

On asynchronous capacity and dispersion

(Invited paper)

Yury Polyanskiy

Abstract—Recently Tchamkerten et al. proposed a mathematical formulation of the problem of joint synchronization and error-correction in noisy channels. A variation of their formulation in this paper considers a strengthened requirement that the decoder estimate both the message and the location of the codeword exactly. It is shown that the capacity region remains unchanged and that the strong converse holds. The finite blocklength regime is investigated and it is demonstrated that even for moderate blocklengths, it is possible to construct capacity-achieving codes that tolerate exponential level of asynchronism and experience only a rather small loss in rate compared to the perfectly synchronized setting; in particular, the channel dispersion does not suffer any degradation due to asynchronism.

Index Terms—Shannon theory, channel capacity, channel coding, asynchronous communication, synchronization, strong converse, non-asymptotic analysis, finite blocklength, discrete memoryless channels

I. INTRODUCTION

The traditional approach to the problem of reliable communication in the presence of noise typically assumes that the decoder has access to a corrupted version of the original waveform with the beginning and the end of the waveform being perfectly known. In such setting modern sparse graph codes achieve almost the best possible error correction and continue to improve. It is natural, therefore, to reconsider other sources of suboptimality in a communication system. Namely, notice that the problem of synchronization is typically solved via an additional frontend (or layer) which employs special prefixes, suffixes and other methods, consuming both the energy and the bandwidth. In this paper we discuss the benefits of performing error-correction and synchronization jointly.

Recently, motivated in part by the sensor networks in which nodes exchange data very infrequently (thus, making constant channel-tracking impractical), Tchamkerten et al [1] formulated the problem in an elegant way and later demonstrated [2] that there are indeed significant advantages in going beyond the conventional synchronization approach.

Mathematically, the formulation of [1] is a generalization of the change point detection problem [3], close in spirit to the so called “detection and isolation” problem introduced in [4], except that in the latter the set of distributions that the original one can switch to is pre-specified whereas [1] allows for an optimal codebook design. Such formulation is quite different from the classical modeling of asynchronism in point-to-point channels as random insertion-deletion [5], [6] or randomly shifting back-to-back codewords [7]. In the context

of multiple-access channels, treatments of both the frame-asynchronism [8]–[10] and symbol-asynchronism [11] focused on the case when the relative time offsets between the users are perfectly known at the decoder (or remain constant across multiple transmissions, which makes them reliably learnable at the decoder). The problem addressed here, therefore, is subsidiary to both of these traditional approaches. We refer the reader to [1, Section II] for further background on the history and motivation of the synchronization problem.

In this paper we consider a variation of the setup of [1], [2]. In particular, we define the rate as the number of data bits k divided by the time n the codeword occupies the channel (as opposed to [1] that defines the rate as the ratio of k and the time it takes the decoder to react to transmission). The definition of rate as in this paper has also been considered in the context of asynchronous communication in [12] and [13]. Unlike that setup, however, we require the decoder to output the message block immediately after the actual transmission terminates or otherwise the error is declared. This requirement is natural since most systems would employ some sort of acknowledgment (Ack) feedback and hence, the transmitter will retransmit the message if the decoder is not sending an Ack signal in time.

In this variation we show that the capacity region is unchanged compared to the one in [12] (for the case when cost of each symbol is 1), we prove the strong converse (with and without the zero delay requirement) and investigate which of the results carry over to finite blocklength. In particular, we demonstrate that even for short blocklengths it is possible to combat a gigantic (exponential) asynchronism while achieving essentially the same performance as for the synchronous setting: namely, the channel dispersion [14] is unchanged. .

The organization of the paper is as follows. Section II defines the problem formally. Section III contains the asymptotic results on the capacity and the strong converse. Section IV presents a non-asymptotic achievability bound, evaluates it and draws conclusions on channel dispersion. With the exception of the non-asymptotic bound in Section IV the discussion focuses on discrete memoryless channels (DMCs).

II. PROBLEM FORMULATION AND NOTATION

Consider a DMC with stochastic matrix $W : \mathcal{X} \rightarrow \mathcal{Y}$ and a preselected symbol $\star \in \mathcal{X}$. We define its blocklength n extension as

$$W^n(y^n|x^n) = \prod_{j=1}^n W(y_j|x_j). \quad (1)$$

Given a number $A_n \geq n$ we define an asynchronous random transformation, denoted (W^n, A_n) , as follows:

- input space is \mathcal{X}^n
- output space is \mathcal{Y}^{A_n}
- the transformation acts as follows:

$$P_{Y^{A_n}|X^n}(\cdot|\cdot) = \sum_t P_\tau(t) P_{Y^{A_n}|X^n, \tau}(\cdot|\cdot, t),$$

where τ is a random variable uniformly distributed on $\{1, \dots, A_n\}$ and

$$P_{Y^{A_n}|X^n, \tau}(y^{A_n}|x^n, t) = W^n(y_t^{t+n-1}|x^n) \prod_{\substack{j < t \\ j \geq t+n}} W(y_j|\star),$$

where $y_a^b = (y_a, \dots, y_b)$.

