

A Novel Asynchronous Communication Paradigm: Detection, Isolation, and Coding

Venkat Chandar¹, Aslan Tchamkerten², and Gregory Wornell³

¹ Department of EECS, MIT,
77 Massachusetts Avenue, Cambridge, MA 02139-4307, USA
vchandar@mit.edu

² Communications and Electronics Department, Telecom ParisTech
75634 Paris Cedex 13, FR
aslan.tchamkerten@telecom-paristech.fr

³ Department of EECS, MIT,
77 Massachusetts Avenue, Cambridge, MA 02139-4307, USA
gww@mit.edu

Abstract. In this work we develop a simple theoretical framework for asynchronous point-to-point communication which can be seen as a generalization of the detection and isolation model introduced in (Nikiforov, 1995). A notion of capacity is defined and corresponding upper and lower bounds are derived.

Keywords. Detection and isolation problem, digital communication, discrete-time communication, error exponent, point-to-point communication, quickest detection, sequential analysis, stopping times, synchronization

1 Introduction

We describe a simple model for asynchronous point-to-point communication, the investigation of which involves a combination of communication theoretic and sequential analysis. Our contribution consists of characterizations of the capacity region for discrete memoryless channels.

Section 2 describes our model and states the main results, and Section 3 provides a sketch of the proofs. All the material presented here refers to the works (Tchamkerten et al., 2007) and (Tchamkerten et al., 2009).

2 Problem formulation and results

We first introduce our model then discuss its connections with the detection and isolation model (Nikiforov, 1995).

We consider discrete-time communication over a discrete memoryless channel characterized by its finite input and output alphabets \mathcal{X} and \mathcal{Y} , respectively, and transition probability matrix $Q(y|x)$, for all $y \in \mathcal{Y}$ and $x \in \mathcal{X}$ (Csiszar and Korner, 1981). Throughout the paper we assume that, for all $y \in \mathcal{Y}$, there is some $x \in \mathcal{X}$ for which $Q(y|x) > 0$.

There are $M \geq 2$ messages $m \in \{1, 2, \dots, M\}$. For each message m there is an associated codeword $c^N(m) \triangleq c_1(m)c_2(m)\dots c_N(m)$, a string of length N composed of symbols from \mathcal{X} . The M codewords form a codebook \mathcal{C} . Communication takes place as follows. The transmitter selects a message m randomly and uniformly over the message set and starts sending the corresponding codeword $c^N(m)$ at a random time ν , unknown to the receiver, independent of $c^N(m)$, and uniformly distributed in $\{1, 2, \dots, A\}$. The transmitter and the receiver know the integer $A \geq 1$, which is interpreted as their ‘asynchronism level.’ If $A = 1$ the channel is said to be synchronized. It should be emphasized that only one codeword is sent within the time interval $\{1, 2, \dots, A + N - 1\}$.

During the information transmission the receiver observes a noisy version of the sent codeword, while before and after the information transmission it observes only noise; conditioned on the value of ν and on the message m to be conveyed, the receiver observes independent symbols Y_1, Y_2, \dots distributed as follows. If $i \in \{1, 2, \dots, \nu - 1\}$ or $i \in \{\nu + N, \nu + N + 1, \dots, A + N - 1\}$, the distribution is $Q(\cdot|\star)$ for some fixed $\star \in \mathcal{X}$. At any time $i \in \{\nu, \nu + 1, \dots, \nu + N - 1\}$ the distribution is $Q(\cdot|c_{i-\nu+1}(m))$. It

should be emphasized that the transition probability matrix $Q(\cdot|\cdot)$ together with the ‘noise’ symbol ‘ \star ’ characterizes the communication channel. In particular, the ‘ \star ’ is not a parameter of the transmitter. This symbol can, however, be used for the codebook design.

The decoder consists of a sequential test (τ, ϕ) , where τ is a stopping time — bounded by $A + N - 1$ — with respect to the output sequence Y_1, Y_2, \dots indicating when decoding happens, and where ϕ denotes a terminal decision rule that declares the decoded message. Formally, ϕ is any \mathcal{F}_τ -measurable map taking values in $\{1, 2, \dots, M\}$, where $\mathcal{F}_1, \mathcal{F}_2, \dots$ is the natural filtration induced by the process Y_1, Y_2, \dots .

