Communication Under Strong Asynchronism

Aslan Tchamkerten, Venkat Chandar, and Gregory W. Wornell, *Fellow, IEEE*

***Abstract*—A formulation of the problem of asynchronous point-to-point communication is developed. In the system model of interest, the message codeword is transmitted over a channel starting at a randomly chosen time within a prescribed window. The length of the window scales exponentially with the codeword length, where the scaling parameter is referred to as the asynchronism exponent. The receiver knows the transmission window, but not the transmission time.**

Communication rate is defined as the ratio between the message size and the elapsed time between when transmission commences and when the decoder makes a decision. Under this model, several aspects of the achievable tradeoff between the rate of reliable communication and the asynchronism exponent are quantified. First, the use of generalized constant-composition codebooks and sequential decoding is shown to be sufficient for achieving reliable communication under strictly positive asynchronism exponents at all rates less than the capacity of the synchronized channel. Second, the largest asynchronism exponent under which reliable communication is possible, regardless of rate, is characterized. In contrast to traditional communication architectures, there is no separate synchronization phase in the coding scheme. Rather, synchronization and communication are implemented jointly.

The results are relevant to a variety of sensor network and other applications in which intermittent communication is involved.

***Index Terms*—Change detection, detection and isolation problem, error exponents, intermittent communication, quickest detection, sensor networks, sequential decoding, synchronization.**

I. Introduction

In the traditional communication system architecture, the subsystem that encodes and decodes the bits to be communicated is designed and implemented separately from the subsystem that establishes a synchronized channel (e.g., by training). Such a separation is convenient, simplifying system design, and allowing code designers to focus on coding for synchronized channels. As a result, much information-theoretic analysis starts with the assumption of perfect synchronization between the transmitter and the receiver, and, indeed, many key quantities, including channel capacity, are defined accordingly [1].

However, in a variety of emerging applications involving intermittent or bursty communication, this architectural separation is less easily justified, and a suitable analysis of the overall problem of asynchronous communication—in which synchronization and communication aspects are combined—is required. In this paper, we propose and analyze a model for this problem, and use information-theoretic analysis to quantify appropriate notions of fundamental limits.

As a motivating application, consider, for example, a monitoring system in which a sensor will, on occasion, emit an alarm message to a command center, but will otherwise remain idle. The time at which an alarm is sent is determined by external events (the phenomenon being monitored), and thus the underlying communication channel is inherently asynchronous. In such an application, important parameters of the communication system are the message size (in bits), the "reaction delay" in detecting the sent message, and the probability of a decoding error.

In our simple point-to-point model, the message is encoded into a codeword of fixed length, and this codeword starts being sent at a time instant that is uniformly distributed over some predefined transmission window. The size of this window is known to both the transmitter and receiver, and governs the level of asynchronism in the system. The receiver uses a sequential decoder to detect and identify the sent message.

In our model, the transmission window size scales exponentially with the codeword length, where the scaling parameter is referred to as the asynchronism exponent. This scaling is rather natural. Indeed, if the window size scales subexponentially, then the price of asynchronism is negligible. By contrast, if the window size scales superexponentially, then the asynchrony is generally catastrophic. Hence, exponential asynchronism is the interesting regime.

In designing a suitable communication system, the goal is to deliver as large a message as possible, as quickly as possible, and as reliably as possible. These are, however, conflicting objectives in general, and thus we quantify the fundamental tradeoffs involved. Specifically, we first define communication rate as the ratio between the message size and the delay between when transmission starts and when the message is detected and decoded. We then describe the capacity region of an asynchronous channel as the efficient frontier of fundamental tradeoffs between achievable rates and the asynchronism exponents for reliable communication, i.e., subject to the constraint of a vanishing error probability in the limit of long codeword lengths.

In our analysis, we focus on discrete memoryless channels, for simplicity of exposition. Using a coding scheme comprising a random generalized constant-composition codebook and sequential decoding, we first develop sufficient conditions on the parameters of the scheme for a rate–exponent pair to be achievable. As an application of this result, we show that any rate below the capacity of the synchronized system can be achieved under some strictly positive asynchronism exponent.

Manuscript received July 27, 2007; revised February 11, 2008 and April 22, 2009. Current version published September 23, 2009. This work was supported in part by the National Science Foundation under Grant CCF-0635191, and by a University IR&D Grant from Draper Laboratory. The material in this paper was presented in part at the IEEE International Symposium on Information Theory (ISIT), Toronto, ON, Canada, July 2008.

A. Tchamkerten is with the Department of Communications and Electronics, Telecom ParisTech, 75634 Paris Cedex 13, France (e-mail: alsan.tchamkerten@telecom-paristech.fr).

V. Chandar and G. W. Wornell are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: vchandar@mit.edu; gww@mit.edu).

Communicated by H. Yamamoto, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2009.2027484

As our second result, we show there exists a synchronization threshold for such channels. This threshold is the largest possible asynchronism exponent that can yield reliable communication when the message consists of at least one bit. We characterize this threshold, which is a function of the underlying channel law. As illustrations, we quantify the synchronization threshold for both a basic binary symmetric channel, and for antipodal signaling over an additive white Gaussian noise channel with hard decision decoding.

An outline of the paper is as follows. Section II describes some background and related work, to put the present contributions in context, and Section III summarizes some notation and other conventions we make use of throughout the paper. Section IV formally develops our system, channel, and problem model of interest, and Section V introduces a particular coding scheme for asynchronous channels that we use to establish our achievability results. Section VI summarizes and interprets our main results, which take the form of two theorems, and a development of these results is provided in Section VII. Finally, Section VIII contains some concluding remarks.

II. Related Models and Problems

There have been a variety of attempts to model and analyze different types of asynchronism in communication systems. Some of the earliest work dates back to the 1960s. In one model from that time [2], a stream of messages is encoded into a sequence of fixed-length codewords that are transmitted, one immediately after the other. The receiver obtains channel outputs beginning at a random point in time in the overall transmission. The goal of the receiver is to detect the location of the next codeword boundary and begin decoding all subsequent messages.

A somewhat different line of inquiry from that time period is represented by the well-known insertion, deletion, and substitution (IDS) channel model of Dobrushin [3], which has seen renewed interest recently [4], [5]. The IDS channel is aimed at modeling a different phenomenon than that of this paper—namely, timing error and irregularity in a communication medium and transceiver hardware. As such it is complementary to our model. In particular, 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. As such, the channel is characterized by the set of all conditional output distributions $Q(\boldsymbol{y}|x)$ for each x in a finite alphabet $\mathcal{X}$, where $\boldsymbol{y}$ is a string of some length (even zero) of symbols from a finite alphabet $\mathcal{Y}$.

With the IDS channel model, the duration of the transmission is random, but the receiver implicitly knows the timing of the last output symbol. For instance, in the special case of the deletion channel, where each input symbol is deleted with some probability, if the codeword $c^N(m)$ produces the output $\boldsymbol{y} \in \mathcal{Y}^k$, $k \leq N$, the receiver knows that *nothing* comes after time k.[1]

By contrast, in our model the receiver knows neither the time at which transmission starts, nor the timing of the last information symbol. However, we do not model timing uncertainty during the information transmission—the duration of the transmission is always equal to a codeword length. 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.

A second kind of asynchronism is that between users in a multiuser communication setting, a particular example of which is the multiple-access problem. Examples of information-theoretic analysis of the effects of such asynchronism in multiple-access communication include [6]–[9], which focus on quantifying the capacity region under various assumptions on the asynchronism among users.

With respect to other work, perhaps the sequential decision problem most closely related to our problem formulation is a generalization of the change-point problem [10] often referred to as the "detection and isolation problem"—see [11]–[13] for a survey. In this problem, introduced in [11], a process starts with some initial distribution and then changes at some unknown time. The post-change distribution can be any of a given set of distributions. From the sequence of observations, the goal is to quickly react to the statistical change and isolate its cause, i.e., identify the post-change distribution, subject to a false-alarm constraint.

While in both the synchronization problem and the detection and isolation problem the goal is to quickly identify the cause of a change in distribution, there are important distinctions between these two problems as well. First, in the detection and isolation problem it is assumed that, once the observed process changes distributions, it remains in the 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 codeword transmission the distribution reverts to its pre-change state.

Second, it is also important to note that the synchronization problem has a codebook design component that the detection and isolation problem does not. In particular, since the changes in distribution are controlled by the number and choice of codewords, the ease and quickness with which change can be detected and isolated depends strongly on the codebook design. Moreover, the best choice of codebook, in turn, depends strongly on the channel parameters.

Finally, in the language of the synchronization problem, the detection and isolation problem is focussed 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. By contrast, the synchronization problem examines the effects of scaling the number of messages.

III. Notation

In general, we reserve capital letters for random variables (e.g., Y) and lower case letters to denote their corresponding realizations (e.g., y), though as is customary, we make a variety of exceptions. Any potential confusion is generally avoided by context. In addition, we use x_i^j to denote the sequence $x_i, x_{i+1}, \ldots, x_j$, for $i \leq j$. Moreover, when $i = 1$, we use the usual simpler notation x^n as an alternative to x_1^n.

[1] In [3], Dobrushin discusses the assumption that the receiver implicitly knows the length of the received sequence; see the discussion after Theorem 1. To avoid this assumption, beyond "one-shot" communication, Dobrushin also analyzes the situation where an infinite sequence of messages are sent and sequentially decoded on the basis of stopping times.

Events (e.g., $\mathcal{E}$) and sets (e.g., $\mathcal{S}$) are denoted using the calligraphic fonts, and if $\mathcal{E}$ represents an event, $\mathcal{E}^{\mathrm{c}}$ denotes its complement. As additional notation, $\mathbb{P}(\cdot)$ and $\mathbb{E}(\cdot)$ denote the probability and expectation of their arguments, respectively, $\|\cdot\|$ denotes the L_1 norm of its argument, $|\cdot|$ denotes absolute value if its argument is numeric, or cardinality if its argument is a set, $\lfloor\cdot\rfloor$ denotes the integer part of its argument, $a \wedge b = \min\{a, b\}$, and $x^+ = \max\{0, x\}$. Furthermore, we denote the Kronecker function using

$$\mathbb{1}_x(x') = \begin{cases} 1, & \text{if } x' = x \\ 0, & \text{otherwise.} \end{cases}$$

We also make use of some familiar order notation for asymptotics. We use $\mathrm{poly}(\cdot)$ to denote a term that grows no faster than polynomially in its argument. We use $o(\cdot)$ to denote a (positive or negative) quantity that grows more slowly than its argument; e.g., $o(1)$ denotes a term that vanishes in the limit. Finally, we use $\Theta(\cdot)$ to denote a nonnegative quantity that is asymptotically bounded above and below by its argument, to within constants of proportionality.

We denote by $\mathcal{P}^{\mathcal{X}}$, $\mathcal{P}^{\mathcal{Y}}$, and $\mathcal{P}^{\mathcal{X}\times\mathcal{Y}}$ the set of all distributions on $\mathcal{X}, \mathcal{Y}$, and $\mathcal{X} \times \mathcal{Y}$, respectively. The set of all conditional distributions of the form $V(y|x)$, $(x,y) \in \mathcal{X} \times \mathcal{Y}$, we denote using $\mathcal{P}^{\mathcal{Y}|\mathcal{X}}$.

For $P \in \mathcal{P}^{\mathcal{X}}$, we use P^n to denote the product distribution induced by P over $\mathcal{X}^n$ for some $n \geq 1$, i.e.,

$$P^n(x^n) = \prod_{i=1}^{n} P(x_i).$$

Likewise, for a memoryless channel characterized by channel law $Q \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}}$, the probability of the output sequence $y^n \in \mathcal{Y}^n$ given an input sequence $x^n \in \mathcal{X}^n$ is

$$Q^n(y^n|x^n) = \prod_{i=1}^{n} Q(y_i|x_i).$$

Additionally, for a distribution $V \in \mathcal{P}^{\mathcal{X}\times\mathcal{Y}}$, we use $V_X \in \mathcal{P}^{\mathcal{X}}$ and $V_Y \in \mathcal{P}^{\mathcal{Y}}$ to denote its left and right marginals, respectively; specifically, for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$

$$V_X(x) = \sum_{y\in\mathcal{Y}} V(x,y) \quad \text{and} \quad V_Y(y) = \sum_{x\in\mathcal{X}} V(x,y).$$

We use $D(\cdot\|\cdot)$ to denote the usual information divergence with respect to the natural logarithm, so for distributions $Q, Q' \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}}$, we have for $x \in \mathcal{X}$

$$D(Q(\cdot|x)\|Q'(\cdot|x)) = \sum_{y} Q(y|x) \ln \frac{Q(y|x)}{Q'(y|x)}.$$

Moreover, the expectation of this divergence with respect to the distribution P over $\mathcal{X}$ is denoted using

$$\begin{aligned} D(PQ\|PQ') &= \mathbb{E}[D(Q(\cdot|X)\|Q'(\cdot|X))] \\ &= \sum_{x\in\mathcal{X}} P(x)\, D(Q(\cdot|x)\|Q'(\cdot|x)). \end{aligned}$$

We likewise denote by $I(PQ)$ the mutual information induced by the joint distribution $P(\cdot)\,Q(\cdot|\cdot)$, i.e.,

$$I(PQ) = \sum_{x\in\mathcal{X}} P(x) \sum_{y\in\mathcal{Y}} Q(y|x) \ln \frac{Q(y|x)}{(PQ)_Y(y)}.$$

Additionally, we use $H_{\mathrm{B}}(\cdot)$ to denote the binary entropy function, i.e., for $p \in [0,1]$

$$H_{\mathrm{B}}(p) \triangleq -p\ln p - (1-p)\ln(1-p).$$

In our analysis, we make use of the usual notion of strong typicality [15]. In particular, a sequence y^N is strongly typical with respect to the distribution $P \in \mathcal{P}^{\mathcal{Y}}$ for some (implicit) parameter $\mu \in (0,1)$ if $|\hat{P}_{y^N}(b) - P(b)| < \mu$ for all $b \in \mathcal{Y}$.

More generally, we make frequent use of the method of types, and rely on the familiar notation for types. In particular, $\hat{P}_{x^n}$ denotes the empirical distribution (or type) of a sequence $x^n \in \mathcal{X}^n$, i.e.,

$$\hat{P}_{x^n}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_x(x_i).$$

The joint empirical distribution $\hat{P}_{(x^n,y^n)}$ for a sequence pair (x^n, y^n) is defined analogously. In turn, a distribution P over some alphabet $\mathcal{X}$ is said to be an n-type if $nP(x)$ is an integer for all $x \in \mathcal{X}$. The set of all n-types over an alphabet $\mathcal{X}$ is denoted using $\mathcal{P}_n^{\mathcal{X}}$, and that over the alphabet $\mathcal{X} \times \mathcal{Y}$ is denoted by $\mathcal{P}_n^{\mathcal{X}\times\mathcal{Y}}$, etc. Finally, the n-type class $\mathcal{T}(P)$ of P is defined to be the set of all sequences x^n that have type P, i.e., such that $\hat{P}_{x^n} = P$.

IV. Problem Formulation

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 the transition probabilities $Q(y|x)$, for all $y \in \mathcal{Y}$ and $x \in \mathcal{X}$. 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 $m \in \{1, 2, \ldots, M\}$.[2] For each message m there is an associated codeword

$$c^N(m) \triangleq c_1(m)c_2(m)\cdots c_N(m)$$

which is a string of N symbols drawn 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 over $\{1, 2, \ldots, A\}$. The transmitter and the receiver know the integer parameter $A \geq 1$, which we refer to as the *asynchronism level* of the channel. Note that the special case $A = 1$ corresponds to the classical synchronous communication scenario.