Definition 1: An M -code for the random transformation (W^n, A_n) is a triplet

- An encoder function $f: \{1, \dots, M\} \rightarrow \mathcal{X}^n$
- A stopping time $\theta \geq n$ of the filtration generated by $\{Y_j, j = 1, \dots, A_n\}$. For convenience, we set

$$\hat{\tau} \triangleq \theta - n + 1,$$

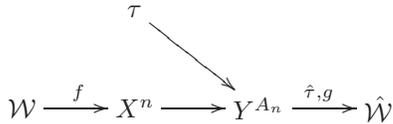
which marks the decoder's estimate of τ .

- A decoder function $g: \mathcal{Y}^{\hat{\tau}+n-1} \rightarrow \{1, \dots, M\}$

A code is said to be an (M, ϵ) code if

$$\mathbb{P}[\hat{\mathcal{W}} = \mathcal{W}, \hat{\tau} \leq \tau] \geq 1 - \epsilon, \quad (2)$$

where $\hat{\mathcal{W}} = g(Y^{\hat{\tau}+n-1})$ and the probability space is constructed by taking \mathcal{W} to be uniform on $\{1, \dots, M\}$ and chaining all random transformations according to the directed graphical model:



Including $\{\hat{\tau} > \tau\}$ in the error event serves the purpose of penalizing the code for declaring the decision late. In [12] the delay conditions on the decoder were weaker and the probability of error was defined (in essence) as

$$\mathbb{P}[\hat{\mathcal{W}} = \mathcal{W}, \hat{\tau} \leq \tau + L_n] \geq 1 - \epsilon, \quad L_n = \exp\{o(n)\} \quad (3)$$

that is a delay L_n is allowed to be non-zero but required to be pre-specified and sub-exponential in n . One of the results of this paper is that codes with $L_n = 0$ exist and achieve the same (asymptotic) performance as the best codes with weaker $L_n = \exp\{o(n)\}$.

Definition 2: A pair (R, \mathcal{A}) is called ϵ -achievable if there exist sequences of numbers $A_n \geq n$ and $M_n \geq 2$ satisfying

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log A_n \geq \mathcal{A}, \quad (4)$$

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log M_n \geq R \quad (5)$$

and a sequence of (M_n, ϵ) codes for random transformations (W^n, A_n) . The asynchronous ϵ -capacity at asynchronism \mathcal{A} is defined as

$$C_\epsilon(\mathcal{A}) \triangleq \sup\{R : (R, \mathcal{A}) \text{ is } \epsilon\text{-achievable}\}.$$

The asynchronous capacity at asynchronism \mathcal{A} is defined as

$$C(\mathcal{A}) \triangleq \lim_{\epsilon \rightarrow 0} C_\epsilon(\mathcal{A}).$$

The ϵ -synchronization threshold $\mathcal{A}_{o, \epsilon}$ is defined as

$$\mathcal{A}_{o, \epsilon} \triangleq \sup\{\mathcal{A} : (0, \mathcal{A}) \text{ is } \epsilon\text{-achievable}\}$$

and the synchronization threshold is

$$\mathcal{A}_o \triangleq \lim_{\epsilon \rightarrow 0} \mathcal{A}_{o, \epsilon}.$$

Remark: Note that $(0, \mathcal{A})$ is ϵ -achievable if and only if there exist a sequence of $(n, 2, \epsilon)$ codes for random transformations $(W^n, 2^{n\mathcal{A}+o(n)})$.

