Information Theoretic and Statistical Perspectives on Asynchronism

Aslan Tchamkerten

Digital Communication Group

Communications and Electronics Department

Telecom ParisTech

submitted in partial fulfillment of the requirements for the

'Habilitation à diriger des recherches'

Jury:
Prof. Venkat Anantharam (UC Berkeley, CA, US)
Prof. François Baccelli (ENS, Paris)
Prof. Giuseppe Caire (USC, CA, US)
Prof. Philippe Loubaton (Université Marne-la-Vallée)
Prof. Sergio Verdù (Princeton University, NJ, US)

April 2011

Foreword

This report describes some of my main research contributions since my Ph.D. in 2005. The unifying theme is ‘asynchronism,’ which I have been investigating from both a communication and a statistical perspective. The results are presented without proofs. The proofs can be found in related papers mentioned at the beginning of each chapter.

Contents

1

Introduction

Consider communication with perfect feedback over a binary erasure channel as depicted in Fig. 1.1. This is a channel where each input bit gets erased with some fixed probability p, and remains unchanged with probability $1 - p$. Perfect feedback means that the encoder gets to observe every channel output in a causal fashion—at time t the encoder knows exactly what was received at time $t - 1$.

Suppose we have a 1-bit message that we transmit using a repetition code. To communicate message '0,' the encoder sends 0000... and similarly for message '1.' The receiver decodes the first time a non-erasure occurs. Because of perfect feedback, the transmitter knows when decoding happens.

This simple communication strategy is 'universally optimal' for the family of binary erasure channels in the following sense. First, it is 'universal' since neither the encoding nor the decoding rules depend on the channel parameters—here the erasure probability p. Second, its long-term communication rate is equal to the channel capacity. Third, it is optimal in the tradeoff between error probability and delay: for the same communication delay, no strategy can achieve a lower error probability—here equal to zero.

The objective of my Ph.D. thesis was to investigate whether there exists universally optimal feedback schemes for families of channels other than the binary erasure. Surprisingly perhaps, the answer turns out to be 'yes' for at least two other families, the binary symmetric and the Z families [TT06]. This is one of the main results of my Ph.D. thesis.

Key to the above result is the (strong) assumption that the feedback channel is *perfect*. On the one hand, this allows the transmitter to adjust its encoding strategy according to the channel behavior; if the channel behaves badly, the encoder may locally add some redundancy. On the other hand, perfect feedback often allows to

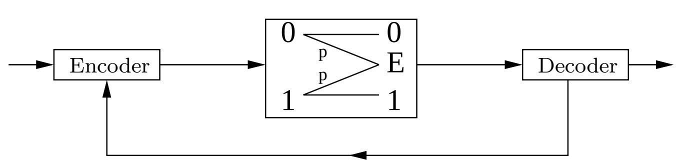

Figure 1.1: The BEC with perfect feedback.

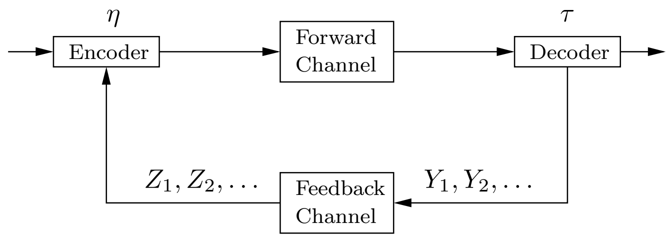

Figure 1.2: The decoding time τ of the current message depends on the outputs of the forward channel. The encoder decides to start sending the next message at time $\eta+1$, based on the outputs of the feedback channel. If the feedback channel is noisy, η and τ need not coincide, which results in a loss of synchronization between the transmitter and the receiver.

reduce communication delay by having the decoding instant depend on the channel outputs. For instance, by decoding as soon as the 'level of confidence' about the sent message is high enough, the communication delay gets shorter whenever the channel behaves well. Because of perfect feedback, the transmitter can look over the shoulders of the receiver, know when decoding happens, and thereby know when it can start sending the next message. That is, perfect feedback guarantees synchronization between the transmitter and the receiver when a variable length code is used. In turn, variable length coding can be universally optimal, similarly as the 'send until a non-erasure' coding strategy that we described above for the family of binary erasure channels.

What happens when feedback is non-perfect, is it still possible to communicate universally as well as the best feedback schemes tuned for the specific channel under use? This natural question also points to a broader concern in information theory: how practically meaningful are theoretical results stated under ideal assumptions—in practice feedback is ubiquitous but hardly 'perfect.'

When there is noise in the feedback channel, we have a situation like the one depicted in Fig. 1.2, where τ denotes the decoding time of the current message, and where $\eta + 1$ denotes the instant when the encoder starts sending the next message. Since feedback is noisy, the encoder can now at best observe only a noisy version of the symbols received by the decoder. Hence, if τ depends on the forward channel outputs, we may have $\eta \neq \tau$, which results in a loss of *synchronization* between the transmitter and the receiver. For instance, if $\eta < \tau$, the encoder starts sending the next message while the current message has not yet been decoded. This synchronization problem imposes to restrict communication to block coding where decoding always happens at a fixed time known by both the transmitter and the receiver. In turn, block coding is much less efficient than variable length coding in the tradeoff between error probability and delay (in particular in the high rate regime [Dob67, Bur76]).

The penalty caused by a potential lack of synchronization in the above setting motivated me to investigate information theoretic aspects of asynchronous communication after my Ph.D. It turned out this was very much an open area. The theoretical literature on asynchronous communication was (and, to a large extent still is) sparse. Asynchronous communication has been representing one of my main research efforts since my Ph.D., and is the focus of Chapter 2. There, fundamental communication limits in terms of delay, energy, rate, and error probability are provided for point-to-point communication.

The other main research direction I have been pursuing since my Ph.D., is the investigation of a statistical decision problem, also motivated by communication with noisy feedback. Let us again consider the channel in Fig. 1.2, and suppose the decoder simply forwards its received symbols to the encoder via the feedback channel. That is, $Y_1, Y_2, \ldots$ also represent the outputs of the forward channel, and the encoder observes a noisy version of these symbols that are denoted by $Z_1, Z_2, \ldots$. Now suppose the decoding instant is a stopping time τ defined over $Y_1, Y_2, \ldots$. What is the best transmitter estimate for τ?

This motivates the following general statistical decision problem. Suppose there is a stopping time τ defined on a stochastic process $Y = Y_1, Y_2, \ldots$. Statistician doesn't have access to Y directly but, instead, has access to correlated observations $Z = Z_1, Z_2, \ldots$. Statistician wants to find a stopping time η with respect to Z that best tracks τ, e.g., so that to minimize the mean absolute deviation $\mathbb{E}|\eta - \tau|$. Interestingly, this problem formulation applies in several areas including monitoring, forecasting, and finance, in addition to generalizing the well known change-point detection problem, a central problem in quality control that dates back to the early 1940's. Chapter 3 is devoted to the TST problem.

This report ends with Chapter 4 where I provide a brief description of Ph.D. thesis for which I have been involved as a (co-)supervisor.

Chapters 2 and 3 can be read independently. Chapter 2 supposes that the reader has some background in Information Theory, Probability, and Statistics, while Chapter 3 assumes some background in Probability and Statistics.

2

Fundamental Limits on Asynchronous Communication

In this chapter, we first present the asynchronous communication model proposed in [TCW09]. Then, in Sections 2.2 and 2.3, we present related capacity and capacity per unit cost results, respectively. These limits characterize the maximum number of bits per channel use that can reliably be transmitted as well as the minimum energy needed to transmit one bit of information reliably. The results presented in this Chapter are from [CTW08], [TCW09], [TCW08], [CTW09], [TCW], and [CTT10].

2.1 Communication model

Shannon's original point-to-point communication model [Sha48] assumes that information is available at the transmitter at time 1 and that communication takes place between time 1 and time N, for some $N \geq 1$ known to both the transmitter and the receiver. We extend this model to address asynchronism between the transmitter and the receiver. The proposed model captures the following features:

- Information is available at the transmitter at a random time;
- Outside the information transmission period, the transmitter stays idle and the receiver observes noise;
- The receiver decodes without knowing the information arrival time at the transmitter.

