Information Theoretic Perspectives on Synchronization

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

***Abstract*— We study the information theoretic limits of communication over asynchronous discrete memoryless channels. The transmitter starts sending a block codeword of length N at a time ν uniformly distributed within the interval $[1, 2, \ldots, L]$. We assume that the receiver knows L but not ν. We give a scaling law of L with respect to N for which reliable communication can be achieved. Specifically, we propose a communication scheme with the property that, unless the asynchrony level L grows at least as e^{NC}, where C denotes the capacity of the synchronized channel, arbitrary low error probability can be achieved. If L grows sub-exponentially in N, the capacity is the same as that of the ordinary synchronized channel. Further, we provide a lower bound to the error probability given a certain channel, codebook, and asynchrony level. This bound together with our scheme shows that, in certain cases, the condition $L \leq e^{NC(1-\delta)}$ for any $\delta > 0$ is an asymptotic necessary and sufficient condition for reliable communication. Finally we extend our analysis to a simple scenario where communication is carried over a Gaussian channel with antipodal signaling $+\sqrt{P}$ and $-\sqrt{P}$. We show that a necessary condition on the amount of power needed in order to guarantee reliable communication is that P must scale as $\frac{1}{N} \log L$ when $L \to \infty$.**

I. Introduction

In information theoretic analysis, a common assumption 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 and coding delay, are defined under this hypothesis. In practice this assumption is rarely fulfilled; due to the bursty nature of the information source, the transmitter may start emitting at random moments, and the receiver needs a certain period of time to realize that the transmitter has started to emit information.

Here we consider communication over general discrete memoryless channels. The basic question we address is *how does a lack of synchronization between the transmitter and receiver affect the set of achievable rates?* In particular, we study a communication setting where the transmitter starts emitting a block codeword of length N at a time ν that is unknown to the receiver and uniformly distributed over the interval $[1, 2, \ldots, L]$. The parameter L defines the level of asynchrony between the transmitter and receiver and is assumed to be known to the receiver. Before time ν and after time $\nu + N - 1$ (i.e., before and after the message is sent) we assume that a constant sequence of known symbols is sent or, alternatively, that an independent identically distributed (i.i.d.) random input sequence is sent. The receiver, using a sequential decoder, makes a decision so as to minimize the delay between ν and the decoding time τ, while keeping the error probability below a certain value.

This work was supported in part by NSF under Grant No. CCF-0515122, and by a University IR&D Grant from Draper Laboratory.

Our main result shows that the capacity of the asynchronous channel is the same as the capacity of the ordinary channel provided that L is sub-exponential in the block length. Further, an achievable error exponent is computed. Our analysis reveals that reliable communication is possible even if L grows exponentially in N, provided that the growth rate is below capacity and that the rate is below a certain threshold. Conversely, we establish a lower bound on the error probability in terms of the channel, the codebook, and the asynchrony level, and evaluate a precise condition on L for the binary symmetric and the Gaussian channels such that the probability of error is asymptotically bounded away from zero.

The problem we address in this paper is closely related to the problem of detection and isolation of abrupt changes in a stochastic system (see, e.g., [2], [3]). An instance of that problem is as follows. Let $\{P^m\}_{m=0}^M$ be a set of probability measures on $\mathcal{Y}^\infty$, each P^m being characterized by its family of conditional probabilities $\{P^m_{Y_1}, P^m_{Y_2|Y_1}, P^m_{Y_3|Y_1^2}, \cdots\}$. A decoder observes a sequence $Y_1, Y_2, \ldots$ where $Y_1, Y_2, \ldots, Y_{\nu-1}$ are generated according to P^0 while $Y_\nu, Y_{\nu+1}, \ldots$ are generated according to a certain P^m with $m \in \{1, 2, \ldots, M\}$. The decoder consists of a sequential test (τ, ϕ), where τ is a stopping rule whose decision to stop at time n depends on Y_1^n, and where ϕ denotes a terminal decision rule that accepts one of the post-change distributions. For example, $(\tau, \phi) = (n, m)$ indicates that the post-change distribution P^m is decoded at time n. Without knowing the value of the change-point ν, the goal of the decoder is to react as quickly as possible after the change point, i.e., to minimize the detection delay $(\tau - \nu)^+$, while keeping the probability of a decoding error below a certain threshold. Thus, our synchronization problem takes the form of a detection and isolation problem in which the change in distribution is induced by the transmitted message.