The receiver begins observing data starting at time $i = 1$. When a codeword is transmitted, a noise-corrupted version of the codeword is obtained at the receiver. When the transmitter is silent, the receiver observes only noise. To formally characterize the output distribution when no input is provided to the channel, it is notationally convenient to make use of a specially designated "no-input" symbol $\star$ in the input alphabet $\mathcal{X}$, as depicted in Figs. 1 and 2. Specifically, $Q(\cdot|\star)$ characterizes

[2]See the companion paper [16] for analysis of the case $M = 1$, corresponding to the pure synchronization problem.

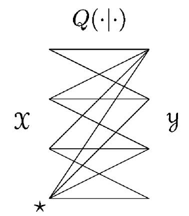

Fig. 1. Graphical depiction of the transition matrix for a discrete memoryless channel. The "no-input" symbol $\star$ is used in characterizing the channel output when the transmitter is silent. In this example, there are three possible non-silence input symbols, and four possible output symbols. The no-input symbol can be used like any other input symbol in codeword construction.

the noise distribution of the channel when there is no channel input. Hence, conditioned on the value of ν and on the message m to be conveyed, the receiver observes independent symbols $Y_1, Y_2, \ldots$ distributed as follows. If $i \in \{1, 2, \ldots, \nu - 1\}$ or $i \in \{\nu + N, \nu + N + 1, \ldots, A + N - 1\}$, the distribution at time i is $Q(\cdot|\star)$. However, if $i \in \{\nu, \nu + 1, \ldots, \nu + N - 1\}$, the distribution at time i is $Q(\cdot|c_{i-\nu+1}(m))$. Note that since the transmitter can choose to be silent for arbitrary portions of its length-N transmission as part of its message-encoding strategy, the symbol $\star$ is eligible for use in the codebook design.[3]

The decoder takes the form 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. Recall that a stopping time τ (deterministic or randomized) is an integer-valued random variable with respect to a sequence of random variables $\{Y_i\}_{i=1}^{\infty}$ so that the event $\{\tau = n\}$, conditioned on $\{Y_i\}_{i=1}^{n}$, is independent of $\{Y_i\}_{i=n+1}^{\infty}$ for all $n \geq 1$. The function ϕ is then defined as any $\mathcal{F}_\tau$-measurable map taking values in $\{1, 2, \ldots, M\}$, where $\mathcal{F}_1, \mathcal{F}_2, \ldots$ is the natural filtration induced by the process $Y_1, Y_2, \ldots$.

We are interested in systems that convey as many message bits as possible, and such that they can be detected and decoded as quickly and reliably as possible. Given these competing objectives, we formulate the system design problem as follows.

First, we define the average probability of a decoding error (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.

Second, we define the average communication rate with respect to the receiver's average delay in reacting to the sent message, i.e.,

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

where[4]

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

with $\mathbb{E}_{m,\nu}$ denoting expectation with respect to $\mathbb{P}_{m,\nu}$.

Defining rate as the "message-size-to-reaction-delay ratio" as in (1) combines message size and reaction delay into a single, physically meaningful figure of merit. Thus, large communication rates R are achieved via large messages sizes and/or small reaction delays.

Additional insight is obtained by rewriting (1) in the form

$$R = \frac{\bar{R}}{\bar{\Delta}} \tag{2}$$

where[5]

$$\bar{R} \triangleq \frac{\ln M}{N} \tag{3}$$

is the normalized message size, and where

$$\bar{\Delta} \triangleq \frac{\mathbb{E}(\tau - \nu)^+}{N} \tag{4}$$

is the normalized reaction delay at the decoder. We refer to the normalized message size (3) as the *code* rate, and it is measured in nats per channel use, i.e., there are $M = e^{N\bar{R}}$ messages. We also emphasize that $\bar{\Delta}$ may be either greater or less than one, and thus R may be greater or less than $\bar{R}$.

With the above definitions, we formulate our system design problem for a given discrete memoryless channel Q and asynchronism level A as one of maximizing the communication rate R subject to the constraint of a small decoding error probability $\mathbb{P}(\mathcal{E})$.

In our analysis, we allow the block length N to be arbitrarily large. If the asynchronism level A is subexponential in N, then there is no rate loss on the asynchronous channel (relative to the capacity of the corresponding synchronous channel), as we will discuss. But if the level of asynchronism grows at least exponentially in the block length, i.e., $A = e^{\alpha N}$ for some constant $\alpha > 0$, a rate loss can be experienced.

With such exponential scaling, as is the focus in the paper, the parameter α, which we refer to as the *asynchronism exponent*, can be interpreted as the number of nats per channel use required to describe the starting time ν of communication. From this perspective, we see that in such asynchronous communication, a total of $N(\alpha + \bar{R})$ nats of information is effectively conveyed over the time interval of size $A + N - 1$. It should be emphasized, however, that such reasoning is rather loose—there is no requirement in our system that the decoder be able to reliably recover ν, only the message.

We also emphasize that incorporating reaction delay into the performance criteria is important with our communication

[3]However, it should be emphasized that which symbol in the alphabet is the no-input symbol is a characteristic of the channel, and therefore beyond the control of the code designer.

[4]Note that the true reaction delay is actually $\tau - \nu + 1$ rather than $\tau - \nu$. However, in our asymptotic analysis, the distinction is negligible, and thus we use the former simply for aesthetic reasons.

[5]Note that we suppress the implicit dependence of $\bar{R}$ on N in our notation.

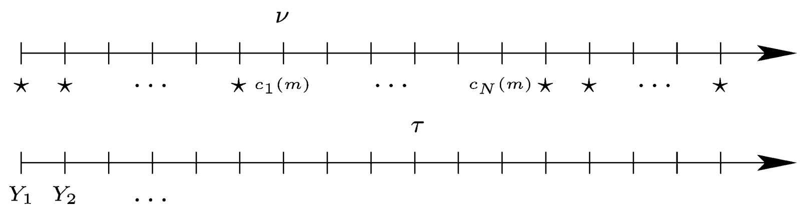

Fig. 2. Temporal representation of the channel input sequence (upper axis) and channel output sequence (lower axis). At time ν message m starts being sent and decoding occurs at time τ. The symbol $\star$ indicates when there is no input to the channel, i.e., when the transmitter is silent.

model. In particular, the channel is being used for communication (and thus unavailable for subsequent use) whenever the receiver is listening for the message. So, for example, without an explicit constraint on reaction delay, larger messages sizes and/or smaller decoding error probabilities will always favor the decoder making decisions as late as possible, i.e., at time $\tau = A+N-1$, which grows exponentially with N. In a practical sense, such a communication system would be particularly inefficient, since the receiver would be interacting with the channel for a time interval that is exponential in the transmission block length N, and thus the effective rate of communication would be vanishingly small.

To develop our results, we introduce the following formal definitions. First, we have the following natural notion of a coding scheme.

Definition 1 (An (R, α) Coding Scheme): Given a channel Q, a rate–exponent pair (R, α) is *achievable* if there exists a sequence $\{(\mathcal{C}_N, (\tau_N, \phi_N)\}$, for $N = 1, 2, \ldots$, of codebook/decoder pairs, indexed by the codebook length N, such that for any $\epsilon > 0$ and N sufficiently large, the pair $(\mathcal{C}_N, (\tau_N, \phi_N))$

i) operates under asynchronism level $A = e^{(\alpha-\epsilon)N}$;

ii) yields an average rate at least equal to $R - \epsilon$;

iii) achieves an average error probability $\mathbb{P}(\mathcal{E})$ of at most 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 the rate–exponent pair (R, α).

The capacity region characterizes the performance of the best coding schemes.

Definition 2 (Asynchronous Capacity Region): The capacity region of an asynchronous discrete memoryless channel Q is the set of rate–exponent pairs $(R, \alpha(R, Q))$ for $0 \leq R \leq C(Q)$, i.e.,

$$[0, C(Q)] \to \mathbb{R}_+$$
$$R \mapsto \alpha(R, Q)$$

where $C(Q)$ is the capacity of the corresponding synchronous channel, and where $\alpha(R, Q)$ is the supremum of the set of asynchronism exponents that are achievable at rate R.

Note that since $C(Q)$ is the highest achievable rate over the synchronous channel, the rate over the corresponding asynchronous channel cannot be higher. Hence, in developing the capacity region, it suffices to restrict attention to the rates in the interval $[0, C(Q))$.

For a given channel Q, the asynchronism exponent function $\alpha(R, Q)$ is nonincreasing in R. Hence, the highest asynchronism level for which reliable communication is possible is obtained when the communication rate is zero. This motivates the following definition.

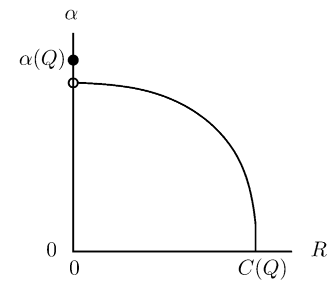

Fig. 3. The (hypothetical) capacity region of a discrete memoryless asynchronous channel with channel law Q. The nonincreasing curve $\alpha(R, Q)$ defines the upper boundary (efficient frontier) of achievable (R, α) pairs, where R is the rate and α is the asynchronism exponent. Moreover, $\alpha(Q)$ is the synchronization threshold, and $C(Q)$ is the capacity of the associated synchronous channel. In this example, $\alpha(R, Q)$ is not continuous at $R = 0$.

Definition 3 (Synchronization Threshold): The *synchronization threshold* of a channel Q, denoted using $\alpha(Q)$, is the supremum of the set of achievable asynchronism exponents at all rates, i.e., $\alpha(Q) \triangleq \alpha(R = 0, Q)$.

For the purposes of illustration, a hypothetical capacity region is depicted in Fig. 3. In this example, the region has a discontinuity at $R = 0$.

Our main results, developed in Section VI, take the form of properties of the asynchronous capacity region of Definition 2. In particular, we provide a simple characterization of the synchronization threshold $\alpha(Q)$, and more generally develop a nontrivial inner bound on the asynchronous capacity region.

V. A Coding Scheme for Asynchronous Channels

The following coding scheme will be used in the development of the main results of the paper. A significant feature of the resulting communication system is that it does not consist of separate transmission detection and message identification subsystems. Rather, detection and identification are treated jointly.

We use a random code construction. In particular, the $e^{N\bar{R}}$ codewords, each of length N, are drawn randomly from a suitable ensemble governed by a distribution $P \in \mathcal{P}^{\mathcal{X}}$. As is typical, this codebook is fixed for the duration of system operation. Before describing the relevant ensemble in more detail, we first describe the decoder.

A. Decoder Design

During decoding, there are two sources of error. The first comes from atypical channel behavior at times when no codeword is being sent, which may result in what we refer to as a false alarm—detecting the presence of a codeword before one has been sent. The second comes from atypical channel behavior during codeword transmission, which may result in what we refer to as a misidentification—detecting the wrong codeword after one has been sent. The probabilities of these kinds of errors depend on the asynchronism level and communication rate. In particular, the higher the asynchronism level α, the more likely a false alarm, while the higher the code rate $\bar{R}$, the more likely a misidentification.

Accordingly, our decoder takes into account both sources of error, and is parameterized by $\bar{R}$ and the target α. More specifically, from the sequence of symbols it observes, the decoder makes a decision as soon as a block of $i \in \{1, 2, \ldots, N\}$ consecutive such symbols satisfy two conditions: 1) the block of symbols must be sufficiently different, in a divergence sense, from the noise; and 2) the block of symbols must be sufficiently correlated, in a mutual information sense, with one of the codewords.

A formal description of our decoder is as follows. Decoding occurs at time

$$\tau = \min_{m \in \{1,2,\ldots,M\}} \tau_m \tag{5}$$

where τ_m is a stopping time associated with message m, and the declared message $\hat{m}$ is any that satisfies $\tau_{\hat{m}} = \tau$. The stopping time τ_m takes the form

$$\tau_m = \inf\{n \geq 1 : \exists i \in \{1, \ldots, N\} \text{ s.t. } D_{n,i} \geq t_D \bar{R} \quad \text{and} \quad I_{n,i}(m) \geq t_I \bar{R}\} \tag{6}$$

where

$$D_{n,i} \triangleq \frac{i}{N} D\left(\hat{P}_{y^n_{n-i+1}} \,\|\, Q(\cdot|\star)\right) \tag{7}$$

and

$$I_{n,i}(m) \triangleq \min_{k \in \{1,\ldots,i\}} I_{n,i,k}(m) \tag{8}$$

with

$$I_{n,i,k}(m) \triangleq \frac{k}{N} I\left(\hat{P}_{c^k(m), y^{n-i+k}_{n-i+1}}\right) + \frac{i-k}{N} I\left(\hat{P}_{c^i_{k+1}(m), y^n_{n-i+k+1}}\right) \tag{9}$$

and with the parameters $t_D \geq 0$ and $t_I > 1$ appropriately chosen as a function of $\bar{R}$ and the target α.

Remarks on Decoding Rule

First, we note that the choice of thresholds is determined by the communication rate of interest. In general, to ensure a small probability of misidentification, the thresholds $t_D \bar{R}$ and $t_I \bar{R}$ must be chosen below $D((PQ)_Y \| Q(\cdot|\star))$ and $I(PQ)$, respectively. As we will see, at zero rate, they need only be slightly below.

We also note that in [17] a simpler decoding rule is considered in which the stopping times are of the form (6), but without the divergence condition $D_{n,i} \geq t_D \bar{R}$, i.e., a decision is made as soon as for some m and i we have $I_{n,i}(m) \geq t_I \bar{R}$. With such a decoder, it is possible to achieve asynchronism exponents α as large as the capacity $C(Q)$ of the synchronized channel. However, it is unclear whether asynchronism exponents beyond $C(Q)$ are achievable with this simplification. By contrast, with the decoder (6), asynchronism exponents larger than $C(Q)$ (and indeed all the way up to $\alpha(Q)$) can be achieved.

Also, it might seem that the term $I_{n,i}(m)$ used in the decoder is unnecessarily complicated, and could be replaced with, for instance, the simpler term

$$I'_{n,i}(m) \triangleq \frac{i}{N} I\left(\hat{P}_{c^i(m), y^n_{n-i+1}}\right)$$

corresponding to fixing $k = i$. However, system performance with this simplified decoder is more difficult to analyze in the scenario when the i symbols being observed by the decoder lie partly inside and partly outside the transmission interval. As such, our particular choice of decoder is one of convenience.

Finally, it should be emphasized that while other sequential decoder designs may achieve the performance levels established in the sequel, a noteworthy feature of our decoder is that it is also nearly universal. In particular, the rule does not depend on the channel statistics, except for the noise distribution $Q(\cdot|\star)$. In fact, this decoder can be viewed as an extension to the asynchronous channel of a sequential universal decoder introduced in [18, eq. (10)] for the synchronized setting.

B. Codebook Design

Our random code construction is based on a natural generalization of the constant-composition codebooks described in, e.g., [15, p. 117]. Specifically, a code of generalized constant composition with respect to a distribution P is one in which the codewords and all their prefixes of significant size have an empirical type close to P. Formally, we have the following definitions.

Definition 4 (Generalized Constant-Composition Code): A codeword c^N is said to have *sequential constant composition* with respect to distribution P, denoted $c^N \in \mathcal{K}_N(P)$, if[6]

$$\|P - \hat{P}_{c^i}\| \leq \frac{1}{\ln N} \quad \text{whenever } \frac{N}{\ln N} < i \leq N. \tag{10}$$

Furthermore, a *generalized constant-composition codebook* is one in which all codewords have sequential constant composition.

