Achievable Rates and Exponents for Asynchronous Communication with ML Decoding

Seçkin Anıl Yıldırım[1], Alfonso Martinez[1], and Albert Guillén i Fàbregas[1,2,3]
[1]Universitat Pompeu Fabra, [2]ICREA, [3]University of Cambridge
seckin.yildirim@upf.edu, alfonso.martinez@ieee.org, guillen@ieee.org

***Abstract*—The asynchronous-communication model is studied by means of i.i.d. codes and ML decoding. A random-coding bound to the joint probability of decoding and synchronization error is determined and used to recover the region of achievable information rates and asynchrony exponents.**

I. Introduction and Main Results

Recently, Tchamkerten *et al.* [1] studied an information-theoretic model that combines synchronization and coding. The transmitter sends a message starting at a random time unknown to the receiver and a known dummy symbol when idle. The channel output length is assumed to scale exponentially with the codeword length at the rate of some asynchrony exponent. The authors investigated the tradeoff between the achievable rates and the asynchrony exponent. In previous work, Chase [2] studied a similar problem where the channel output length is proportional to the codeword length. Polyanskiy [3] found that the requirement to exactly locate the codeword does not reduce the largest achievable rate. Wang [4] studied the tradeoff between the missed detection and false-alarm exponents in the detection process. More recently, Weinberger and Merhav [5] provided the optimal tradeoff between these exponents and that of decoding error.

A common point of these works is the use of the constant-composition codes and sequential threshold decoders. In this paper, we study the problem from a complementary point of view. Firstly, instead of constant-composition codes, we consider a random-coding ensemble where all codeword symbols are independent and identically distributed (i.i.d). Secondly, rather than a sequential threshold decoder, we consider a maximum-likelihood (ML) decoder that observes the whole output sequence and makes a joint decision on the transmitted codeword and on its starting time. Our contribution is threefold: 1) we present upper bounds on the average error probability and lower bounds on the corresponding error exponent by adapting standard random-coding techniques [6] to the error events that appear in this joint decision problem; 2) we find the corresponding achievable rates and study the conditions under which infinite asynchrony exponents are permitted; and 3) our analysis does not require finite alphabets, and the results are directly applicable to continuous channels.

This work was funded in part by the European Research Council under ERC grant agreement 259663, by the European Union's 7th Framework Programme under grant agreement 303633 and by the Spanish Ministry of Economy and Competitiveness under grants RYC-2011-08150 and TEC2012-38800-C03-03.

II. Problem Formulation

We consider a source that selects a message m and a transmission time t which are respectively equiprobably distributed over the sets $\mathcal{M} = \{1, \dots, M\}$ and $\mathcal{T} = \{1, \dots, \nu+1\}$. The encoder maps a message m to a length-n codeword $\boldsymbol{c}^{(m)} = \big[c_1^{(m)}, \ c_2^{(m)}, \ \dots \ , c_n^{(m)}\big]$ consisting of letters $c_i^{(m)}$ from an input alphabet $\mathcal{X}$. The codebook contains M such codewords $\boldsymbol{c}^{(m)}$, $m = 1, \dots, M$. From the codeword $\boldsymbol{c}^{(m)}$ and the transmission time t, the encoder forms a channel input sequence $\boldsymbol{x}_{m,t}$ of length $n+\nu$ by appending dummy symbols $* \in \mathcal{X}$ before and after the length-n codeword $\boldsymbol{c}^{(m)}$ as

$$\boldsymbol{x}_{m,t} = \Big[\underbrace{*, \dots, *}_{t-1}, \ \underbrace{c_1^{(m)}, \dots, c_n^{(m)}}_{n}, \ \underbrace{*, \dots, *}_{\nu-t+1}\Big]. \qquad (1)$$

We let $\mathcal{S} = \big\{\boldsymbol{x}_{1,1}, \dots, \boldsymbol{x}_{M,\nu+1}\big\}$ denote the set of all possible channel input sequences.

We assume communication over a discrete memoryless channel (DMC) described by a probabilistic map $W(y|x)$ between the input and the channel output y. We denote the output alphabet by $\mathcal{Y}$ and the output sequence by $\boldsymbol{y}$. Since the channel is memoryless, the conditional distribution of an output sequence $\boldsymbol{y}_{i+1}^{i+\ell} = (y_{i+1}, \dots, y_{i+\ell})$ of length ℓ given the input sequence $\boldsymbol{x}_{i+1}^{i+\ell} = (x_{i+1}, \dots, x_{i+\ell})$ is given by

$$W^{\ell}(\boldsymbol{y}_{i+1}^{i+\ell}|\boldsymbol{x}_{i+1}^{i+\ell}) \triangleq \prod_{j=i+1}^{i+\ell} W(y_j|x_j). \qquad (2)$$

When $i = 0$, we drop both the subscript and superscript from the corresponding vector notation, e.g., $W^{n+\nu}(\boldsymbol{y}|\boldsymbol{x})$.