Communication is assumed to be discrete-time and carried 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 $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.

The message to be transmitted is available at the transmitter at a random time ν, uniformly distributed over $\{1, 2, \ldots, A\}$, where the integer $A \geq 1$ characterizes the *asynchronism level* between the transmitter and the receiver. Only one message arrives over the period $[1, 2, \ldots, A + N - 1]$. If $A = 1$, the channel is said to be synchronous. The capacity of the synchronous channel is denoted by C.

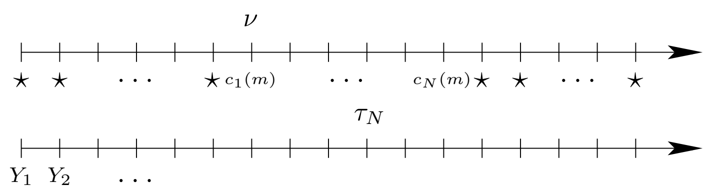

Figure 2.1: Time representation of what is sent (upper arrow) and what is received (lower arrow). The '$\star$' represents the 'idle' symbol. Message m is being sent from time ν up to time time $\nu + N - 1$. Decoding occurs at time τ.

For a given message $m \in \{1, 2, \ldots, M\}$ to be transmitted, the transmitter sends a codeword $c^N(m)$, a sequence of length N composed of symbols from $\mathcal{X}$. Hence, from time ν up to time $\nu + N - 1$, the receiver observes a noisy version of $c^N(m)$. Outside this transmission period, the receiver observes pure noise, which we model as independent channel outputs distributed according to $Q(\cdot|\star)$, where $\star \in \mathcal{X}$ denotes the special 'idle' symbol.

Knowing the asynchronism level A, but not the value of ν, the receiver decodes by means of a sequential test (τ, ϕ), where τ is a stopping time, bounded by $A+N-1$, with respect to the output sequence $Y_1, Y_2, \ldots$ indicating when decoding happens, and where ϕ denotes a decision rule that declares the decoded message (see Fig. 2.1).[1]

The above asynchronous communication model complements the well-known insertion, deletion, and substitution (IDS) channel model proposed by Dobrushin [Dob67]. In the IDS model, the time at which transmission begins is known to the receiver. However, each time a symbol from the codeword is transmitted, a string of symbols of variable length (possibly even length zero) is received.[2] Hence, the duration of the information transmission is random. By contrast, in our model the receiver doesn't know the time at which transmission starts, but the information transmission period is constant—equal to the codeword length. Hence, the IDS channel is suitable for modeling asynchronism caused by the communication medium itself, whereas our model is suitable for modeling asynchronism caused by a bursty source of information. It is also worth remarking that, in contrast to our model, the intuitive notion of 'asynchronism level' for a channel is more difficult to capture succinctly with the IDS model since any reasonable such notion would depend on the associated channel transition probabilities.

[1]To be more precise, ϕ is any $\mathcal{F}_\tau$-measurable function that takes value in the message set, where $\mathcal{F}_t$ is the sigma field generated by $Y_1, Y_2, \ldots, Y_t$.

[2]The IDS channel is characterized by the set of all conditional output distributions $Q(\mathbf{y}|x)$ for each $x \in \mathcal{X}$, where $\mathbf{y}$ is a string of some length (even zero) of symbols from $\mathcal{Y}$.

2.2 Bits per channel use

In this section we derive limits on the maximum number of bits per channel use that can be communicated over an asynchronous channel. First, we consider the simple problem of 'communicating' a single message, the decoding problem being only to detect it accurately. Second, we consider the transmission of one out of many possible messages, the decoding problem being now to detect and locate the transmitted message reliably. At the end of this section we discuss the performance of training based schemes where codewords start with a 'sync' preamble to help the decoder detect the codeword. We show that these practically widely used communication architectures need not be optimal.

Throughout this chapter, X always refers to a random channel input and Y its corresponding ouput, i.e., if $X \sim P$ then $XY \sim P(\cdot)Q(\cdot|\cdot)$. Also, Y_x denotes the ouput of the channel when the channel input symbol is $x \in \mathcal{X}$. Thus, for instance, $Y_\star$ denotes the random 'pure noise' output of the channel when the transmitter is idle.

2.2.1 The single message case: sequential 'frame synchronization'

Assume there is just one possible message, i.e., $M = 1$, and that the receiver wants to quickly locate this message on the basis of sequential observations. This is one of the most basic synchronization problems which Massey refers to as the 'one-shot' frame synchronization problem in [Mas72].

A key parameter we shall be concerned with is

$$\alpha \triangleq \frac{\log A}{N} ,$$

which we call the asynchronism exponent. Theorem 1 below characterizes the largest value of α for which it is possible to perfectly locate the codeword at the channel output with arbitrarily high probability.[3]

For a given codeword $c^N = c_1, c_2, \ldots, c_N$ and decoding detection rule τ_N, we first define the error probability as

$$\mathbb{P}(\tau_N \neq \nu + N - 1) ,$$

i.e., the probability of not stopping exactly at the end of the codeword transmission. Second, we define the *synchronization threshold*.

Definition 1. *An* asynchronism exponent α is achievable *if there exists a sequence of pairs codeword/decoder* $\{(c^N, \tau_N)\}_{N\geq 1}$ *such that, for any* $\varepsilon > 0$ *and all* N *large enough,* (c^N, τ_N) *operates under asynchronism level* $A = 2^{(\alpha-\varepsilon)N}$*, and so that*

$$\mathbb{P}(\tau_N \neq \nu - N + 1) \leq \varepsilon .$$

[3] In this setting where there is a unique codeword, it is custom to call it '(sync) preamble' instead of 'codeword' as this sequence of bits carries no information and is usually used for synchronization purposes only.

The synchronization threshold*, denoted by α^T, is the supremum of the set of achievable asynchronism exponents.*

THEOREM 1 ([CTW08, TCW09]). *The synchronization threshold is given by*

$$\alpha^T = \max_{x \in \mathcal{X}} D(Y_x || Y_\star) \,, \tag{2.1}$$

where $D(Y_x||Y_\star)$ is the Kullback-Leibler distance between $Q(\cdot|x)$ and $Q(\cdot|\star)$. 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 the theorem is that a sequential decoder can (asymptotically) locate the sync pattern as well as the optimal maximum likelihood decoder that operates on a non-sequential basis having access to sequences of maximum size $A + N - 1$. The last part of the theorem indicates that the synchronization threshold refers to a phase transition: for $\alpha < \alpha^{\mathrm{T}}$ the error probability tends to zero whereas for $\alpha > \alpha^{\mathrm{T}}$ the error probability tends to one.

Finally note that $D(Y_x||Q_\star)$ in (2.1) is simply the Chernoff exponent for discriminating pure noise from a string composed of symbol x at the output of the channel. Thus, not too surprisingly, an optimal codeword is mostly composed of symbols that maximize this exponent.[4]

2.2.2 Multiple messages: information transmission

In this section, we consider the case of multiple possible messages, i.e., where $M \geq 2$. For a given set of codewords $\mathcal{C} = \{c^N(m)\}_{m=1}^M$, which we refer to as 'codebook,' and for a given decoder, we first define the maximum over messages, average over time, decoding error probability as

$$\mathbb{P}(\mathcal{E}|\mathcal{C}) = \max_m \frac{1}{A} \sum_{t=1}^{A} \mathbb{P}_{m,t}(\mathcal{E}), \tag{2.2}$$

where $\mathcal{E}$ indicates the event that the decoded message does not correspond to the sent message, and where the subscripts '$_{m,t}$' indicate conditioning on the event that message m starts being sent at time $\nu = t$. In the expression (2.2), with a slight abuse of notation, we let $\mathcal{C}$ denote the codebook together with the decoder

[4] Certain readers may wonder why optimal codewords are 'mostly' and not 'only' composed of symbols that maximize the exponent. If a constant codeword is used, with non-vanishing probability the receiver will observe a sequence of length N, typical with the codeword, and that is shifted by one unit from the transmitted codeword (and in fact this remains true for any fixed length shift). Because of this, a constant codeword cannot be pinpointed (asymptotically) with arbitrarily high probability. This imposes to have a codeword with sufficiently small Hamming distance with any of its (circular) shifts. Interestingly, it is possible to design a sufficiently 'constant' codeword that has low enough 'autocorrelation' so that to achieve the synchronization threshold [CTW08].

