}, \mathbf{D}) + Nn\varepsilon_n && (79)\\
&= \mathrm{H}(\mathbf{X}_S^N|\mathbf{D}) - \mathrm{H}(\mathbf{X}_S^N|\mathbf{Y}^{N+1}, \mathbf{D}) + \mathrm{H}(\mathbf{X}_S^N|\mathbf{Y}^{N+1}, \mathbf{D}) - \mathrm{H}(\mathbf{X}_S^N|\mathbf{X}^{sw}, \mathbf{Y}^{N+1}, \mathbf{D}) + Nn\varepsilon_n && (80)\\
&\le \mathrm{H}(\mathbf{X}_S^N|\mathbf{X}_{S^c}^N, \mathbf{D}) - \mathrm{H}(\mathbf{X}_S^N|\mathbf{Y}^{N+1}, \mathbf{X}_{S^c}^N, \mathbf{D}) + \mathrm{I}(\mathbf{X}^{sw} \wedge \mathbf{X}_S^N|\mathbf{Y}^{N+1}, \mathbf{D}) + Nn\varepsilon_n && (81)\\
&= \mathrm{I}(\mathbf{X}_S^N \wedge \mathbf{Y}^{N+1}|\mathbf{X}_{S^c}^N, \mathbf{D}) + Kn\log|\mathcal{X}| + Nn\varepsilon_n && (82)\\
&= \mathrm{H}(\mathbf{Y}^{N+1}|\mathbf{X}_{S^c}^N, \mathbf{D}) - \mathrm{H}(\mathbf{Y}^{N+1}|\mathbf{X}_S^N, \mathbf{X}_{S^c}^N, \mathbf{D}) + Kn\log|\mathcal{X}| + Nn\varepsilon_n && (83)\\
&= \mathrm{H}(\mathbf{Y}^{N+1}|\mathbf{X}_{S^c}^N, \mathbf{D}) + Kn\log|\mathcal{X}| + Nn\varepsilon_n - \sum_{j=0}^{N}\sum_{i=0}^{n-1} \mathrm{H}(Y_{nj+i}|\mathbf{Y}_1^{nj+i-1}, \mathbf{X}_S^N, \mathbf{X}_{S^c}^N, \mathbf{D}) && (84)\\
&\le \mathrm{H}(\mathbf{Y}^{N+1}|\mathbf{X}_{S^c}^N, \mathbf{D}) - \sum_{j=1}^{N-1}\sum_{i=0}^{n-1} \mathrm{H}(Y_{nj+i}|\mathbf{X}_{[K],nj+i+D_{[K]}}, \mathbf{D}) + Kn\log|\mathcal{X}| + Nn\varepsilon_n. && (85)
\end{aligned}
$$

Now introduce the a random variable $\tilde{Y}_i$ linked to the random variables $X_{1,i\oplus D_1}$,$X_{2,i\oplus D_2}$,

$\dots, X_{K,i\oplus D_K}$ by the channel W for all $i \in \{0, 1, \dots, n-1\}$. Then (85) is continued as

$$\leq \sum_{i=0}^{n-1} \Big[(N-1)\,\mathrm{H}(\tilde{Y}_i | \mathbf{X}_{S^c, i\oplus D_{S^c}}, \mathbf{D}) + \sum_{i=0}^{n-1} \mathrm{H}(Y_i)$$

$$+ \sum_{i=Nn}^{Nn+n-1} \mathrm{H}(Y_i) - (N-1)\,\mathrm{H}(\tilde{Y}_i | \mathbf{X}_{[K], i\oplus D_{[K]}}, \mathbf{D}) \Big] + Kn \log |\mathcal{X}| + Nn\varepsilon_n \tag{86}$$

$$\leq (N-1) \sum_{i=0}^{n-1} \mathrm{I}(\mathbf{X}_{S, i\oplus D_S} \wedge \tilde{Y}_Q | \mathbf{X}_{S^c, i\oplus D_{S^c}} | \mathbf{D}) + 2n \log |\mathcal{Y}| + Kn \log |\mathcal{X}| + Nn\varepsilon_n. \tag{87}$$