The ML decoder receives the channel output $\boldsymbol{y}$ of length $n+\nu$, and forms the estimate

$$(\hat{m}, \hat{t}) = \operatorname*{arg\,max}_{(k,l) \in \mathcal{M}\times\mathcal{T}} W^{n+\nu}(\boldsymbol{y}|\boldsymbol{x}_{k,l}), \qquad (3)$$

where ties are broken at random. An error occurs whenever $(\hat{m}, \hat{t}) \neq (m, t)$. The average error probability is defined by

$$P_e(\mathcal{C}) = \mathbb{P}\big((\hat{m}, \hat{t}) \neq (m, t)\big). \qquad (4)$$

A rate pair (R, A) is said to be *achievable* if there exists a sequence of codebooks $\mathcal{C}_n$ with $M = \lfloor e^{nR} \rfloor$ codewords whose average error probability satisfies $P_e(\mathcal{C}_n) \to 0$ as $n \to \infty$ for $\nu = \lfloor e^{nA} \rfloor$, where A is the asynchrony exponent. An error exponent $E(R, A)$ is said to be *achievable* if there exists a sequence of codebooks $\mathcal{C}_n$ such that the average error probability satisfies

$$E(R,A) \leq \liminf_{n\to\infty} -\frac{1}{n}\log P_e(\mathcal{C}_n). \tag{5}$$

III. An Upper Bound on the Error Probability

We consider an ensemble of codebooks each with M pairwise independently generated codewords of length n. In such an ensemble, the probability of a particular codebook is

$$P(\mathcal{C}) = \prod_{m=1}^{M}\prod_{i=1}^{n} Q\big(c_i^{(m)}\big), \tag{6}$$

where $Q(\cdot)$ is an arbitrary probability assignment on $\mathcal{X}$.

By assuming that ties are broken as errors, the average error probability averaged over the ensemble is bounded by

$$\overline{P}_e \leq \frac{1}{(\nu+1)M}\sum_{m,t}\overline{P}_{m,t}\ , \tag{7}$$

where, for $\mathcal{S}_{m,t} \triangleq \mathcal{S}\setminus\{(m,t)\}$, $\overline{P}_{m,t}$ is given by

$$\overline{P}_{m,t} = \mathbb{P}\Bigg(\bigcup_{(k,l)\in\mathcal{S}_{m,t}}\left\{\frac{W^{n+\nu}(\boldsymbol{Y}|\boldsymbol{X}_{k,l})}{W^{n+\nu}(\boldsymbol{Y}|\boldsymbol{X}_{m,t})}\geq 1\right\}\Bigg), \tag{8}$$

and the probability in (8) is over the joint distribution of the channel and the codebook.

We partition the set of channel input sequences $\mathcal{S}_{m,t}$ into several subsets $\mathcal{S}^i_{m,t}$ according to: 1) the degree of overlap between the codewords in the transmitted and alternative sequences; and 2) whether the messages in the transmitted and alternative sequences coincide or not. Such partitioning eases the probability of error analysis and allows to gain further insight on the interplay between the information rate and the asynchrony exponent. As each $\mathcal{S}^i_{m,t}$ induces a bound to the achievable exponent, the overall error probability is dominated by the lowest of these exponents. As we shall see, five subsets are sufficient for our purposes. We next describe these subsets.

1) Full overlap, different message: The set $\mathcal{S}^1_{m,t}$ contains the channel input sequences $\overline{\boldsymbol{x}}$ that have the transmission time $l=t$ for messages $k\neq m$,

$$\mathcal{S}^1_{m,t} = \Big\{(k,l)\in\mathcal{M}\times\mathcal{T} : k\neq m, l=t\Big\}. \tag{9}$$

Its cardinality is bounded as $|\mathcal{S}^1_{m,t}| = M-1 \leq M$.

2) No overlap, same message: The set $\mathcal{S}^2_{m,t}$ contains non-overlapping shifts of the transmitted sequence $\boldsymbol{x}$,

$$\mathcal{S}^2_{m,t} = \Big\{(k,l)\in\mathcal{M}\times\mathcal{T} : k=m, |l-t|\geq n\Big\}. \tag{10}$$

Let $\omega_t \triangleq \min(\nu+n, t+2n-2) - \max(t,n)$. Its cardinality is bounded as $|\mathcal{S}^2_{m,t}| = \nu-\omega_t \leq \nu$.

3) No overlap, different message: The set $\mathcal{S}^3_{m,t}$ contains non-overlapping shifts of the sequences $\overline{\boldsymbol{x}}$ of messages different from the transmitted one,

$$\mathcal{S}^3_{m,t} = \Big\{(k,l)\in\mathcal{M}\times\mathcal{T} : k\neq m, |l-t|\geq n\Big\}. \tag{11}$$

Its cardinality is bounded as $|\mathcal{S}^3_{m,t}| = (M-1)|\mathcal{S}^2_{m,t}| \leq M\nu$.

