# Frame Alignment and Communication under Strong Asynchronism

Aslan Tchamkerten, Venkat Chandar, and Gregory Wornell
Electrical Engineering and Computer Science Department
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
Email: {tcham,vchandar,gww}@mit.edu

***Abstract*— We consider asynchronous communication over point-to-point discrete memoryless channels without feedback. The transmitter starts sending one block codeword at an instant that is uniformly distributed within a certain time period, which represents the level of asynchronism between the transmitter and the receiver. The receiver, by means of a sequential decoder, must isolate the message without knowing when the codeword transmission starts but being cognizant of the asynchronism level. We are interested in quick detection and isolation of the sent message, particularly in the regime where the asynchronism level is exponentially larger than the codeword length, which we refer to as 'strong asynchronism.'**

**This model of sparse communication might represent, for instance, the situation of a sensor that remains idle most of the time and, only occasionally, transmits information to a remote base station which needs to quickly take action. Because of the limited amount of energy the sensor possesses, assuming the same cost per transmitted symbol, it is of interest to consider minimum size codewords given the asynchronism level.**

**The first result is an asymptotic characterization of the largest asynchronism level, in terms of the codeword length, for which reliable communication can be achieved: vanishing error probability can be guaranteed as the codeword length $N$ tends to infinity while the asynchronism level grows as $e^{N\alpha}$ if and only if $\alpha$ does not exceed the *synchronization threshold*, a constant that admits a simple closed form expression, and is at least as large as the capacity of the synchronized channel.**

**The second result is the characterization of a set of achievable strictly positive rates in the regime where the asynchronism level is exponential in the codeword length, and where the rate is defined with respect to the expected (random) delay between the time information starts being emitted until the time the receiver makes a decision. Interestingly, this achievability result is obtained by a coding strategy whose decoder not only operates asynchronously, but has also an almost universal decision rule, in the sense that it is almost independent of the channel statistics.**

**As an application of the first result we consider antipodal signaling over a Gaussian additive channel and derive a simple necessary condition between blocklength, asynchronism level, and SNR for achieving reliable communication.**

**Finally we note that the communcation model we study can be seen as a complement to the insertion, deletion, and substitution channel model introduced by Dobrushin in 1967. The chief difference is that this channel models timing uncertainty that result from the channel, whereas our setting models timing uncertainty caused by the users (or by a bursty source of information).**

***Index Terms*— Asynchronous communication, detection and isolation problem, discrete-time communication, error exponent, low probability of detection, point-to-point communication, quickest detection, sequential analysis, sparse communication, stopping times**

## I. Introduction

A common assumption in information theory is that 'whenever the transmitter speaks the receiver listens.' In other words, in general, there is the assumption of perfect synchronization between the transmitter and the receiver and, basic quantities, such as the channel capacity, are defined under this hypothesis [14]. In practice this assumption is rarely fulfilled. Time uncertainty due, for instance, to bursty sources of information often causes asynchronous communication, i.e., communication for which the receiver has only a partial knowledge of *when* information is sent.

This work was supported in part by NSF under Grant No. CCF-0515122, and by a University IR&D Grant from Draper Laboratory. The present paper is a partial summary of the results obtained by the authors in [15] and [3].

There are, however, notable channels for which asynchronism effects have been studied from an information theoretic standpoint. An example is the multiple access channel (see, e.g., [4], [10], [13], [18]) for which the capacity region has been computed under various assumptions on the users' asynchronism. Another important example is the insertion, deletion, and substitution (IDS) channel for which only bounds on the capacity are known (see, e.g., [1], [8], [9], [7]).

In this paper we propose an information theoretic framework that models users' asynchronism for point-to-point discrete-time communication without feedback. We consider the situation where the transmitter may start sending information at a time unknown to the receiver. The time transmission starts is assumed to be uniformly distributed within a certain interval, which defines the asynchronism level between the transmitter and the receiver. A suitable notion of rate is introduced and scaling laws between block message size and asynchronism level are given for which reliable communication can or cannot be achieved.[1] Our first result is the characterization of the highest asynchronism level with respect to the codeword length under which reliable communication can still be achieved. This limit is attained by a coding strategy that operates at vanishing rate. This strategy also allows for communication at positive rates while operating at asynchronism levels that are exponentially larger than the codeword length.

Note that the above channel setting can be considered as a complement to the IDS channel model since these two channels model different types of asynchronism. Specifically, the IDS channel models asynchronism effects during information transmission, whereas our setting models users' asynchronism. We will return to this issue in Section II.