Dividing by Nn and going with N to infinity give

$$R(S) \leq \mathrm{I}(\mathbf{X}_{S, Q\oplus D_S} \wedge \tilde{Y}_Q | \mathbf{X}_{S^c, Q\oplus D_{S^c}}, Q, \mathbf{D}) + \varepsilon_n.$$

This proves Theorem 3.1.

10. APPENDIX C – SOME CALCULATIONS TO THEOREM 6.1.

Let us address the coding/decoding task of sender 3. The random codebook of the third sender consists of i.i.d symbols with distribution p_{X_3}. This codebook contains $2^{\left(n\alpha I(X_3 \wedge Y | X_{1a}, X_2) + (1-\alpha) I(X_3 \wedge \tilde{Y} | \tilde{X}_{1a}, \tilde{X}_2) - 2\delta\right)}$ codewords. αn consecutive symbols of an input codeword go through channel W^3, while $(1-\alpha)n$ consecutive symbols of the codeword go through channel $\tilde{W}^3$. Here 'consecutive' is understood modulo n. Let $T \subset \{0, \dots, n-1\}$ denote the set of indices when W^3 was used. T will be called separating pattern. Note that $|T| = \alpha n$ and T contains consecutive numbers. The separating pattern depends on the relative delay D between the synchronized senders $1a$, $1b$, 2 and the unsynchronized sender 3. The decoder sees an output flow (note that the symbols of senders $1a$, 2 are also the part of the output). The decoder checks the same output n tuples as the decoder of Section 2 when looking for joint typicality, but when it examines an output n tuple Y^n the decoder checks every possible separating pattern. We say that the s'th codeword is typical in window Y^n relative to separating pattern T if parts of the codewords consisting of the coordinates in T of $\mathbf{X}^n(s)$ and Y^n are jointly typical according to channel W^3, and the same holds for the coordinates in $T^c = \{0, \dots, n-1\} \setminus T$ according to channel $\tilde{W}^3$. If s is the only codeword which is typical in all the examined output windows relative to all possible separation patterns, then the decoder's estimation is s for the 0'th message. It is crucial that the number of joint typicality examinations is n^2 (polynomial in n).

The error analysis can be done similarly as in Section 4a. Let us consider first the case when the examined output window corresponds to a full codeword of sender 3, say of the r'th one. Then in this window the r'th codeword will be typical relative to the true separating pattern T with probability exponentially close to 1 by classical arguments. First we show that no other codewords will be typical in this window. Let $T^{'}$ be any separation pattern ($T^{'} = T$ is not excluded). We will estimate the probability that the $s \neq r$'th codeword will be typical in this window relative to $T^{'}$.

Let $P_{X_3}^{T^{'}}(x^n, y^n)$ be the joint distribution on $\mathcal{X}^{|T^{'}|} \times \mathcal{Y}^{|T^{'}|}$ induced by the $|T^{'}|$th power of p_{X_3} and by the memoryless channel W^3. Let $q_{X_3}^{T^{'}}$ be the marginal of $P_{X_3}^{T^{'}}$ on $\mathcal{Y}^{|T^{'}|}$.