The value of the generalized constant-composition codebook is as follows. With an independent and identically distributed (i.i.d.) codebook, in which each of the N elements of each of the $M = e^{N\bar{R}}$ codewords is i.i.d. according to the prescribed distribution P, there is a small probability that any codeword of interest will be atypical, i.e., have an empirical distribution that is not close to the P. This effect ultimately contributes to the overall error probability of the coding scheme. By contrast, in a generalized constant-composition codebook, this additional source of error is eliminated as all codewords are guaranteed to have an empirical distribution sufficiently close to P. Moreover, requiring that prefixes of the codewords also have their empirical distributions constrained in this manner takes into account

[6]Our choices of $1/\ln N$ for the type match accuracy and $N/\ln N$ for the minimum prefix size are convenient but not unique, as will become apparent.

that the decoder makes decisions based on blocks of data that may be smaller than the full codeword length.

Remarks on Codebook Construction

First, it is conceptually straightforward to convert an i.i.d. codebook into a generalized constant-composition one, i.e., to generate the generalized constant-composition ensemble from the i.i.d. ensemble. In particular, given a message m, the codeword $c^N(m)$ is generated so that all of its symbols are i.i.d. according to the distribution P. If the obtained codeword does not satisfy the sequential constant-composition property (10), it is discarded and a new codeword is regenerated until the sequential constant-composition condition is satisfied.

In practice, very little regeneration is required to generate each codeword. Indeed, if N is large enough, with overwhelming probability a random codeword satisfies the sequential constant-composition property. Specifically, we have the following.

Lemma 1: The probability that a sequence $C_1, C_2, \dots, C_N$ of random variables i.i.d. according to P satisfies the sequential constant-composition condition tends to one as $N \to \infty$.

Proof: By the union bound, the probability of generating a sequence $c^N(m)$ that does not satisfy the sequential constant-composition condition is upper-bounded by $N \exp[-\Omega(N/(\ln N)^3)]$, which tends to zero as $N \to \infty$. □

Note that with generalized constant-composition codebooks, it simplifies the analysis to impose the mild constraint that

$$\bar{R} \geq \frac{\ln |\mathcal{X}|}{\ln N} \tag{11}$$

in order to ensure that the decoder only operates on codeword prefixes large enough that the sequential constant-composition property holds. To verify this, referring to (6) in the description of our coding scheme, we see that if i_* denotes the minimum value of i for which decoding can occur, then

$$\begin{aligned} \frac{i_*}{N} \ln |\mathcal{X}| &\geq I_{n,i_*}(m) && (12)\\ &= \min_{k \in \{1,\dots,i_*\}} I_{n,i_*,k}(m) && (13)\\ &\geq t_I \bar{R} \\ &> \bar{R} && (14)\\ &\geq \frac{\ln |\mathcal{X}|}{\ln N} && (15) \end{aligned}$$

where to obtain (12) we have used that $I(V) \leq H(V_X) \leq \ln |\mathcal{X}|$ for all V, where to obtain (14) we have used that $t_I > 1$, and where to obtain (15) we have used (11). Hence

$$i_* \geq \frac{N}{\ln N}, \tag{16}$$

from which we see that the message will only ever be decoded from a prefix of the codeword that is sufficiently long that its behavior is controlled by the sequential constant-composition condition as defined in Definition 4. Note that any codebook with exponentially many codewords, as will generally be adequate for our purposes, will meet this condition.

In general, the codebook distribution needs to be tailored to the target rate of interest. Some useful insight is gained by examining the zero-rate regime. As will become apparent, in this regime, to ensure a sufficiently small false-alarm probability, given the threshold choices needed to control the probability of misidentification, P must be chosen so as to ensure both $I(PQ) > \bar{R}$ and

$$(D((PQ)_Y \| Q(\cdot|\star)) - \alpha) + (I(PQ) - \bar{R}) > 0. \tag{17}$$

By choosing $\bar{R}$ small enough, the first of these conditions is readily satisfied for a rich class of distributions P. Thus, with such an $\bar{R}$, it suffices to choose a P in this class such that $D((PQ)_Y \| Q(\cdot|\star)) > \alpha$ to ensure reliable communication.

Evidently, to accommodate the largest possible levels of asynchronism, we should choose a codebook such that the induced output distribution is as far as possible (in a divergence sense) from the noise distribution. But

$$D((PQ)_Y \| Q(\cdot|\star)) \leq \arg\max_x D(Q(\cdot|x) \| Q(\cdot|\star)) \tag{18}$$

via the convexity of divergence, with equality if P is the distribution in which some maximizing symbol x_* in (18) is used with probability one. Hence, assuming, as is the case, the class of distributions P such that $I(PQ) > \bar{R}$ includes distributions arbitrarily close to this maximizing one, reliable communication is possible whenever

$$\alpha < D(Q(\cdot|x_*) \| Q(\cdot|\star)).$$

As we will see, this turns out to be precisely the synchronization threshold of the channel. Thus, using a codebook distribution in which codewords are composed primarily of the symbol x_* accommodates the largest possible asynchronism exponent.

Beyond the zero-rate regime, finding optimal choices for P is more complicated, though we find useful choices in the sequel. Nevertheless, the general strategy is the same: among all P that allow the target rate to be achieved, we choose that which is as different as possible from the noise distribution of the channel.

VI. Results

In this section, we summarize, interpret, and discuss our main results. We also present a couple of representative examples.

We begin with the following useful inner bound on the capacity region $\alpha(R, Q)$, a proof of which is given in Section VII-A.

Theorem 1: Let Q be a discrete memoryless channel such that $Q(y|\star) > 0$ for all $y \in \mathcal{Y}$. If for some constants $\alpha \geq 0, \Delta \in (0,1), t_D \geq 0, t_I > 1$, and input distribution P such that $I(PQ) > 0$, the following conditions are satisfied:

$$\alpha = \left(\frac{t_D + t_I - 1}{t_I}\right) I(PQ)\Delta^2 \tag{I}$$

$$\frac{t_D}{t_I} < \frac{D((PQ)_Y \| Q(\cdot|\star))}{I(PQ)} \tag{II}$$

$$0 < \alpha_D^+ \triangleq \inf_{\substack{\{V \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}}: \\ D((PV)_Y \| Q(\cdot|\star)) \\ < (t_D/t_I) I(PQ)\Delta\}}} D((PV)_Y \| (PQ)_Y) \tag{III}$$

then the pair $(R = 0, \alpha)$ is achievable, where the infimum in (III) is defined to be $+\infty$ whenever the set over which it is defined is empty. If, in addition, the following conditions are also satisfied:

$$\alpha < \alpha_D^+ \tag{IV}$$

$$\alpha < \alpha_I^+ \triangleq \min_{\substack{\{V \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}}: \\ I(PV) \leq I(PQ)\Delta\}}} D(PV \| PQ), \tag{V}$$

then the pair $(R = I(PQ)/t_I, \alpha)$ is achievable. Moreover, in both cases, communication is achieved using a codebook of rate $\bar{R} = I(PQ)\Delta/t_I$.

Note that the conditions (III)–(V) in Theorem 1 are easy to check numerically since they involve only convex optimizations.

As our proof reveals, the identified rate pairs are achieved by a combination of generalized constant-composition codebooks and our sequential decoding rule. Some additional comments on this construction are worthwhile. In particular, we begin by noting that the right-hand side of condition (V) is $E_{\mathrm{sp}}(I(PQ)\Delta, P, Q)$, where

$$E_{\mathrm{sp}}(r, P, Q) = \min_{\substack{\{V\in\mathcal{P}^{\mathcal{Y}|\mathcal{X}}:\\ I(PV)\le r\}}} D(PV\|PQ) \tag{19}$$

is the sphere-packing exponent function at rate r for a channel Q with input distribution P. Since the sphere-packing exponent at rate r is (see, e.g., [15, p. 166])

$$E_{\mathrm{sp}}(r, Q) = \max_{P\in\mathcal{P}^{\mathcal{X}}} E_{\mathrm{sp}}(r, P, Q)$$

we conclude that at strictly positive rates, the asynchronism exponents achieved in Theorem 1 cannot exceed the zero-rate sphere-packing exponent for the channel, i.e.,

$$E_{\mathrm{sp}}(Q) = E_{\mathrm{sp}}(0, Q) = \max_{P\in\mathcal{P}^{\mathcal{X}}} \min_{V\in\mathcal{P}^{\mathcal{Y}}} D(PV\|PQ).$$

A key implication of Theorem 1 is given by the following corollary, which establishes that the asynchronous capacity region is nondegenerate—reliable communication is possible at all rates below the synchronous capacity even with exponentially large levels of asynchronism.

Corollary 1: For every channel Q with synchronous capacity $C(Q) > 0$, any rate $R \in [0, C(Q))$ can be achieved at a strictly positive asynchronism exponent.

Proof of Corollary 1: For $R \in (0, C(Q))$, consider conditions (I)–(IV) in Theorem 1. First set $t_D = 0$ and choose $t_I > 1$ and the input distribution P so that $I(PQ)/t_I \ge R$ and $(PQ)_Y \ne Q(\cdot|\star)$, which is always possible since $C(Q) > 0$. With this choice conditions (II), (IV), and (V) are satisfied for any $\Delta \in (0, 1)$ and any $\alpha > 0$ small enough. Picking such a small $\alpha > 0$, condition (I) is satisfied for a sufficiently small $\Delta \in (0, 1)$. Finally, since $\alpha(R, Q)$ is a nonincreasing function of R, it follows that $\alpha(0, Q) = \alpha(Q) > 0$ as well. □

Note that the corollary, together with the fact that the capacity region is nonincreasing, imply that a rate loss (relative to the capacity of the synchronized channel) is experienced only if the asynchronism level is at least exponential in the codeword length.

As our second main result, we characterize the synchronization threshold (for any discrete memoryless channel).

Theorem 2: For any discrete memoryless channel Q, the synchronization threshold of Definition 3 is given by

$$\alpha(Q) = \max_{x\in\mathcal{X}} D(Q(\cdot|x)\|Q(\cdot|\star)). \tag{20}$$

The proof, developed in Section VII-B, consists of two parts. The converse part establishes that no coding scheme can achieve an arbitrarily low error probability if the asynchronism level grows at least as fast as $e^{N\alpha}$ with $\alpha > \max_{x\in\mathcal{X}} D(Q(\cdot|x)\|Q(\cdot|\star))$. The direct part, which follows from the first part of Theorem 1, establishes the existence of a coding scheme with vanishing error probability as $N\to\infty$ when the asynchronism level grows no faster than $e^{N\alpha}$ with $\alpha < \max_{x\in\mathcal{X}} D(Q(\cdot|x)\|Q(\cdot|\star))$.

As a special case, note that if $Q(y|\star) = 0$ for some $y\in\mathcal{Y}$, then since $Q(y|x) > 0$ for some other $x\in\mathcal{X}$, it follows that the right-hand side of (20) is infinite, i.e., reliable communication is possible regardless of the rate at which the asynchronism level grows exponentially with the block length.

We also note that the capacity $C(Q)$ of the synchronized channel and the synchronization threshold $\alpha(Q)$ represent opposing extremal points on the capacity region for the asynchronous channel, and each characterizes a distinct limit on hypothesis discrimination at the output of the channel. The synchronous channel capacity characterizes the maximum number M of message sequences with respect to the block length N that can be discriminated at the output of the channel. By contrast, the synchronization threshold characterizes the largest value of A with respect to the block length N such that two sequences of length $A+N-1$, each constrained to use the symbol $\star$ except over an arbitrarily placed block of length N in the sequence, can be discriminated at the output of the channel.

Nevertheless, although $C(Q)$ and $\alpha(Q)$ do not appear to be related in any other more fundamental ways, it is noteworthy that one is bounded by the other. Indeed, with P denoting a capacity-achieving distribution of the (synchronous) channel Q, we have

$$\begin{aligned} C(Q) &\triangleq D(PQ\|P(PQ)_Y) \\ &\le D(PQ\|PQ(\cdot|\star)) \qquad (21)\\ &= \sum_x P(x)\, D(Q(\cdot|x)\|Q(\cdot|\star)) \\ &\le \max_x D(Q(\cdot|x)\|Q(\cdot|\star)) \\ &= \alpha(Q) \end{aligned}$$

where (21) follows from using the fact [14, Lemma 13.8.1]

$$D(PQ\|P(PQ)_Y) \le D(PQ\|PV), \quad \text{for all } V\in\mathcal{P}^{\mathcal{Y}}$$

with $V = Q(\cdot|\star)$. Moreover, it can be checked that if $C(Q) = 0$ then $\alpha(Q) = 0$.

As applications of Theorem 2, we have the following simple examples.

Example: Binary Symmetric Channel

Consider the binary symmetric channel of Fig. 4, where $\star = 0$, and the crossover probability is ϵ. The synchronous capacity of this channel is $C(Q) = \ln 2 - H_{\mathrm{B}}(\epsilon)$, while the synchronization threshold is

$$\alpha(Q) = D(Q(\cdot|1)\|Q(\cdot|\star)) = (1-2\epsilon)\ln\frac{1-\epsilon}{\epsilon},$$

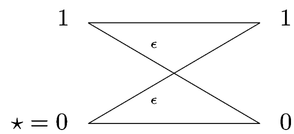

Fig. 4. A binary symmetric channel with crossover probability ϵ. The capacity of the synchronized channel is $C(Q) = \ln 2 - H_{\rm B}(\epsilon)$, and the zero-rate sphere-packing exponent is $E_{\rm sp}(Q) = -\ln(2\sqrt{\epsilon(1-\epsilon)})$. With $\star = 0$, the synchronization threshold is $\alpha(Q) = (1-2\epsilon)\ln((1-\epsilon)/\epsilon)$.

so $\alpha(Q) \gg C(Q)$ for sufficiently small ϵ. Moreover, the zero-rate sphere-packing exponent is $E_{\rm sp}(Q) = -\ln(2\sqrt{\epsilon(1-\epsilon)})$. Therefore

$$E_{\rm sp}(Q) \leq (1/2)\,\alpha(Q)(1+o(1)), \quad \text{as } \epsilon \to 0$$

so this channel is an example of one for which there is a discontinuity at zero rate in the asynchronism exponent achieved by generalized constant-composition codebooks and our sequential decoding rule.

For the special case $\epsilon = 0$, the channel is error free. In this case, the synchronous capacity is $C(Q) = \ln(2)$, while the synchronization threshold is

$$\alpha(Q) = D(Q(\cdot|1)\|Q(\cdot|0)) = \infty,$$

which expresses that reliable communication is possible for this channel no matter how large the asynchronism level, as we would expect. Indeed, a suitable codebook for this channel consists of $M = 2^{N-1}$ codewords, where the mth codeword consists of the prefix 1, followed by the $(N-1)$-bit binary representation of m. The rate of this code is $\bar{R} = (\ln 2)(N-1)/N$, so $\lim_{N\to\infty} \bar{R} = C(Q)$. Moreover, the decoder locates the start of the codeword transmission by finding the first 1 in the output stream, and decoding happens after collecting the next $N-1$ bits. Hence, the communication rate R and code rate $\bar{R}$ coincide asymptotically, i.e., $\lim_{N\to\infty} R = C(Q)$ as well. Evidently, this channel has the largest possible asynchronous capacity region: $\alpha(R,Q) = \infty$ for $R \in [0, C(Q))$.

Example: Gaussian Channel

Next, consider antipodal signaling over an additive white Gaussian noise channel with hard decision decoding. With this model, the channel output at any particular time is $Y = X + Z$, where X is the corresponding channel input, and where the noise Z is a Gaussian random variable, independent of X, with zero-mean and variance $1/\mathrm{SNR}$, where SNR represents the (peak) signal-to-noise ratio in the channel. Before decoding, the receiver makes a hard decision on each received symbol Y and declares $+1$ if $Y \geq 0$ and -1 if $Y < 0$.