4) Partial overlap, same message: The set $\mathcal{S}^4_{m,t}$ contains the overlapping shifts of the transmitted sequence $\boldsymbol{x}$,

$$\mathcal{S}^4_{m,t} = \Big\{(k,l)\in\mathcal{M}\times\mathcal{T} : k=m, |l-t|<n, l\neq t\Big\}. \tag{12}$$

Its cardinality is bounded as $|\mathcal{S}^4_{m,t}| = \omega_t \leq 2n$.

The amount of overlap between $\boldsymbol{x}$ and $\overline{\boldsymbol{x}}$, which we denote by δ_n, is determined by l and t as $\delta_n = |l+n-t|$. Note that, as l varies over the range defined by (12), δ_n varies over the range 1 to $n-1$ twice for the cases $l<t$ and $l>t$.

5) Partial overlap, different message: The set $\mathcal{S}^5_{m,t}$ contains the overlapping shifts of the sequences $\overline{\boldsymbol{x}}$ corresponding to other messages than the transmitted one,

$$\mathcal{S}^5_{m,t} = \Big\{(k,l)\in\mathcal{M}\times\mathcal{T} : k\neq m, |l-t|<n, l\neq t\Big\}. \tag{13}$$

Its cardinality is bounded as $|\mathcal{S}^5_{m,t}| = (M-1)|\mathcal{S}^4_{m,t}| \leq 2Mn$. As for $\mathcal{S}^4_{m,t}$, we let δ_n denote the overlap between $\boldsymbol{x}$ and $\overline{\boldsymbol{x}}$.

A. Decomposition of Error Probabilities

Having defined the different subsets $\mathcal{S}^i_{m,t}$, we now apply the union bound to the union over elements in $\mathcal{S}_{m,t} = \bigcup_{i=1}^{5}\bigcup_{(k,l)\in\mathcal{S}^i_{m,t}}$ in (8) to obtain $\overline{P}_{m,t} \leq \sum_{i=1}^{5}\overline{P}^{(i)}_{m,t}$, where

$$\overline{P}^{(i)}_{m,t} \triangleq \mathbb{P}\Bigg(\bigcup_{(k,l)\in\mathcal{S}^i_{m,t}}\left\{\frac{W^{n+\nu}(\boldsymbol{Y}|\boldsymbol{X}_{k,l})}{W^{n+\nu}(\boldsymbol{Y}|\boldsymbol{X}_{m,t})}\geq 1\right\}\Bigg). \tag{14}$$

Due to space limitations, we merely sketch our error probability analysis within each of the subsets $\mathcal{S}^i_{m,t}$.

1) Error Probability Analysis for $\mathcal{S}^1_{m,t}$: The error events within $\mathcal{S}^1_{m,t}$ correspond to those of standard channel decoding [6, Ch. 5, pp. 135–138]. The average error probability is then bounded as

$$\overline{P}^{(1)}_{m,t} \leq M^\rho\,\mathbb{E}\left[\mathbb{E}\left[\left(\frac{W(Y|\overline{X})}{W(Y|X)}\right)^s\Big|X,Y\right]^\rho\right]^n. \tag{15}$$

2) Error Probability Analysis for $\mathcal{S}^2_{m,t}$: In these error events, the receiver confuses a noise sequence over the range t to $t+n-1$ with the transmitted codeword (i. e. false alarm) and the transmitted codeword with a noise sequence over the range l to $l+n-1$ (i. e. missed detection).

The analysis starts with the expression

$$\overline{P}^{(2)}_{m,t} = \mathbb{E}\Bigg[\mathbb{P}\Bigg(\bigcup_{(m,l)\in\mathcal{S}^2_{m,t}}\left\{\frac{W^{n+\nu}(\boldsymbol{Y}|\boldsymbol{X}_{m,l})}{W^{n+\nu}(\boldsymbol{Y}|\boldsymbol{X}_{m,t})}\geq 1\right\}\Big|\boldsymbol{X}_{m,t},\boldsymbol{Y}\Bigg)\Bigg]. \tag{16}$$

Then, we apply the bound $\mathbb{P}\left(\bigcup_l A_l\right) \leq \left(\sum_l \mathbb{P}(A_l)\right)^\rho$ for any $0\leq\rho\leq 1$, to (16), and apply Markov's inequality for $s\geq 0$ to the inner probability, to obtain

$$\overline{P}^{(2)}_{m,t} \leq \mathbb{E}\Bigg[\Bigg(\sum_{(m,l)\in\mathcal{S}^2_{m,t}}\mathbb{P}\Bigg(\frac{W^{n+\nu}(\boldsymbol{Y}|\boldsymbol{X}_{m,l})}{W^{n+\nu}(\boldsymbol{Y}|\boldsymbol{X}_{m,t})}\geq 1\Big|\boldsymbol{X}_{m,t},\boldsymbol{Y}\Bigg)\Bigg)^\rho\Bigg] \tag{17}$$

$$\leq \mathbb{E}\Bigg[\Bigg(\sum_{(m,l)\in\mathcal{S}^2_{m,t}}\bigg(\frac{W^{n+\nu}(\boldsymbol{Y}|\boldsymbol{X}_{m,l})}{W^{n+\nu}(\boldsymbol{Y}|\boldsymbol{X}_{m,t})}\bigg)^s\Bigg)^\rho\Bigg], \tag{18}$$

where we used that the probability distribution over $\boldsymbol{X}_{m,l}$ has mass one at the shifted transmitted sequence for the given l.