In Section II we formally introduce our model and draw connections with the related 'detection and isolation' problem in sequential analysis, and Section III contains our main results.

## II. Problem formulation and background

We consider discrete-time communication over a discrete memoryless channel characterized by its finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively, transition probability matrix $Q(y|x)$, for all $y \in \mathcal{Y}$ and $x \in \mathcal{X}$, and 'noise' symbol $\star \in \mathcal{X}$ (see Fig. 1).[2] The codebook consists of $M \geq 2$ equally likely codewords of length $N$ composed of symbols from $\mathcal{X}$ — possibly also the $\star$ symbol. The transmission of a particular codeword starts at a random time $\nu$, independent of the codeword to be sent, uniformly distributed in $[1, 2, \ldots, A]$, where the integer $A \geq 1$ characterizes the asynchronism level. We assume that the receiver knows $A$ but not $\nu$. If $A = 1$ the channel is said to be synchronized. Throughout the paper, whenever we refer to the capacity of a channel, it is intended to be the capacity of the synchronized channel. Throughout the paper we only consider channels $Q$ with strictly positive capacity $C(Q)$.

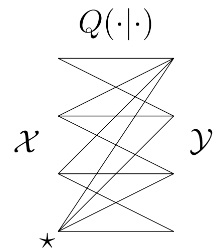


Fig. 1. Communication is carried over a discrete memoryless channel. When 'no information' is sent the input of the channel is the '$\star$' symbol.

Before and after the transmission of the information, i.e., before time $\nu$ and after time $\nu+N-1$, the receiver observes noise. Specifically, conditioned on the value of $\nu$ and on the message to be conveyed $m$, the receiver observes independent symbols $Y_1, Y_2, \ldots$ distributed as follows. If $i \leq \nu - 1$ or $i \geq \nu + N$, the distribution is $Q(\cdot|\star)$. At any time $i \in [\nu, \nu+1, \ldots, \nu+N-1]$ the distribution is $Q(\cdot|c_{i-\nu+1}(m))$, where $c_n(m)$ denotes the $n$th symbol of the codeword $c^N(m)$ assigned to message $m$.

The decoder consists of a sequential test $(\tau, \phi)$, where $\tau$ is a stopping time with respect to the output sequence $Y_1, Y_2, \ldots$[3] indicating when decoding

[1] We refer to 'reliable communication' whenever arbitrary low error probability can be achieved.

[2] Throughout the paper we always assume that for all $y \in \mathcal{Y}$ there is some $x \in \mathcal{X}$ for which $Q(y|x) > 0$.

[3] Recall that a stopping time $\tau$ is an integer-valued random variable with respect to a sequence of random variables $\{Y_i\}_{i=1}^{\infty}$ so that the event $\{\tau = n\}$, conditioned on $\{Y_i\}_{i=1}^{n}$, is independent of $\{Y_i\}_{i=n+1}^{\infty}$ for all $n \geq 1$.

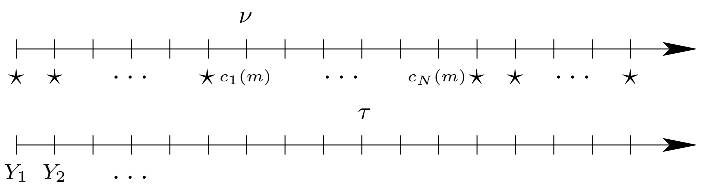


Fig. 2. Time representation of what is sent (upper arrow) and what is received (lower arrow). The '$\star$' represents the 'noise' symbol. At time $\nu$ message $m$ starts being sent and decoding occurs at time $\tau$.

happens, and where $\phi$ denotes a decision rule[4] that declares the decoded message (see Fig. 2).

We are interested in *reliable and quick decoding*. To that aim we first define the average decoding error probability as

$$\mathbb{P}(\mathcal{E}) = \frac{1}{A}\frac{1}{M}\sum_{m=1}^{M}\sum_{l=1}^{A}\mathbb{P}_{m,l}(\mathcal{E}),$$

where $\mathcal{E}$ indicates the event that the decoded message does not correspond to the sent message, and where the subscripts $_{m,l}$ indicate the conditioning on the event that message $m$ starts being sent at time $l$. Second, we define the average communication rate with respect to the average delay it takes the receiver to react to a sent message, i.e.

$$R = \frac{\ln M}{\mathbb{E}(\tau-\nu)^{+}} \tag{1}$$

with

$$\mathbb{E}(\tau-\nu)^{+} \triangleq \frac{1}{A}\frac{1}{M}\sum_{m=1}^{M}\sum_{l=1}^{A}\mathbb{E}_{m,l}(\tau-l)^{+}$$

where $x^{+}$ denotes $\max\{0,x\}$, and where $\mathbb{E}_{m,l}$ denotes the expectation with respect to $\mathbb{P}_{m,l}$.[5] With the above definitions we now introduce the notion of achievable rate with respect to a certain asynchronism level as well as the notion of *synchronization threshold*.