The antipodal channel inputs are $X = +1$ and $X = -1$. In addition, $X = 0$ corresponds to there being no input to the channel, so this represents the silence symbol $\star$. When $X = 0$, the hard decision is $+1$ or -1 with equal probability. For each of the antipodal inputs, the corresponding hard decision has the opposite sign with probability ϵ where

$$\epsilon = e^{-(\mathrm{SNR}/2)(1+o(1))}, \quad \text{as SNR} \to \infty. \tag{22}$$

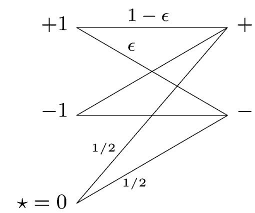

Fig. 5. A ternary-input, binary-output channel, with crossover probability parameter ϵ. Among other applications, this models antipodal signaling over a Gaussian channel with hard decisions at the decoder. The synchronous capacity and the synchronization threshold coincide for this channel; specifically, $\alpha(Q) = C(Q) = \ln 2 - H_{\rm B}(\epsilon)$.

The equivalent ternary-input, binary-output discrete memoryless channel Q is depicted in Fig. 5. For this channel, the synchronization threshold is

$$\begin{aligned}\alpha(Q) &= \max_x D(Q(\cdot|x)\|Q(\cdot|\star)) \\ &= \ln 2 - H_{\rm B}(\epsilon),\end{aligned} \tag{23}$$

which is the same as the capacity $C(Q)$ of the synchronized channel.

From Theorem 2, in order to achieve vanishing error probability it is necessary that $\alpha \leq \alpha(Q)$. Substituting for ϵ in (23) using (22), we then see that in order to achieve reliable communication it is necessary that

$$\alpha \leq \ln 2 - H_{\rm B}(e^{-(\mathrm{SNR}/2)(1+o(1))}), \quad \text{as SNR} \to \infty. \tag{24}$$

Via the right-hand side of (24) we have

$$\lim_{\mathrm{SNR}\to\infty} \alpha(Q) = \ln 2,$$

so at high-signal-to-noise ratio, increasing SNR results in a negligible increase in the level of asynchronism for which reliable communication is possible. This means that to exploit power in the high-signal-to-noise ratio regime it is necessary to employ a finer quantization (i.e., decisions that are "less hard") at the channel output.

Finally, note that the high-signal-to-noise ratio limit corresponds to the special case of Fig. 5 in which $\epsilon = 0$. This special case is an example of a channel for which $\lim_{R\to 0} \alpha(R,Q) = \alpha(Q)$. To see this, in Theorem 1, let us choose $t_D = 0, t_I = 1/(1-\Delta), P$ such that $P(+1) = p = 1 - P(-1)$ for some fixed $p \in (0, 1/2)$, and $\Delta \in (0,1)$ arbitrarily. Moreover, note that $C(Q) = \alpha(Q) = 1 - H_{\rm B}(0) = \ln 2$. Now we examine the conditions of the theorem. First, conditions (III) and (IV) are satisfied since with our choice of t_D the infimum is infinite. Second, condition (V) is satisfied because $D(PV\|PQ) = \infty$ unless $V = Q$, but Q is not in the set over which the minimization is taken since $\Delta < 1$. Third, condition (II) is satisfied because the left-hand side is zero, but the right-hand side is strictly positive since $I(PQ) = H(P) = H_{\rm B}(p) > 0$ and $(PQ)_Y = P \neq Q(\cdot|\star)$ so $D((PQ)_Y\|Q(\cdot|\star)) > 0$. Finally, we choose p sufficiently close to $1/2$ that $I(PQ) = \alpha(Q) + \Delta - 1$, so our rate is

$$R = (\alpha(Q) + \Delta - 1)(1 - \Delta). \tag{25a}$$

Then, via condition (I) we have that the achievable exponent is

$$\alpha = (\alpha(Q) + \Delta - 1)\Delta^3. \tag{25b}$$

Hence, from (25) we have $(R, \alpha) \to (0, \alpha(Q))$ as $\Delta \to 1$.

VII. Analysis

We first prove Theorem 1, then prove the converse and the direct parts of Theorem 2, using Theorem 1 for the latter.

Our proofs exploit large-deviations bounding techniques for finite-alphabet random variables, as described in, e.g., [15, Secs. 1.1 and 1.2], [14, Ch. 12]. In particular, we make extensive use of [14, Theorems 12.1.1 and 12.1.4] on the cardinality of N-types and on the probability of an N-type class.

The following identity, which follows immediately from the fact that $I(V) = D(V \| V_X V_Y)$ for $V \in \mathcal{P}^{\mathcal{X}\times\mathcal{Y}}$, will be useful in our analysis.

Fact 1: For any distributions $V \in \mathcal{P}^{\mathcal{X}\times\mathcal{Y}}, P_X \in \mathcal{P}^{\mathcal{X}}$, and $P_Y \in \mathcal{P}^{\mathcal{Y}}$

$$D(V\|P_X P_Y) = I(V) + D(V_X\|P_X) + D(V_Y\|P_Y). \tag{26}$$

In our proofs (of achievability), it will be convenient to define the following events. Let the message m start being emitted at time ν, and let $\mathcal{E}_{\text{ie}}$ be the event that an incorrect codeword is detected early; specifically

$$\mathcal{E}_{\text{ie}} \triangleq \{\min_{m'\neq m} \tau_{m'} < \nu + N - 1\} \tag{27}$$

with τ_m as defined in (6). This event happens when there is anomalous behavior in the channel during the time interval preceding transmission.

Similarly, let $\mathcal{E}_{\text{cl}}$ be the event that the correct codeword is detected late; specifically

$$\mathcal{E}_{\text{cl}} \triangleq \{\tau_m \geq \nu + N\}. \tag{28}$$

This event happens when there is anomalous behavior in the channel during the transmission interval.

Now since

$$\mathcal{E} \subset (\mathcal{E}_{\text{ie}} \cup \mathcal{E}_{\text{cl}}), \tag{29}$$

we have

$$\mathbb{P}_{m,\nu}(\mathcal{E}) \leq \mathbb{P}_{m,\nu}(\mathcal{E}_{\text{ie}}) + \mathbb{P}_{m,\nu}(\mathcal{E}_{\text{cl}}). \tag{30}$$

To upper-bound $\mathbb{P}_{m,\nu}(\mathcal{E}_{\text{ie}})$ and $\mathbb{P}_{m,\nu}(\mathcal{E}_{\text{cl}})$, we begin by using $\mathcal{E}(m,n,i)$ to denote the event that message m is declared at time n by observing the last i symbols, i.e.,

$$\mathcal{E}(m,n,i) = \bigcap_{k\in\{1,\ldots,i\}} \mathcal{E}(m,n,i,k) \tag{31}$$

where

$$\mathcal{E}(m,n,i,k) = \{D_{n,i}(m) \geq t_D \bar{R}\} \cap \{I_{n,i,k}(m) \geq t_I \bar{R}\} \tag{32}$$

with $D_{n,i}(m)$ and $I_{n,i,k}(m)$ as defined in (7) and (9).[7]

[7]We emphasize that the relevant observations Y_{n-i+1}^n and the codeword symbols $C^i(m)$ implicit in $D_{n,i}(m)$ and $I_{n,i,k}(m)$ in (32) are random variables.

With this notation, we then have, via the union bound

$$\mathbb{P}_{m,\nu}(\mathcal{E}_{\text{ie}}) \leq P_{\text{ie}}^{+} \triangleq \sum_{\substack{m'\in\{1,\ldots,M\}\setminus\{m\},\\ n\in\{1,\ldots,A+N-1\},\\ i\in\{1,\ldots,N\wedge n\}}} \mathbb{P}_{m,\nu}(\mathcal{E}(m',n,i)) \tag{33}$$

and

$$\mathbb{P}_{m,\nu}(\mathcal{E}_{\text{cl}}) \leq P_{\text{cl}}^{+} \triangleq \mathbb{P}_{m,\nu}\left(\mathcal{E}(m,\nu+N-1,N)^{\text{c}}\right). \tag{34}$$

Finally, we emphasize that we average over the relevant codebook ensemble in our analysis of both rate and error probability. By the usual random coding argument, when these averaged quantities meet their targets, some particular codebooks in the ensemble must also meet these targets.

A. Proof of Theorem 1

For our development, we use the basic coding scheme described in Section V.

Our proof is obtained by suitably bounding both the probability of error and, for the second part of the theorem, the average reaction delay in decoding. For this purpose, we require the following two lemmas. Proofs immediately follow the proof of the theorem.

Lemma 2: Let the codebook be random, of generalized constant composition with respect to P, and have code rate $\bar{R}$ satisfying (11), and let the decoding rule have (constant) thresholds $t_D, t_I \in \mathbb{R}$. Then for any (nonnegative) asynchronism exponent α, the bound P_{ie}^{+} as defined in (33) satisfies, as $N \to \infty$

$$P_{\text{ie}}^{+} \leq \text{poly}(N)\left[e^{-N(\bar{R}(t_D+t_I-1+o(1))-\alpha)} + e^{-N\bar{R}(t_I-1+o(1))}\right]. \tag{35}$$

Lemma 3: Let the codebook be random, of generalized constant composition with respect to P, and have code rate $\bar{R}$ satisfying (11), and let the decoding rule have (constant) thresholds satisfying $t_D \geq 0, t_I > 0$. Then for any fixed $\delta \in (0,1]$ we have, as $N \to \infty$

$$\begin{aligned}\mathbb{P}_{m,\nu}&(\tau_m - \nu \geq N\delta)\\ &\leq \mathbb{P}_{m,\nu}(\mathcal{E}(m,\nu+N\delta-1,N\delta)^{\text{c}})\\ &\leq \text{poly}(N)\left[e^{-N\delta(E_D(\delta)+o(1))} + e^{-N\delta(E_I(\delta)+o(1))}\right],\end{aligned} \tag{36}$$

where

$$E_D(\delta) \triangleq \inf_{\substack{\{V\in\mathcal{P}^{\mathcal{Y}|\mathcal{X}}:\\ D((PV)_Y\|Q(\cdot|\star))<t_D\bar{R}/\delta\}}} D((PV)_Y\|(PQ)_Y) \tag{37}$$

$$E_I(\delta) \triangleq \min_{\substack{\{V\in\mathcal{P}^{\mathcal{Y}|\mathcal{X}}:\\ I(PV)\leq t_I\bar{R}/\delta\}}} D(PV\|PQ), \tag{38}$$

and where we note that for the special case $\delta = 1$, (36) is an upper bound on P_{cl}^{+} as defined in (34).

For our proof, we begin by choosing the code rate to be

$$\bar{R} = \frac{I(PQ)\Delta}{t_I} \tag{39}$$

and analyzing the error probability of the scheme.

Probability of Error Analysis: We show that the probability of error vanishes. Bounding $\mathbb{P}_{m,\nu}(\mathcal{E})$ via (30) with (27) and (28), we obtain

$$\mathbb{P}_{m,\nu}(\mathcal{E}) \leq P_{\text{ie}}^{+} + \mathbb{P}_{m,\nu}(\tau_m - \nu \geq N), \tag{40}$$

where P_{ie}^{+} is as defined in (33).

From Lemma 3 with $\delta = 1$, it follows that the second term in (40) vanishes. Indeed, with $\bar{R}$ as in (39), we have that $E_D(1) = \alpha_D^{+}$, which is strictly positive according to condition (III). Moreover, again with $\bar{R}$ as in (39), $E_I(1) > 0$ whenever $I(PQ) > 0$ since $\Delta < 1$.

Thus, it remains only to show that the first term in (40), i.e., P_{ie}^{+}, also vanishes. From Lemma 2, the second term in (35) vanishes since $t_I > 1$. Substituting for $\bar{R}$ from (39) into (35), we see that condition (I) ensures that the first term also vanishes.

This establishes the first part of Theorem 1. To establish the second part we must analyze the associated communication rate.

Rate Analysis: We first bound the (conditional) normalized average reaction delay, which can be written in the form

$$\frac{\mathbb{E}_{m,\nu}(\tau - \nu)^{+}}{N} \leq \frac{\mathbb{E}_{m,\nu}(\tau_m - \nu)^{+}}{N} \triangleq \bar{\Delta}_{m,\nu} \tag{41}$$

$$= \bar{\Delta}_{m,\nu}^{(0)} + \bar{\Delta}_{m,\nu}^{(1)} + \bar{\Delta}_{m,\nu}^{(2)} \tag{42}$$

where to obtain (41) we have used that $\tau_m \leq \tau$, and where in (42)

$$\begin{aligned}\bar{\Delta}_{m,\nu}^{(0)} &= \frac{1}{N}\mathbb{E}_{m,\nu}[(\tau_m - \nu)^{+} \,|\, \tau_m - \nu < N\bar{\Delta}_*] \\ &\qquad \cdot \mathbb{P}_{m,\nu}(\tau_m - \nu < N\bar{\Delta}_*) \\ &\leq \bar{\Delta}_* \end{aligned} \tag{43}$$

$$\begin{aligned}\bar{\Delta}_{m,\nu}^{(1)} &= \frac{1}{N}\mathbb{E}_{m,\nu}[(\tau_m - \nu)^{+} \,|\, N\bar{\Delta}_* \leq \tau_m - \nu < N] \\ &\qquad \cdot \mathbb{P}_{m,\nu}(N\bar{\Delta}_* \leq \tau_m - \nu < N) \\ &\leq \mathbb{P}_{m,\nu}(\tau_m - \nu \geq N\bar{\Delta}_*) \end{aligned} \tag{44}$$

$$\begin{aligned}\bar{\Delta}_{m,\nu}^{(2)} &= \frac{1}{N}\mathbb{E}_{m,\nu}[(\tau_m - \nu)^{+} \,|\, \tau_m - \nu \geq N] \\ &\qquad \cdot \mathbb{P}_{m,\nu}(\tau_m - \nu \geq N) \\ &\leq (A + N)\,\mathbb{P}_{m,\nu}(\tau_m - \nu \geq N) \end{aligned} \tag{45}$$

where $\bar{\Delta}_* \in [0, 1]$ is arbitrary.

Under the conditions of the theorem, both $\bar{\Delta}_{m,\nu}^{(1)}$ and $\bar{\Delta}_{m,\nu}^{(2)}$ vanish, and thus $\bar{\Delta}_{m,\nu}^{(0)}$ determines the overall reaction delay (and hence communication rate).

Focusing first on the nonvanishing term $\bar{\Delta}_{m,\nu}^{(0)}$, we judiciously choose[8]

$$\bar{\Delta}_* = \Delta(1 + e^{-N\bar{R}})\, d(\eta) \tag{46}$$

where

$$d(\eta) \triangleq \frac{I(PQ)}{\displaystyle\min_{\substack{\{V \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}}:\\ D(PV\|PQ) \leq \eta\}}} I(PV)} \tag{47}$$

with

$$\eta = \eta(N) = \frac{1}{\sqrt{N\bar{R}}}. \tag{48}$$

[8]The term $e^{-N\bar{R}}$ in the definition of $\bar{\Delta}_*$ can be replaced by any positive strictly decreasing function of N.

As we verify at the end of the proof, $\bar{\Delta}_*$ behaves asymptotically as follows.