(i.e., τ and ϕ). When $\mathcal{C}$ represents a codebook and a decoder we refer to it as 'code' instead of 'codebook.' This convention carries throughout this chapter.

Second, we define communication rate with respect to the average elapsed time between the instant the codeword starts being sent and the instant when the decoder makes a decision, i.e.,

$$R = \frac{\log |\mathcal{C}|}{\mathcal{D}} \quad \text{bits per channel use,} \tag{2.3}$$

where

$$\mathcal{D} = \max_m \frac{1}{A} \sum_{t=1}^{A} \mathbb{E}_{m,t}(\tau - t)^+ \ ,$$

x^+ denotes $\max\{0, x\}$, and where $\mathbb{E}_{m,t}$ denotes the expectation with respect to $\mathbb{P}_{m,t}$. (In the above delay expression, expectation comes after taking the positive value of $(\tau - t)$, i.e., by $\mathbb{E}_{m,t}(\tau - t)^+$ we mean $\mathbb{E}_{m,t}((\tau - t)^+)$.)

The following definition extends Shannon's original definition of an achievable transmission rate to the situation where transmitter and receiver are not a priori synchronized.

DEFINITION 2 ((R, α) CODING SCHEME). *A pair (R, α) with $R \geq 0$ and $\alpha \geq 0$ is achievable if there exists a sequence of codes $\{\mathcal{C}_N\}_{N \geq 1}$, labeled by the codebook length N, that asymptotically achieves rate R at asynchronism exponent α. This means that for any $\varepsilon > 0$ and every N large enough, the code $\mathcal{C}_N$*

- *operates under asynchronism level $A = 2^{(\alpha - \varepsilon)N}$;*
- *has a rate at least equal to $R - \varepsilon$;*
- *achieves a maximum error probability at most equal to ε.*

An (R, α) coding scheme is a sequence of codes $\{\mathcal{C}_N\}_{N \geq 1}$ that achieves a rate R at an asynchronism exponent α.

In Definition 2, we choose to grow A exponentially with N. Indeed, when A grows sub-exponentially with N, the capacity turns out to be the same as for the synchronous channel [TCW09]. When A grows faster than exponentially with N, reliable communication is not possible in general. In fact, Theorem 1 says that when asynchronism is exponential in N with an exponent higher than the synchronization threshold, message location is not possible. This, in turn, can be shown to imply that the decoding error probability can't be made negligible. Hence, the interesting asynchronism regime is exponential in the blocklength N with an exponent not exceeding the synchronization threshold.

DEFINITION 3 (ASYNCHRONOUS CAPACITY). *For given $\alpha \geq 0$, the asynchronous capacity $R(\alpha)$ is the supremum of the set of rates that are achievable at asynchronism exponent α. Equivalently, the asynchronous capacity is characterized by $\alpha(R)$, defined as the supremum of the set of asynchronism exponents that are achievable at rate $R \geq 0$.*

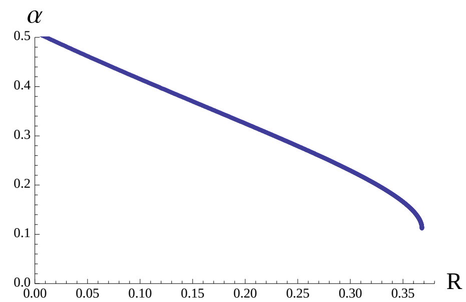

Figure 2.2: Lower bound on asynchronous capacity, given by expression (2.5), for the binary symmetric channel with crossover probability 0.1 and $\star = 0$. Any pair (R, α) below the curve is achievable.

While $R(\alpha)$ may have a more natural interpretation than $\alpha(R)$, most of our results are more conveniently stated in terms of $\alpha(R)$.

A decoder at the output of an asynchronous channel should discriminate between noise and a typical channel output when a message is sent. The following Theorem provides a lower bound to the asynchronous capacity by considering the (Chernoff) error exponent for discriminating hypothesis 'noise' from hypothesis 'message' when codewords are randomly generated according to a certain distribution. This random coding bound can be shown to be achievable with deterministic codes using standard expurgation techniques (see, e.g., [Gal68, p. 151]).

Theorem 2 (Lower Bound on Asynchronous Capacity [TCW08]). *Let X be a random input to the channel and Y the corresponding output. Then, $(R = I(X;Y), \alpha)$ is achievable, where $I(X;Y)$ denotes the input-output channel mutual information, and where*

$$\alpha = \min_V \max\{D(V||Y), D(V||Y_\star)\}, \tag{2.4}$$

with the minimization being over all random variables V defined over the output alphabet $\mathcal{Y}$. Thus, maximizing over all possible input distributions, a lower bound to $\alpha(R)$ in Definition 3 is

$$\max_{X:I(X;Y)\geq R} \min_V \max\{D(V||Y), D(V||Y_\star)\}. \tag{2.5}$$

In Fig. 2.2, we plot the expression (2.5) as a function of the rate for a binary symmetric channel with crossover probability 0.1 and where $\star = 0$ (i.e., when the transmitter is idle the receiver observes 0's and 1's with probability 0.9 and 0.1, respectively). Note that, at rate equal to the synchronous capacity (equal to 0.368...), the asynchronism exponent is strictly positive: one doesn't need to back off from the synchronous capacity in order to achieve a strictly positive asynchronism exponent. It turns out that this somewhat surprising property is shared by most channels (notice that we clearly have $\alpha(R) = 0$ for $R > C$):

Theorem 3 (Discontinuity of $\alpha(R)$ at $R = C$ [CTW09]). *$\alpha(R = C) > 0$ if and only if $Q_\star$ differs from the (unique) capacity achieving output distribution of the synchronous channel.*

Theorem 3 in [TCW08] provides an upper bound on asynchronous capacity by stating necessary conditions that must be fulfilled by any (R, α) coding scheme. However, this theorem has an unwieldy form which is why we prefer to omit it here. Instead, we provide the simple capacity expression for any channel with infinite synchronization threshold, i.e., whose noise distribution $Q(\cdot|\star)$ cannot produce all possible channel outputs.

Theorem 4 (Capacity when $\alpha^{\mathrm{T}} = \infty$ [TCW]). *If $\alpha^T = \infty$, then*

$$\alpha(R) = \max_X \min_{\tilde{Y}|X} \max\{D(X\tilde{Y}||XY), D(X\tilde{Y}||X, Y_\star)\} \quad R \in [0, C],$$

where $D(X\tilde{Y}||XY)$ refers to the Kullback-Leibler distance between the joint distributions of $X\tilde{Y}$ and XY; where $D(X\tilde{Y}||X, Y_\star)$ refers to the Kullback-Leibler distance between the joint distributions of $X\tilde{Y}$ and the product distribution of X and $Y_\star$;[5] and where the minimization is over all conditional distributions of $\tilde{Y}$ given X.

Therefore, when $\alpha^{\mathrm{T}} = \infty$, $\alpha(R)$ is actually a constant that doesn't depend on the rate. Put it the other way around, the asynchronous capacity is the same as the synchronous capacity up to a certain level of asynchronism. Above this level, capacity is zero.

2.2.3 Suboptimality of training based schemes

The usual practical approach for communicating over asynchronous channels is a training-based architecture. In such schemes, each codeword is composed of two parts. The first part, the sync preamble, is a sequence of symbols common to all the codewords. As such, this sequence carries no information, and its only purpose is to help the decoder locate the sent message. The second part carries information. The decoder operates according to a two-step procedure. First, it tries to locate the codeword by seeking the sync preamble. Once the sync preamble is located, it declares a message based on the subsequent symbols.

Below we first formally define training-based schemes. It can be shown that this definition is natural in the sense that the properties that characterize it are satisfied by most practical training based strategies [CTW09]. Second, we characterize the set of rate/asynchronism exponent pairs (R, α) that are achievable with such schemes. As a consequence of this result, we shall deduce the important conclusion that training based schemes can be suboptimal in certain communication regimes.

[5] To differentiate product distributions we put a comma in the argument of the Kullback-Leibler distance, such as for $D(X\tilde{Y}||X, Y_\star)$.