The main difference with the model studied in [1], [2] is that the definition of rate there was

$$\tilde{R} \triangleq \frac{\log M}{\mathbb{E}[|\hat{\tau} - \tau + n|+]} \quad (6)$$

and correspondingly the error event was defined as just $\{\hat{\mathcal{W}} \neq \mathcal{W}\}$. With such modifications, one defines the capacity $\tilde{C}(\mathcal{A})$ in exactly the same manner as $C(\mathcal{A})$; the key results of [1], [2] provide upper and lower bounds on $\tilde{C}(\mathcal{A})$ (but not $\tilde{C}_\epsilon(\mathcal{A})$). The definition (6) was chosen, perhaps, to model the situation when one wants to assess the minimal number of channel uses (per data bit) that the channel remains under the scrutiny of the decoder, whereas our definition

$$R \triangleq \frac{\log M}{n} \quad (7)$$

serves the purpose of studying the minimal number of channel uses (per data bit) that the channel remains occupied by the transmitter. With such definition, our model can be interpreted as the problem of communicating both the data \mathcal{W} and the state τ as in [15], except that the state is no longer a realization of the discrete memoryless process and it enters the channel law (W^n, A_n) in a different way.

The notation in this paper follows that of [16] and [14, Section IV.A], in particular, $D(P||Q)$ denotes the relative entropy between distributions P and Q ; $W_x(\cdot) = W(\cdot|x)$; for a distribution P on \mathcal{X} a distribution PW on \mathcal{Y} is defined as $PW(y) = \sum_x W(y|x)P(x)$; we agree to identify distribution Q on \mathcal{Y} with a stochastic kernel $Q: \mathcal{X} \rightarrow \mathcal{Y}$ which is constant on \mathcal{X} , so under this agreement $PW_x = W_x$; and $I(P, W)$ is a mutual information between $X \sim P$ and $Y \sim PW$ and coupled via $P_{Y|X} = W$: $I(P, W) = D(W||PW|P)$. We also denote by P^n the product distribution on \mathcal{X}^n and similarly for \mathcal{Y}^n .

III. CAPACITY AND STRONG CONVERSE

We summarize the previously known results:

Theorem 1 ([1], [12]): For any DMC W we have

$$\mathcal{A}_o = \max_{x \in \mathcal{X}} D(W_x || W_\star). \quad (8)$$

The asynchronous capacity of the DMC W under the probability of error criterion (3) is:

$$C(\mathcal{A}) = \max_{P: D(PW || W_\star) \geq \mathcal{A}} I(P, W), \quad (9)$$

where we agree that the maximum is zero whenever $\mathcal{A} > \mathcal{A}_o$.

Our main asymptotic results are the following:

Theorem 2: For any DMC W we have under either (2) or (3) definitions of probability of error

$$\mathcal{A}_{o,\epsilon} = \mathcal{A}_o, \quad (10)$$

where \mathcal{A}_o is given by (8). The asynchronous capacity of the DMC W under either (2) or (3) definitions of probability of error is

$$C_\epsilon(\mathcal{A}) = C(\mathcal{A}), \quad (11)$$

where $C(\mathcal{A})$ is given by (9).

Remark: As shown in [12, Theorem 5] the weak converse in Theorem 1 is unchanged if τ is not precisely uniform on $\exp\{n\mathcal{A}\}$ atoms but rather is “essentially” such: namely, the length ℓ_n of the optimal binary lossless compressor of τ satisfies:

$$\frac{1}{n} \ell_n \rightarrow \mathcal{A},$$

where the convergence is in probability. In the proof below we will show that the strong converse of Theorem 2 is also unchanged if τ is non-uniform and satisfies a stronger condition:

$$\max_i \mathbb{P}[\tau = i] = \frac{1}{A_n} \exp\{o(n)\}. \quad (12)$$

A. Discussion and comparison of results

Note that if $\mathcal{A}_o = \infty$ then according to (9)

$$C(\mathcal{A}) = \max_P I(P, W) \triangleq C, \quad \forall \mathcal{A} \geq 0, \quad (13)$$

i.e. capacity can be achieved for all exponents $\mathcal{A} \geq 0$.

For example, consider the binary symmetric channel $BSC(\delta)$ with $\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{0, 1\}$, $\star = 0$ and

$$W(y|x) = \begin{cases} 1 - \delta, & y = x \\ \delta, & y \neq x, \end{cases}$$

For such a model, computation of (8)-(9) yield

$$\mathcal{A}_o = d(\delta || 1 - \delta), \quad (14)$$

$$C(d(p * \delta || 1 - \delta)) = h(p * \delta) - h(\delta), \quad p \in [0, \frac{1}{2}] \quad (15)$$

where the latter is presented in parametric form and we have defined

$$d(x||y) = x \log \frac{x}{y} + (1-x) \log \frac{1-x}{1-y}, \quad (16)$$

$$h(x) = x \log \frac{1}{x} + (1-x) \log \frac{1}{1-x}, \quad (17)$$

$$p * \delta = (1-p)\delta + p(1-\delta). \quad (18)$$

First, we compare results in Theorems 1 and 2:

- Theorem 2 proves achievability part under a more stringent condition (2). Unlike [12] (and [2]) our proof relies on showing a variant of the packing lemma, which among other things should be useful for future investigations of universality and error-exponent questions.
- Theorem 2 shows that a strong converse for the $C(\mathcal{A})$ holds under both conditions (2) and (3), and also under non-uniform τ 's as in (12). To that end we employ the meta-converse framework [14, Section III.E] and [17, Section 2.7], which results in a short proof and is known to be quite tight non-asymptotically too, e.g. [14, Section III.J4]. It is possible that our methods would also prove useful for improving the bounds on the capacity $\tilde{C}(\mathcal{A})$ in the model (6).