Fact 2: The quantity $\bar{\Delta}_*$ as defined in (46) satisfies

$$\bar{\Delta}_* = \Delta(1 + o(1)), \quad \text{as } N \to \infty. \tag{49}$$

From (49), it follows that, as required, $\bar{\Delta}_* \leq 1$ for N sufficiently large. Moreover, provided that both $\bar{\Delta}_{m,\nu}^{(1)}$ and $\bar{\Delta}_{m,\nu}^{(2)}$ vanish, then using (49) with (43) in (42), we obtain $\bar{\Delta}_{m,\nu} \leq \Delta(1 + o(1))$, and thus (4) satisfies

$$\bar{\Delta} \leq \Delta(1 + o(1)), \quad \text{as } N \to \infty \tag{50}$$

where we have exploited that the bound on $\bar{\Delta}_{m,\nu}$ is uniform in m and ν. Hence, using (39) and (50) in (2), we obtain that the communication rate satisfies

$$R \geq \frac{I(PQ)}{t_I}, \quad \text{as } N \to \infty. \tag{51}$$

We now verify that the two required terms in (42) vanish. Focusing on $\bar{\Delta}_{m,\nu}^{(1)}$, starting from (44), and using Lemma 3 with $\delta = \bar{\Delta}_*$ we have

$$\begin{aligned}&\mathbb{P}_{m,\nu}(\tau_m - \nu \geq N\bar{\Delta}_*) \\ &\quad \leq \text{poly}(N)\left[e^{-N\bar{\Delta}_*(E_D(\bar{\Delta}_*)+o(1))} + e^{-N\bar{\Delta}_*(E_I(\bar{\Delta}_*)+o(1))}\right].\end{aligned} \tag{52}$$

Hence, it suffices to show that $E_D(\bar{\Delta}_*)$ is strictly positive and $E_I(\bar{\Delta}_*)$ does not decay too quickly as $N \to \infty$.

For $E_I(\bar{\Delta}_*)$, rewriting (46) with (47) as

$$\frac{t_I\bar{R}}{\bar{\Delta}_*}(1 + e^{-N\bar{R}}) = \min_{\{V \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}}:\, D(PV\|PQ) \leq \eta\}} I(PV) \tag{53}$$

implies that

$$E_I(\bar{\Delta}_*) = \min_{\{V \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}}:\, I(PV) \leq t_I\bar{R}/\bar{\Delta}_*\}} D(PV\|PQ) \geq \eta \tag{54}$$

where we have used the following simple fact.

Fact 3: If

$$\min_{\{x:\, g(x) \leq g_*\}} f(x) = f_* + \epsilon$$

for some f_*, g_*, and $\epsilon > 0$, then

$$\min_{\{x:\, f(x) \leq f_*\}} g(x) \geq g_*.$$

Using (48), (49), and (54) in the second term in (52), we obtain

$$e^{-N\bar{\Delta}_*(E_I(\bar{\Delta}_*)+o(1))} = e^{-\Theta(\sqrt{N})}. \tag{55}$$

Turning now to $E_D(\bar{\Delta}_*)$, using condition (II), together with (39) and (49), we have

$$\frac{t_D\bar{R}}{\bar{\Delta}_*} < D((PQ)_Y\|Q(\cdot|\star))(1 + o(1))$$

so for some $\mu > 0$

$$\frac{t_D \bar{R}}{\bar{\Delta}_*} < (1-\mu) D((PQ)_Y \| Q(\cdot|\star))(1+o(1))$$

whence

$$E_D(\bar{\Delta}_*) \geq \inf_{\substack{\{V \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}}: \\ D((PV)_Y \| Q(\cdot|\star)) < (1-\mu) D((PQ)_Y \| Q(\cdot|\star))\}}} D((PV)_Y \| (PQ)_Y) > 0. \tag{56}$$

Using (56) and (49) in the first term in (52), we obtain

$$e^{-N\bar{\Delta}_*(E_D(\bar{\Delta}_*)+o(1))} \leq e^{-\Theta(N)}. \tag{57}$$

Substituting (55) and (57) into (52), we see that $\mathbb{P}_{m,\nu}(\tau_m - \nu \geq N\bar{\Delta}_*)$—and hence $\bar{\Delta}^{(1)}_{m,\nu}$—vanishes as $N \to \infty$ as claimed.

Focusing now on $\bar{\Delta}^{(2)}_{m,\nu}$, starting from (45) and using Lemma 3 with $\delta = 1$, we obtain

$$\bar{\Delta}^{(2)}_{m,\nu} \leq \text{poly}(N) \left[e^{-N(E_D(1)-\alpha+o(1))} + e^{-N(E_I(1)-\alpha+o(1))} \right]. \tag{58}$$

But with $\bar{R}$ chosen according to (39), it follows from (IV) and (V) that $E_D(1) = \alpha_D^+$, as before, and $E_I(1) = \alpha_I^+$, so

$$\alpha < E_D(1) = \inf_{\substack{\{V \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}}: \\ D((PV)_Y \| Q(\cdot|\star)) < t_D \bar{R}\}}} D((PV)_Y \| (PQ)_Y)$$

$$\alpha < E_I(1) = \min_{\substack{\{V \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}}: \\ I(PV) \leq t_I \bar{R}\}}} D(PV \| PQ)$$

from which we conclude that $\bar{\Delta}^{(2)}_{m,\nu}$ vanishes as $N \to \infty$, as required.

To conclude the proof we verify (49), and establish Lemmas 2 and 3.

Proof of Fact 2: It suffices to show that $d(\eta)$ as defined in (47) satisfies $d(\eta) = 1 + o(1)$ as $\eta \to 0$. To establish this result, we exploit that a monotonic bounded sequence converges [19, Theorem 3.14]. To apply this result, we first observe that since $I(PV)$ is a continuous function over the compact set

$$\{V \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}} : D(PV \| PQ) \leq \eta\} \tag{59}$$

in which $D(PV\|PQ)$ is a continuous function of V that is zero if and only if $V = Q$, the minimum in the denominator of (47) is a well-defined and continuous function, and thus so is $d(\eta)$ except where it is infinite. Furthermore, we note that $d(\eta)$ is a monotonically decreasing function as $\eta \to 0$, and that $d(\eta) \geq d(0) = 1$.

It remains only to verify that for all $\eta > 0$ sufficiently small, $d(\eta) < \infty$, which is equivalent to showing that for $\eta > 0$ small enough, the set in (59) contains no trivial $V \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}}$ such that $I(PV) = 0$, i.e., such that $V(\cdot|x)$ is the same for all $x \in \mathcal{X}$.

Consider the set $\mathcal{P}^\pi$ of product measures in $\mathcal{P}^{\mathcal{X}\times\mathcal{Y}}$ whose left marginal is P, i.e., $W \in \mathcal{P}^\pi \subset \mathcal{P}^{\mathcal{X}\times\mathcal{Y}}$ if $W = PW_Y$. Since $\mathcal{P}^\pi$ is compact and $D(\cdot\|PQ)$ is continuous over $\mathcal{P}^\pi$

$$\min_{W \in \mathcal{P}^\pi} D(W\|PQ)$$

is well defined. To complete the proof, we must show that this minimum is positive. If the minimum is 0, then $W = PQ$ must be contained in $\mathcal{P}^\pi$, because $D(J_1\|J_2) = 0$ if and only if $J_1 = J_2$. But $PQ \in \mathcal{P}^\pi$ implies that $I(PQ) = 0$, which is a contradiction. Hence, the set (59) contains no trivial conditional probability. Therefore, for η small enough, the denominator in the definition (47) is strictly positive, implying that $d(\eta)$ is finite. □

To prove Lemmas 2 and 3 we make use of the following additional two small lemmas.

Lemma 4: Let $C^N \in \mathcal{X}^N$, and let $\mathbb{P}(\cdot)$ denote probability with respect to the measure by which C^N is generated in an i.i.d. manner according to distribution $P \in \mathcal{P}^{\mathcal{X}}$. Let $\mathbb{P}_{\mathcal{K}_N(P)}(\cdot)$ denote probability when C^N is instead drawn from the random generalized constant-composition ensemble with respect to the distribution P. Then for any events $\mathcal{A}_N$, $N = 1, 2, \ldots$, such that

$$\lim_{N\to\infty} \mathbb{P}(\mathcal{A}_N) < 1 - \delta \tag{60}$$

for some constant $\delta > 0$, we have

$$\mathbb{P}_{\mathcal{K}_N(P)}(\mathcal{A}_N) \leq \mathbb{P}(\mathcal{A}_N)^{1+o(1)}, \quad \text{as } N \to \infty. \tag{61}$$

Proof: First, note that

$$\mathbb{P}(\mathcal{A}_N) \geq \mathbb{P}(\mathcal{A}_N \,|\, C^N \in \mathcal{K}_N(P))\, \mathbb{P}(C^N \in \mathcal{K}_N(P))$$

where we recognize that the first term on the right-hand side is $\mathbb{P}_{\mathcal{K}_N(P)}(\mathcal{A}_N)$, and note that the second term is $1 + o(1)$ via Lemma 1. Finally, since $\mathbb{P}(\mathcal{A}_N)(1+o(1))$ is also $\mathbb{P}(\mathcal{A}_N)^{1+o(1)}$ as $N \to \infty$ when (60) holds, we obtain (61). □

When two distributions V_1 and V_2 satisfy $\|V_1 - V_2\| \leq 1/\ln N$, we express this relation via the notation $V_1 \overset{N}{\approx} V_2$. In such cases we have the following.

Lemma 5: Let $J \in \mathcal{P}^{\mathcal{X}\times\mathcal{Y}}$, $V \in \mathcal{P}^{\mathcal{X}}$, and $W_N \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}}$ for $N = 1, 2, \ldots$ be arbitrary. Moreover, let $V_N \in \mathcal{P}^{\mathcal{X}}$ be such that $V_N \overset{N}{\approx} V$ for $N = 1, 2, \ldots$. Then, as $N \to \infty$

$$D(V_N W_N \| J) \geq D(V W_N \| J)(1 + o(1)) \tag{62a}$$

$$I(V_N W_N) \geq I(V W_N)(1 + o(1)) \tag{62b}$$

and

$$D((V_N W_N)_Y \| J_Y) \geq D((V W_N)_Y \| J_Y)(1 + o(1)). \tag{62c}$$

Proof: First consider (62a). When $D(VW_N\|J)$ is zero, the inequality holds trivially. When $D(VW_N\|J)$ is infinite, the inequality also holds trivially since in that case $D(V_N W_N\|J)$

is also infinite for N large enough. When $D(VW_N\|J)$ is finite, we can write

$$\begin{aligned} D(V_N W_N\|J) - D(VW_N\|J) &= D(V_N\|V) + \sum_{x,y}(V_N(x)W(y|x) - V(x)W(y|x)) \\ &\quad \times \log\left(\frac{V(x)W(y|x)}{J(x,y)}\right) \end{aligned} \tag{63}$$

where the sum extends over $(x,y) \in \mathcal{X}\times\mathcal{Y}$ such that $J(x,y) \neq 0$. Thus, for any fixed J, as $N \to \infty$, both terms in (63) go to 0, which proves (62a). Expressions (62b) and (62c) follow via similar reasoning, so we omit the details. □

Proof of Lemma 2: We separately consider two cases, corresponding to whether $\mathcal{E}(m',n,i)$ occurs outside the transmission interval (Case I), or partly inside and partly outside the transmission interval (Case II). Recall from (16) that $i > N/\ln N$.

Case I: $n < \nu$ or $n - i + 1 \geq \nu + N$: We have, as $N \to \infty$

$$\begin{aligned} &\mathbb{P}_{m,\nu}(\mathcal{E}(m',n,i)) \\ &\leq \mathbb{P}_{m,\nu}(\mathcal{E}(m',n,i,i)) \\ &\leq \sum_{\substack{\{V\in\mathcal{P}_i^{\mathcal{X}\times\mathcal{Y}}: \\ (i/N)I(V)\geq t_I\bar{R}, \\ (i/N)D(V_Y\|Q(\cdot|\star))\geq t_D\bar{R}\}}} e^{-iD(V\|PQ(\cdot|\star))(1+o(1))} &(64)\\ &\leq \sum_{\substack{\{V\in\mathcal{P}_i^{\mathcal{X}\times\mathcal{Y}}: \\ (i/N)I(V)\geq t_I\bar{R}, \\ (i/N)D(V_Y\|Q(\cdot|\star))\geq t_D\bar{R}\}}} e^{-i(I(V)+D(V_Y\|Q(\cdot|\star))(1+o(1))} &(65)\\ &\leq (i+1)^{|\mathcal{X}||\mathcal{Y}|}e^{-N\bar{R}(t_I+t_D+o(1))} &(66)\\ &\leq \mathrm{poly}(N)\, e^{-N\bar{R}(t_I+t_D+o(1))} &(67)\end{aligned}$$

where to obtain (64) we have used (61) of Lemma 4 in addition to [14, Theorem 12.1.4], where to obtain (65) we have used the identity (26), where to obtain (66) we have used the identity $|\mathcal{P}_n^{\mathcal{X}\times\mathcal{Y}}| \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|}$ [14, Theorem 12.1.1], and where to obtain (67) we have used that $i \leq N$.

Case II: $n \geq \nu$ and $n - i + 1 \leq \nu + N - 1$: The event $\mathcal{E}(m',n,i)$ involves the observations $Y_{n-i+1}, Y_{n-i+2}, \ldots, Y_n$, the first k being distributed according to the noise distribution, and the remaining $i-k$ according to the distribution induced by the sent codeword.[9] In order to deal with the discrepancy that results because codeword prefixes shorter than $N/\ln N$ do not satisfy the sequential constant-composition property, we distinguish two subcases.

Case II-A: $k \geq N/\ln N$ and $i - k \geq N/\ln N$: We have

$$\begin{aligned} &\mathbb{P}_{m,\nu}\left(\mathcal{E}(m',n,i)\right) \\ &\leq \mathbb{P}_{m,\nu}\left(\mathcal{E}(m',n,i,k)\right) &(68)\\ &\leq \sum\sum_{\substack{\{V\in\mathcal{P}_k^{\mathcal{X}\times\mathcal{Y}},W\in\mathcal{P}_{i-k}^{\mathcal{X}\times\mathcal{Y}}: \\ (k/N)I(V)+(i-k)/NI(W)\geq t_I\bar{R}\}}} \Big(e^{-kD(V\|PQ(\cdot|\star))(1+o(1))} \\ &\qquad \cdot e^{-(i-k)D(W\|P(PQ)_Y)(1+o(1))}\Big) &(69)\\ &\leq \sum\sum_{\substack{\{V\in\mathcal{P}_k^{\mathcal{X}\times\mathcal{Y}},W\in\mathcal{P}_{i-k}^{\mathcal{X}\times\mathcal{Y}}: \\ (k/N)I(V)+(i-k)/NI(W)\geq t_I\bar{R}\}}} e^{-(kI(V)+(i-k)I(W))(1+o(1))} &(70)\\ &\leq \mathrm{poly}(N)\, e^{-N\bar{R}(t_I+o(1))} &(71)\end{aligned}$$

where to obtain (68) we have used that $\mathcal{E}(m',n,i) \subset \mathcal{E}(m',n,i,k)$ for any $k \in \{1,\ldots,i\}$ due to (31), where to obtain (69), we have used, via (32), $E(m,n,i,k) \subset \{I_{n,i,k}(m) \geq t_I\bar{R}\}$ in addition to [14, Theorem 12.1.4] and (61) of Lemma 4, where to obtain (70) we have used the identity (26), and where to obtain (71) we have used both [14, Theorem 12.1.1] and that $i \leq N$.

Case II-B: $k < N/\ln N$ but $i - k \geq N/\ln N$: We consider only the case $k < N/\ln N$; by symmetry, the case $i - k < N/\ln N$ yields the identical result. We have, as $N \to \infty$

$$\begin{aligned} &\mathbb{P}_{m,\nu}(\mathcal{E}(m',n,i)) \\ &\leq \mathbb{P}_{m,\nu}\left(\mathcal{E}(m',n,i,k)\right) \\ &\leq \sum_{\substack{\{W\in\mathcal{P}_{i-k}^{\mathcal{X}\times\mathcal{Y}}: \\ \ln|\mathcal{X}|/\ln N+(i-k)/NI(W)\geq t_I\bar{R}\}}} e^{-(i-k)(D(W\|P(PQ)_Y)(1+o(1))} &(72)\\ &\leq \sum_{\substack{\{W\in\mathcal{P}_{i-k}^{\mathcal{X}\times\mathcal{Y}}: \\ \ln|\mathcal{X}|/\ln N+(i-k)/NI(W)\geq t_I\bar{R}\}}} e^{-(i-k)I(W)(1+o(1))} &(73)\\ &\leq \mathrm{poly}(N)\, e^{-N\bar{R}(t_I+o(1))} &(74)\end{aligned}$$

where to obtain (72) we have used, in addition to [14, Theorem 12.1.4] and (61) of Lemma 4, that $(k/N)D(V\|P(PQ)_Y) \geq 0$ and $I(V) \leq \ln|\mathcal{X}|$ for any $V \in \mathcal{P}_k^{\mathcal{X}\times\mathcal{Y}}$, where to obtain (73) we have used the identity (26), and where to obtain (74) we have used both [14, Theorem 12.1.1] and that $i \leq N$.