Definition 4 (Training Based Schemes). *A training based scheme is a coding scheme* $\{(\mathcal{C}_N, (\tau_N, \phi_N))\}_{N\geq 1}$ *with the following properties. There exist* $\varepsilon > 0$ *and* $\eta \in [0,1]$ *such that*

i. *For all* $N \geq 1$ *large enough, there is a common preamble across codewords of size* ηN*;*

ii. *The empirical distribution* P_N *of the preamble of the codebook* $\mathcal{C}_N$ *converges to some fixed distribution as* $N \to \infty$ *(say,* L_1*-norm convergence);*

iii. *For all* N *large enough, the decoding time* τ_N *is such that the event* $\{\tau_N = t\}$*, conditioned on the* ηN *observations* $Y_{t-N+1}^{t-N+\eta N}$*, is independent of all other observations (i.e.,* Y_1^{t-N} *and* $Y_{t-N+\eta N+1}^{A+N-1}$*);*

iv. *For all* N *large enough, the codebook* $\mathcal{C}_N$ *and the decoding time* τ_N *satisfy*

$$\mathbb{P}(\tau_N \geq t + 2N - 1 | \tau_N \geq t + N, \nu = t) \geq \varepsilon$$

for all $t \in \{1, 2, \ldots, A - 2N + 1\}$*.*

Property ii. is a mild technical requirement. Property iii. says that the decoder of a training based scheme should stop based on a 'sliding window' sequential rule that seeks the sync preamble. Property iv. is a key requirement. The intuition behind it is that the codeword information symbols shouldn't help the decoder stop at the right time. If the decoder missed the sync preamble, which is captured by the event $\{\tau_N \geq \nu + N\}$, then with some non vanishing probability the decoder will actually miss the entire codeword, which is captured by the event $\{\tau_N \geq \nu + 2N - 1\}$. Note that, without Property iv., one could envision 'information' symbols that actually start with a second preamble. The overall codeword would be composed of two successive sync preambles followed by information symbols. The second preamble would clearly help the decoder stop at the right moment, even if the decoder tries to locate the first preamble.

The following theorem characterizes the asynchronous capacity restricted to training based schemes:

Theorem 5 (Capacity of Training Based Schemes [CTW09]). *The capacity restricted to training based schemes is given by*

$$\alpha(R) = \left(1 - \frac{R}{C}\right) \max_{X} \min_{\tilde{Y}|X} \max\{D(X\tilde{Y}||XY), D(X\tilde{Y}||X, Y_\star)\} \quad R \in (0, C]\,.$$

If the synchronization threshold α^{T} is finite, the max-min-max term in the above expression is finite, which yields the following result:

Corollary 1. *If* $\alpha^T < \infty$*, then* $\alpha(R) \overset{R\to C}{\longrightarrow} 0$ *for training based schemes.*

In Theorem 3 we saw that $\alpha(R = C) > 0$ if and only if the capacity achieving output distribution of the synchronous channel differs from $Q(\cdot|\star)$. In contrast, Corollary 1 says that for training based schemes we have $\alpha(R) \to 0$ as $R \to C$, whenever $\alpha^{\mathrm{T}} < \infty$. Thus, we are lead to the following somewhat surprising conclusion: training based are sub-optimal in the high rate communication regime for the broad class of channels for which $\alpha^{\mathrm{T}} < \infty$ and for which the capacity achieving output distribution of the synchronous channel differs from $Q(\cdot|\star)$. This suggests that for these channels, to achieve a high rate under strong asynchronism (i.e., exponential in the codeword length), synchronization and information transmission must be jointly performed. Each transmitted bit should carry information while also acting as an information 'flag' to help the decoder locate the codeword.

2.2.4 Open problems

Two immediate open questions are the characterization of the asynchronous capacity in the general case, and finding explicit (non-random) code constructions that perform well asynchronously. Some insight for the latter problem has recently been obtained in [CST10]. Another interesting research direction to pursue would be to consider a multi-user setting, say a multiple-access setting where each transmitter communicates in a bursty fashion. Note that this setting is different than the asynchronous multiple-access setting considered in [CMP81, HH85] where the senders operate asynchronously but not in a bursty manner—each sender transmits messages on a continuous basis, without idle periods between them.

2.3 Bits per unit cost

In Section 2.2, the performance metric is data rate: the number of information bits divided by the average elapsed time between the instant information starts being sent and the instant it is decoded.

The data rate is a sensible performance metric for bursty communication if the information to be communicated is delay-sensitive. Then, maximizing the data rate is equivalent to minimizing the time to transmit the burst of data. In many applications, however, the allowable delay may not be tightly constrained, and data rate is less relevant a measure than the *energy* needed to transmit the information. In this case, the minimum energy needed to transmit one bit of information is an appropriate fundamental measure. Thus, we are led to ask the following question: what is the impact of asynchronism on the minimum energy needed to transmit one bit of information?

The main result in this section is a single-letter characterization of the asynchronous capacity per unit cost, or, equivalently, the minimum cost to transmit one bit of information. The results in this section can be found in [CTT10].

2.3.1 Performance criterion

We use the communication model described in Section 2.1, with one small modification that makes the current model somewhat more general. Instead of requiring that information transmission starts exactly at the moment when information is available, i.e., time ν, we now let the transmitter choose when it actually wants to start transmitting information. More precisely, if message m is available at time ν, information transmission starts at a time $\sigma(\nu, m)$ with the only constraint that

$$\nu \leq \sigma(\nu, m) \leq A\,,$$

i.e., the transmitter cannot start transmitting before the message arrives or after the end of the uncertainty window. The reason for allowing here to delay information transmission is given after Definition 7.

In addition to the error probability as defined in (2.2), we are interested in the communication cost:

Definition 5 (Cost of a Code). *The (maximum) cost of a code $\mathcal{C}$ is defined as*

$$\mathcal{K}(\mathcal{C}) \triangleq \max_m \sum_{i=1}^{N} k(c_i(m))$$

where $k : \mathcal{X} \longrightarrow [0, \infty]$ is a function that assigns a non-negative value to each channel input.

The alphabet $\mathcal{X}$ is the set of symbols that can be used for codebook design. In this section, however, we shall distinguish the cases where $\mathcal{X}$ contains $\star$ from the case where $\star$ can't be used for codebook design and thus lies outside $\mathcal{X}$.[6] The reason for not always including $\star$ into $\mathcal{X}$ is practically motivated. The $\star$ symbol corresponds to 'pure noise,' and so there is no point in assigning to it a cost other than zero. On the other hand, the transmitter may, in certain cases, not be able to stay 'idle' at no cost (just being 'on' may incur some energy cost). To model such scenarios, we allow the possibility for $\star$ not being in $\mathcal{X}$.

Definition 6 (Delay of a Code). *Given $\varepsilon > 0$, the (maximum) delay of a code $\mathcal{C}$, denoted by $\boldsymbol{D}(\mathcal{C}, \varepsilon)$, is defined as the smallest d such that*

$$\min_m \mathbb{P}_m(\tau - \nu \leq d) \geq 1 - \varepsilon,$$

where $\mathbb{P}_m$ denotes the output distribution conditioned on the sending of message m.[7]

[6] Hence, in this section, the overall 'input alphabet' is $\mathcal{X} \cup \{\star\}$.

[7] Hence, by definition we have

$$\mathbb{P}_m(\cdot) = \frac{1}{A} \sum_{t=1}^{A} \mathbb{P}_{m,t}(\cdot)\,.$$

Letting $B = \log M$, a key parameter we shall be concerned with is

$$\beta \triangleq \frac{\log A}{B},$$

which we call the timing uncertainty per information bit.

Next, we define the asynchronous capacity per unit cost in the asymptotic regime where $B \to \infty$ while β is kept fixed.