**Definition 1.** *An asynchronism exponent $\alpha$ is* achievable *at a rate $R$ if, for any $\varepsilon > 0$, there exists a block code with (sufficiently large) codeword length $N$, operating under asynchronism level $A = e^{(\alpha-\varepsilon)N}$, while yielding a rate at least as large as $R-\varepsilon$ and an error probability $\mathbb{P}(\mathcal{E}) \leq \varepsilon$. The supremum of the set of asynchronism exponents that are achievable at rate $R$ is denoted $\alpha(R,Q)$.*

Note that, for a given channel $Q$, the asynchronism exponent function $\alpha(R,Q)$ is non-increasing in $R$.

**Definition 2.** *The* synchronization threshold *of a channel $Q$, denoted by $\alpha(Q)$, is the supremum of the set of achievable asynchronism exponents at all rates, i.e., $\alpha(Q) = \alpha(R=0,Q)$.*

Throughout the paper we often use the terminology 'coding strategy' or 'coding scheme' to denote an infinite sequence of pairs codebook/decoder labeled by the blocklength. In particular, whenever we refer to a coding strategy that 'achieves a certain rate,' it is intended to be asymptotically in the limit $N \to \infty$.

Let us comment on the above bursty communication model and its associated notions of rate and synchronization threshold. We first compare the insertion, deletion, and substitution (IDS) channel model setting with ours. Recall that the IDS channel is specified by the set of all conditional output distributions $Q(\mathbf{y}|x)$ where $x$ belongs to some finite alphabet $\mathcal{X}$ and $\mathbf{y}$ is a string, possibly the empty one, where each element belongs to the same finite alphabet $\mathcal{Y}$. The IDS channel models timing uncertainty that result from the channel itself and not from the users'. Instead, our setting models timing uncertainty that result from the users' and not from the channel. Indeed, in the IDS channel there is the implicit assumption that information tries to be conveyed from time one, and that the receiver knows the timing of the last output symbol. For instance, in the purely deletion channel where each input symbol result deleted with a certain probability, if the codeword $c^{N}$ produces the output $\mathbf{y} \in \mathcal{Y}^{l}$, $l \leq N$, the receiver knows that *nothing* comes after time $l$. In contrast, in our setting information does not necessary start to be conveyed from time one and the receiver is not informed of the timing of the last output symbol. However, in contrast with the IDS channel model, our setting does not model timing uncertainty during transmission: conditioned on the time $\nu$ when information starts being conveyed, information transmission ends at the fixed time $\nu+$

[4]Formally $\phi$ is an $\mathcal{F}_{\tau}$-measurable map where $\mathcal{F}_1, \mathcal{F}_2, \ldots$ is the natural filtration induced by the process $Y_1, Y_2, \ldots$

[5]Here $\ln$ denotes the natural logarithm.

$N-1$.[6] Finally note that in our setting, in contrast with IDS channel, the level of asynchronism is captured through a single parameter ($A$ or $\alpha$).

First observe that we do not introduce a feedback channel from the receiver to the transmitter. With a noiseless feedback it is possible to inform the transmitter of the receiver's decoding time, say in the form of ack/nack, therefore allowing the sending of multiple messages instead of just one as in our model. Here the noiseless assumption is crucial. If the feedback is noisy, the receiver's decision may be wrongly recognized by the transmitter, which possibly may result in a loss of message synchronization between transmitter and receiver (say the receiver hasn't yet decoded the first message while the transmitter has already started to emit the second one). Therefore, in order to avoid a potential second source of asynchronism, we omit feedback in our study and limit transmission to only one message.