We are interested in *reliable and quick decoding*. To that aim we first define the average decoding error probability (given a codebook and a decoder) as

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

where \mathcal{E} indicates the event that the decoded message does not correspond to the sent message, and where $_{m,\nu}$ indicates the conditioning on the event that message m starts being sent at time ν . With a slight abuse of notation m and ν represent random variables or their realizations. From the context no confusion should occur.

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.

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

where \ln denotes the natural logarithm. The expectation $\mathbb{E}(\tau - \nu)^+$ is defined as

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

where x^+ denotes $\max\{0, x\}$, and where $\mathbb{E}_{m,\nu}$ denotes the expectation with respect to $\mathbb{P}_{m,\nu}$. With the above definitions we now introduce the notion of capacity region and the notion of synchronization threshold associated to an asynchronous discrete memoryless channel. We first define an (R, α) coding scheme.

Definition 1 ((R, α) coding scheme) *Given a channel Q , a pair (R, α) is achievable if there exists a sequence $\{(\mathcal{C}_N, (\tau_N, \phi_N))\}_{N \geq 1}$ of pairs codebook/decoder, labeled by the codebook length N , that asymptotically achieves a rate R at an asynchronism exponent α . This means that, for any $\varepsilon > 0$ and any N large enough, the pair $(\mathcal{C}_N, (\tau_N, \phi_N))$*

- operates under asynchronism level $A = e^{(\alpha - \varepsilon)N}$;
- yields an average rate at least equal to $R - \varepsilon$;
- achieves an average error probability $\mathbb{P}(\mathcal{E})$ at most equal to ε .

Given a channel Q , an (R, α) coding scheme is a sequence $\{(\mathcal{C}_N, (\tau_N, \phi_N))\}_{N \geq 1}$ that achieves a rate R at an asynchronism exponent α as $N \rightarrow \infty$.

Definition 2 (capacity region of an asynchronous discrete memoryless channel) *The capacity region of an asynchronous discrete memoryless channel with (synchronized) capacity $C(Q)$ is the function*

$$\begin{aligned} [0, C(Q)] &\rightarrow \mathbb{R}_+ \\ R &\mapsto \alpha(R, Q) \end{aligned}$$

where $\alpha(R, Q)$ is the supremum of the set of asynchronism exponents that are achievable at rate R .

In some sense, the closest sequential decision problem our model relates to is the ‘detection and isolation problem,’ introduced by Nikiforov in 1995. A process Y_1, Y_2, \dots 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, \dots the goal is to quickly react to the statistical change and isolate its cause, i.e., the post-change distribution, with a false-alarm constraint.

In the synchronization problem and in the detection and isolation problem the goal is to quickly identify the cause of a change in distribution. There are, however, important distinctions between these two problems. In the detection and isolation problem it is assumed that, once the observed process changes distributions, it remains in that post-change state forever. Hence, with arbitrarily high probability a correct decision can be made simply by waiting long enough. This is not possible in the synchronization problem since the transmitted message induces only a *local* change in distribution — after the codeword transmission the distribution returns to its pre-change distribution. We also note that the synchronization problem is a quickest detection and isolation decoding problem coupled with a coding problem. Each codeword defines a ‘signature of change’ whose goal is to allow a quick detection and isolation by the receiver — the codebook design is made as a function of its cardinality, the desired reaction delay, the channel, and the asynchronism level. Instead, the detection and isolation problem involves only the decoding part since the post-change distributions are given.

Finally, in the synchronization problem the goal is to compute the capacity region. This region characterizes (in the limit of long codewords) an optimal tradeoff between the number of messages, the expected reaction delay, the asynchronism level, and the error probability. In contrast, for the detection and isolation problem, the analysis has focused on the ‘zero-rate regime,’ i.e., on the minimum reaction delay in the limit of small error probabilities, the number of messages being kept fixed.

Note that, for a given channel Q , the asynchronism exponent function $\alpha(R, Q)$ as defined above is non-increasing in R . Therefore the highest asynchronism level for which reliable communication is still possible is obtained for zero-rate communication. This motivates the following definition.