The main difference between our problem and the detection and isolation problem is that in our setting the change in distribution occurring at time ν has limited duration (only up to time

$\nu+N-1$). Further, in the detection and isolation problem, the pre- and post-change distributions are given and the problem is only a decoding problem. In the synchronization setting, the pre-change distribution is given but, in addition, there is the problem of finding codewords inducing post-change distributions with "good detection and isolation properties." Also, and to the best of our knowledge, optimal decoding rules for the detection and isolation problem where obtained only in the case where the number M is fixed, and in the limit where the probability of error tends to zero.[1] Informally, we may say that optimal decoding rules for the detection and isolation problem have been obtained only in the "zero rate" regime.

II. The problem and main result

We consider discrete-time communication over a discrete memoryless channel. We implicitly assume that the time scale of the transmitter and the receiver are "synchronized," i.e., time zero is the same for both of them. We assume that the transmission of a particular codeword of length N may start at any time ν, uniformly distributed in the interval $[1, 2, \ldots, L]$. We assume that the choice of the message to be conveyed is independent of ν and that the receiver does not know the value of ν but knows L. At time i, the receiver gets a symbol Y_i that, conditioned on ν and L, is distributed according to the conditional probability of the channel $Q(\cdot|x_{i,\nu}(m))$, where $x_{i,\nu}(m)$ is defined as

$$\begin{aligned} x_{i,\nu}(m) &= a \quad \text{for } i \in [1, \nu-1] \\ x_{i,\nu}(m) &= c_{i-\nu+1}(m) \quad \text{for } i \in [\nu, \nu+N-1] \\ x_{i,\nu}(m) &= a \quad \text{for } i \geq \nu+n\ . \end{aligned} \tag{1}$$

The symbol a is known to the receiver. The symbol $c_i(m)$ denotes the i-th symbol of the block codeword $c^N(m)$ for message m. When no message is sent we may also consider that an i.i.d. random input sequence is sent, rather than a constant sequence of symbols (see Section III-B for an example). As in the detection and isolation problem discussed the previous section, the decoder consists of a sequential test (τ, ϕ), where τ is a stopping rule and where ϕ denotes a terminal decision rule.

Transmission is restricted to only one message. Indeed, since there is no feedback, the receiver has no means to inform the transmitter when a decision has been made. If feedback were available it would allow the sending of multiple messages, also using variable length codes. However, including feedback in our setting introduces significant subtleties to the analysis. In fact, the simplest setting would be to have a noiseless feedback since, if the feedback were noisy, the use of variable length codes would become problematic precisely because of the problem of synchronization between the transmitter and the receiver.[2] To avoid this circular problem we omit feedback in our study. Nonetheless, and as in the rateless coding setting, we may suppose in practice that a one-bit perfect feedback link is available so that multiple messages can be sent.

Our goal can be phrased as one of characterizing the maximum achievable error exponent with respect to the time it takes the decoder to react to a sent message, i.e., $(\tau-\nu)^+$. In this direction, we first define the decoding error probability as

$$\mathbb{P}(\mathcal{E}) = \frac{1}{L}\frac{1}{M}\sum_{m=1}^{M}\sum_{l=1}^{L}\mathbb{P}_{m,l}(\mathcal{E}), \tag{2}$$

where the subscripts $_{m,l}$ indicate that the probability is conditioned on the event that message m starts being sent at time l. Then we define the average rate

$$R = \frac{\ln M}{\mathbb{E}(\tau-\nu)^+}, \tag{3}$$

where

$$\mathbb{E}(\tau-\nu)^+ = \frac{1}{L}\frac{1}{M}\sum_{m=1}^{M}\sum_{l=1}^{L}\mathbb{E}_{m,l}((\tau-\nu)^+)\ . \tag{4}$$