After substituting the likelihood ratio, which is given by

$$\frac{W^{n+\nu}(\boldsymbol{y}|\overline{\boldsymbol{x}})}{W^{n+\nu}(\boldsymbol{y}|\boldsymbol{x})} = \frac{W^n\big(\boldsymbol{y}_t^{t+n-1}|*^n\big)}{W^n\big(\boldsymbol{y}_t^{t+n-1}|\boldsymbol{x}_t^{t+n-1}\big)} \frac{W^n\big(\boldsymbol{y}_l^{l+n-1}|\boldsymbol{x}_t^{t+n-1}\big)}{W^n\big(\boldsymbol{y}_l^{l+n-1}|*^n\big)}, \tag{19}$$

where $*^n$ is a length-n string of dummy symbols $*$, and replacing $\boldsymbol{X}_t^{t+n-1}, \boldsymbol{Y}_t^{t+n-1}$ and $\boldsymbol{Y}_l^{l+n-1}$ by $\boldsymbol{X}, \boldsymbol{Y}$ and $\overline{\boldsymbol{Y}}$ respectively, the error probability in (18) becomes

$$\overline{P}_{m,t}^{(2)} \le \mathbb{E}\left[\left(\sum_{(m,l)\in\mathcal{S}_{m,t}^2}\left(\frac{W^n\big(\boldsymbol{Y}|*^n\big)}{W^n\big(\boldsymbol{Y}|\boldsymbol{X}\big)}\frac{W^n\big(\overline{\boldsymbol{Y}}|\boldsymbol{X}\big)}{W^n\big(\overline{\boldsymbol{Y}}|*^n\big)}\right)^s\right)^\rho\right]. \tag{20}$$

Conditioned on $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\overline{\boldsymbol{Y}}$, the likelihood ratio becomes a deterministic quantity which is not necessarily the same for all values of l such that $(m,l) \in \mathcal{S}_{m,t}^2$. We circumvent this problem by applying Jensen's inequality to (20) to get

$$\overline{P}_{m,t}^{(2)} \le \mathbb{E}\left[\left(\sum_{(m,l)\in\mathcal{S}_{m,t}^2}\mathbb{E}\left[\left(\frac{W^n\big(\boldsymbol{Y}|*^n\big)}{W^n\big(\boldsymbol{Y}|\boldsymbol{X}\big)}\frac{W^n\big(\overline{\boldsymbol{Y}}|\boldsymbol{X}\big)}{W^n\big(\overline{\boldsymbol{Y}}|*^n\big)}\right)^s\right]\right)^\rho\right] \tag{21}$$

$$\le |\mathcal{S}_{m,t}^2|^\rho\mathbb{E}\left[\left(\mathbb{E}\left[\left(\frac{W^n\big(\boldsymbol{Y}|*^n\big)}{W^n\big(\boldsymbol{Y}|\boldsymbol{X}\big)}\frac{W^n\big(\overline{\boldsymbol{Y}}|\boldsymbol{X}\big)}{W^n\big(\overline{\boldsymbol{Y}}|*^n\big)}\right)^s\right]\right)^\rho\right]. \tag{22}$$

where the inner expectation is over $\overline{\boldsymbol{Y}}$ and the outer expectation is over $\boldsymbol{X}$ and $\boldsymbol{Y}$ and in (22) we used that the inner expectation in (21) becomes independent of the index l.

Finally, by following standard arguments and bounding the cardinality as $|\mathcal{S}_{m,t}^2| \le \nu$, (22) can be factorized as

$$\overline{P}_{m,t}^{(2)} \le \nu^\rho\left(\mathbb{E}\left[\left(\mathbb{E}\left[\left(\frac{W(Y|*)}{W(Y|X)}\frac{W(\overline{Y}|X)}{W(\overline{Y}|*)}\right)^s\,\middle|\,X,Y\right]\right)^\rho\right]\right)^n. \tag{23}$$

3) Error Probability Analysis for $\mathcal{S}_{m,t}^3$: In these error events, the receiver simultaneously confuses the transmitted codeword with a noise sequence in the range t to $t+n-1$ (i. e. missed detection) and a noise sequence with another codeword than the transmitted one in the range l to $l+n-1$ (i. e. false alarm and decoding error).