Similarly let $\tilde{P}^{T'}_{X_3}(x^n, y^n)$ be the joint distribution on $\mathcal{X}^{|T'|} \times \mathcal{Y}^{|T'|}$ induced by the $|T'|$th power of p_{X_3} and by the memoryless channel $\tilde{W}^3$. Let $\tilde{q}^{T'}_{X_3}$ be the marginal of $\tilde{P}^{T'}_{X_3}$ on $\mathcal{Y}^{|T'|}$. Furthermore, if $\mathbf{x}$ is an n-length sequence, then $\mathbf{x}^{T'}$ will denote the vector of length $|T'|$ consisting of those coordinates of $\mathbf{x}$ which are in $|T'|$. We have

$$\mathrm{Pr}_{cond}\left\{(X_0(s), \dots, X_{n-1}(s), Y^n) \in S^\delta_n(T')\right\} \tag{88}$$

$$= \sum_{(\mathbf{x}^n(s), \mathbf{y}^n) \in S^\delta_n(T')} p^n_{X_3}(\mathbf{x}^n(s)) \mathrm{Pr}_{cond}\left\{Y^n = \mathbf{y}^n | (X_0(s), \dots, X_{n-1}(s)) = \mathbf{x}^n(s)\right\} \tag{89}$$

$$= \sum_{(\mathbf{x}^n(s), \mathbf{y}^n) \in S^\delta_n(T')} p^n_{X_3}(\mathbf{x}^n(s)) \frac{q^{T'}_{X_3}(\mathbf{y}^{T'}) \tilde{q}^{T'^c}_{X_3}(\mathbf{y}^{T'^c})}{q^{T'}_{X_3}(\mathbf{y}^{T'}) \tilde{q}^{T'^c}_{X_3}(\mathbf{y}^{T'^c})}$$

$$\cdot \mathrm{Pr}_{cond}\left\{Y^n = \mathbf{y}^n | (X_0(s), \dots, X_{n-1}(s)) = \mathbf{x}^n(s)\right\} \tag{90}$$

$$\leq \sum_{(\mathbf{x}^n(s), \mathbf{y}^n) \in S^\delta_n(T')} 2^{-n\alpha(I(X_3 \wedge Y | X_{1a}, X_2) - \delta)} 2^{-n(1-\alpha)\left(I(X_3 \wedge \tilde{Y} | \tilde{X}_{1a}, \tilde{X}_2) - \delta\right)}$$

$$\cdot \frac{P^{T'}_{X_3}(\mathbf{x}^{T'}(s), \mathbf{y}^{T'}) \tilde{P}^{T'^c}_{X_3}(\mathbf{x}^{T'^c}(s), \mathbf{y}^{T'^c})}{q^{T'}_{X_3}(\mathbf{y}^{T'}) \tilde{q}^{T'^c}_{X_3}(\mathbf{y}^{T'^c})} \cdot \mathrm{Pr}_{cond}\left\{Y^n = \mathbf{y}^n | (X_0(s), \dots, X_{n-1}(s)) = \mathbf{x}^n(s)\right\} \tag{91}$$

$$\leq 2^{-n\alpha(I(X_3 \wedge Y | X_{1a}, X_2) - \delta)} 2^{-n(1-\alpha)\left(I(X_3 \wedge \tilde{Y} | \tilde{X}_{1a}, \tilde{X}_2) - \delta\right)}$$

$$\cdot \sum_{(\mathbf{x}^n(s), \mathbf{y}^n) \in S^\delta_n(T')} P^{T'}_{X_3}(\mathbf{x}^{T'}(s) | \mathbf{y}^{T'}) \tilde{P}^{T'^c}_{X_3}(\mathbf{x}^{T'^c}(s) | \mathbf{y}^{T'^c})$$

$$\cdot \mathrm{Pr}_{cond}\left\{Y^n = \mathbf{y}^n | (X_0(s), \dots, X_{n-1}(s)) = \mathbf{x}^n(s)\right\}. \tag{92}$$

As $(X_0(s), \dots, X_{n-1}(s))$ is independent of Y^n (since $s \neq r$),the sum in (92) is bounded above by 1.

If the examined output window does not correspond to a full codeword but rather to parts of two codewords, say of the r'th and l'th codewords. We may assume that $r \neq l$ as in the proof of Theorem 4.1. Then the argument works if $s \neq l$ and $s \neq r$. If one of r and l is equal to s then Gray's summing technique works (as in (41), (42)).