We say that the error exponent E and the rate R are asymptotically achievable if there exists a sequence (labeled by the number of messages) of block codes and sequential tests such that

$$R = \lim_{M\to\infty}\frac{\ln M}{\mathbb{E}(\tau-\nu)^+} \tag{5}$$

and

$$E \geq \liminf_{n\to\infty} -\frac{1}{\mathbb{E}(\tau-\nu)^+}\ln\mathbb{P}(\mathcal{E})\ . \tag{6}$$

Implicitly in (5) and (6) we also allow the asynchrony level L to increase with M. Indeed, when L is fixed the problem is not particularly interesting: a decoder that makes a decision at time $L+N-1$ incurs no loss in either R or E due to asynchrony. Thus, we instead consider the situation where L scales with the number of messages M and write $L = L(M)$.

Our main result stands in the following theorem:

Theorem 1 *Suppose that* $\delta \triangleq \limsup_{M\to\infty}\frac{\ln L(M)}{\ln M} < \infty$. *For any discrete memoryless channel Q with capacity C, and for any rate $R \in [0, C]$, the error exponent $C - R(1+\delta)$ is achievable.* □

The key implication of this theorem is that, unless the asynchrony level L has an exponential growth rate with respect to the block length that is higher or equal than $C(Q)$, reliable communication can be guaranteed at sufficiently low rate.

[1] Here optimal decoding rules refer to sequential tests yielding minimum error probability given a certain delay.

[2] Suppose that the forward and the reverse channel have zero error capacity equal to zero and that multiple messages have to be sequentially sent. Since the reverse channel is noisy, if the receiver makes a decision on the basis of a stopping time τ, it is not clear how the transmitter and the receiver can agree upon what has to be sent at time $\tau+1$: starting a new message or continue to send bits with respect to the previous message.

Also notice that, even though the transmitter emits a block codeword, an error exponent higher than the sphere packing bound can be achieved at rates close to capacity (provided that $\delta = 0$). This seems to violate the fact that the sphere-packing exponent cannot be surpassed when the transmitter and the receiver are synchronized. This apparent contradiction is resolved by observing that even though the transmitter emits a block codeword of length N from time ν, the receiver uses a sequential decoder. This in turn allows us to achieve decoding delays $\tau - \nu$ smaller on average than N while keeping the probability of error close to that obtained by maximum likelihood decoding at the end of the message transmission, i.e., at time $\nu + N - 1$.

We now describe the communication scheme that asymptotically yields the error exponent claimed in the theorem. The codewords $\{c^N(m)\}_{m=1}^M$ are randomly generated so that the $c_i(m)$'s ($i \in \{1,\dots,N\}$ and $m \in \{1,\dots,M\}$) are i.i.d. samples according to the capacity-achieving distribution of the channel. For 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., $\hat{P}_{(x^n,y^n)}(x,y) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{(x,y)}(x_i, y_i)$. Also, given a distribution P on $\mathcal{X} \times \mathcal{Y}$ we write $I(P)$ for the mutual information induced by P. The decoding rule is chosen according to the following stopping time

$$\tau = \inf \Bigg\{ n \geq 1 \,:\, \exists m \in \mathcal{M} \text{ and } l \in \{1,\dots,N\} \text{ so that } \min_{k\in[1,\dots,l]} \Big[kI(\hat{P}_{c^k(m),y_{n-l+1}^{n-l+k}}) + (l-k) I(\hat{P}_{c_{k+1}^l(m),y_{n-l+k+1}^n}) \Big] \geq \alpha \ln M \Bigg\} \tag{7}$$

where $\alpha > 1$ is some fixed constant. At time τ the decoder declares the message for which the minimum of the sum of the mutual informations that defines τ exceeds $\alpha \ln M$. If multiple messages satisfy this condition the decoder picks the one with the smallest index. Loosely speaking, the decoder makes a decision as soon as there exist l last symbols ($l \in \{1,\dots,N\}$) that have "empirical mutual information" with one of the codewords that exceeds a certain threshold. Our choice of the decision rule in (7) is a consequence of our analysis of the error event occurring when Y_{n-l+1}^n is induced partly by noise and partly by the transmitted codeword.