Case II-C: $k \geq N/\ln N$ but $i-k < N/\ln N$: By symmetry, we have, from Case II-B

$$\mathbb{P}_{m,\nu}(\mathcal{E}(m',n,i)) \leq \mathrm{poly}(N)\, e^{-N\bar{R}(t_I+o(1))}. \tag{75}$$

We note that (74), and thus (75), is identical to (71), and thus Cases II-A, II-B, and II-C all yield the same bound.

Finally, a bound on P_{ie}^{+} is obtained by summing the upper bounds in (67) and (71), recognizing that in the defining summation (33) there are less than $M(A+N)N$ of the former terms, and less than $2MN^2$ of the latter ones, yielding (35) as desired. □

Proof of Lemma 3: For now, we restrict our attention to the case in which $N\delta$ is an integer. We remove this restriction at the end of the proof.

Applying the union bound to the complement of (32) for the particular case of interest yields

$$\begin{aligned} &\mathbb{P}_{m,\nu}(\tau_m - \nu \geq N\delta) \\ &\leq \mathbb{P}_{m,\nu}(\mathcal{E}(m,\nu+N\delta-1,N\delta)^{\mathrm{c}}) \\ &\leq P_D(\delta) + P_I^{\{1,\ldots,N\delta\}}(\delta) \end{aligned} \tag{76}$$

[9] Actually, this case also captures the uninteresting scenario when the first portion of the i symbols is a codeword, and the second portion is noise. We could have avoided this by more accurately defining the range of n in (33) to be $\{1,\ldots,\nu+N-1\}$, though our results are unaffected.

where

$$P_D(\delta) = \mathbb{P}_{m,\nu}\left(D\left(\hat{P}_{Y_\nu^{\nu+N\delta-1}}\|Q(\cdot|\star)\right) < \frac{t_D\bar{R}}{\delta}\right) \quad (77)$$

and where

$$P_I^{\mathcal{I}}(\delta) = \sum_{k\in\mathcal{I}} \mathbb{P}_{m,\nu}\left(\frac{k}{N\delta} I\left(\hat{P}_{C^k(m),Y_\nu^{\nu+k-1}}\right) + \left(1-\frac{k}{N\delta}\right) I\left(\hat{P}_{C_{k+1}^{N\delta}(m),Y_{\nu+k}^{\nu+N\delta-1}}\right) \le \frac{t_I\bar{R}}{\delta}\right) \quad (78)$$

for $\mathcal{I} \subset \{1,2,\ldots,N\delta\}$.

We upper-bound (77) via

$$P_D(\delta) \le \sum_{\substack{\{V\in\mathcal{P}_{N\delta}^{\mathcal{X}\times\mathcal{Y}}: V_X \overset{N}{\approx} P,\\ D(V_Y\|Q(\cdot|\star))<t_D\bar{R}/\delta\}}} e^{-N\delta D(V_Y\|(PQ)_Y)(1+o(1))} \quad (79)$$

$$\le \sum_{\substack{\{W\in\mathcal{P}_{N\delta}^{\mathcal{Y}|\mathcal{X}}:\\ D((PW)_Y\|Q(\cdot|\star))\\ <t_D\bar{R}/\delta(1+o(1))\}}} e^{-N\delta D((PW)_Y\|(PQ)_Y)(1+o(1))} \quad (80)$$

$$\le \text{poly}(N)\, e^{-N\delta E_D(\delta(1+o(1)))(1+o(1))} \quad (81)$$

$$= \text{poly}(N)\, e^{-N\delta(E_D(\delta)+o(1))} \quad (82)$$

where to obtain (79) we have used [14, Theorem 12.1.4] and (61) of Lemma 4, where to obtain (80) we have used (62c) of Lemma 5, where to obtain (81) we have used [14, Theorem 12.1.1], and where to obtain (82) we have used the continuity of $E_D(\cdot)$.

We next upper-bound $P_I^{\{1,\ldots,N\delta\}}(\delta)$ as defined in (78), breaking the summation into three parts, to deal with effects of codeword prefixes that are shorter than $N/\ln N$. In particular, we write

$$P_I^{\{1,\ldots,N\delta\}} = P_I^{\mathcal{I}_1} + P_I^{\mathcal{I}_{2a}} + P_I^{\mathcal{I}_{2b}} \quad (83)$$

with

$$\mathcal{I}_1 = \{k\in\{1,\ldots,N\delta\}: k \ge N/\ln N, N\delta-k \ge N/\ln N\}$$
$$\mathcal{I}_{2a} = \{k\in\{1,\ldots,N\delta\}: k < N/\ln N, N\delta-k \ge N/\ln N\}$$
$$\mathcal{I}_{2b} = \{k\in\{1,\ldots,N\delta\}: k \ge N/\ln N, N\delta-k < N/\ln N\}.$$

Case I: $k \ge N/\ln N$ and $N\bar{\Delta} - K \ge N/\ln N$: We have, as $N\to\infty$

$$P_I^{\mathcal{I}_1} \le \sum_{k=1}^{N\delta} \sum\sum_{V,W\in\mathcal{S}_{N\delta,k}^{\mathcal{X}\times\mathcal{Y}}(\delta)} e^{-(kD(V\|PQ)+(N\delta-k)D(W\|PQ))(1+o(1))} \quad (84)$$

$$\le N\delta \max_{\lambda\in[0,1]} \sum\sum_{V,W\in\mathcal{S}_{N\delta,\lambda(1+o(1))}^{\mathcal{X}\times\mathcal{Y}}(\delta)} \left(e^{-N\delta\lambda D(V\|PQ)(1+o(1))} \cdot e^{N\delta(1-\lambda)D(W\|PQ)(1+o(1))}\right) \quad (85)$$

$$\le N\delta \max_{\lambda\in[0,1]} \sum\sum_{\substack{V,W\in\\ \mathcal{S}_{N\delta,\lambda(1+o(1))}^{\mathcal{Y}|\mathcal{X}}(\delta(1+o(1)))}} \left(e^{-N\delta\lambda D(V\|PQ)(1+o(1))} \cdot e^{N\delta(1-\lambda)D(W\|PQ)(1+o(1))}\right) \quad (86)$$

$$\le \text{poly}(N)\, e^{-N\delta E_*(\delta(1+o(1)))(1+o(1))} \quad (87)$$

$$= \text{poly}(N)\, e^{-N\delta E_I(\delta(1+o(1)))(1+o(1))} \quad (88)$$

$$= \text{poly}(N)\, e^{-N\delta(E_I(\delta)+o(1))} \quad (89)$$

where

$$\mathcal{S}_{N\delta,k}^{\mathcal{X}\times\mathcal{Y}}(\delta) = \left\{V\in\mathcal{P}_k^{\mathcal{X}\times\mathcal{Y}}, W\in\mathcal{P}_{N\delta-k}^{\mathcal{X}\times\mathcal{Y}}: V_X \overset{N}{\approx} P,\, W_X \overset{N}{\approx} P, \frac{k}{N\delta}I(V) + \left(1-\frac{k}{N\delta}\right)I(W) \le \frac{t_I\bar{R}}{\delta}\right\},$$

$$\mathcal{S}_{N\delta,k}^{\mathcal{Y}|\mathcal{X}}(\delta) = \left\{V\in\mathcal{P}_k^{\mathcal{Y}|\mathcal{X}}, W\in\mathcal{P}_{N\delta-k}^{\mathcal{Y}|\mathcal{X}}: \frac{k}{N\delta}I(PV) + \left(1-\frac{k}{N\delta}\right)I(PW) \le \frac{t_I\bar{R}}{\delta}\right\}$$

and

$$E_*(\delta) = \min_{\lambda\in[0,1]} \min_{\substack{\{V,W\in\mathcal{P}^{\mathcal{Y}|\mathcal{X}}:\\ \lambda I(PV)+(1-\lambda)I(PW)\le t_I\bar{R}/\delta\}}} [\lambda D(PV\|PQ) + (1-\lambda)D(PW\|PQ)].$$

To obtain (84) we have used [14, Theorem 12.1.4] and (61) of Lemma 4, to obtain (85) we have used that there is a constant $\lambda\in[0,1]$ such that

$$k = \lambda N\delta(1+o(1)),$$

to obtain (86) we have used (62a) and (62b) of Lemma 5, to obtain (88) we have used Lemma 7 in Appendix II to show that $E_*(\delta) = E_I(\delta)$, and to obtain (89) we have used the continuity of $E_I(\cdot)$.

Case II-A: $k < N/\ln N$ but $N\delta - k \ge N/\ln N$: We have, as $N\to\infty$

$$P_I^{\mathcal{I}_{2a}} \le \sum_{k=1}^{N/\ln N} \sum\sum_{\substack{\{V\in\mathcal{P}_k^{\mathcal{X}\times\mathcal{Y}}, W\in\mathcal{P}_{N\delta}^{\mathcal{X}\times\mathcal{Y}}: W_X \overset{N^{1/(|\mathcal{X}|+1)}}{\approx} P,\\ (k/(N\delta))I(V)\\ +(1-k/(N\delta))I(W)\le t_I\bar{R}/\delta\}}} \left(e^{-kD(V\|PQ)(1+o(1))} \cdot e^{-(N\delta-k)D(W\|PQ)(1+o(1))}\right) \quad (90)$$

$$\le \sum_{k=1}^{N/\ln N} \sum\sum_{\substack{\{W\in\mathcal{P}_{N\delta}^{\mathcal{X}\times\mathcal{Y}}: W_X \overset{N^{1/(|\mathcal{X}|+1)}}{\approx} P,\\ (1-k/(N\delta))I(W)\le t_I\bar{R}/\delta\}}} e^{-(N\delta-k)D(W\|PQ)(1+o(1))} \quad (91)$$

$$\le \sum_{k=1}^{N/\ln N} \sum\sum_{\substack{\{W\in\mathcal{P}_{N\delta}^{\mathcal{X}\times\mathcal{Y}}: W_X \overset{N^{1/(|\mathcal{X}|+1)}}{\approx} P,\\ I(W)\le t_I\bar{R}/\delta(1+o(1))\}}} e^{-N\delta D(W\|PQ)(1+o(1))} \quad (92)$$

$$\le \sum_{k=1}^{N/\ln N} \sum\sum_{\substack{\{V\in\mathcal{P}_{N\delta}^{\mathcal{Y}|\mathcal{X}}:\\ I(PV)\le t_I\bar{R}/\delta(1+o(1))\}}} e^{-N\delta D(PV\|PQ)(1+o(1))} \quad (93)$$

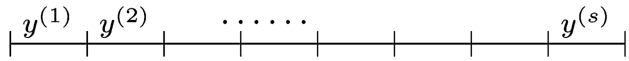

Fig. 6. Parsing of the received sequence of maximal length $A+N-1$ into s blocks $y^{(1)}, y^{(2)}, \ldots, y^{(s)}$ of length N, where $s = \lfloor (A+N-1)/N \rfloor$.

$$\leq \text{poly}(N)\, e^{-N\delta E_I(\delta(1+o(1)))(1+o(1))} \tag{94}$$
$$= \text{poly}(N)\, e^{-N\delta(E_I(\delta)+o(1))} \tag{95}$$

where to obtain (90) we have used [14, Theorem 12.1.4] and (61) of Lemma 4,[10] where to obtain (91) we have used that $I(V) \geq 0$ and $D(V\|PQ) \geq 0$, where to obtain (92) we have used that $1 - k/(N\delta) = 1 + o(1)$ since $k < N/\ln N$, where to obtain (93) we have used (62a) and (62b) of Lemma 5, and where to obtain (95) we have used the continuity of $E_I(\cdot)$.

Case II-B: $k \geq N/\ln N$ But $N\delta - k < N/\ln N$: By symmetry, we have, from Case II-A,

$$P_I^{\mathcal{I}_{2b}} \leq \text{poly}(N)\, e^{-N\delta(E_I(\delta)+o(1))}. \tag{96}$$

Finally, summing (88), (95), and (96) in accordance with (83), we obtain (36), as desired, for the case in which $N\delta$ is an integer. When $N\delta$ is not an integer, note that

$$\mathbb{P}_{m,\nu}(\tau_m - \nu \geq N\delta) \leq \mathbb{P}_{m,\nu}(\tau_m - \nu \geq N\delta')$$

where $N\delta' = \lfloor N\delta \rfloor$. Now $\delta' = \delta(1+o(1))$ as $N \to \infty$, so $\delta' E(\delta')(1+o(1)) = \delta(E(\delta)+o(1))$ whenever $E(\cdot)$ is continuous, which is the case for both $E = E_D$ and $E = E_I$, so our results do not change. □

B. Proof of Theorem 2

Proposition 1 (Converse): Given a channel Q, no coding scheme can achieve an asynchronism exponent α strictly greater than $\alpha(Q)$ as defined in (20).

Proof of Proposition 1: It suffices to restrict our attention to the case in which $Q(y|\star) > 0$ for all $y \in \mathcal{Y}$; otherwise, the proposition holds trivially since $\alpha(Q)$ is infinite.

Our approach is to lower-bound the error probability associated with the (optimistic) scenario in which

1) there are only $M = 2$ possible messages, so $m \in \{1, 2\}$;
2) as depicted in Fig. 6, the chosen message is transmitted in one of s distinct time slots of duration N, where

$$s = \left\lfloor \frac{A+N-1}{N} \right\rfloor = e^{\alpha N(1+o(1))}, \quad \text{as } N \to \infty; \tag{97}$$

3) the receiver is cognizant of these s possible time slots; and
4) the decoder makes a decision at the end of the uncertainty window, i.e., at time $A+N-1$.

We show that when

$$\alpha = \alpha(Q) + \epsilon \tag{98}$$

for any $\epsilon > 0$, the probability of error of the best communication systems for this scenario is asymptotically bounded away from zero.

We let $y^{(l)}, l = 1, \ldots, s$, denote the lth received block of size N, and we use $Q^N(\cdot|x^N)$ to denote the distribution of a received block when the input is x^N.

The received data y^{A+N-1} is distributed according to

$$\frac{1}{s}\sum_{l=1}^{s} Q^N\left(y^{(l)}|c^N(m)\right) \prod_{j\in\{1,\ldots,s\}\setminus\{l\}} Q^N(y^{(j)}|\star^N)$$
$$= \left[\prod_{j=1}^{s} Q^N\left(y^{(j)}|\star^N\right)\right] \frac{1}{s}\sum_{l=1}^{s} \frac{Q^N(y^{(l)}|c^N(m))}{Q^N(y^{(l)}|\star^N)} \tag{99}$$

when message m is sent, where $\star^N$ is the sequence of N consecutive $\star$ symbols. Using (99), it is straightforward to verify that a maximum-likelihood decoder, which minimizes the probability of a decoding error, declares message $m = 1$ or $m = 2$ depending on whether the sufficient statistic

$$\sum_{l=1}^{s} z\left(y^{(l)}\right) \tag{100}$$

is positive or negative, respectively, where

$$z\left(y^N\right) \triangleq \frac{Q^N(y^N|c^N(1))}{Q^N(y^N|\star^N)} - \frac{Q^N(y^N|c^N(2))}{Q^N(y^N|\star^N)} \tag{101}$$

and where if (100) is zero the decoder declares one of the two messages at random.