The reason for defining the rate with respect to the average delay $\mathbb{E}(\tau-\nu)^+$ (see (1)) is motivated by the following considerations. At first sight, a natural measure of delay may be the codeword length $N$. However, in light of the use of sequential decoding, the codeword length does not provide a measure of the delay needed for the information to be reliably decoded. Another candidate for the delay one might consider is $\mathbb{E}(\tau)$ or, equivalently, $\mathbb{E}\nu+\mathbb{E}(\tau-\nu)$. The fact that this delay takes into account the initial offset $\mathbb{E}\nu$ can be regarded as a weakness since this offset can be influenced neither by the transmitter nor by the receiver. Also, with such a delay measure, it can be shown that, in the regime of positive asynchronism exponents we are interested in, the rate is always (asymptotically) vanishing for any reliable coding strategy. Instead, we propose to consider $\mathbb{E}(\tau-\nu)^+$, the average time the transmitter needs to wait until the receiver makes a decision. Also note that, in the definition of achievable rate (Definition 1), we choose to grow $A$ with $N$. Indeed, when $A$ is fixed the problem becomes trivial. By using sufficiently long codewords and simply decoding at the (fixed) time $A+N-1$ the asynchronism effect on the rate can be made negligible.

We now briefly discuss the notion of synchronization threshold. This threshold is defined with respect to zero rate coding strategies, that is strategies for which $\ln M/\mathbb{E}(\tau-\nu)^+$ tends to zero (as $N\to\infty$). However, because $\mathbb{E}(\tau-\nu)^+$ and $N$ need not coincide in general, zero rate coding strategies need not, in general, yield a vanishing fraction $\ln M/N$ as $N$ tends to infinity. Indeed, as we will see, one can operate arbitrarily closely to the synchronization threshold while having $\ln M/N$ asymptotically bounded away from zero.

Perhaps the closest sequential decision problem our model relates to is a generalization of the change-point problem, often called the 'detection and isolation problem,' introduced by Nikiforov in $1995$ (see [12], [11] and [2] for a survey). A process $Y_1, Y_2, \ldots$ starts with some initial distribution and changes it at some unknown time. The post change distribution can be any of a given set of $M$ distributions. By sequentially observing $Y_1, Y_2, \ldots$ the goal is to quickly react to the statistical change and isolate its cause, i.e., the post-change distribution. Hence, our synchronization problem takes the form of a detection and isolation problem where the change in distribution is induced by the transmitted message. However, to the best of our knowledge studies related to the detection and isolation problem usually assume that once the observed process jumps into one of its post-change distributions, it remains in that state forever. This means that, eventually, if we wait long enough, a correct decision is be possible. Instead, in the synchronization problem the change in distribution is *local* since it only lasts the duration of a codeword length. In particular once the codeword is 'missed' no recovery is possible. Finally, optimal decoding rules for the detection and isolation problem seem to have been obtained only in the limit of small error probabilities $\mathbb{P}(\mathcal{E})$ while keeping $M$, the number of post-change distributions, fixed.[7] In our case we typically let $M$ grow as $(1/\mathbb{P}(\mathcal{E}))^\xi$, for some $\xi>0$.

[6]In [8] (see discussion after Theorem 1) Dobrushin discusses the assumption that the receiver implicitely knows the length of the received sequence. To avoid this assumption, instead of 'one-shot' communication, Dobrushin proposes to consider the sending of an infinite sequence of messages.

[7]Here optimal decoding rules refer to sequential tests yielding minimum reaction delay, usually a function of $\tau-\nu$, given a certain error probability.

## III. RESULTS

Our first result is the characterization of the synchronization threshold.

**Theorem 1.** *For any discrete memoryless channel $Q$, the synchronization threshold as given in Definition 2 is given by*

$$\alpha(Q) = \max_x D(Q(\cdot|x)||Q(\cdot|\star))$$

*where $D(Q(\cdot|x)||Q(\cdot|\star))$ is the divergence (Kullback-Leibler distance) between $Q(\cdot|x)$ and $Q(\cdot|\star)$. Furthermore, any synchronization threshold $\alpha < \alpha(Q)$ can be achieved by a coding strategy that yields $\lim_{N\to\infty} \ln M/N > 0$.* □

The theorem says that vanishing error probability can be achieved as the blocklength $N$ tends to infinity if the asynchronism level grows as $e^{N\alpha}$ where $\alpha < D(Q(\cdot|x)||Q(\cdot|\star))$. Conversely, any coding strategy that operates at an asynchronism exponent $\alpha > D(Q(\cdot|x)||Q(\cdot|\star))$ cannot achieve arbitrary low error probability. The second part of the theorem shows the distinction between the delay measured by the codeword length $N$ and by the expected 'reaction time' $\mathbb{E}(\tau-\nu)^+$. Arbitrary closely to the synchronization threshold one can (asymptotically) guarantee $\ln M/N$ to be strictly positive, while the question remains open for the rate $\ln M/\mathbb{E}(\tau-\nu)^+$.