Observe that our communication scheme does not subdivide the synchronization problem into a detection problem followed by a message isolation problem: detection and isolation are treated jointly. Also note that the above decoder is universal in the sense that its decision does not rely on the statistics of the channel under use. In fact this decoder is an extension of a universal decoder introduced in [4, eq. (10)] for the ordinary synchronized setting. As for that decoder, it can be shown that our new decoder (asymptotically) achieves a rate R equal to $C(Q)/\alpha$, where α is the constant appearing in (7).

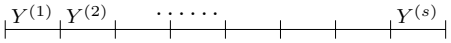

Fig. 1. Parsing of the received sequence into s blocks of length N.

III. Lower bound on the error probability without synchronization

Given a channel Q, a codebook $\{c^N(m)\}_{m=1}^M$, and a level of asynchrony L, we first derive a general lower bound on the probability of error. This lower bound appears to be quite complicated but can in principle be numerically evaluated. We then turn to a simple example — the binary symmetric channel with only two possible messages — and derive a simple sufficient condition on the growth rate of L as function of the block length N for the error probability to be asymptotically bounded away from zero. This sufficient condition, in the limit of vanishing crossover probability in the channel, also becomes necessary. Extending our analysis to the Gaussian channel with peak power constraint P, we derive a necessary condition on P as function of N and L so that the error probability is asymptotically vanishing.

A. General case

It might be tempting to conjecture that "if L grows faster than e^{NC} then the error probability is asymptotically bounded away from zero." To convince oneself that this claim is wrong in general, it suffices to consider a channel and a noise input symbol a that doesn't produce all the possible channel output symbols (e.g., consider the Z channel). In this case, by starting each message with a short prefix (say of length $\sqrt{N}$) that does not use any of the noise symbols that appear before and after transmission, one can easily show that there is no rate loss by lack of synchronization, no matter how large L is.

In order to derive a lower bound on the probability of error, we assume that the decoder makes a decision by observing a sequence Y of maximal length $L + N - 1$. Further, let us optimistically assume that the decoder knows that the message was sent in one of the $s = \lfloor (L+N-1)/N \rfloor$ time slots of duration N as shown in Fig. 1. We have

$$\mathbb{P}(\mathcal{E}) \geq \frac{1}{M}\frac{1}{s}\sum_{m=1}^{M}\sum_{l=1}^{s} \mathbb{P}_{m,l}(\mathcal{E}), \tag{8}$$

where $\mathbb{P}_{m,l}(\mathcal{E})$ denotes the probabilities of error given that message m is sent in the l-th slot of size N. The maximum likelihood decoder yields for any m'

$$\mathbb{P}_{m,l}(\mathcal{E}) \geq \mathbb{P}_{m,l}\left(\frac{\mathbb{P}_{m'}(Y)}{\mathbb{P}_m(Y)} > 1 \right). \tag{9}$$

Now letting $Y^{(i)}$ denote the sequence of length N occurring in the i-th slot in Y, we obtain

$$\mathbb{P}_{m,i}(Y) = Q(Y^{(i)}|c^N(m)) \prod_{j\neq i} Q(Y^{(j)}|a^N), \tag{10}$$

where a denotes the symbol that is sent when no message is sent. Therefore we obtain

$$\begin{aligned}&\mathbb{P}_{m,l}\left(\frac{\mathbb{P}_{m'}(Y)}{\mathbb{P}_m(Y)}>1\right)\\&=\mathbb{P}_{m,l}\left(\frac{\frac{1}{s}\sum_{i=1}^{s}Q(Y^{(i)}|c^N(m'))\prod_{j\neq i}Q(Y^{(j)}|a^N)}{\frac{1}{s}\sum_{i=1}^{s}Q(Y^{(i)}|c^N(m))\prod_{j\neq i}Q(Y^{(j)}|a^N)}>1\right)\\&=\mathbb{P}_{m,l}\left(\frac{\sum_{i=1}^{s}\frac{Q(Y^{(i)}|c^N(m'))}{Q(Y^{(i)}|a^N)}}{\sum_{i=1}^{s}\frac{Q(Y^{(i)}|c^N(m))}{Q(Y^{(i)}|a^N)}}>1\right),\end{aligned}\tag{11}$$