Definition 3 (synchronization threshold) *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., $\alpha(Q) = \alpha(R = 0, Q)$.*

Theorem 1 *For any discrete memoryless channel Q , the synchronization threshold is given by*

$$\alpha(Q) = \max_{x \in \mathcal{X}} D(Q(\cdot|x) || Q(\cdot|\star))$$

where $D(Q(\cdot|x) || Q(\cdot|\star))$ is the divergence (Kullback-Leibler distance) between $Q(\cdot|x)$ and $Q(\cdot|\star)$.

The next two theorems provide inner and outer bounds on the capacity region of asynchronous channels. Before stating these results we introduce some notation. We denote by \mathcal{P}_X and \mathcal{P}_Y the set of probability measures defined over the finite alphabet \mathcal{X} and \mathcal{Y} , respectively. The set $\mathcal{P}_{Y|X}$ consists of all conditional distributions of the form $Q(y|x)$ with $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. The notation $I(PQ)$, for $P \in \mathcal{P}_X$ and $Q \in \mathcal{P}_{Y|X}$, stands for the mutual information induced by the joint distribution $P(\cdot)Q(\cdot|\cdot)$, i.e.

$$I(PQ) \triangleq \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x)Q(y|x) \log \left(\frac{P(x)Q(y|x)}{P(x)(PQ)_Y(y)} \right)$$

where $(PQ)_Y(y) \triangleq \sum_{x \in \mathcal{X}} P(x)Q(y|x)$. Given $W_1, W_2 \in \mathcal{P}_{Y|X}$ and $P \in \mathcal{P}_X$, $D(W_1 || W_2 | P)$ denotes the Kullback-Leibler distance between $P(\cdot)W_1(\cdot|\cdot)$ and $P(\cdot)W_2(\cdot|\cdot)$. Throughout the paper we often use the shorthand notation Q_\star to denote $Q(\cdot|\star)$. The cardinality of a set S is denoted $|S|$. Finally, a collection of sequences all of the same length is said to have constant composition if the relative frequency of occurrence of symbols in a sequence is the same for all the sequences, i.e., the sequences all have the same type (empirical distribution).

Theorem 2 Let $\alpha \geq 0$ and let $P \in \mathcal{P}_X$ be such that, for any $V \in \mathcal{P}_Y$, at least one of the following inequalities

$$\begin{aligned} D(V||(\mathcal{P}Q)_Y) &> \alpha \\ D(V||Q_\star) &> \alpha \end{aligned}$$

holds, i.e.

$$\min_{V \in \mathcal{P}_Y} \max\{D(V||(\mathcal{P}Q)_Y), D(V||Q_\star)\} > \alpha.$$

Then, the pair $(R = I(\mathcal{P}Q), \alpha)$ is achievable over the channel Q .

Theorem 3 Let $\{(\mathcal{C}_N, (\tau_N, \phi_N))\}_{N \geq 1}$ achieve (R, α) with $\alpha > 0$, $R > 0$ and fix $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, and $\varepsilon_3 > 0$. There exist distributions $P_1, P_2 \in \mathcal{P}_X$, such that

- i. $\frac{R - \varepsilon_2}{1 + \varepsilon_1} \leq I(P_1Q)$.
- ii. $\alpha \leq \varepsilon_3 + \min_{W \in \mathcal{P}_{Y|X}} \max\{D(W||Q_\star|P_2), D(W||Q|P_2)\}$.
- iii. $\alpha \leq \frac{\min\{[(1 + \varepsilon_1)\mathbb{E}(\tau_N - \nu)^+], N\}}{N} (I(P_1Q) - R + D((P_1Q)_Y||Q_\star))$.

3 Proof Sketches

In this section, we sketch the main ideas involved in the proofs of Theorems 1, 2, and 3. Full proofs can be found in (Tchamkerten et al., 2007) and (Tchamkerten et al., 2009).

3.1 Proof Sketch of Theorem 1

Theorem 1 characterizes the synchronization threshold, i.e., the highest asynchronism level under which reliable communication is still possible, regardless of the communication rate (i.e., regardless of the number messages). As such it is achieved with codebooks with a minimal number of codewords, i.e., two. When there are only two codewords, in the high asynchronism level regime the probability of error is mainly governed by the probability of mislocating the sent codeword. (Given that the sent codeword has been correctly located, a correct message isolation can be performed with high probability since two messages can, in principle, be designed very differently — say, their Hamming distance can be made very large.) To minimize the probability of mislocating the sent codeword, it should be clear that the best codewords to use are ones that induce output symbols that look as ‘far as possible’ to typical noise symbols.