Definition 7 (Asynchronous Capacity per Unit Cost). *$\boldsymbol{R}$ is an achievable rate per unit cost at timing uncertainty per information bit β and delay exponent δ if there exists a sequence of codes $\{\mathcal{C}_B\}$, and a sequence of numbers $\{\varepsilon_B\}$ with $\varepsilon_B \overset{B\to\infty}{\longrightarrow} 0$, such that*

$$\mathbb{P}(\mathcal{E}|\mathcal{C}_B) \leq \varepsilon_B ,$$

$$\limsup_{B\to\infty} \log(\boldsymbol{D}(\mathcal{C}_B, \varepsilon_B))/B \leq \delta ,$$

and

$$\liminf_{B\to\infty} \frac{B}{\mathcal{K}(\mathcal{C}_B)} \geq \boldsymbol{R}.$$

The asynchronous capacity per unit cost, denoted by $\boldsymbol{C}(\beta, \delta)$, is the largest achievable rate per unit cost. In the important case when $\delta = 0$, we define $\boldsymbol{C}(\beta) \triangleq \boldsymbol{C}(\beta, 0)$.[8]

Note that asynchronism is captured here by the parameter $\beta = (\log A)/B$, instead of by $\alpha = (\log A)/N$ as in Section 2.2. The motivation for this is that, in the context of capacity per unit cost, N is an implicit parameter that can be optimized,[9] and hence is less a fundamental quantity than B, the number of bits to be transmitted.

Recall that in the current asychronous channel model we allow the transmitter to delay information transmission after it is available; transmission may start at any time σ with $\nu \leq \sigma \leq A$. The reason for allowing this is that when the performance metric is the cost, it may be sub-optimal to start transmitting information right when it is available. As an extreme case, note that if there is no delay constraint, the transmitter can choose to send information at time A, which nullifies the impact of asynchronism. More generally, when a large delay is tolerated, i.e., when $\delta > 0$, it is suitable to delay information transmission as this contributes to increase reliability via asynchronism reduction [CTT10]. In contrast, when data rate is the performance metric (as in Section 2.2), and thus delay is important, delaying information transmission is sub-optimal.

In the next section we characterize the capacity per unit cost for an arbitrary asynchronism parameter $\beta \geq 0$ and an arbitrary delay parameter $\delta \geq 0$. Similarly as for the synchronous case, the results simplify when there is a zero cost symbol, specifically when $\star \in \mathcal{X}$ and $k(\star) = 0$.

[8] Throughout this chapter, we assume that Q has non-zero capacity, for otherwise the capacity per unit cost and, a fortiori, the asynchronous per unit cost, equal to zero.

[9] Just as for the synchronous capacity per unit cost (see comment after [Ver90, Definition 2]).

2.3.2 Results

Our first result gives the asynchronous capacity per unit cost when $\delta = 0$. It can be viewed as the asynchronous analogue of [Ver90, Theorem 2], which states that the synchronous capacity per unit cost is

$$\max_X \frac{I(X;Y)}{\mathbb{E}[k(X)]}, \tag{2.6}$$

where the maximization is over random variables X defined over the channel input alphabet $\mathcal{X}$.

Theorem 6 (Asynchronous Capacity per Unit Cost: Sub-exponential Delay Constraint). *The asynchronous capacity per unit cost at delay exponent $\delta = 0$ is given by*

$$\boldsymbol{C}(\beta) = \max_X \min\left\{\frac{I(X;Y)}{\mathbb{E}[k(X)]}, \frac{I(X;Y) + D(Y||Y_\star)}{\mathbb{E}[k(X)](1+\beta)}\right\} \tag{2.7}$$

where the maximization is over all random variables X defined over $\mathcal{X}$. Furthermore, capacity can be achieved by codes whose delay grows linearly with B.

The two terms in (2.7) reflect the two constraints on reliable communication. The first term corresponds to the standard constraint that the number of bits that can reliably be transmitted per channel use cannot exceed the input-output mutual information. This constraint applies when the channel is synchronous, hence also in the absence of synchrony. To see this, note that by swapping the max and the min in (2.7), we deduce that $\boldsymbol{C}(\beta)$ is less than (2.6), the synchronous capacity per unit cost.

The second term in (2.7) corresponds to the receiver's ability to determine the arrival time ν of the data. Indeed, even though the decoder is required only to produce a message estimate, because of the delay constraint there is no loss in terms of capacity per unit cost to also require the decoder to (approximately) locate the time codeword transmission begins—the delay constraint imposes the decoder to locate the sent message within a time window that is negligible compared to A. The quantity

$$I(X;Y) + D(Y||Y_\star)$$

measures how difficult it is for the receiver to discern a data-carrying transmitted symbol from pure noise and thus determines how difficult it is for the receiver to get the timing correct.

When the alphabet $\mathcal{X}$ contains a zero-cost symbol 0, the synchronous result (2.6) simplifies, and Theorem 3 in [Ver90] says that the synchronous capacity per unit cost becomes

$$\max_{x\in\mathcal{X}} \frac{D(Y_x||Y_0)}{k(x)}, \tag{2.8}$$

an optimization over the input alphabet instead of over the set of all input distributions, where Y_x refers to is the output distribution given that x is transmitted.

We find an analogous simplification in the asynchronous setting when $\star$ is in $\mathcal{X}$ and has zero cost:

Theorem 7 (Asynchronous Capacity per Unit Cost With Zero Cost Symbol: Sub-exponential Delay Constraint). *If $\star$ is in $\mathcal{X}$ and has zero cost, the asynchronous capacity per unit cost at delay exponent $\delta = 0$ is given by*

$$\boldsymbol{C}(\beta) = \frac{1}{1+\beta} \max_{x \in \mathcal{X}} \frac{D(Y_x || Y_\star)}{k(x)}, \tag{2.9}$$

and capacity can be achieved by codes whose delay grows linearly with B.

Hence, a lack of synchronization multiplies the cost of sending one bit of information by $1+\beta$. An intuitive justification for this is as follows. Suppose there exists an optimal coding scheme that can both isolate and locate the sent message with high probability—as alluded to above, the ability to 'locate' the message is a consequence of the decoder's delay constraint. This allows us to consider message/location pairs as inducing a code of size $\approx \frac{A}{N} 2^B$ used for communication across the *synchronous channel*. Hence, if, say, N grows sub-exponentially with B, we are effectively communicating $\approx \beta B + B = B(1+\beta)$ bits reliably over the synchronous channel. Therefore, sending B bits of information at asynchronism level β is at least as costly as sending $B(1+\beta)$ bits over the synchronous channel. Flipping this reasoning around, the asynchronous channel effectively induces a codebook for message/location pairs where the location is encoded via PPM. From [Ver90], optimal coding schemes are similar to PPM in that the codewords consist almost entirely of the zero cost symbol. This provides an intuitive justification for why asynchronism multiplies the cost of sending one bit of information by a factor $1+\beta$.

Theorem 7 can be extended to the (continuous-valued) Gaussian channel, where the idle symbol $\star$ is the 0-symbol:

Theorem 8 (Asynchronous Capacity per Unit Cost for the Gaussian Channel: Sub-exponential Delay Constraint). *The asynchronous capacity per unit cost for the Gaussian channel with variance $N_0/2$, quadratic cost function (i.e., $k(x) = x^2$), and delay exponent $\delta = 0$, is given by*

$$\mathcal{C}(\beta) = \frac{1}{1+\beta} \frac{\log e}{N_0} \quad \beta \geq 0 .$$

Theorem 6 can be extended to the case of a large delay constraint, i.e., when $0 < \delta < \beta$. As for Theorem 6, the following result holds irrespectively of whether or not $\mathcal{X}$ contains $\star$. A simplification similar to Theorem 7 applies if $\mathcal{X}$ contains $\star$ and it has zero cost.

Theorem 9 (Asynchronous Capacity per Unit Cost: Exponential Delay Constraint). *The asynchronous capacity per unit cost at delay constraint δ, with $0 < \delta < \beta$, is given by*

$$\boldsymbol{C}(\beta, \delta) = \boldsymbol{C}(\beta - \delta),$$

i.e., it is the same as the capacity per unit cost with delay exponent $\delta = 0$, but with asynchronism exponent β reduced to $\beta - \delta$.

The uniform distribution on ν in our model is not critical. The next result extends Theorem 6 to the case where ν is non-uniform. For a non-uniform distribution on ν, what is important turns out to be its 'smallest' set of mass points that contains 'most' of the probability.

Below, ν^B denotes the arrival time random variable when B bits of information have to be transmitted (In Theorem 6, ν^B has the uniform distribution over $\{1, 2, \ldots, 2^{\beta B}\}$).