As a result, we have

$$\mathbb{P}(\mathcal{E}) \geq \frac{1}{2}\left[\mathbb{P}_1\left(\sum_{l=1}^{s} z\left(Y^{(l)}\right) < 0\right)\right.$$
$$\left. + \mathbb{P}_2\left(\sum_{l=1}^{s} z\left(Y^{(l)}\right) > 0\right)\right] \tag{102}$$

where $\mathbb{P}_m$ denotes probability conditioned on message m being sent.

Let l_* denote the block during which the selected message is transmitted. Then under $\mathbb{P}_m$, and given l_*, the $Y^{(l)}$ for $l \in \{1, \ldots, s\} \setminus \{l_*\}$ are all i.i.d. according to $Q^N(\cdot|\star^N)$, and $Y^{(l_*)}$ has distribution $Q^N(\cdot|c^N(m))$.

To simplify the exposition, we let the codeword $c^N(m)$ consist of a symbol $c(m)$ repeated N times, with $c(1) \neq c(2)$. The generalization to the case where the codewords each comprise multiple symbols is obtained by a simple (if notationally more cumbersome) extension. With our simplification, $z(y^N)$ depends only on the type $\hat{P}_{y^N}$ of its argument y^N, so at the expense of a slight abuse of notation we equivalently write $z(y^N) = z(\hat{P}_{y^N})$, i.e.,

$$z\left(\hat{P}_{y^N}\right)$$
$$= \prod_{b\in\mathcal{Y}} \left(\frac{Q(b|c(1))}{Q(b|\star)}\right)^{N\hat{P}_{y^N}(b)} - \prod_{b\in\mathcal{Y}} \left(\frac{Q(b|c(2))}{Q(b|\star)}\right)^{N\hat{P}_{y^N}(b)}$$
$$= \exp\left[N\sum_{b\in\mathcal{Y}} \hat{P}_{y^N}(b) \ln\frac{Q(b|c(1))}{Q(b|\star)}\right]$$
$$- \exp\left[N\sum_{b\in\mathcal{Y}} \hat{P}_{y^N}(b) \ln\frac{Q(b|c(2))}{Q(b|\star)}\right]. \tag{103}$$

[10]Note that the empirical distribution of a codeword of length N and the empirical distribution of its suffix of size $(1 - 1/\ln N)N$ have an L_1 distance of at most $(1/\ln N) * |\mathcal{X}|$. Hence, via the triangle inequality, if the empirical distribution of the codeword is $1/\ln N$ close to P, its suffix is $((1+|\mathcal{X}|)/\ln N$ close to P.

Let $\mathcal{T}_m$ be the set of sequences y^N that are strongly typical with respect to $Q^N(\cdot|c^N(m))$ for some implicit parameter $\mu \in (0,1)$, i.e.,

$$\mathcal{T}_m = \{y^N \in \mathcal{Y}^N : |\hat{P}_{y^N}(b) - Q(b|c(m))| < \mu, \quad \forall b \in \mathcal{Y}\}. \tag{104}$$

Accordingly, for N sufficiently large, we have

$$\begin{aligned}
&\mathbb{P}_1\left(\sum_{l=1}^{s} z\left(Y^{(l)}\right) < 0\right)\\
&\geq \mathbb{P}_1\left(\left\{z\left(Y^{(l_*)}\right) < h\right\} \bigcap \left\{\sum_{l\neq l_*} z\left(Y^{(l)}\right) < -h\right\}\right) &(105)\\
&= \mathbb{P}_1\left(z\left(Y^{(l_*)}\right) < h\right) \mathbb{P}_1\left(\sum_{l\neq l_*} z\left(Y^{(l)}\right) < -h\right) &(106)\\
&\geq \mathbb{P}_1\left(Y^{(l_*)} \in \mathcal{T}_1\right) \mathbb{P}_1\left(\sum_{l\neq l_*} z\left(Y^{(l)}\right) < -h\right) &(107)\\
&\geq (1-\mu)\, \mathbb{P}\left(\sum_{l\neq l_*} z\left(Y^{(l)}\right) < -h\right) &(108)\\
&= (1-\mu)\, \mathbb{P}\left(\sum_{l=1}^{s-1} Z_l < -h\right) &(109)
\end{aligned}$$

where (105) holds for any constant h, where to obtain (106) we have used the independence of $z(Y^{(l_*)})$ and $\sum_{l\neq l_*} z(Y^{(l)})$ under $\mathbb{P}_1$, where to obtain (107) we have made the particular choice

$$h \triangleq \max_{y^N \in \mathcal{T}_1 \cup \mathcal{T}_2} |z(y^N)|, \tag{110}$$

where to obtain (108) we have used that $\mathbb{P}_1(Y^{(l_*)} \in \mathcal{T}_1) \geq 1-\mu$ when the block length N is chosen sufficiently large, and where in (109), $Z_1, \ldots, Z_{s-1}$ are a set of i.i.d. random variables, each distributed according to $z(Y^N)$ where Y^N is distributed according to $Q^N(\cdot|\star^N)$.

Substituting (109), and the corresponding expression

$$\mathbb{P}_2\left(\sum_{l=1}^{s} z\left(Y^{(l)}\right) > 0\right) \geq (1-\mu)\, \mathbb{P}\left(\sum_{l=1}^{s-1} Z_l > h\right)$$

obtained by symmetry, into (102), we obtain

$$\mathbb{P}(\mathcal{E}) \geq \left(\frac{1-\mu}{2}\right) \mathbb{P}(\mathcal{E}_h) \tag{111}$$

where

$$\mathcal{E}_h \triangleq \left\{Z^{s-1} : \left|\sum_{l=1}^{s-1} Z_l\right| > h\right\}. \tag{112}$$

From (110) we have

$$h = \max(h_1, h_2), \tag{113}$$

where

$$\begin{aligned}
h_m &= \max_{y^N \in \mathcal{T}_m} |z(y^N)|\\
&= \max_{\substack{\{V \in \mathcal{P}_N^{\mathcal{Y}}:\\ |V(b) - Q(b|c(m))| < \mu,\ \forall b \in \mathcal{Y}\}}} |z(V)| &(114)\\
&= |z(Q(\cdot|c(m)) + o_\mu(1))| \quad \text{as } \mu \to 0 &(115)\\
&= e^{N(D(Q(\cdot|c(m))\|Q(\cdot|\star)) + o_\mu(1))}\\
&\quad \cdot \left(1 - e^{-ND(Q(\cdot|c(m))\|Q(\cdot|c(3-m)))}\right) &(116)\\
&= e^{N(D(Q(\cdot|c(m))\|Q(\cdot|\star)) + o_\mu(1) + o_N(1))} &(117)
\end{aligned}$$

as $\mu \to 0$ and $N \to \infty$, where $o_\mu(1)$ and $o_N(1)$ denote vanishing terms in μ and N, respectively. To obtain (114) we have used the definition (104) of $\mathcal{T}_m$, to obtain (115) we have used that the admissible V are of the form[11]

$$V = Q(\cdot|c(m)) + o_\mu(1), \quad \text{as } \mu \to 0, \tag{118}$$

and to obtain (116) we have used (103) and the continuity of $D(\cdot\|Q(\cdot|\star))$. In turn, letting

$$c_* \triangleq \arg \max_{x \in \{c(1), c(2)\}} D(Q(\cdot|x)\|Q(\cdot|\star)) \tag{119}$$

we then have, using (117) with (113)

$$h = e^{N(D(Q(\cdot|c_*)\|Q(\cdot|\star)) + o_\mu(1) + o_N(1))} \tag{120}$$

as $\mu \to 0$ and $N \to \infty$.

Hence, to establish our proposition, via (97), it remains only to verify that the random walk $\sum_{i=1}^{s} Z_l$ crosses $\pm h$ with finite probability as $N \to \infty$. Our argument makes use of the following lemma, whose proof we defer to Appendix I.

Lemma 6: Let $P \in \mathcal{P}^{\mathcal{A}}$ be a distribution over a finite alphabet $\mathcal{A} = \{a_1, a_2, \ldots, a_{|\mathcal{A}|}\}$ such that, for some integer $n \geq 1$ and $\delta_0 \in (0,1)$

$$P(a_2) \geq \frac{1}{n}. \tag{121}$$

Let $\hat{P}$ be an n-type over $\mathcal{A}$ so that

$$\min\left\{\frac{\hat{P}(a_1)}{P(a_1)}, \frac{P(a_2)}{\hat{P}(a_2)}\right\} \geq \delta_0 \quad \text{and} \quad \hat{P}(a_1) \geq \frac{3}{n}. \tag{122}$$

Let

$$\bar{P}(a) \triangleq \varphi(\hat{P}) = \begin{cases} \hat{P}(a) - 3/n, & a = a_1\\ \hat{P}(a) + 3/n, & a = a_2\\ \hat{P}(a), & a \neq a_1, a_2. \end{cases} \tag{123}$$

Then

$$P^n(\mathcal{T}(\bar{P})) \geq \delta P^n(\mathcal{T}(\hat{P})) \tag{124}$$

for some $\delta > 0$ (that depends on δ_0).

To apply Lemma 6, we first let $\mathcal{A}$ be the alphabet for Z_l, i.e.,

$$\mathcal{A} = \{a : a = z(y^N) \text{ for some } y^N \in \mathcal{Y}^N\}$$

[11]For the purposes of interpreting such expressions, the distributions involved can be viewed as vectors, and the order factor applies to the vector as a whole.

and let

$$a_1 = h, \tag{125}$$
$$a_2 = z(V_*), \quad \text{where } V_* = \arg\max_{V \in \mathcal{P}_N^{\mathcal{Y}}} Q^N(\mathcal{T}(V)|\star^N). \tag{126}$$

Furthermore, we let P be the distribution of Z_l when no message is sent in the corresponding block, i.e., for all $a \in \mathcal{A}$

$$P(a) = \sum_{\{y^N \in \mathcal{Y}^N : z(y^N) = a\}} Q^N(y^N|\star^N) \tag{127}$$

and we let, using (97) with (98)

$$n = s - 1 = e^{N(\alpha(Q)+\epsilon)(1+o(1))}, \quad \text{as } N \to \infty. \tag{128}$$

Finally, let $\hat{P}$ be any distribution such that $\mathcal{T}(\hat{P}) \subset \bar{\mathcal{E}}_{h,\mu}$, where

$$\bar{\mathcal{E}}_{h,\mu} \triangleq \mathcal{E}_h^{\mathrm{c}} \cap \tilde{\mathcal{E}}_\mu \tag{129}$$

with

$$\tilde{\mathcal{E}}_\mu \triangleq \left\{ Z^n \in \mathcal{A}^n : \left| \frac{\hat{P}_{Z^n}(a_i)}{P(a_i)} - 1 \right| < \mu, \ i = 1, 2 \right\}. \tag{130}$$

Now $\bar{P} = \varphi(\hat{P})$ as defined in (123) satisfies $\mathcal{T}(\bar{P}) \subset \mathcal{E}_h$, i.e.,

$$\mathcal{T}(\varphi^{-1}(\bar{P})) \subset \bar{\mathcal{E}}_{h,\mu} \Rightarrow \mathcal{T}(\bar{P}) \subset \mathcal{E}_h, \tag{131}$$

as we now show. First, $\mathcal{E}_h$ as defined in (112) depends only on the type of Z^n, and can be equivalently expressed in the form

$$\mathcal{E}_h = \left\{ Z^n \in \mathcal{A}^n : \left| n \sum_{a \in \mathcal{A}} a \hat{P}_{Z^n}(a) \right| > h \right\}.$$

Next, using (123) and (125), we have

$$n \sum_{a \in \mathcal{A}} a \bar{P}(a) = -3h \left(1 - \frac{a_2}{h}\right) + n \sum_{a \in \mathcal{A}} a \hat{P}(a). \tag{132}$$

and, via (129), since $\mathcal{T}(\hat{P}) \in \bar{\mathcal{E}}_{h,\mu}$, we have $\mathcal{T}(\hat{P}) \in \mathcal{E}_h^{\mathrm{c}}$, i.e.,

$$\left| n \sum_{a \in \mathcal{A}} a \hat{P}(a) \right| \le h. \tag{133}$$

But then (132) implies

$$\left| n \sum_{a \in \mathcal{A}} a \bar{P}(a) \right| \ge 2h(1 + o(1)), \quad \text{as } N \to \infty \tag{134}$$

so $\mathcal{T}(\bar{P}) \in \mathcal{E}_h$ as claimed, where we have used that $a_2/h = o(1)$. To verify this latter ratio is vanishing, note that while h grows exponentially with N according to (120), a_2 is bounded. Specifically, V_* in (126) satisfies

$$V_* = Q(\cdot|\star) + o(1), \quad \text{as } N \to \infty \tag{135}$$

which, when substituted into (103), yields

$$z(V_*) = e^{-N(D(Q(\cdot|\star)\|Q(\cdot|c(1)))+o(1))} - e^{-N(D(Q(\cdot|\star)\|Q(\cdot|c(2)))+o(1))}, \quad \text{as } N \to \infty$$

whence $|a_2| \le e^{o(N)}$ as $N \to \infty$, i.e., a_2 grows at most subexponentially with N.

In turn, we have

$$\begin{aligned} P^n(\mathcal{E}_h) &= \sum_{\{\bar{P} \in \mathcal{P}_n^{\mathcal{A}} : \mathcal{T}(\bar{P}) \subset \mathcal{E}_h\}} P^n(\mathcal{T}(\bar{P})) \\ &\ge \sum_{\{\bar{P} \in \mathcal{P}_n^{\mathcal{A}} : \mathcal{T}(\varphi^{-1}(\bar{P})) \subset \bar{\mathcal{E}}_{h,\mu}\}} P^n(\mathcal{T}(\bar{P})) \qquad (136) \\ &= \sum_{\{\hat{P} \in \mathcal{P}_n^{\mathcal{A}} : \mathcal{T}(\hat{P}) \subset \bar{\mathcal{E}}_{h,\mu}\}} P^n(\mathcal{T}(\varphi(\hat{P}))) \\ &\ge \delta \sum_{\{\hat{P} \in \mathcal{P}_n^{\mathcal{A}} : \mathcal{T}(\hat{P}) \subset \bar{\mathcal{E}}_{h,\mu}\}} P^n(\mathcal{T}(\hat{P})) \qquad (137) \\ &= \delta P^n(\bar{\mathcal{E}}_{h,\mu}) \qquad (138) \end{aligned}$$

for some $\delta > 0$, where to obtain (136) we have used that from (123), $\hat{P}$ and $\bar{P}$ are related by the $1 : 1$ transformation φ, and (131), and where to obtain (137) we have used (124) in Lemma 6.

We now proceed to establish our proposition by contradiction. Suppose that the right-hand side of (111) vanishes as $N \to \infty$, i.e.,

$$P^n(\mathcal{E}_h) \le \rho \tag{139}$$

for some arbitrarily small $\rho > 0$ and N sufficiently large. Then it follows from (138) and the fact

$$P^n(\bar{\mathcal{E}}_{h,\mu}) \ge 1 - \mu - \rho \tag{140}$$

that

$$P^n(\mathcal{E}_h) \ge \delta(1 - \mu - \rho) \tag{141}$$

for N large enough, which is in contradiction with (139) for ρ small enough. We conclude that $\mathbb{P}(\mathcal{E}_h)$ must be asymptotically bounded away from zero, and so is the right-hand side of (111).