The analysis is similar to that in (16)–(22) for $\mathcal{S}_{m,t}^2$. Thus, after substituting the likelihood ratio, which is given by

$$\frac{W^{n+\nu}(\boldsymbol{y}|\overline{\boldsymbol{x}})}{W^{n+\nu}(\boldsymbol{y}|\boldsymbol{x})} = \frac{W^n\big(\boldsymbol{y}_t^{t+n-1}|*^n\big)}{W^n\big(\boldsymbol{y}_t^{t+n-1}|\boldsymbol{x}_t^{t+n-1}\big)} \frac{W^n\big(\boldsymbol{y}_l^{l+n-1}|\overline{\boldsymbol{x}}_l^{l+n-1}\big)}{W^n\big(\boldsymbol{y}_l^{l+n-1}|*^n\big)}, \tag{24}$$

and redefining the strings of n letters in the range of the transmitted codeword, i.e. $\boldsymbol{X}_t^{t+n-1}$ and $\boldsymbol{Y}_t^{t+n-1}$, as $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively, and similarly the strings of n letters in the range of alternative codeword, i.e. $\overline{\boldsymbol{X}}_l^{l+n-1}$ and $\boldsymbol{Y}_l^{l+n-1}$, as $\overline{\boldsymbol{X}}$ and $\overline{\boldsymbol{Y}}$, respectively, we obtain

$$\overline{P}_{m,t}^{(3)} \le |\mathcal{S}_{m,t}^3|^\rho\mathbb{E}\left[\left(\mathbb{E}\left[\left(\frac{W^n(\boldsymbol{Y}|*^n)}{W^n(\boldsymbol{Y}|\boldsymbol{X})}\frac{W^n(\overline{\boldsymbol{Y}}|\overline{\boldsymbol{X}})}{W^n(\overline{\boldsymbol{Y}}|*^n)}\right)^s\right]\right)^\rho\right], \tag{25}$$

where the outer expectation in (25) is over $\boldsymbol{X}$ and $\boldsymbol{Y}$, the inner expectation is over $\overline{\boldsymbol{X}}$ and $\overline{\boldsymbol{Y}}$. Remark that, in contrast with the result for $\mathcal{S}_{m,t}^2$, the inner expectation is also performed over the alternative codeword $\overline{\boldsymbol{X}}$.

For a memoryless channel and i.i.d. codewords, (25) can be expressed in single-letter form by bounding the cardinality as

$$\overline{P}_{m,t}^{(3)} \le (M\nu)^\rho\left(\mathbb{E}\left[\left(\mathbb{E}\left[\left(\frac{W(Y|*)}{W(Y|X)}\frac{W(\overline{Y}|\overline{X})}{W(\overline{Y}|*)}\right)^s\,\middle|\,X,Y\right]\right)^\rho\right]\right)^n. \tag{26}$$

4) Error Probability Analysis for $\mathcal{S}_{m,t}^4$: The error events in $\mathcal{S}_{m,t}^4$ correspond to alternative sequences obtained by shifting the transmitted codeword to either direction by a position $n-\delta_n$ with an overlap of $\delta_n = 1, \dots, n-1$ positions.

We first apply the union bound to the union in (14), which gives a summation over δ_n as

$$\overline{P}_{m,t}^{(4)} \le 2\sum_{\delta_n=1}^{n-1}\mathbb{E}\left[\mathbb{P}\left(\frac{W^{n+\nu}(\boldsymbol{Y}|\boldsymbol{X}_{k,t+n-\delta_n})}{W^{n+\nu}(\boldsymbol{Y}|\boldsymbol{X}_{m,t})} \ge 1\,\middle|\,\boldsymbol{X}_{m,t},\boldsymbol{Y}\right)\right], \tag{27}$$

where the factor 2 comes from the fact that, as l varies over the range defined by (12), δ_n varies over the range 1 to $n-1$ twice for the cases $l < t$ and $l > t$.

Substituting the likelihood ratio in (28) and applying Markov's inequality with $s \ge 0$, we obtain (29) where $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$ denote the non-overlapping and overlapping portions of $\boldsymbol{X}_{m,t}$, $\boldsymbol{Y}$ and $\overline{\boldsymbol{Y}}$ denote the channel outputs corresponding to $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$, respectively; similarly, $\overline{\boldsymbol{X}}$ and $\overline{\overline{\boldsymbol{X}}}$ are related to the shifted codeword $\boldsymbol{X}_{m,l}$, and $\overline{\overline{\boldsymbol{Y}}}$ is the output corresponding to $\overline{\overline{\boldsymbol{X}}}$. The error probability is thus expressed as the product of three correlated factors which make further simplification difficult. Nevertheless, it does not depend on neither the number of codewords M nor the channel output length ν.

5) Error Probability Analysis for $\mathcal{S}_{m,t}^5$: The error events in $\mathcal{S}_{m,t}^5$ correspond to partially overlapping codewords for messages different from the transmitted one.