or, equivalently,

$$\mathbb{P}_{m,l}\left(\frac{\mathbb{P}_{m'}(Y)}{\mathbb{P}_m(Y)}>1\right)=\mathbb{P}_{m,l}\left(\sum_{i=1}^{s}Z_i>0\right),\tag{12}$$

where we defined $Z_i=\frac{Q(Y^{(i)}|c^N(m'))}{Q(Y^{(i)}|a^N)}-\frac{Q(Y^{(i)}|c^N(m))}{Q(Y^{(i)}|a^N)}$. By symmetry (12) is the same for any $l\in\{1,\dots,s\}$ and hence, for any $m'\neq m$

$$\mathbb{P}_{m,l}(\mathcal{E})\geq\mathbb{P}_{m,1}\left(\sum_{l=2}^{s}Z_l>-Z_1\right).\tag{13}$$

Under $\mathbb{P}_{m,1}$, all the Z_i's are independent and are all identically distributed with mean zero, except for Z_1 that has a negative mean. Unfortunately the analysis of the random walk $\{\sum_{i=2}^{n}Z_i\}_{n\geq 2}$ is cumbersome and no simple general lower bound on $\mathbb{P}_{m,1}\left(\sum_{i=2}^{s}Z_i>-Z_1\right)$ could be found. However, using the Chernoff bound we obtain

$$\begin{aligned}&\mathbb{P}_{m,1}\left(\sum_{i=2}^{s}Z_i>-Z_1\right)\\&\quad\geq\sup_{t\in\mathbb{R}}\sup_{q>0}[1-(\mathbb{E}_{m,1}e^{-qZ_2})^{s-1}]e^{qt}]\mathbb{P}_{m,1}(Z_1\geq t).\end{aligned}\tag{14}$$

Finally from (8) we obtain

$$\begin{aligned}&\mathbb{P}(\mathcal{E})\geq\\&\frac{1}{M}\sum_{m=1}^{M}\max_{m'\neq m}\sup_{t\in\mathbb{R}}\sup_{q>0}[1-(\mathbb{E}_{m,1}e^{-qZ_2})^{s-1}]e^{qt}]\mathbb{P}_{m,1}(Z_1\geq t).\end{aligned}\tag{15}$$

B. Binary symmetric channel

We consider communication over a binary symmetric channel with only two possible messages, A and B. When no message is sent, the channel outputs are random (i.i.d. Bernoulli $1/2$) bits. One can show that the codewords yielding the lowest error probability are $c^N(A)=(0,0,\dots,0)$ and $c^N(B)=(1,1,\dots,1)$. From Section III-A we get

$$\mathbb{P}(\mathcal{E})\geq\mathbb{P}_{A,1}\left(\sum_{j=2}^{s}Z_j>-Z_1\right).\tag{16}$$

Letting k_i denote the number of ones observed in the i-th slot of size N of the received sequence, an easy computation yields

$$\begin{aligned}&\mathbb{P}_{A,1}\left(\sum_{i=2}^{s}Z_i>-Z_1\right)\\&\qquad\geq\mathbb{P}_{A,1}\left(\sum_{i=2}^{s}u^{k_i}-u^{N(s)-k_i}>u^{N(s)}-1\right),\end{aligned}\tag{17}$$

where $u\triangleq(1-\varepsilon)/\varepsilon$, and where we write $N(s)$ since the block length can be chosen as a function of the uncertainty level. Note that to further lower bound the term on the right hand side of (17) we need to analyze the random walk $\sum_{i=2}^{s}u^{k_i}-u^{N(s)-k_i}$, since under $\mathbb{P}_{A,1}$ the k_i ($i\in\{2,\dots,s\}$) are i.i.d. Applying the Central Limit Theorem for triangular arrays (see, e.g., [1, p. 326]) we have that