Theorem 10 (Asynchronous Capacity per Unit Cost With Non-uniform Arrival Time: Sub-exponential Delay Constraint). *Define*

$$\bar{\beta} = \inf_{\{\epsilon_B\}} \lim_{B \to \infty} \frac{\log(S(\epsilon_B))}{B}, \tag{2.10}$$

where the infimum is with respect to all sequences $\{\epsilon_B\}$ of nonnegative numbers such that $\lim_{B\to\infty} \epsilon_B = 0$, where $S(\epsilon_B)$ denotes the size of the smallest set with probability at least $1 - \epsilon_B$, and it is assumed that the limit in (2.10) exists.

Then, the asynchronous capacity per unit cost at delay exponent 0 is given by

$$\boldsymbol{C}(\bar{\beta}) = \max_X \min \left\{ \frac{I(X;Y)}{\mathbb{E}[k(X)]}, \frac{I(X;Y) + D(Y||Y_\star)}{\mathbb{E}[k(X)](1+\bar{\beta})} \right\}.$$

Although the formula for $\bar{\beta}$ in (2.10) appears unwieldy, in many cases it can easily be evaluated. For example, in many cases, such as for the uniform or the geometric distributions, the formula reduces to the normalized entropy

$$\bar{\beta} = \lim_{B \to \infty} H(\nu^B)/B\,.$$

There are cases, however, where (2.10) doesn't reduce to the normalized entropy. For instance, consider the case when $\nu^B = 1$ with probability 1/2, and $\nu^B = i$ with probability $(1/2)2^{-\beta B}$ for $i = 2, \ldots, 2^{\beta B} + 1$. Then,

$$\bar{\beta} = 2 \lim_{B \to \infty} H(\nu^B)/B\,.$$

3

Tracking a Stopping Time Through Noisy Observations

In Chapter 2, the task of the sequential decoder is to quickly identify the transmitted message. Statistically, this means that the decoder must quickly detect and isolate the cause of the change in distribution with respect to the 'nominal distribution' given by 'pure noise'—before the change-point ν the decoder observes pure noise and from time ν to time $\nu + N - 1$ the decoder observes a noisy version of the transmitted message. In this chapter, we present the tracking stopping time (TST) problem, a statistical decision problem whose formulation applies in several areas, including communication. In the TST problem the goal is also to quickly react to a change in distribution, but not to isolate its cause. The chief difficulty here lies in the distribution of the change-point, which may depend on an unobserved stochastic sequence that is correlated to the observation process. In contrast, in the communication problem there is no such dependency since ν is randomly drawn independently of everything else. This chapter is based on [NT09] and [BT10].

3.1 Introduction

The TST problem is defined as follows. Let $X = \{X_t\}_{t\geq 0}$ be a discrete-time stochastic process and let τ be a stopping time defined over X. Statistician has access to X only through correlated observations $Y = \{Y_t\}_{t\geq 0}$. Knowing the probability distribution of (X, Y) and the stopping rule τ, Statistician wishes to find a stopping η so as to minimize the mean absolute deviation $\mathbb{E}|\eta - \tau|$.

The TST problem formulation, introduced in [NT09], naturally generalizes to continuous time and other delay penalty functions such as $\mathbb{E}(\eta - \tau)_+$ for a fixed 'false-alarm' probability level $\mathbb{P}(\eta < \tau)$. Important situations are when the observation process is a noisy version of X, a delayed version of X, or represents partial information with respect to X — at time t, $X_t = (\widetilde{X}_t, \widetilde{Y}_t)$ and Statistician observes only $Y_t = \widetilde{Y}_t$. For specific examples of applications of the TST problem related to monitoring, forecasting, and communication we refer to [NT09].

In [NT09], an algorithmic approach is proposed for discrete-time settings where all the X_i's and Y_i's take values in a common finite alphabet (otherwise the X and

Y processes are arbitrary), and where τ is bounded by some constant $c \geq 1$. Given the probability distribution of (X, Y) and the stopping rule of τ, the algorithm outputs the minimum reaction delay $\mathbb{E}(\eta - \tau)_+$ together with an optimal stopping rule, for all false-alarm probability levels $\mathbb{P}(\eta < \tau) \leq \alpha$, $\alpha \in [0, 1]$. Under certain conditions on (X, Y) and τ, the computational complexity of this algorithm is polynomial in c.

What motivated an algorithmic approach for the TST problem, is that it generalizes the Bayesian change-point detection problem, a long studied problem with applications to industrial quality control that dates back to the 1940's [AGP47], and for which analytical solutions have been reported only for specific, mostly asymptotic, settings.

In the Bayesian change-point problem, there is a random variable θ, taking values in the positive integers, and two probability distributions P_0 and P_1. Under P_0, the conditional density function of Y_t given $Y_1, Y_2, \ldots, Y_{t-1}$ is $f_0(Y_t|Y_1, Y_2, \ldots, Y_{t-1})$, for every $t \geq 0$. Under P_1, the conditional density function of Y_t given $Y_1, Y_2, \ldots, Y_{t-1}$ is $f_1(Y_t|Y_1, Y_2, \ldots, Y_{t-1})$, for every $t \geq 0$. The observed process is distributed according P^θ which assigns the same conditional density functions as P_0 for all $t < \theta$, and the same conditional density functions as P_1 for all $t \geq \theta$.

The Bayesian change-point problem typically consists in finding a stopping time η, with respect to $\{Y_t\}$, that minimizes some function of the delay $\eta - \tau$. Shiryaev [Shi63, Shi78], for instance, considered minimizing the Lagrangian function

$$\mathbb{E}(\eta - \theta)_+ + \lambda \mathbb{P}(\eta < \theta)$$

for given constant $\lambda \geq 0$. Assuming a geometric prior on the change-point θ, and that before and after θ the observations are independent with common density function f_0, for $t < \theta$, and f_1 for $t \geq \theta$, Shiryaev showed that an optimal η stops as soon as the posterior probability that a change occurred exceeds a certain fixed threshold. Later, Yakir [Yak94] generalized Shiryaev's result by considering finite-state Markov chains. For more general prior distributions on θ, the problem is known to become difficult to handle. However, in the limit of small false-alarm probabilities $\mathbb{P}(\eta < \theta) \to 0$, Lai [Lai98] and, later, Tartakovsky and Veeravalli [TV05], derived asymptotically optimal detection policies for the Bayesian change-point problem under general assumptions on the distributions of the change-point and observed process. (For the non-Bayesian version of the change-point problem we refer the reader to [Lor71, Mou86].)

It can be shown that any Bayesian change-point problem can be formulated as a TST problem, and that a TST problem cannot, in general, be formulated as a Bayesian change-point problem [NT09]. The TST problem therefore generalizes the Bayesian change-point problem, which is analytically tractable only in special cases.

In the next section, we consider the situation where X and Y are correlated Gaussian random walks given by $X_0 = Y_0 = 0$, $X_t = s \cdot t + \sum_{i=1}^{t} V_i$ and $Y_t =$

$X_t + \varepsilon \sum_{i=1}^t W_i$, for $t \geq 1$ and some arbitrary constant $s > 0$ and $\varepsilon > 0$. The V_i's and W_i's are assumed to be independent zero mean unit variance Gaussian random variables. The stopping time to be tracked is the threshold crossing moment $\tau_l = \inf\{t \geq 0 : X_t \geq l\}$ for some arbitrary threshold level $l > 0$. For this setting, we provide upper and lower bounds on $\inf_\eta \mathbb{E}|\eta - \tau_l|$ that imply

$$\inf_\eta \mathbb{E}|\eta - \tau_l| = \sqrt{\frac{2l\varepsilon^2}{\pi s^3(1+\varepsilon^2)}}(1 + o(1)) \quad (l \to \infty) \tag{3.1}$$

for fixed $s > 0$ and $\varepsilon > 0$. Interestingly, (3.1) is still valid if we let η be an estimator of τ that depends on the entire sequence Y_0^∞; causality doesn't come at the expense of increased delay in the above asymptotic regime.

For the particular case where the random walks have no drift, i.e., $s = 0$, we show that $\mathbb{E}|\eta - \tau_l|^r = \infty$ whenever $r \geq 1/2$, $\varepsilon > 0$, and $l > 0$, for any estimate η of τ_l that potentially may also depend on the entire observation process Y_0^∞.