At least some connections between channel capacity and synchronization threshold exist. Although these two quantities are not directly related, both refer to limits on hypothesis discrimination. The first concerns a purely isolation problem whereas the second concerns an almost purely detection problem (since there is no rate constraint). It may be interesting to note that the synchronization threshold $\alpha(Q)$ is always at least as large as $C(Q)$. To see this let $P$ be the capacity achieving distribution of the (synchronized) channel $Q$. It is well known [5, Lemma 13.8.1] that for any distribution $V$ on $\mathcal{Y}$

$$D(PQ||PP_Y) \leq D(PQ||PV)$$

where $P_Y$ is the right marginal of $PQ = P(\cdot)Q(\cdot|\cdot)$. Letting $V = Q(\cdot|\star)$ we get

$$\begin{aligned} C(Q) &\triangleq D(PQ(\cdot|\cdot)||PP_Y) \\ &\leq D(PQ(\cdot|\cdot)||PQ(\cdot|\star)) \\ &= \sum_x P(x) \sum_y Q(y|x) \ln \frac{Q(y|x)}{Q(y|\star)} \\ &\leq \max_x D(Q(\cdot|x)||Q(\cdot|\star)) \\ &= \alpha(Q) \end{aligned}$$

Finally it can be checked that if $C(Q) = 0$ then $\alpha(Q) = 0$.

### *Example: the Gaussian channel*

As an application of Theorem 1 we consider antipodal signaling over a Gaussian channel and derive a necessary condition between asynchronism level, block length, and signal to noise ratio (SNR) for achieving reliable communication. Suppose communication takes place over an additive channel $X \to Y = X + Z$ where $X$ denotes the input, $Y$ the output, and where $Z$ is a normally distributed random variable, independent of $X$, with zero mean and unit variance. We consider antipodal signaling, that is $c_i(m) = \pm\sqrt{\text{SNR}}$ for all $i \in \{1, 2, \ldots, N\}$ and $m \in \{1, \ldots, M\}$, where the SNR is some positive constant. Before decoding, the receiver makes a hard decision on each received symbol and declares $+1$ if $Y_i \geq 0$ and $-1$ if $Y_i < 0$. The noise symbol $\star$ equals zero meaning that when no information is sent the receiver declares $+1$ or $-1$ with probability $1/2$. The inputs $+\sqrt{\text{SNR}}$ and $-\sqrt{\text{SNR}}$ are received correctly with probability $1-\varepsilon$ and are flipped with probability $\varepsilon$, where $\varepsilon = e^{-\frac{\text{SNR}}{2}(1+o(1))}$ as the SNR tends to infinity. The discrete channel $Q$ that results from the hard decision procedure is depicted in Fig. 3. From Theorem 1, any cod-

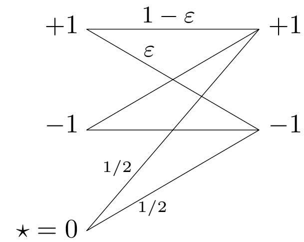


Fig. 3. Antipodal signaling over a Gaussian channel with hard decision at the decoder.

ing strategy that yields vanishing error probability satisfies $\limsup_{N\to\infty} 1/N \ln A \leq \alpha(Q)$ where

$$\begin{aligned}\alpha(Q) &= \max_x D(Q(\cdot|x)||Q(\cdot|\star)) \\ &= \ln 2 - H(\varepsilon) \\ &= \ln 2 - H(e^{-\frac{\text{SNR}}{2}(1+o(1))}) \quad \text{as SNR } \to \infty\end{aligned}$$

with $H(\varepsilon) \triangleq -\varepsilon \ln \varepsilon - (1-\varepsilon)\ln(1-\varepsilon)$. Therefore, as $N$ tends to infinity, in order to achieve reliable communication it is necessary that

$$\frac{1}{N}\ln A \leq \ln 2 - H(e^{-\frac{\text{SNR}}{2}(1+o_1(1))}) + o_2(1)$$

where $o_1(1)$ and $o_2(1)$ are vanishing functions of the SNR and of $N$, respectively. Because of the chosen quantization, in the limit of high SNR we have $\frac{1}{N}\ln A \stackrel{\sim}{\leq} \ln 2$, and an increase in the power results in a negligible increase of the asynchronism level for which reliable communication is possible (for fixed blocklength). To exploit power at high SNR it is necessary to have a finer quantization at the output. Finally notice that for this (quantized) channel the synchronization threshold coincides with the channel capacity. □

While we do not characterize the asynchronism exponent function $\alpha(R,Q)$ for $R > 0$, Theorem 2 provides a non trivial lower bound characterization of $\alpha(R,Q)$, for any $R \in [0, C(Q))$.