Let x^* achieve $\max_{x \in \mathcal{X}} D(Q(\cdot|x)||Q(\cdot|\star))$. Then, consider the codeword that consists of N repetitions of x^* . Intuitively, this codeword should be more distinguishable from the noise than any other possible codeword. Now, from standard large deviations results, with probability $\Theta(e^{-ND(Q(\cdot|x^*)||Q(\cdot|\star))})$, a block of N channel outputs generated by noise (i.e., the \star symbol) will look as if it were generated by the codeword above. Thus, if $A = e^{N(D(Q(\cdot|x^*)||Q(\cdot|\star)) + \varepsilon)}$, $\varepsilon > 0$, we expect that the noise will generate many blocks of N symbols that look as if they were generated by the codeword. In this case it is impossible to locate the codeword reliably, i.e., determine ν . Intuitively, it is plausible that if we cannot determine ν , then we cannot identify the sent codeword. This implies that $\alpha(Q)$ is upper bounded by $D(Q(\cdot|x^*)||Q(\cdot|\star))$.

To prove that any asynchronism exponent lower than $D(Q(\cdot|x^*)||Q(\cdot|\star))$ is achievable, one proceeds as follows. The idea is to consider a codebook where every codeword starts with, say $N - \log(N)$ repetition of x^* , and where the last $\log(N)$ symbols are drawn from a code that achieves a small probability of error when used over the synchronized channel Q . The rate this code achieves over the synchronized channel is irrelevant. It should also be said that $\log N$ is merely a choice, any sublinear growing function of N works as well. A similar argument as above reveals that when $A = e^{(N - \log(N))(D(Q(\cdot|x^*)||Q(\cdot|\star)) - \varepsilon)}$, $\varepsilon > 0$, the noise never produces a block of $N - \log(N)$ channel outputs that look as if they had been generated by the first $N - \log(N)$ symbols of a codeword. Thus, in this asynchronism regime, it is possible to locate ν . Once ν is known, we are essentially in the synchronized setting and can reliably decode. Therefore, communication is possible for asynchronism exponents arbitrarily close to $D(Q(\cdot|x^*)||Q(\cdot|\star))$.

3.2 Proof Sketch of Theorem 2

Like many achievability arguments in information theory, the proof of Theorem 2 uses a random coding argument. Instead of constructing a specific coding scheme that achieves the performance promised by Theorem 2, the idea is to pick a random scheme, from a suitable set, and show that, on average, a randomly selected scheme achieves an (R, α) given by Theorem 2. This implies the existence of a particular (non-random) coding scheme that also achieves this (R, α) .

Specifically, we consider a codebook where each symbol of each codeword is generated i.i.d. according to some probability distribution P on \mathcal{X} .

To decode a sent message from the received symbols, a sequential decoder is used. Intuitively, the idea behind decoding is that at each time, the decoder makes a decision whether to stop or to keep going based on the last N received symbols. Specifically, the decoder looks at the type (empirical distribution) of the last N symbols, and considers how likely this type is to be observed under the noise distribution $Q(\cdot|\star)$. If the probability of this type under the noise distribution is much smaller than $\frac{1}{A}$, this suggests that the last N symbols were probably generated by a codeword, so the decoder should stop and look for the most likely codeword assuming that the last N symbols represent observations of some codeword through the channel Q . For technical reasons, the decoder is more complex than this, but this is the main idea.

To see where the condition

$$\begin{aligned} D(V||(\mathcal{P}Q)_Y) &> \alpha \\ D(V||Q_\star) &> \alpha \end{aligned}$$

comes from, note that under this condition, we can classify empirical distributions V as “noise” or “codeword” distributions. Specifically, we say that V is a noise distribution if $D(V||(\mathcal{P}Q)_Y) < \alpha$, and that V is a codeword distribution otherwise. Then, standard large deviation results are used to show that the probability that the decoder outlined above stops before time ν or after time $\nu + N$ is much less than $e^{-\alpha N}$, and therefore it is reasonable to expect that the probability of error goes to 0 and that the expected delay is about N . Thus, the random code ensemble with the decoder outlined above should be able to achieve rate $I(\mathcal{P}Q)$ if the condition is satisfied.