The above results naturally extend to the continuous time setting where $\sum_{i=1}^t V_i$ and $\sum_{i=1}^t W_i$ are replaced by two independent standard Brownian motions. In particular, (3.1) remains valid for fixed $s > 0$ and $\varepsilon > 0$.

3.2 Problem Formulation and Main Results

The results in this section can be found in [BT10]. We consider the discrete-time processes

$$X: \quad X_0 = 0 \qquad X_t = \sum_{i=1}^t V_i + st \qquad t \geq 1$$

$$Y: \quad Y_0 = 0 \qquad Y_t = X_t + \varepsilon \sum_{i=1}^t W_i \qquad t \geq 1$$

where $V_1, V_2, \ldots$ and $W_1, W_2, \ldots$ are two independent sequences of independent standard (i.e., zero mean unit variance) Gaussian random variables, and where $s > 0$ and $\varepsilon > 0$ are arbitrary constants.

Given the threshold crossing time

$$\tau_l = \inf\{t \geq 0 : X_t \geq l\}$$

for some arbitrary level $l > 0$, we aim at finding a stopping time with respect to observation process Y that best tracks τ_l. Specifically, we consider the optimization problem

$$\inf_\eta \mathbb{E}|\eta - \tau_l|\,, \tag{3.2}$$

where the minimization is over all stopping times η defined with respect to the natural filtration induced by the Y process.

To avoid trivial situations, we restrict l and ε to be strictly positive. When $l = 0$ or $\varepsilon = 0$, (3.2) is equal to zero: for $l = 0$, $\eta = 0$ is optimal, and for $\varepsilon = 0$, $\eta = \tau_l$ is optimal.

The reason for restricting our attention to the case where also s is strictly positive is that, when $s = 0$, (3.2) is infinite for all $l > 0$ and $\varepsilon > 0$. In fact, Proposition 1, given at the end of this section, provides a stronger statement: for $s = 0$, $\varepsilon > 0$, and $l > 0$, we have $\mathbb{E}|\eta - \tau_l|^r = \infty$ for any $r \geq 1/2$ and any estimator $\eta = \eta(Y_0^\infty)$ of τ_l that may depend on the entire observation process Y_0^∞ (i.e., η need not be a stopping time).

The following theorem provides a non-asymptotic upper bound on (3.2) which is achieved by a threshold crossing stopping time applied to a certain estimate of the X process:

Theorem 11 (Upper bound). *Fix $\varepsilon > 0$, $s > 0$, $l > 0$, and define $\hat{X}_t$ as*

$$\hat{X}_0 = 0 \qquad \hat{X}_t = st + \frac{1}{1+\varepsilon^2}(Y_t - st) \quad \text{for } t \geq 1.$$

Then, the stopping time $\eta = \inf\{t \geq 0 : \hat{X}_t \geq l\}$ satisfies

$$\mathbb{E}\,|\eta - \tau_l| \leq \sqrt{\frac{2l\varepsilon^2}{\pi(1+\varepsilon^2)s^3}} + \frac{6}{s}\left(\frac{l}{(2\pi s)^3}\right)^{1/4} + \sqrt{\frac{8(s+2)}{\pi s^3}} + 10 + \frac{20}{s}. \tag{3.3}$$

The next theorem provides a non-asymptotic lower bound on $\mathbb{E}|\eta - \tau_l|$ for any estimate $\eta = \eta(Y_0^\infty)$ of τ_l that has access to the entire sequence Y_0^∞. The function $Q(x)$ is the standard Q-function defined as $Q(x) = (2\pi)^{-1/2}\int_x^\infty \exp(-u^2/2)du$.

Theorem 12 (Lower bound). *Let $\varepsilon > 0$ and $l/s \geq 2$ with $s > 0$. Then, for any integer n such that $1 \leq n < l/s$, the following lower bound holds:*

$$\begin{aligned}\inf_{\eta(Y_0^\infty)} \mathbb{E}|\eta - \tau_l| \geq\ & \sqrt{\frac{2n\varepsilon^2}{\pi s^2(1+\varepsilon^2)}}\left(1 - Q\left(\frac{l - sn}{\sqrt{n(1+\varepsilon)}}\right)\right) \\ & - \sqrt{\frac{2}{\pi s^3}}\left(l - sn + \sqrt{\frac{n}{2\pi}}\right)^{1/2} - 2 - \frac{4}{s}.\end{aligned} \tag{3.4}$$

When n approaches l/s and l/s tends to infinity in a suitable way, the upper and lower bounds (3.3) and (3.4) become tight. The following result can be derived from Theorems 11 and 12:

Theorem 13 (Asymptotics). *Let q be a constant such that $1/2 < q < 1$. In the asymptotic regime where $l/s \geq 2$,*

$$s\left(\frac{l}{s}\right)^{q-1/2} \longrightarrow \infty\,,$$

and

$$\left(\frac{l}{s}\right)^{1-q}\frac{\varepsilon^2}{1+\varepsilon^2} \longrightarrow \infty\,,$$

we have

$$\inf_{\eta(Y_0^\infty)} \mathbb{E}|\eta - \tau_l| = \inf_{\eta} \mathbb{E}|\eta - \tau_l| = \sqrt{\frac{2l\varepsilon^2}{\pi s^3(1+\varepsilon^2)}}\,[1+o(1)]\ . \tag{3.5}$$

In particular, (3.5) *holds in the limit $l \to \infty$ for fixed $s > 0$ and $\varepsilon > 0$.*

To prove Theorem 11, we consider $\eta = \inf\{t \geq 0 : \hat{X}_t^{(c)} \geq l\}$, where $\hat{X}_t^{(c)}$ is the estimate of X_t defined as $\hat{X}_t^{(c)} = st + c(Y_t - st)$, then optimize over $c \geq 0$. It should be noted that, in the asymptotic regime (given by Theorem 13) where the upper and lower bounds on $\inf_\eta \mathbb{E}|\eta - \tau_l|$ coincide, the optimal c (equal to $1/(1+\varepsilon^2)$) is the value for which the variance of $X_t - \hat{X}_t^{(c)}$ is minimized.

Let us now consider the setting where $\sum_{i=1}^t V_i$ and $\sum_{i=1}^t W_i$ are replaced by standard Brownian motions, i.e., with the X and the Y processes being defined as

$$\begin{aligned} X: &\quad X_0 = 0 \qquad X_t = B_t + st \quad \text{for } t > 0 \\ Y: &\quad Y_0 = 0 \qquad Y_t = X_t + \varepsilon N_t \quad \text{for } t > 0 \end{aligned}$$

where $\{B_t\}_{t>0}$ and $\{N_t\}_{t>0}$ are two independent standard Brownian motions. The previous results easily extend to the Brownian motion setting. Indeed, the analysis is simpler than for the Gaussian random walk setting as there is no 'excess over threshold' for a Brownian motion — the value of a Brownian motion the first time it crosses a certain level equals this level.

Theorems 14, 15, and 16 are analogous to Theorems 11, 12, and 13, respectively.

Theorem 14 (Upper bound: Brownian motion with drift). *Fix $\varepsilon > 0$, $s > 0$, $l > 0$, and define $\hat{X}_t$ as*

$$\hat{X}_0 = 0 \qquad \hat{X}_t = st + \frac{1}{1+\varepsilon^2}(Y_t - st) \quad \textit{for } t > 0.$$

Then, the stopping time $\eta = \inf\{t \geq 0 : \hat{X}_t = l\}$ satisfies

$$\mathbb{E}\,|\eta - \tau_l| \leq \sqrt{\frac{2l\varepsilon^2}{\pi(1+\varepsilon^2)s^3}} + \frac{6}{s}\left(\frac{l}{(2\pi s)^3}\right)^{1/4}.$$

Theorem 15 (Lower bound: Brownian motion with drift). *Let $\varepsilon > 0$, $s > 0$, and $l > 0$, and let n be such that $1 \leq n < l/s$. Then,*

$$\inf_{\eta(Y_0^\infty)} \mathbb{E}|\eta - \tau_l| \geq \sqrt{\frac{2n\varepsilon^2}{\pi s^2(1+\varepsilon^2)}}\left(1 - Q\left(\frac{l - sn}{\sqrt{n(1+\varepsilon)}}\right)\right) - \sqrt{\frac{2}{\pi s^3}}\left(l - sn + \sqrt{\frac{n}{2\pi}}\right)^{1/2}.$$

Theorem 16 (Asymptotics : Brownian motion with drift). *Theorem 13 is also valid in the Brownian motion setting.*

We end this section with a proposition related to the particular case where $s = 0$, which we referred to earlier. When $s = 0$, $\varepsilon > 0$, and $l > 0$, it is impossible to finitely track τ_l, even having access to the entire observation process Y_0^∞: for any estimate $\eta = \eta(Y_0^\infty)$, $\mathbb{E}(|\eta - \tau_l|^r) = \infty$ for all $r \geq 1/2$. The proposition is valid in both the Gaussian random walk and the Brownian motion settings.