We use the notation $(PQ)_Y$ to denote the right marginal of a joint distribution $P(\cdot)Q(\cdot|\cdot)$ and, given a joint distribution $J$ on $\mathcal{X}\times\mathcal{Y}$ we denote by $I(J)$ the mutual information induced by $J$. Also we denote by $\mathcal{P}^{\mathcal{Y}|\mathcal{X}}$ the set of conditional distributions of the form $V(y|x)$ with $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.

**Theorem 2.** *Let $Q$ be a discrete memoryless channel. If for some constants $\alpha \geq 0$, $t_1 \geq 0$, $t_2 > 1$, and input distribution $P$, with $I(PQ) > 0$, the following inequalities*

$$a. \quad \alpha < \inf_{\substack{V\in\mathcal{P}^{\mathcal{Y}|\mathcal{X}} \\ D((PV)_Y||Q(\cdot|\star)) < \frac{t_1\alpha}{\delta(t_1+t_2-1)}}} D((PV)_Y||(PQ)_Y)$$

$$b. \quad \alpha < \min_{\substack{V\in\mathcal{P}^{\mathcal{Y}|\mathcal{X}} \\ I(PV) \leq \frac{t_2\alpha}{\delta(t_1+t_2-1)}}} D(PV||PQ)$$

$$c. \quad \frac{t_1}{t_2} < \frac{D((PQ)_Y||Q(\cdot|\star))}{I(PQ)}$$

*are satisfied for some $\delta \in (0,1)$, then the rate $I(PQ)/t_2$ is achievable at an asynchronism exponent $\alpha$.* □

Note that the conditions $a$ and $b$ in Theorem 2 are easy to check numerically since they only involve convex optimizations. Also notice, on the right hand side of the inequality $b$, the sphere packing exponent function — of the channel $Q$ with input distribution $P$ — evaluated at $\frac{t_2\alpha}{\delta(t_1+t_2-1)}$ (see [6, p.166]).

**Corollary.** *For any channel $Q$ with capacity $C(Q) > 0$, any rate $R \in (0, C(Q))$ can be achieved at a strictly positive asynchronism exponent.*

*Proof of the Corollary.* Consider the inequalities $a$, $b$, and $c$ from Theorem 2. First choose some $P$ and $t_2 > 1$ so that $I(PQ)/t_2 \geq R$ and so that $(PQ)_Y \neq Q(\cdot|\star)$ (this is always possible since $C(Q) > 0$). By setting $t_1 = 0$ the inequality $c$ holds (since its right hand side is strictly positive). Also inequality $a$ holds for any finite $\alpha$ (the infimum equals infinity). For the inequality $b$, observe that its right hand side is a decreasing function of $\alpha$ and has a strictly positive value at $\alpha = 0$ (since $I(PQ) > 0$). It follows that inequality $b$ holds for strictly positive and small enough values of $\alpha$. □

### *A. Coding for asynchronous channels*

In this section we present the coding scheme from which one deduces Theorem 2 and the direct part of Theorem 1. As we will see, our scheme does not subdivide the synchronization problem into a detection problem followed by a message isolation problem: detection and isolation are treated jointly.

The codebook is randomly generated according to some distribution $P$. If the aim is only to reliably communicate at a certain asynchronism exponent $\alpha$, there is some degrees of freedom in choosing $P$. One possible choice is to pick a $P$ that satisfies

$$D((PQ)_Y||Q(\cdot|\star)) + I(PQ) - \ln M/N > \alpha$$

with $D((PQ)_Y||Q(\cdot|\star)) > 0$ and $I(PQ) > 0$, where $M$ represents the size of the message set and $N$ the size of the codewords. In the regime where the asynchronism exponent is close to $\alpha(Q)$ the codewords are mainly composed of the symbol $\arg\max_x D(Q(\cdot|x)||Q(\cdot|\star))$. Indeed, in this asynchronism regime, the main source of error comes from a miss detection of the sent codeword, later referred to as 'false-alarm.' We deal with this source of error by distillating information using codewords with (mostly) symbols that induce output distributions that are 'as far as possible' from the output distribution induced by the $\star$ symbol. Finally if the aim is to accommodate both rate and asynchronism constraints, the distribution $P$ has to satisfy the conditions explicitly stated in Theorem 2.