Following similar steps to those in the analysis of $\mathcal{S}_{m,t}^3$ and $\mathcal{S}_{m,t}^4$ above, we obtain (30) where $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$ denote the non-overlapping and overlapping portions of $\boldsymbol{X}_{m,t}$, $\boldsymbol{Y}$ and $\overline{\boldsymbol{Y}}$ denote the channel outputs corresponding to $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$, respectively; similarly, $\overline{\boldsymbol{X}}$ and $\overline{\overline{\boldsymbol{X}}}$ are related to $\boldsymbol{X}_{k,l}$, and $\overline{\overline{\boldsymbol{Y}}}$ is the output corresponding to $\overline{\overline{\boldsymbol{X}}}$. This way, the error probability can be expressed as the product of three independent factors since the codewords are pairwise independent. Moreover, the first and the third factors are reduced to a single expectation, similar to the expression in (25) that appeared in the analysis of $\mathcal{S}_{m,t}^3$ whereas the second factor keeps a double expectation, similar to the expression in (15) that appeared in the analysis

$$\frac{W^{n+\nu}\left(\boldsymbol{y}|\overline{\boldsymbol{x}}\right)}{W^{n+\nu}\left(\boldsymbol{y}|\boldsymbol{x}\right)} = \frac{W^{n-\delta_n}\left(\boldsymbol{y}_t^{t+n-\delta_n-1}|*^{n-\delta_n}\right)}{W^{n-\delta_n}\left(\boldsymbol{y}_t^{t+n-\delta_n-1}|\boldsymbol{x}_t^{t+n-\delta_n-1}\right)} \frac{W^{\delta_n}\left(\boldsymbol{y}_{t+n-\delta_n}^{t+n-1}|\overline{\boldsymbol{x}}_{t+n-\delta_n}^{t+n-1}\right)}{W^{\delta_n}\left(\boldsymbol{y}_{t+n-\delta_n}^{t+n-1}|\boldsymbol{x}_{t+n-\delta_n}^{t+n-1}\right)} \frac{W^{n-\delta_n}\left(\boldsymbol{y}_{t+n}^{t+2n-\delta_n-1}|\overline{\boldsymbol{x}}_{t+n}^{t+2n-\delta_n-1}\right)}{W^{n-\delta_n}\left(\boldsymbol{y}_{t+n}^{t+2n-\delta_n-1}|*^{n-\delta_n}\right)}. \tag{28}$$

$$\overline{P}^{(4)}_{m,t} \le 2\sum_{\delta_n=1}^{n-1}\mathbb{E}\left[\left(\frac{W^{n-\delta_n}\left(\boldsymbol{Y}|*^{n-\delta_n}\right)}{W^{n-\delta_n}\left(\boldsymbol{Y}|\boldsymbol{X}\right)}\frac{W^{\delta_n}\left(\overline{\boldsymbol{Y}}|\overline{\boldsymbol{X}}\right)}{W^{\delta_n}\left(\overline{\boldsymbol{Y}}|\tilde{\boldsymbol{X}}\right)}\frac{W^{n-\delta_n}\left(\overline{\overline{\boldsymbol{Y}}}|\overline{\overline{\boldsymbol{X}}}\right)}{W^{n-\delta_n}\left(\overline{\overline{\boldsymbol{Y}}}|*^{n-\delta_n}\right)}\right)^s\right]. \tag{29}$$

$$\overline{P}^{(5)}_{m,t} \le 2\,M^{\rho}\sum_{\delta_n=1}^{n-1}\mathbb{E}\left[\left(\mathbb{E}\left[\left(\frac{W^{n-\delta_n}\left(\boldsymbol{Y}|*^{n-\delta_n}\right)}{W^{n-\delta_n}\left(\boldsymbol{Y}|\boldsymbol{X}\right)}\frac{W^{\delta_n}\left(\overline{\boldsymbol{Y}}|\overline{\boldsymbol{X}}\right)}{W^{\delta_n}\left(\overline{\boldsymbol{Y}}|\tilde{\boldsymbol{X}}\right)}\frac{W^{n-\delta_n}\left(\overline{\overline{\boldsymbol{Y}}}|\overline{\overline{\boldsymbol{X}}}\right)}{W^{n-\delta_n}\left(\overline{\overline{\boldsymbol{Y}}}|*^{n-\delta_n}\right)}\right)^s \,\middle|\, \boldsymbol{X},\tilde{\boldsymbol{X}},\boldsymbol{Y},\overline{\boldsymbol{Y}}\right]\right)^{\rho}\right]. \tag{30}$$

---

of $\mathcal{S}^1_{m,t}$. Accordingly, Eq. (30) can be bounded as

$$\overline{P}^{(5)}_{m,t} \le \sum_{\delta_n=1}^{n-1}(2M)^{\rho}\,\mathbb{E}\left[\mathbb{E}\left[\left(\frac{W(Y|\overline{X})}{W(Y|X)}\right)^s \middle| X,Y\right]^{\rho}\right]^{\delta_n} \times \mathbb{E}\left[\mathbb{E}\left[\left(\frac{W(Y|*)}{W(Y|X)}\frac{W(\overline{Y}|\overline{X})}{W(\overline{Y}|*)}\right)^s \middle| X,Y\right]^{\rho}\right]^{n-\delta_n}. \tag{31}$$