ACKNOWLEDGEMENT

The preparation of this article would not have been possible without the support of Prof. Imre Csiszár. We would like to thank him for his help and advice within this subject area.

(Received October 8, 2013)

REFERENCES

[1] R. Ahlswede: Multi-way communication channels. In: Proc. 2nd International Symposium on Information Theory, Tsahkadsor, Armenian SSR (1971), Akadémiai Kiadó, Budapest, pp. 23–52.

[2] M. Bierbaum and H. M. Wallmeier: A note on the capacity region of the multi-access channel. IEEE Trans. Inform. Theory *25* (1979), 484.

[3] T. M. Cover, R. J. McEliece, and E. C. Posner: Asynchronous multiple-acces channel capacity. IEEE Trans. Inform. Theory *27* (1981), 409–413.

[4] I. Csiszár and J. Körner: Information theory, Coding theorems for Discrete Memoryless Systems Second edition. Cambridge University Press, Cambridge 2011.

[5] L. Farkas and T. Kói: Capacity region of discrete asynchronous multiple access channels. Int. Symp. Inform. Theory Proc. (ISIT) *19* (2011), 2273–2277.

[6] L. Farkas and T. Kói: Capacity regions of partly asynchronous multiple access channels. Int. Symp. Inform. Theory Proc. (ISIT) *20* (2012), 3018–3022.

[7] A. El Gamal and Y.-H. Kim: Network Information Theory. Cambridge University Press, Cambridge 2012

[8] A. J. Grant, B. Rimoldi, R. L. Urbanke, and P. A. Whiting: Rate-splitting multiple acces for discrete memoryless channels. IEEE Trans. Inform. Theory *47* (2001), 873–890.

[9] R. M. Gray: Sliding-block joint source/noisy-channel coding theorems. IEEE Trans. Inform. Theory *22* (1976), 682–690.

[10] S. Hanly and P. Whiting: Constraints on capacity in a multi-user channel. Int. Symp. Inform. Theory Proc. (ISIT) *4* (1994), 54.

[11] J. Y. N. Hui and P. A. Humblet: The capacity region of the totally asynchronous multiple-access channel. IEEE Trans. Inform. Theory *31* (1985), 207–216.

[12] H. Liao: Multiple Access Channels. Ph.D. Dissertation, Dept. Elec. Eng., Univ. Hawai, Honolulu 1972.

[13] B. Rimoldi: Generalized time sharing: A low-complexity capacity-achieving multiple-access technique. IEEE Trans. Inform. Theory *47* (2001), 2432–2442.

[14] D. Tse and S. Hanly: Multi-access fading channels – Part I: Polymatroid structure, optimal resource allocation and throughput capacities. IEEE Trans. Inform. Theory *44* (1998), 2796–2815.

[15] G. Sh. Poltyrev: Coding in an asynchronous multiple-access channel. Problemy Peredachi Informatsii *19* (1983), 12–21.

[16] Y. Polyanskiy: On asynchronous capacity and dispersion. In: 46th Annual Conference on Information Sciences and Systems (CISS) (2012), pp. 1–6.

[17] A. Tchamkerten, V. Chandar, and G. W. Wornell: Communication under strong asynchronism. IEEE Trans. Inform. Theory *55* (2009), 4508–4528.

[18] S. Verdu: Multiple-access channels with memory with and without frame synchronism. IEEE Trans. Inform. Theory *35* (1989), 605–619.

Lóránt Farkas, Budapest University of Technology and Economics, Department of Analysis, 1111 Egry József street 1, Budapest. Hungary.
e-mail: lfarkas@math.bme.hu

Tamás Kói, Budapest University of Technology and Economics, Department of Stochastics and MTA-BME Stochastics Research Group, 1111 Egry József street 1, Budapest. Hungary.
e-mail: koitomi@math.bme.hu