$$\frac{1}{\sqrt{(s-1)\mathrm{Var}(u^{k_2}-u^{N(s)-k_2})}}\sum_{i=2}^{s}u^{k_i}-u^{N(s)-k_i}\tag{18}$$

converges to a normal standard random variable provided that the Lindeberg condition

$$\mathbb{E}_{A,1}\left(\frac{(u^{k_2}-u^{N(s)-k_2})^2}{\mathrm{Var}(u^{k_2}-u^{N(s)-k_2})}\mathbb{1}_{\frac{|u^{k_2}-u^{N(s)-k_2}|}{\sqrt{(\alpha-1)\mathrm{Var}(u^{k_2}-u^{N(s)-k_2})}}>\varepsilon}\right)=o(1)\tag{19}$$

is satisfied as $s\to\infty$. A straightforward computation shows that (19) is equivalent to

$$\frac{1}{s}\frac{\left(\frac{2}{1+\frac{1}{u^2}}\right)^{N(s)}}{1-\left(\frac{2}{u+\frac{1}{u}}\right)^{N(s)}}=o(1)\tag{20}$$

as $s\to\infty$. It can then be easily checked that under the same condition (20), the quantity

$$\frac{1}{\sqrt{(s-1)\mathrm{Var}(u^{k_2}-u^{N(s)-k_2})}}(u^{N(s)}-1)\tag{21}$$

tends to zero as $s\to\infty$. Therefore from (16), (17), (18), (20), and (21) we conclude that if (20) is satisfied then $\mathbb{P}(\mathcal{E})$ is lower bounded by a quantity that tends to $1/2$ as $s\to\infty$. One can finally check that, in the limit where $\varepsilon\to 0$, the condition (20) implies that if s grows faster than $e^{N(1+o(1))C}$, vanishing error probability cannot be asymptotically guaranteed. One then further deduces that, in the limit $\varepsilon\to 0$, our theorem is tight in the sense that, given our noise model, reliable communication can be achieved if and only if $L=o(e^{N(1-\delta)C})$ for some $\delta>0$.

C. The Gaussian channel

A similar approach as for the binary symmetric channel can be applied to the Gaussian channel where we now allow also the power to scale with the asynchrony level. Suppose that there is only two possible messages with codewords $c^N(A)=\sqrt{P}(1,1,\dots,1)$ and $c^N(B)=-\sqrt{P}(1,1,\dots,1)$. We assume

that the noise symbol a equals to zero. One can show that, for $P = P(L) > \frac{1}{2}\log 3$, as $L \to \infty$

$$\frac{1}{L} e^{N(L)(P(L)-\frac{1}{2}\log 3)} = o(1) \tag{22}$$

the probability of error is asymptotically bounded away from zero. Condition (22) is perhaps most useful in the regime when $(\log L)/N$ tends to a constant, in which case it gives a lower bound on the minimum amount of power needed for reliable communication given by $\frac{1}{N}\log L$.

IV. Conclusion

We have shown that a sufficient condition under which the lack of synchronization does not affect the set of achievable rates for discrete memoryless channels is that $\ln L$ is negligible compared to $\ln M$. In certain cases this condition is also necessary.

By extending our analysis to the Gaussian channel, we showed that a proper scaling of the power as a function of the asynchrony level and of the block length is needed in order to guarantee reliable communication.

Acknowledgment

A. Tchamkerten wishes to thank Emre Telatar for stimulating discussions.

References

[1] A.N. Shiryayev, "Probability," *Springer-Verlag*, 1984.

[2] T.L. Lai, "Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems," *IEEE Trans. Inform. Th.*, vol. 46, , pp. 595–607, 2000.

[3] I.V. Nikiforov, "A generalized change detection problem," *IEEE Trans. Inform. Th.*, vol. 41, , pp. 171–187, 1995.

[4] A. Tchamkerten and I.E. Telatar, "Variable length coding over an unknown channel," *IEEE Trans. Inform. Th.*, vol. 52, , pp. 2126–2145, 2006.