PROPOSITION 1. *Let $s = 0$ and let $f(x)$, $x \geq 0$, be a non-negative and non-decreasing function such that*

$$\mathbb{E}f(\tau_h/2) = \infty \tag{3.6}$$

for some constant $h > 0$. Then,

i. $\mathbb{E}f(|\tau_l - \eta|) = \infty$ *for any estimate $\eta = \eta\,(Y_0^\infty)$, whenever $\varepsilon > 0$ and $l > 0$.*

ii. If $f(x) = x^r$, $r \geq 1/2$, then (3.6) holds for all $h > 0$, whenever $\varepsilon > 0$ and $l > 0$. (Hence, $\mathbb{E}|\tau_l - \eta|^r = \infty$ for any estimate $\eta = \eta\,(Y_0^\infty)$ of τ_l whenever $r \geq 1/2$, $s = 0$, $\varepsilon > 0$, and $l > 0$.)

3.3 Open problems

In a further collaboration with M.V. Burnashev from the Russian Academy of Sciences, we are extending the above results to higher order penalty functions, i.e., $\mathbb{E}|\eta - \tau|^r$ with $r > 1$. Another research direction we are pursuing is to investigate the situation where the Y process is a delayed version of the X process. This setting applies, for instance, in high-frequency trading in finance and leads to a collaboration with L. Kogan from the Sloan School of Management at MIT.

4

Thesis supervisions

4.0.1 Ashish Khisti: secrecy communication

I was a co-phd-supervisor with Prof. Gregory Wornell for Ashish Khisti while I was at MIT as a postdoctoral associate between 2005 and 2008. The Ph.D thesis was defended in 2008. My involvement was only in first part of the thesis whose main results appeared in [KTG08].

In the first part of the thesis, Ashish considers secret communication in the presence of eavesdroppers. The goal is to reliably communicate to a set of intended receivers, while keeping the eavesdroppers in ignorance. Ashish first considers a time varying fading channel. Both the scenarios when each legitimate receiver wants a common message as well as the scenario when they all want separate messages are studied and capacity results are established.

4.0.2 Milad Sefidgaran: communication and computing

Milad started his Ph.D. under my supervision in October 2009. Milad is investigating fundamentals of cooperation in the context of multi-user computation. For instance, consider the multiple-access channel where the receiver wants to compute a certain function of two correlated inputs. For a broad class of multiple-access channels, Milad recently derived the minimum number of bits that each sender needs to transmit in order for the receiver to compute the function with high probability. This work extends the work of [OR98] where point-to-point communication is considered.

Bibliography

[AGP47] F. J. Anscombe, H. J. Godwin, and R. L. Plackett. Methods of deferred sentencing in testing the fraction defective of a continuous output. *Supplement to the Journal of the Royal Statistical Society*, 9(2):198–217, 1947.

[BT10] M. V. Burnashev and A. Tchamkerten. Tracking a threshold crossing time of a gaussian random walk through correlated observations (invited paper). In *Inter. Work. on App. Prob.*, 2010.

[Bur76] M. V. Burnashev. Data transmission over a discrete channel with feedback: Random transmission time. *Probl. Inform. Transm.*, 12(4):250–265, 1976.

[CMP81] T. M. Cover, R. J. McEliece, and E. C. Posner. Asynchronous multiple-access channel capacity. *IEEE Trans. Inform. Th.*, 4:409–413, July 1981.

[CST10] G. Cohen, P. Solé, and A. Tchamkerten. Heavy weight codes. In *IEEE Intl. Sympo. on Info. Th. (ISIT)*, 2010.

[CTT10] V. Chandar, A. Tchamkerten, and D.N.C. Tse. Asynchronous capacity per unit cost. In *IEEE Intl. Sympo. on Info. Th. (ISIT)*, 2010.

[CTW08] V. Chandar, A. Tchamkerten, and G.W. Wornell. Optimal sequential frame synchronization. *IEEE Trans. Inform. Th.*, 54(8):3725–3728, 2008.

[CTW09] V. Chandar, A. Tchamkerten, and G.W. Wornell. Training-based schemes are suboptimal for high rate asynchronous communication. In *Info. Th. Work. (ITW)*, Taormina, October 2009.

[Dob67] R. L. Dobrushin. Shannon's theorems for channels with synchronization errors. *Problems Information transmission*, 3(4):11–26, 1967.

[Gal68] R. G. Gallager. *Information Theory and Reliable Communication*. Wiley, Budapest, 1968.

[HH85] J.Y.N. Hui and P.A. Humblet. The capacity of the totally asynchronous multiple-access channel. *IEEE Trans. Inform. Th.*, 31(2):207–216, March 1985.

[KTG08] A Khisti, A. Tchamkerten, and G.Wornell. Secure broadcasting over fading channels. *IEEE Trans. Inform. Th.*, 54(6):471–480, June 2008.

[Lai98] T.Z. Lai. Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Trans. Inform. Th.*, 44:2917–2929, November 1998.

[Lor71] G. Lorden. Procedures for reacting to a change in distribution. *Ann. Math. Stat.*, 42(2):1897–1908, 1971.

[Mas72] J. L. Massey. Optimum frame synchronization. *IEEE Trans. Comm.*, 2:115–119, april 1972.

[Mou86] G. Moustakides. Optimal procedures for detecting changes in distribution. *Ann. Stat.*, 14:1379–1387, 1986.

[NT09] U. Niesen and A. Tchamkerten. Tracking stopping times through noisy observations. *IEEE Trans. Inform. Th.*, 55(1):422–432, January 2009.

[OR98] Alon Orlitsky and James R. Roche. Coding for computing. *IEEE Transactions on Information Theory*, 47:903–917, 1998.

[Sha48] C. E. Shannon. A mathematical theory of communication. *The Bell Sys. Tech. Journal*, 27:379–423, October 1948.

[Shi63] A. N. Shiryaev. On optimum methods in quickest detection problems. *Th. Prob. and its App.*, 8(1):22–46, 1963.

[Shi78] A. N. Shiryayev. *Optimal Stopping rules.* Springer-Verlag, 1978.

[TCW] A. Tchamkerten, V. Chandar, and G.W. Wornell. Basic limits on asynchronous communication over memoryless channels. *to be submitted for publication.*

[TCW08] A. Tchamkerten, V. Chandar, and G.W. Wornell. On the capacity region of asynchronous channels. In *IEEE Intl. Sympo. on Info. Th. (ISIT)*, 2008.

[TCW09] A. Tchamkerten, V. Chandar, and G.W. Wornell. Communication under strong asynchronism. *IEEE Trans. Inform. Th.*, 55(10):4508–4528, October 2009.

[TT06] A. Tchamkerten and I. E. Telatar. Variable length coding over an unknown channel. *IEEE Trans. Inform. Th.*, 52(5):2126–2145, May 2006.

[TV05] A. G. Tartakovsky and V. V. Veeravalli. General asymptotic Bayesian theory of qickest change detection. *Th. Prob. Appl.*, 49(3):458–497, 2005.

[Ver90] S. Verdú. On capacity per unit cost. *IEEE Trans. Inform. Th.*, 36(5):1019–1030, September 1990.

[Yak94] B. Yakir. Optimal detection of a change in distribution when the observations form a Markov chain with a finite state space. In *Change-point problems*, volume 23, pages 346–358. Institute of Mathematical Statistics, Lecture Notes, Monograph Series, 1994.