B. Exponents and Achievable Rates

Next, we present the random coding exponents and achievable rates derived from the previous analysis. For a given distribution Q, we can write the bounds on the previously given error probabilities in exponential form,

$$\overline{P}^{(i)}_{m,t} \le e^{-nE_{r,i}(A,R)+o(n)}, \tag{32}$$

where $o(n)$ is a term that does not affect the exponential decay and the exponent $E_{r,i}(R,A)$ is given by

$$E_{r,i}(R,A) = \max_{0\le\rho\le1} E_{0,i}(\rho,Q) - \rho(a_iA + b_iR) \tag{33}$$

in terms of some Gallager-type functions $E_{0,i}(\rho,Q)$ and the binary parameters a_i and b_i that we define next for $i = 1,\dots,5$. In all cases, the optimum s is $s = \frac{1}{1+\rho}$ and the functions $E_{0,i}(\rho,Q)$ are concave in ρ.

From (15) (see also [6, Ch. 5, eq. (5.6.14)]), we have that $a_1 = 0$, $b_1 = 1$, and that the $E_{0,1}(\cdot)$ function is given by

$$E_{0,1}(\rho,Q) = -\log\left(\sum_y\left(\sum_x Q(x)W(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right). \tag{34}$$

Optimizing $E_{r,1}(R,A)$ over ρ provides an upper bound on the rate R in terms of the mutual information $I(Q;W)$ as

$$R < \lim_{\rho\to0}\frac{E_{0,1}(\rho,Q)}{\rho} = \left.\frac{\partial E_{0,1}(\rho,Q)}{\partial\rho}\right|_{\rho=0} = I(Q;W). \tag{35}$$

For (23), the binary parameters are $a_2 = 1$ and $b_2 = 0$ and the corresponding $E_{0,2}(\cdot)$ function can be given as

$$E_{0,2}(\rho,Q) = -\log\left(\sum_x Q(x)\Big(\sum_y W(y|*)^{\frac{\rho}{1+\rho}}W(y|x)^{\frac{1}{1+\rho}}\Big)^{1+\rho}\right). \tag{36}$$

Optimizing $E_{r,2}(R,A)$ over the parameter ρ yields an upper bound on the asynchrony rate A as

$$A < \lim_{\rho\to0}\frac{E_{0,2}(\rho,Q)}{\rho}. \tag{37}$$

If $E_{0,2}(0,Q) = 0$, i.e. there is no output y that is unreachable from the dummy symbol $*$ yet reachable from some input x, the bound of the asynchrony rate A can be expressed in terms of a conditional divergence

$$A < \left.\frac{\partial E_{0,2}(\rho,Q)}{\partial\rho}\right|_{\rho=0} = D(W_x\|W_*|Q), \tag{38}$$

where W_x and W_* stand for $W(y|x)$ and $W(y|*)$, respectively. An alternative expression of the divergence is $D(W_x\|W_*|Q) = I(Q;W) + D(QW\|W_*)$. Otherwise, if $E_{0,2}(0,Q) > 0$, we have no constraints, i. e. $A < \infty$. Note that $D(W_x\|W_*|Q) = \infty$ in this case too.

For (26), the corresponding binary parameters are $a_3 = b_3 = 1$ and the $E_{0,3}(\cdot)$ function can be given as

$$E_{0,3}(\rho,Q) = -\log\left(\sum_y\sum_x Q(x)W(y|*)^{\frac{\rho}{1+\rho}}W(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}. \tag{39}$$

Optimizing $E_{r,3}(R,A)$ over ρ yields an upper bound on the asynchronous exponent A and the information rate R as

$$R + A < \lim_{\rho\to0}\frac{E_{0,3}(\rho,Q)}{\rho}. \tag{40}$$

If $E_{0,3}(0,Q) = 0$, which again happens if there is no output y which is unreachable from the dummy symbol $*$ yet reachable from some other input x, the bound becomes

$$R + A < \left.\frac{\partial E_{0,3}(\rho,Q)}{\partial\rho}\right|_{\rho=0} = D(W_x\|W_*|Q). \tag{41}$$

Finally, it is possible to prove that this constraint dominates over the one arising from the $E_{0,2}(\cdot)$ function, so we may safely ignore the former.

The function $E_{0,4}(\cdot)$ does not depend on ρ, only on Q and s. Although we have not found a simple single-letter expression, we conjecture that the best possible choice is $s = \frac{1}{2}$ and that

$E_{0,4}(\cdot)$ lies between the values of $E_{0,1}(1,Q)$ and $E_{0,2}(1,Q)$. To any extent, as the bound (29) does not depend on neither the number of codewords M ($b_4=0$) nor the channel output length ν ($a_4=0$), this case does not impose any constraint on the achievable rates and, if the conjecture is true, it does not impose any constraints on the exponents neither.