Next, we compare to the results to the results in [1], [2], which concern a different definition of rate (6):

- In both cases the synchronization threshold is given by (8); see [1]. This is not surprising since (as remarked above) \mathcal{A}_o is determined by the ability to communicate with $M = 2$ codewords, for which the precise definition of rate is immaterial.
- In both cases, there is a “discontinuity at $R = C$ ” in the sense that $C(\mathcal{A}) = C$ for all $\mathcal{A} \leq \mathcal{A}_1$ with $\mathcal{A}_1 > 0$ if and only if

$$D(P_Y^* || W_\star) > 0,$$

where P_Y^* denotes the unique capacity achieving output distribution. However, the precise value of this critical exponent \mathcal{A}_1 is unknown for the model (6) even for the BSC, whereas in the model (7) we always have

$$\mathcal{A}_1 = D(P_Y^* || W_\star). \quad (19)$$

- In both cases, for a certain natural class of synchronization schemes based on preambles, see [2, Definition 3], we have $\mathcal{A}_1 = 0$, that is restricting communication system design prevents achieving capacity with positive asynchronism exponent. For the model (6) this is shown in [2, Corollary 3], while for the model (7) this is simply trivial: to combat a positive asynchronism exponent one would require preamble of the size δn , but this penalizes the rate to be at most $C - \delta$.
- According to [2] there exist channels (and BSC is one of them – see below) for which the capacity $\tilde{C}(\mathcal{A}) = 0$ for some range of $\mathcal{A} < \mathcal{A}_o$. In such regime there exist codes reliably sending $M = \exp\{nR\}$ codewords, but the rate \tilde{R} , as defined in (6), remains zero. This strange behavior, called “discontinuity at $R = 0$ ” in [2, Corollary 2] does not occur in the definition of rate (7): the capacity is positive for all $\mathcal{A} < \mathcal{A}_o$.
- Somewhat counter-intuitively although in our model we impose a seemingly strong condition $\{\hat{\tau} \leq \tau\}$ absent in [1], [2], it turns out that the capacity vs. asynchronous exponent region is larger. This is explained by noticing that if $\hat{\tau} > \tau$ then one typically has $\hat{\tau} = \tau + \exp\{n\epsilon\}$. Thus in the model (6), to avoid significant penalty in

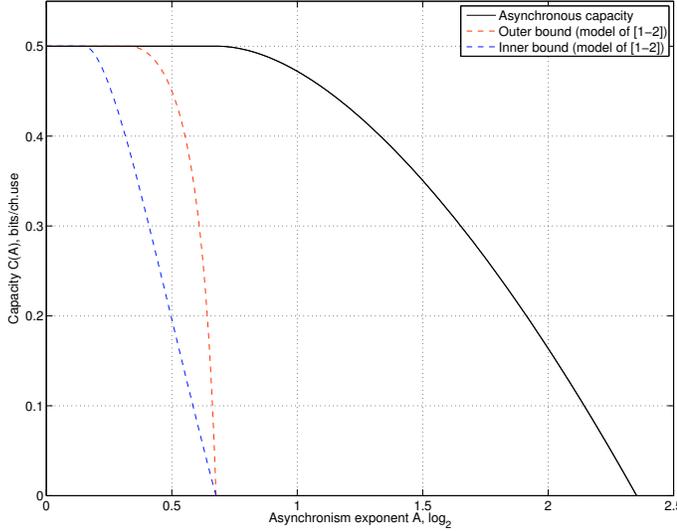


Fig. 1. $BSC(0.11)$: The asynchronous capacity (15) compared with inner-outer bounds of [2] for the different model (6).

rate the occurrence of $\hat{\tau} > \tau$ should happen with exponentially small probability.

Additionally, [13] considers the definition of rate as in (7) but models asynchronism differently and restricts the decoders to operate solely on the basis of each observed n -block. Curiously, however, their region of rate vs. false alarm error-exponent coincides with the region (9) of rate vs. asynchronism exponent; see [13, Theorem 1].