For the decoder, let us observe first that our communication model admits two sources of error. The first comes from an atypical behavior of the noise during the period when no information is conveyed, which may result in a false-alarm. The second comes from an atypical behavior of the channel during information transmission, which may result in a miss-isolation of the sent codeword. These two sources of error depend on the asynchronism level as well as on the communication rate: the higher the asynchronism the higher the first source of error, the higher the communication rate the higher the second source of error. Accordingly, our decoder is the combination of two criteria parameterized by constants that are chosen based on the level of asynchronism and according to the rate we aim at.

More specifically, the decoder observes the channel outputs $Y_1, Y_2, \ldots$ and makes a decision as soon as it observes $i$ consecutive output symbols, with $i \in [1, 2, \ldots, N]$, that simultaneously satisfy two conditions. The first condition is that these symbols should look 'sufficiently different' from the noise, as measured by the divergence. The second condition is that these symbols must be sufficiently correlated, in a mutual information sense, with one of the codewords. We formalize this below.

For $j \geq i$ we write $x_i^j$ for $x_i, x_{i+1}, \ldots, x_j$. If $i = 1$ we use the shorthand notation $x^j$ instead of $x_i^j$. Given a pair $(x^n, y^n)$ let us denote by $\hat{P}_{(x^n,y^n)}$ the empirical distribution of $(x^n, y^n)$, i.e., $\hat{P}_{(x^n,y^n)}(x,y) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{(x,y)}(x_i, y_i)$ where $\mathbb{1}_{(x,y)}(x_i,y_i) = 1$ if $(x_i,y_i) = (x,y)$, else equals zero. To each message $m \in [1, 2, \ldots, M]$ associate the stopping time

$$\begin{aligned}\tau_m = \inf\Big\{ & n \geq 1 \ : \ \exists i \in \{1,\ldots,N\} \text{ so that} \\ & iD(\hat{P}_{Y^n_{n-i+1}}||Q(\cdot|\star)) \geq t_1 \ln M \quad \text{and} \\ & \min_{k\in[1,\ldots,i]} \Big[kI(\hat{P}_{c^k(m),y^{n-i+k}_{n-i+1}}) \\ & + (i-k)I(\hat{P}_{c^i_{k+1}(m),y^n_{n-i+k+1}})\Big] \geq t_2 \ln M \Big\} \end{aligned} \tag{2}$$

where $t_1 \geq 0$ and $t_2 > 1$ are some fixed threshold constants to be appropriately chosen according to the asynchronism level and desired communication rate. The decoding is made at time

$$\tau = \min_{m\in[1,2,\ldots,M]} \tau_m$$

and the message $\bar{m}$ that is declared is any that satisfies $\tau_{\bar{m}} = \tau$.

It should be emphasized that there may be other sequential decoders that also achieve the synchronization threshold. The one we propose has the property that it also allows for communication at positive rates and positive asynchronism exponents. Also, an interesting feature of the above decoder is that, in addition to operating in an asynchronous setting, it is also almost universal in the sense that its rule does not depend of the channel statistics, except for the noise distribution $Q(\cdot|\star)$. In fact this decoder is an extension of a sequential universal decoder introduced in [17, eq. (10)] for the synchronized setting.

In the context of asynchronous communication, the same decoding rule as above is considered in [16], but without the divergence condition, i.e., a decision is made as soon as for some $m$ and $i$ the condition

$$\min_{k\in[1,\ldots,i]} \Big[kI(\hat{P}_{c^k(m),y^{n-i+k}_{n-i+1}}) + (i-k)I(\hat{P}_{c^i_{k+1}(m),y^n_{n-i+k+1}})\Big] \geq t_2 \ln M$$

holds. With the mutual information condition alone, however, it was not possible to prove that reliable communication can be achieved for asynchronism exponents higher than the capacity of the channel.

### B. Optimal sequential frame synchronization

In this section we consider the situation where there is only one codeword, $c^N$, that needs to be detected at the output of the channel on the basis of sequential observations. In the litterature this setting is referred to as the sequential frame synchronization problem. To identify the instant when the message starts being emitted, the receiver uses a sequential decoder in the form of a stopping time $\tau$ with respect to the output sequence $Y_1, Y_2, \ldots$ If $\tau = n$ the receiver declares that the message started being sent at time $n-N+1$. The associated error probability is defined as

$$\mathbb{P}(\tau \neq \nu + N - 1) \ .$$

We now define the *frame synchronization threshold*.