3.3 Proof Sketch of Theorem 3

Theorem 3 has several parts. The first step consists of extracting a constant composition codebook, i.e., given C_N , we construct a codebook C'_N by expurgation, i.e., by choosing a subset of the codewords in C_N . C'_N is chosen to have constant composition P_1 over the first m_N symbols, and constant composition P_2 over all N symbols. Part *i* of Theorem 3 is proved by showing that $I(P_1Q)(1 + o(1)) \geq \frac{\ln|C'_N|}{m_N}$ and that $\frac{\ln|C'_N|}{m_N} \geq \frac{R - \epsilon_2}{1 + \epsilon_1}(1 + o(1))$. The first inequality is proved by using the fact that if a codebook with constant composition P is used over a (synchronized) channel Q , the maximum possible rate is $I(PQ)$. The second inequality essentially states that the expurgation procedure does not reduce the number of codewords substantially.

Part *iii* of Theorem 3 is a technical lemma used to formally prove that the expurgation procedure does not reduce the number of codewords substantially. The proof is similar to the proof of Theorem 1, so we will not elaborate on it here.

The most interesting part of Theorem 3 is part *ii*. The intuition behind the proof is that the decoder used to prove Theorem 2 should in some sense be close to optimal. Specifically, it seems plausible that since codewords are only N symbols long, a decoder that is constrained to make a decision based on just the last N observed symbols, as opposed to the entire sequence of past observations, should do nearly as well as the unconstrained decoder. If we constrain ourselves to such decoders, and then suitably modify the notion of noise distribution and codeword distribution to apply to constant composition codes instead of random codes, the proof technique described in section 3.2 suggests that to achieve an asynchronism

exponent of α , we need that for any $W \in \mathcal{P}_{Y|X}$ at least one of the following two inequalities holds:

$$\begin{aligned} D(W||Q_{\star}|P_2) &> \alpha - \varepsilon_3 \\ D(W||Q|P_2) &> \alpha - \varepsilon_3 ; \end{aligned}$$

The proof in some sense tries to reverse this reasoning. Specifically, the formal proof proceeds by contradiction, i.e., we assume that the probability of error goes to 0, and that the condition above is violated. Then, we use this to show that the expected reaction delay grows exponentially with N . Since $\ln |C'_N|$ can grow at most linearly with N , this proves that the rate goes to 0.

To show that the expected reaction delay is so large, we use a change of measure argument. Intuitively, the idea is that if the divergence condition is violated, then the chance that the channel behaves so atypically during the time period $\nu, \dots, \nu + N - 1$ that the channel output during this time period looks like it could have been generated by noise is more than $e^{-(\alpha - \varepsilon_3)N}$. Now, because the probability of error is small, with high probability the decoder stops after time ν . Armed with these two facts, the proof is completed by considering the channel output distribution when the input distribution is \star for all time, i.e., the input is \star even during the time period $\nu, \dots, \nu + N - 1$. Roughly stated, the second fact says that for this distribution, the decoder will not stop until it reaches the end of the output sequence, i.e., the expected delay is $\Omega(e^{\alpha N})$. On the other hand, the first fact says that the actual output distribution is not very different than the modified distribution, so it is reasonable to expect that the expected reaction delay for the actual output distribution is

$$\Omega(e^{\alpha N})e^{-(\alpha - \varepsilon_3)N} = \Omega(e^{\varepsilon_3 N}).$$

This rough argument suggests that the expected delay grows exponentially with N .

References

- Csiszar, I. and Korner, J. (1981). *Information Theory: Coding Theorems for Discrete Memoryless Systems*
- Nikiforov, I. V. (1995). *A generalized change detection problem*, IEEE Trans. Inform. Th., Vol. 41, 171187
- Tchamkerten, A., Chandar, V., and Wornell, G. W. (2007) *Communication Under Strong Asynchronism*, available at <http://arxiv.org/abs/0707.4656>
- Tchamkerten, A., Chandar, V., and Wornell, G. W. (2009) *Fundamental Tradeoffs for One-Shot Asynchronous Communication* preprint