To illustrate these points, in Fig. 1 we compare the region (9) with inner (achievability) and outer (converse) bounds found in [2, Theorem 2 and Theorem 3], respectively, which for the case of the $BSC(\delta)$ can be shown to be

$$\tilde{C}_{in}(d(q||\delta)) = h(p * \delta) - h(\delta) \quad (20)$$

$$\tilde{C}_{out} \left(\frac{d(p * \delta || 1 - \delta) d(\frac{1}{2} || \delta)}{d(p * \delta || 1 - \delta) + p d(\frac{1}{2} || \delta)} \right) = h(p * \delta) - h(\delta) \quad (21)$$

where parameter runs over $p \in [0, \frac{1}{2}]$ and in (20) q solves

$$d(q||p * \delta) = d(q||\delta).$$

Note that according to the \tilde{C}_{out} bound the capacity in the model (6) is zero between $d(\frac{1}{2} || \delta)$ and $\mathcal{A}_o = d(1 - \delta || \delta)$. This demonstrates the above mentioned discontinuity at $R = 0$ for the BSC and therefore closes the open question mentioned after [2, Corollary 2].

B. Achievability

We omit a detailed proof in this extended abstract, but mention the key ingredient.

The main problem in achieving a good error-correction performance in the presence of asynchronism is the ability to resolve partially aligned codewords. For example, suppose that a codeword $x \in \mathcal{X}^n$ is being transmitted. Then, if there is a k -symbol misalignment, $0 \leq k < n$, the decoder observes

outputs effectively generated by a shifted codeword x^{*k} :

$$x^{*k} \triangleq (\underbrace{*, \dots, *}_k, x_1, \dots, x_{n-k}) \in \mathcal{X}^n. \quad (22)$$

Thus, a good codebook for asynchronous communication must be such that not only a given codeword x is far away from all other codewords, but also all of its k -shifts x^{*k} are. The existence of such codebooks follows from a simple generalization of a packing lemma [16, Lemma 2.5.1]:

Lemma 3: For every $R > 0$, $\delta > 0$ and every type P of sequences in \mathcal{X}^n satisfying $H(P) > R$, there exist at least $M = \exp\{n(R - \delta)\}$ distinct sequences $c_i \in \mathcal{X}^n$ of type P such that for every pair of stochastic matrices $V : \mathcal{X} \rightarrow \mathcal{Y}$, $\hat{V} : \mathcal{X} \rightarrow \mathcal{Y}$, every i and every $0 \leq k < n$ we have

$$\left| T_V(c_i^{*k}) \cap \bigcup_{j \neq i} T_{\hat{V}}(c_j) \right| \leq |T_V(c_i^{*k})| \exp\{-n|I(P, \hat{V}) - R|^+\} \quad (23)$$

provided that $n \geq n_0(|\mathcal{X}|, |\mathcal{Y}|, \delta)$.

Remark: In fact, there is nothing special about the transformations $c \mapsto c^{*k}$. The lemma and the proof hold verbatim if c_i^{*k} is replaced by $f(c_i)$, and clause “every $0 \leq k < n$ ” with “every $f \in \mathcal{F}_n$ ”, where \mathcal{F}_n is an arbitrary collection of maps $f : \mathcal{X}^n \rightarrow \mathcal{X}^n$ of subexponential size: $|\mathcal{F}_n| = e^{o(n)}$.

C. Converse

Detailed proofs are omitted but we provide sketch of the main steps.

First, we introduce the performance of the optimal binary hypothesis test. Consider a \mathcal{X} -valued random variable X which can take probability measures P or Q . A randomized test between those two distributions is defined by a random transformation $P_{Z|X} : \mathcal{X} \mapsto \{0, 1\}$ where 0 indicates that the test chooses Q . The best performance achievable among those randomized tests is given by

$$\beta_\alpha(P, Q) = \min_{x \in \mathcal{X}} \sum_{x \in \mathcal{X}} Q(x) P_{Z|X}(1|x), \quad (24)$$

where the minimum is over all probability distributions $P_{Z|X}$ satisfying

$$P_{Z|X} : \sum_{a \in \mathcal{X}} P(x) P_{Z|X}(1|x) \geq \alpha. \quad (25)$$

The minimum in (24) is guaranteed to be achieved by the Neyman-Pearson lemma. Thus, $\beta_\alpha(P, Q)$ gives the minimum probability of error under hypothesis Q if the probability of error under hypothesis P is not larger than $1 - \alpha$.