Finally, from (31), we conclude that the binary parameters are $a_5=1$ and $b_5=0$ and, after summing the geometric series over δ_n, that the $E_{0,5}(\cdot)$ function can be given as

$$E_{0,5}(\rho,Q)=\min\big\{E_{0,1}(\rho,Q),E_{0,3}(\rho,Q)\big\}. \tag{42}$$

It is possible to prove that optimizing $E_{r,5}(R,A)$ over ρ yields an upper bound on the information rate R as

$$R<\lim_{\rho\to 0}\frac{E_{0,5}(\rho,Q)}{\rho}=I(Q;W), \tag{43}$$

IV. Example: Z Channel

This section illustrates the application of our bounds to a Z-channel with crossover probability ε. The input and output alphabets are $\mathcal{X}=\mathcal{Y}=\{0,1\}$. The probability of transmitting $X=0$ is $q\triangleq Q(0)$. We also consider two possible assignments of the dummy symbol, namely $*=0$ and $*=1$. Following the earlier discussion, the only relevant constraints to the rate pair region come from the first and third sets. Therefore, the corresponding two E_0 functions suffice to characterize the rate region for the Z-channel.

For $*=0$, we have that

$$E_{0,1}(\rho,Q) = -\log\Big(\big(q+(1-q)\varepsilon^{\frac{1}{1+\rho}}\big)^{1+\rho}+(1-q)^{1+\rho}(1-\varepsilon)\Big). \tag{44}$$

$$E_{0,3}(\rho,Q)=-\log\big(q+(1-q)\varepsilon^{\frac{1}{1+\rho}}\big)^{1+\rho}. \tag{45}$$

It can be easily verified that $E_{0,3}(0,Q)>0$. According to (35) and (40) the bounds on the achievable rates are

$$R<H\big((1-q)(1-\varepsilon)\big)-(1-q)H(\varepsilon) \tag{46}$$

$$A+R<\infty, \tag{47}$$

where $H(p)=-p\log(p)-(1-p)\log(1-p)$ is the binary entropy function.

For $*=1$, $E_{0,1}(\rho,Q)$ is still given by (44), as this function does not depend on the assignment of the dummy symbol. The remaining Gallager-type function is given by

$$E_{0,3}(\rho,Q)=-\log\Big(1-q+q\varepsilon^{\frac{\rho}{1+\rho}}\Big)^{1+\rho} \tag{48}$$

In this case, $E_{0,3}(0,Q)=0$. Therefore, according to (35) and (41) the bounds on the achievable rates are

$$R<H\big((1-q)(1-\varepsilon)\big)-(1-q)H(\varepsilon) \tag{49}$$

$$A+R<-q\log\varepsilon. \tag{50}$$

This simple example illustrates the importance of the condition that there is no output y which is unreachable from the dummy symbol $*$ yet reachable from some other input x. For $*=0$, arbitrarily large values of A are allowed, whereas for $*=1$, we have a constraint $A+R<-q\log\varepsilon$.

V. Discussion

We have derived a random-coding bound to the joint probability of decoding and synchronization error in asynchronous communication and used it to recover the region of achievable information rates and asynchrony exponents.

As defined in the present work, various exponents appear in (33). These exponents are related to a partitioning of the set of possible error events into several subsets. As each subset provides an achievable exponent, the overall error probability behaviour is determined by the minimum of these exponents. Moreover, these subsets serve to illustrate connections to different detection and coding problems. For instance, the decoding error exponent is determined by (34). The minimum of false alarm and missed detection exponents is determined by (36). In contrast to other detection problems, false alarm and missed detection happen simultaneously since the ML decoder always outputs an estimate of a codeword and its starting time.

References

[1] A. Tchamkerten, V. Chandar, and G. Wornell, "Communication under strong asynchronism," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4508–4528, Oct. 2009.

[2] D. Chase, "Coding theorems for the nonsynchronized channel," *IEEE Trans. Inf. Theory*, vol. 16, no. 1, pp. 55–73, Jan. 1970.

[3] Y. Polyanskiy, "Asynchronous communication: Exact synchronization, universality, and dispersion," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1256–1270, Mar. 2013.

[4] D. Wang, "Distinguishing codes from noise: fundamental limits and applications to sparse communication," Master's thesis, Dept. Elec. Eng. Comput. Sci., MIT, Cambridge, MA, Jun. 2010.

[5] N. Weinberger and N. Merhav, "Codeword or noise? Exact random coding exponents for joint detection and decoding," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5077–5094, Sep. 2014.

[6] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley and Sons, Inc., 1968.

[7] A. Tchamkerten, V. Chandar, and G. Wornell, "Asynchronous communication: Capacity bounds and suboptimality of training," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1227–1255, Mar. 2013.