**Definition 3.** *An* asynchronism exponent $\alpha$ is achievable *if there exists a sequence of pairs codeword/decoder* $\{(c^N, \tau_N)\}_{N\geq 1}$ *such that* $c^N$ *and* $\tau_N$ *operate under asynchronism level* $A = e^{\alpha N}$*, and so that*

$$\mathbb{P}(\tau_N \neq \nu - N + 1) \stackrel{N\to\infty}{\longrightarrow} 0 \ .$$

*The* frame synchronization threshold, *denoted* $\alpha(Q)$*, is the supremum of the set of achievable asynchronism exponents.*

The following theorem shows that the frame synchronization threshold corresponds to the maximum achievable asynchronism exponent at rate $R = 0$. This should not come as a surprise since the limit of asynchronous communication is obtained in the zero rate regime where decoding errors are mainly due to a miss location of the transmitted message.

**Theorem 3.** *The synchronization threshold as defined in Definition 3 is given by*

$$\alpha(Q) = \alpha(R = 0, Q)$$

*where* $\alpha(R, Q)$ *is defined in Definition 1. Furthermore, if the asynchronism exponent is above the synchronization threshold, a maximum likelihood decoder that is revealed the maximum length sequence of size* $A + N - 1$ *makes an error with a probability that tends to one as* $N \to \infty$. □

A direct consequence of Theorem 3 is that a sequential decoder can (asymptotically) locate the sync pattern as well as the optimal maximum likelihood decoder that has access to sequences of maximum size $A + N - 1$, but with much fewer observations.

## REFERENCES


[1] R. Ahlswede and J. Wolfowitz, *Channels without synchronization*, Advances in applied probability **3** (1971), no. 2, 383–403.

[2] Michèle Basseville and Igor Nikiforov, *Fault isolation for diagnosis: nuisance rejection and multiple hypothesis testing*, rapport de recherche 4438, INRIA, 2002.

[3] V. Chandar, A. Tchamkerten, and G.W. Wornell, *Optimal sequential frame synchronization*, submitted to IEEE Trans. Inform. Th. (http://arxiv.org/abs/0708.2911).

[4] T. M. Cover, R. J. McEliece, and E. C. Posner, *Asynchronous multiple-access channel capacity*, IEEE Trans. Inform. Th. **4** (1981), 409–413.

[5] T.M. Cover and J.A. Thomas, *Elements of information theory*, Wiley, New York, 1991.

[6] I. Csiszàr and J. Körner, *Information theory: Coding theorems for discrete memoryless channels*, Academic Press, New York, 1981.

[7] S. Diggavi and M. Grossglauser, *On transmission over deletion channels*, Allerton Conference, Monticello, Illinois, October, 2001.

[8] R. L. Dobrushin, *Shannon's theorems for channels with synchronization errors*, Problems Information transmission **3** (1967), no. 4, 11–26.

[9] E. Drinea and M. Mitzenmacher, *A simple lower bound for the capacity of the deletion channel*, IEEE Trans. Inform. Th. **52** (2006), 4657–4660.

[10] J.Y.N. Hui and P.A. Humblet, *The capacity of the totally asynchronous multiple-access channel*, IEEE Trans. Inform. Th. **31** (1985), no. 2, 207–216.

[11] T.Z. Lai, *Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems*, IEEE Trans. Inform. Th. **46** (2000), 595–607.

[12] I. V. Nikiforov, *A generalized change detection problem*, IEEE Trans. Inform. Th. **41** (1995), 171–187.

[13] G.S. Poltyrev, *Coding in an asynchronous multiple-access channel*, Problems Inform. Trans. **36** (1983), 12–21.

[14] C. E. Shannon, *A mathematical theory of communication*, The Bell Sys. Tech. Journal **27** (1948), 379–423.

[15] A. Tchamkerten, V. Chandar, and G.W. Wornell, *Communication under strong asynchronism*, submitted to IEEE Trans. Inform. Th. (http://arxiv.org/abs/0707.4656).

[16] A. Tchamkerten, A. Khisti, and G.W. Wornell, *Information theoretic perspectives on synchronization*, IEEE Intl. Sympo. on Info. Th. (ISIT), 2006, pp. 371–375.

[17] A. Tchamkerten and I. E. Telatar, *Variable length coding over an unknown channel*, IEEE Trans. Inform. Th. **52** (2006), no. 5, 2126–2145.

[18] S. Verdú, *The capacity region of the symbol-asynchronous gaussian multiple-access channel*, IEEE Trans. Inform. Th. **35** (1989), no. 4, 733–751.