The proof of the converse relies on a pair of simple lemmas of separate interest:

Lemma 4: If $\mathcal{A}_o < \infty$ then there exists V_1 such that for any input x^n we have

$$\beta_\alpha(P_{Y^n|X^n=x^n}, W_\star^n) \geq \frac{\alpha}{2} \exp \left\{ -nD(W||W_\star|\hat{P}_{x^n}) - \sqrt{\frac{2nV_1}{\alpha}} \right\}, \quad (26)$$

where \hat{P}_{x^n} is the composition of x^n and

$$P_{Y^n|X^n=x^n}(\cdot) \triangleq W^n(\cdot|x^n),$$

with W^n defined in (1).

Lemma 5: Consider a DMC W . If $\mathcal{A}_o < \infty$ then there exists V_1 such that for any synchronous (n, M, ϵ) code (maximal probability of error) with codewords $\{c_i, i = 1, \dots, M\}$ of constant composition P_0 we have

$$\beta_\alpha(P_{Y^n}, W_\star^n) \geq M \frac{\alpha}{4} \frac{\alpha - 2\epsilon}{2 - \alpha} \times \exp \left\{ -nD(W||W_\star|P_0) - \sqrt{\frac{4nV_1}{\alpha - 2\epsilon}} \right\} \quad (27)$$

provided that $\alpha > 2\epsilon$, where in (27) P_{Y^n} denotes the output distribution induced by the code:

$$P_{Y^n}[\cdot] = \frac{1}{M} \sum_{j=1}^M W^n(\cdot|c_j).$$

Converse part in Theorem 2 (sketch): Assume $\epsilon < \frac{1}{3}$ and that

$$\mathbb{P}[\hat{\mathcal{W}} = \mathcal{W}, \hat{\tau} \leq \tau | \mathcal{W} = j] \geq 1 - \epsilon, \quad j = 1, \dots, M_n. \quad (28)$$

Clearly, such a code must be synchronously decodable over DMC W with maximal probability of error at most ϵ . By a standard argument we may restrict to a subcode with a constant composition P_n . Then for some constant $b_3 > 0$ (depending only on W and ϵ)

$$\log M_n \leq nI(P_n, W) + b_3\sqrt{n}. \quad (29)$$

We now apply the meta-converse principle [14, Section III.E], which consists of changing the channel and using the event $\{\hat{\mathcal{W}} = \mathcal{W}\}$ as a binary hypothesis test between the two channels. Namely, in addition to the true channel $P_{Y^{A_n}|X^n, \tau}$ we consider an auxiliary channel

$$Q_{Y^{A_n}|X^n, \tau} = W_\star^{A_n}$$

which outputs W_\star -distributed noise in all of A_n symbols, regardless of X^n and τ . Obviously, under the Q -channel we have

$$\mathbb{Q}[\tau - n < \hat{\tau} \leq \tau] = \frac{n}{A_n} \quad (30)$$

by independence of $\hat{\tau}$ and τ , whereas under the P -channel we have

$$\mathbb{P}[\tau - n < \hat{\tau} \leq \tau] \geq 1 - \epsilon - \frac{1}{M'_n}.$$

Using the test $\{\tau - n < \hat{\tau} \leq \tau\}$ we show

$$\beta_{1-\epsilon'}(P_{Y^{A_n \tau}}, Q_{Y^{A_n \tau}}) \leq \frac{n}{A_n}, \quad (31)$$

where we denoted for convenience $\epsilon' = \epsilon + \frac{1}{M'_n}$. On the other hand,

$$\begin{aligned} & \beta_{1-\epsilon'}(P_{Y^{A_n \tau}}, Q_{Y^{A_n \tau}}) \\ &= \beta_{1-\epsilon'}(P_{Y^n|\tau=1}, W_\star^n) \end{aligned} \quad (32)$$

$$\geq M'_n \exp \left\{ -nD(W||W_\star|P_n) - b_2\sqrt{n} \right\}, \quad (33)$$

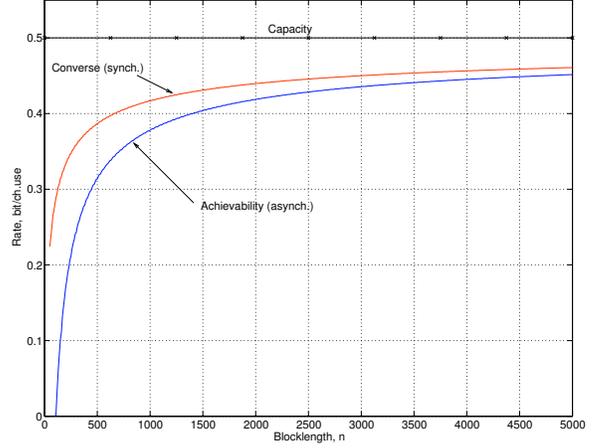


Fig. 2. BSC(0.11): Non-asymptotic performance of asynchronous codes compared with the upper (converse) bound for the synchronous channel. Probability of error $\epsilon = 10^{-3}$, asynchronism level $A_n = 2^{0.68n - 5.25\sqrt{n}}$.

and by Lemma 5

$$\begin{aligned} & \log \beta_{1-\epsilon'}(P_{Y^n|\tau=1}, W_\star^n) \geq \\ & \log M_n - nD(W||W_\star|P_n) + O(\sqrt{n}). \end{aligned} \quad (34)$$

We conclude that

$$R + \mathcal{A} \leq D(W||W_\star|P_n) \quad (35)$$

$$R \leq I(P_n, W), \quad (36)$$

which after trivial manipulations results in (9). ■

IV. NON-ASYMPTOTIC BOUND AND CHANNEL DISPERSION

One of the important conclusions is that the function $C(\mathcal{A})$ is constant on the interval $[0; \mathcal{A}_1]$, where \mathcal{A}_1 is given by (19). In other words, a certain level of asynchronism (up to $\exp\{n\mathcal{A}_1\}$) is completely unharmed to the capacity of the channel. This surprising result has also been noticed in [2] (the value of \mathcal{A}_1 is not known exactly for their model).

All the arguments so far were asymptotical and it is very natural to doubt whether such effect is actually possible for blocklengths of interest. To show that it does indeed happen for practical lengths we will prove a non-asymptotic achievability bound and show that it implies that the channel dispersion is unchanged. First, however, we recall some of the results of [14].

Let $M^*(n, \epsilon)$ be the maximal cardinality of a codebook of blocklength n which can be (synchronously) decoded with block error probability no greater than ϵ over the DMC defined by (1). By Shannon's theorem asymptotically we have

$$\log M^*(n, \epsilon) \approx nC \quad (37)$$

It has been shown in [14] that a much tighter approximation can be obtained by defining an additional figure of merit referred to as the channel dispersion:

Definition 3: The dispersion V (measured in squared information units per channel use) of a channel with capacity C is equal to

$$V = \lim_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \frac{(nC - \log M^*(n, \epsilon))^2}{2 \ln \frac{1}{\epsilon}}. \quad (38)$$

For example, the minimal blocklength required to achieve a given fraction η of capacity with a given error probability ϵ can be estimated as:¹

$$n \gtrsim \left(\frac{Q^{-1}(\epsilon)}{1 - \eta} \right)^2 \frac{V}{C^2}. \quad (39)$$

The motivation for Definition 3 and estimate (39) is the following expansion for $n \rightarrow \infty$

$$\log M^*(n, \epsilon) = nC - \sqrt{nV}Q^{-1}(\epsilon) + O(\log n). \quad (40)$$

As shown in [14] in the context of memoryless channels, (40) gives an excellent approximation for blocklengths and error probabilities of practical interest.

Theorem 6: Consider arbitrary random transformation $P_{Y^n|X^n} : \mathcal{X}^n \rightarrow \mathcal{Y}^n$. Then for any $\gamma \geq 0$ and any input distribution P_{X^n} on \mathcal{X}^n there exists an (M, ϵ) code for the random transformation $(P_{Y^n|X^n}, A)$ with

$$\epsilon \leq \mathbb{E} [\exp\{-|r(Y^n) - \log A|^+\}] + P[i(X^n; Y^n) \leq \gamma] + nM \exp\{-\gamma\}, \quad (41)$$

where P denotes probability with respect to the distribution $P_{X^n Y^n}(x, y) = P_{Y^n|X^n}(y|x)P_{X^n}(x)$, \mathbb{E} is the expectation with respect to P and we also defined

$$r(y^n) \triangleq \log \frac{P_{Y^n}(y^n)}{W_*(y^n)} \quad (42)$$

$$i(x^n; y^n) \triangleq \log \frac{P_{Y^n|X^n}(y^n|x^n)}{P_{Y^n}(y^n)}. \quad (43)$$

Proof is omitted due to space constraints.

An interesting qualitative conclusion from Theorem 6 is the following:

Corollary 7: Consider a DMC W with (synchronous) capacity C and dispersion V . Then for every $0 < \epsilon < 1$ there exist capacity-dispersion optimal codes for the asynchronous DMC at asynchronism $A_n = 2^{nA_1 + o(n)}$. More precisely the number of messages M_n for such codes satisfies

$$\log M_n = nC - \sqrt{nV}Q^{-1}(\epsilon) + O(\log n), n \rightarrow \infty \quad (44)$$

The proof is a simple application of Theorem 6 with a capacity-achieving input distribution and Berry-Esseen estimates.

Remark: As (40) demonstrates, it is not possible to improve the second term in expansion (44) even in the synchronous setting, see also [14, Theorem 48]. Corollary 7 demonstrates that not only it is possible to communicate with rates close to capacity and still handle an exponential asynchronism (up to 2^{nA_1}), but in fact one can even do so using codes which are capacity-dispersion optimal.

Finally, in Fig. 2 we illustrate this last point numerically by computing the bound of Theorem 6 for the $BSC(\delta)$ and comparing it with the converse for the corresponding synchronous channel [14, Theorem 35]. For the purpose of this illustration we have chosen $\epsilon = 10^{-3}$, $\delta = 0.11$ and, somewhat arbitrarily,

$$A_n = \exp \left\{ nD(P_Y^* || W_*) + \sqrt{nV(P_Y^* || W_*)}Q^{-1}\left(\frac{\epsilon}{4}\right) \right\} \quad (45)$$

$$\approx 2^{0.68n - 5.25\sqrt{n}}. \quad (46)$$

In particular, the plot shows that it is possible to construct asynchronous codes that do not lose much compared to the best possible synchronous codes in terms of rate, but which at the same time are capable of tremendous tolerance to asynchronism. For example, already at $n = 500$ the decoder is able to find and error-correct the codeword inside a noisy binary string of unimaginable length $2^{221} \approx 10^{66}$.

REFERENCES

- [1] A. Tchamkerten, V. Chandar, and G. W. Wornell, "Communication under strong asynchronism," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4508–4528, Oct. 2009.
- [2] —, "Asynchronous communication: Capacity bounds and suboptimality of training," 2011. [Online]. Available: <http://arxiv.org/abs/1105.5639v1>
- [3] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Stat.*, pp. 1897–1908, 1971.
- [4] I. Nikiforov, "A generalized change detection problem," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 171–187, 1995.
- [5] V. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Soviet Physics Doklady*, vol. 10, no. 8, pp. 707–710, 1966.
- [6] R. L. Dobrushin, "Shannon's theorems for channels with synchronization errors," *Prob. Peredachi Inform.*, vol. 3, no. 4, pp. 11–26, Dec. 1967.
- [7] V. I. Levenshtein, "One method of constructing quasilinear codes providing synchronization in the presence of errors," *Prob. Peredachi Inform.*, vol. 7, no. 3, pp. 30–40, 1971.
- [8] G. Poltyrev, "Coding in an asynchronous multiple-access channel," *Prob. Peredachi Inform.*, vol. 19, no. 3, pp. 12–21, Sep. 1983.
- [9] J. Hui and P. Humblet, "The capacity region of the totally asynchronous multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 31, no. 3, pp. 207–216, Mar. 1985.
- [10] L. Farkas and T. Koi, "Capacity regions of discrete asynchronous multiple access channels," in *Proc. 2011 IEEE Int. Symp. Inf. Theory (ISIT)*, Aug. 2011, pp. 2273–2277.
- [11] S. Verdú, "The capacity region of the symbol-asynchronous gaussian multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 35, no. 4, pp. 733–751, Jul. 1989.
- [12] V. Chandar, A. Tchamkerten, and D. N. C. Tse, "Asynchronous capacity per unit cost," *IEEE Trans. Inf. Theory*, 2011, submitted. [Online]. Available: <http://arxiv.org/abs/1007.4872v1>
- [13] D. Wang, V. Chandar, S.-Y. Chung, and G. Wornell, "Error exponents in asynchronous communication," in *Proc. 2011 IEEE Int. Symp. Inf. Theory (ISIT)*. St. Petersburg, Russia: IEEE, 2011.
- [14] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [15] T. Cover, Y. Kim, and A. Sutivong, "Simultaneous communication of data and state," in *Proc. 2007 IEEE Int. Symp. Inf. Theory (ISIT)*. Nice, France: IEEE, 2007, pp. 916–920.
- [16] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
- [17] Y. Polyanskiy, "Channel coding: non-asymptotic fundamental limits," Ph.D. dissertation, Princeton Univ., Princeton, NJ, USA, 2010, available: <http://www.princeton.edu/~ypolyans>.

¹As usual, $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$.