To verify (140), we first note

$$\begin{aligned} P^n(\bar{\mathcal{E}}_{h,\mu}) &= P^n(\mathcal{E}_h^{\mathrm{c}} \cap \tilde{\mathcal{E}}_\mu) \\ &\ge P^n(\mathcal{E}_h^{\mathrm{c}}) + P^n(\tilde{\mathcal{E}}_\mu) - 1 \qquad (142) \\ &\ge P^n(\tilde{\mathcal{E}}_\mu) - \rho \qquad (143) \\ &\ge 1 - \mu - \rho \qquad (144) \end{aligned}$$

where to obtain (142) we have used the union bound, where to obtain (143) we have used (139), and where to obtain (144) we used the fact that

$$P^n(\tilde{\mathcal{E}}_\mu) \ge 1 - \mu \tag{145}$$

for any sufficiently small $\mu > 0$, as $N \to \infty$.

To verify (145), we first use Chebyshev's inequality and the fact that the variance of a binomial random variable is upper-bounded by its mean to obtain, for $i \in \{1, 2\}$

$$P^n \left(\left| \sum_{l=1}^n \mathbb{1}_{a_1}(Z_l) - nP(a_i) \right| \ge n\mu P(a_i) \right) \le \frac{1}{n\mu^2 P(a_i)} \tag{146}$$

whence

$$P^n(\tilde{\mathcal{E}}_\mu^{\mathrm{c}}) \le \frac{1}{n\mu^2 P(a_1)} + \frac{1}{n\mu^2 P(a_2)}.$$

Thus, it suffices to show that $nP(a_i)$ grows with N for $i \in \{1,2\}$.

Considering first, $i=1$, we have

$$\begin{aligned} P(a_1) &\triangleq \sum_{\{y^N\in\mathcal{Y}^N:\, z(y^N)=h\}} Q^N(y^N|\star^N) \\ &\geq \sum_{\{y^N\in\mathcal{T}_m:\, z(y^N)=h\}} Q^N(y^N|\star^N) \\ &\geq \frac{1}{\mathrm{poly}(N)} \\ &\quad\cdot \sum_{\substack{\{V\in\mathcal{P}_N^{\mathcal{Y}}:\, z(V)=h,\\ |V(b)-Q(b|c(m))|<\mu,\ \forall b\in\mathcal{Y}\}}} e^{-ND(V\|Q(\cdot|\star))} \qquad (147) \\ &\geq e^{-ND(Q(\cdot|c(m))+o_\mu(1)\|Q(\cdot|\star))(1+o_N(1))} \qquad (148) \\ &\geq e^{-N(D(Q(\cdot|c(m))\|Q(\cdot|\star))+o_\mu(1))(1+o_N(1))} \qquad (149) \\ &\geq e^{-N(D(Q(\cdot|c_*)\|Q(\cdot|\star))+o_\mu(1)+o_N(1))} \qquad (150) \end{aligned}$$

where to obtain (147) we have used the usual lower bound on the probability of a type class [14, Theorem 12.1.4], where to obtain (148) we have used that the admissible V are of the form (118), where to obtain (149) we have used the continuity of $D(\cdot\|Q(\cdot|\star))$, and where to obtain (150) we have used (119). Combining (128) and (150) we thus obtain

$$nP(a_1) \geq e^{N(\epsilon+\alpha(Q)-D(Q(\cdot|c_*)\|Q(\cdot|\star))+o_\mu+o_N(1))} \qquad (151)$$

which grows exponentially with N provided μ is small enough, since $\alpha(Q) \geq D(Q(\cdot|c_*)\|Q(\cdot|\star))$ via (20).

For the case $i=2$, we note that by [14, Theorem 12.1.4]

$$P(a_2) \geq \frac{1}{\mathrm{poly}(N)} \qquad (152)$$

which when combined with (128) yields

$$nP(a_2) \geq \frac{n}{\mathrm{poly}(N)} > e^{N(\alpha(Q)+\epsilon)(1+o(1))} \qquad (153)$$

which also grows exponentially with N.

To conclude the proof we need only verify that P and $\hat{P}$ satisfy the conditions (121) and (122) of Lemma 6 for our choices (125) and (126) of a_1 and a_2, respectively.

First, that (121) is satisfied follows immediately from (153).

Next, by the definition of $\hat{P}$ it follows that

$$\left|\frac{\hat{P}(a_1)}{P(a_1)} - 1\right| < \mu$$

so

$$\frac{\hat{P}(a_1)}{P(a_1)} \geq 1-\mu. \qquad (154)$$

Likewise, from the definition of $\hat{P}$ it also follows that

$$\frac{P(a_2)}{\hat{P}(a_2)} \geq \frac{1}{1+\mu}. \qquad (155)$$

Hence, the left-hand inequality in (122) is satisfied with $\delta_0 = \min(1-\mu, 1/(1+\mu)) > 0$.

Finally, the definition of $\hat{P}$ equivalently implies that

$$\left|\frac{n\hat{P}(a_1)}{nP(a_1)} - 1\right| < \mu$$

i.e.,

$$n\hat{P}(a_1) \geq (1-\mu)nP(a_1). \qquad (156)$$

But, as we showed, $nP(a_1)$ grows exponentially with N, so (156) implies that $n\hat{P}(a_1)$ grows without bound, so the second requirement of (122) is satisfied. □

Proposition 2 (Achievability): Given a channel Q, any asynchronism exponent α strictly less than $\alpha(Q)$ as defined in (20) is achievable.

Proof of Proposition 2: When $C(Q)=0$, Proposition 1 establishes that $\alpha(Q) \leq 0$, and thus $\alpha(Q)=0$ in this case. Hence, in the sequel, we consider $C(Q) > 0$.

Also, if $Q(b|\star)=0$ for some $b \in \mathcal{Y}$, then arbitrarily large asynchronism exponents can be achieved, i.e., $\alpha(Q)=\infty$. Indeed, in such a scenario, it suffices to use a code for synchronous channels of block length $N-\ln N$ together with a prefix of length $\ln N$ consisting solely of the symbol b. Then the probability that the codeword start cannot be detected vanishes with N, so the resulting code achieves at least $(0,\alpha)$ for any α.

Hence, in the sequel we need only consider the case in which $Q(y|\star) > 0$ for all $y \in \mathcal{Y}$, to which Theorem 1 can be applied. Accordingly, we establish our proposition by showing that P, t_D, t_I, and Δ can be appropriately chosen in Theorem 1 so that α is arbitrarily close to $\alpha(Q)$.

First, we pick an input distribution P_0 so that $I(P_0Q) > 0$ and $D((P_0Q)_Y\|Q(\cdot|\star)) > 0$; this is possible since $C(Q) > 0$. Next, let

$$P = (1-\Delta)P_0 + \Delta P_*$$

where P_* is the distribution in which a maximizing symbol in (20) occurs with probability 1. Note that from the concavity of $I(PQ)$ in P, we have, for all $\Delta \in (0,1)$

$$I(PQ) \geq (1-\Delta)I(P_0Q) + \Delta I(P_*Q) > 0.$$

Proceeding, fix Δ, and let $t_I = 1/(1-\Delta)$. Moreover, let

$$t_D = \frac{D((PQ)_Y\|Q(\cdot|\star))}{I(PQ)}\frac{\Delta}{(1-\Delta)}$$

so that conditions (II) and (III) are satisfied. From condition (I) the achieved exponent is

$$\begin{aligned} \alpha_\Delta &= \frac{t_D}{t_I}I(PQ)\Delta^2 + I(PQ)\Delta^2 - \frac{1}{t_I}I(PQ)\Delta^2 \\ &= D((PQ)_Y\|Q(\cdot|\star))\Delta^3 + I(PQ)\Delta^3 \qquad (157) \\ &\geq D((PQ)_Y\|Q(\cdot|\star))\Delta^3. \qquad (158) \end{aligned}$$

Taking limits, we have

$$\begin{aligned} \lim_{\Delta\to1} D((PQ)_Y\|Q(\cdot|\star)) &= D((P_*Q)_Y\|Q(\cdot|\star)) \qquad (159) \\ &= \max_x D(Q(\cdot|x)\|Q(\cdot|\star)) \qquad (160) \\ &= \alpha(Q) \qquad (161) \end{aligned}$$

where to obtain (159) we have used, via (18), that $D((PQ)_Y \| Q(\cdot|\star))$ is bounded and therefore continuous in P, and where to obtain (160) we have used the definition of P_*.

Finally, using (161) with (158), we see that

$$\lim_{\Delta\to 1} \alpha_\Delta \geq \alpha(Q).$$

from which we conclude all asynchronism exponents strictly less than $\alpha(Q)$ can be achieved. □

VIII. Concluding Remarks

Our main contribution is a simple but meaningful model for the kinds of highly sporadic communication characteristic of emerging sensor network and related applications. Two key features of our model are that: 1) message transmission commences at a random time within some window, which characterizes the level of asynchronism, and 2) communication rate is the number of message bits relative to the (average) receiver reaction delay, i.e., the elapsed time between when the transmission commences and the decoder makes a decision.

Under this model, when the asynchronism level scales subexponentially in the length of the codeword used to represent the message, the rate loss relative to synchronous communication is zero—the capacity of the synchronized channel can be achieved—and when it scales superexponentially, reliable communication is generally not possible at all. As such, the exponential regime is the interesting one. As we show, there is a sharp phase transition phenomenon: reliable communication is possible if and only if the scaling exponent is below a particular channel-dependent critical value. When, in addition, there is a particular rate requirement, the critical value decreases, i.e., less asynchronism can be tolerated. However, we show that at any rate below the capacity of the synchronized channel, reliable communication is generally possible with at least some level of exponential asynchronism.

There are several natural directions for further research. First, characterizing the asynchronous capacity region $\alpha(R, Q)$ for all $R > 0$ would be useful, or at least obtaining good inner and outer bounds. Recent preliminary results along these lines appear in [20], [21]. More generally, there is much to be done in the development of practical codes that both approach these fundamental limits, and can be decoded with low complexity.

There are also important architectural questions. For example, while existing communication systems make use of separate synchronization and communication phases in the transmission, this is not a constraint in our formulation. Indeed, our coding schemes did not impose such separation. It will be useful to quantify the rate loss inherent in schemes with separate synchronization, and understand the regimes in which such losses are and are not significant. Recent preliminary results on these issues appear in [22].

There are many extensions of the present model that warrant investigation. One example is the extension to continuous-time channels, an important example of which is the general Gaussian channel. Another is the extension to channels with memory, such as finite-state channels.

Another extension involves incorporating channel state uncertainty into the problem. In the current model, the parameters of the channel law are fixed and known *a priori* to both transmitter and receiver. Indeed, our codebook and decoding rule depend on them. In practice, however, such side information is often time-varying, and only partially or imperfectly available at the transmitter and/or receiver.

Still another important extension involves incorporating feedback into the model, and the transmission of sequences of messages. Among other questions, there is a need to understand the impact of feedback in such asynchronous settings, and any qualitative differences from the synchronous setting. Naturally, there are many possible feedback mechanism models. In one simple model, the receiver is able to send to the transmitter—without error or delay—a single acknowledgment (ACK) bit when it has successfully decoded a message. Other models allow more extensive feedback. In any such analysis, it will be important to understand the degree to which performance is sensitive to the assumption of noiseless feedback. Indeed, if the feedback is noisy, the receiver's decision may be wrongly recognized by the transmitter, which can result in a loss of message synchronization between transmitter and receiver (e.g., the receiver has not yet decoded the first message while the transmitter has already started to send the second one). Ultimately, this potential additional source of asynchronism needs to be taken into account.

It is also worth exploring extensions of the model to the case in which the transmitter may have no message to send in the designated transmission window. For instance, a message is sent with probability $1 - p$; otherwise, no message is sent. For this setting, natural scalings between p and the asynchronism level remain to be investigated.

Finally, exploring variations on our basic model is also worthwhile. As one example, one might consider other ways to capture the requirement of quick decoding. In some sense, our formulation investigates aspects of the tradeoff between the code rate $\bar{R}$ and the average reaction delay $\bar{\Delta} = \mathbb{E}(\tau - \nu)^+$. Equivalently, it examines code rates $\bar{R}$ achievable under the expectation constraint $\mathbb{E}(\tau-\nu)^+ \leq d$, as a function of d. In such a formulation, the communication rates $R = \ln M/d$ obtained for a given d are comparatively low under exponential asynchronism. This is because even though the probability of missing the codeword is exponentially small, once the codeword is missed we pay a penalty in reaction delay that is on the order of the asynchronism level, i.e., exponentially large. As a result, it may be useful to examine code rates achievable under a typicality constraint of the form $\mathbb{P}((\tau - \nu)^+ \leq d) \approx 1$, which may yield higher effective communication rates $R = \ln M/d$.

Appendix I
Proof of Lemma 6

The binomial expansion for $P^n(\mathcal{T}(\hat{P}))$ takes the form (see, e.g., [14, eq. (12.25)])

$$P^n(\mathcal{T}(\hat{P})) = \binom{n}{n\hat{P}(a_1), \ldots, n\hat{P}(a_{|\mathcal{A}|})} \prod_{a\in\mathcal{A}} P(a)^{n\hat{P}(a)}.$$

Using the hypotheses on $P, \hat{P}$, and $\bar{P}$ gives $\hat{P}(a_i) \geq 3/n$ for $i \in \{1,2\}$, whence

$$\begin{aligned}
&\frac{P^n(\mathcal{T}(\bar{P}))}{P^n(\mathcal{T}(\hat{P}))} \\
&\quad= \left(\frac{P(a_2)}{P(a_1)}\right)^3 \frac{(n\hat{P}(a_1)-2)(n\hat{P}(a_1)-1)(n\hat{P}(a_1))}{(n\hat{P}(a_2)+1)(n\hat{P}(a_2)+2)(n\hat{P}(a_2)+3)} \\
&\quad= \left[\frac{P(a_2)^3}{(\hat{P}(a_2)+1/n)(\hat{P}(a_2)+2/n)(\hat{P}(a_2)+3/n)}\right] \\
&\qquad \cdot \left[\frac{(\hat{P}(a_1)-2/n)(\hat{P}(a_1)-1/n)\hat{P}(a_1)}{P(a_1)^3}\right] \\
&\quad\geq \delta
\end{aligned}$$

for some $\delta = \delta(\delta_0) > 0$. □

Appendix II

Lemma 7: For any distributions $P \in \mathcal{P}^{\mathcal{X}}$ and $Q \in \mathcal{P}^{\mathcal{Y}|\mathcal{X}}$, and any $r \geq 0$

$$\begin{aligned}
&\min_{\lambda\in[0,1]} \min_{\substack{\{V,W\in\mathcal{P}^{\mathcal{Y}|\mathcal{X}}:\\ \lambda I(PV)+(1-\lambda)I(PW)\leq r\}}} \lambda D(PV\|PQ)+(1-\lambda)D(PW\|PQ) \\
&\quad= \min_{\{V\in\mathcal{P}^{\mathcal{Y}|\mathcal{X}}: I(PV)\leq r\}} D(PV\|PQ). \qquad (162)
\end{aligned}$$

Proof: If $r \geq I(PQ)$, the claim holds trivially, since the left- and right-hand sides of (162) are both zero. Hence, it suffices to restrict our attention to the case $r < I(PQ)$ in the sequel. Let us use $E_{\mathrm{L}}(r)$ and $E_{\mathrm{R}}(r)$ to denote the left- and right-hand sides of (162), respectively.

Proceeding, we define

$$\begin{aligned}
E'_{\mathrm{L}}(r) &= \min_{\lambda\in[0,1]} \inf_{\substack{\{V,W\in\mathcal{P}^{\mathcal{Y}|\mathcal{X}}:\\ \lambda I(PV)+(1-\lambda)I(PW)=r,\\ I(PV)>I(PW)\}}} (\lambda D(PV\|PQ) \\
&\qquad + (1-\lambda)D(PW\|PQ)) \\
E'_{\mathrm{R}}(r) &= \min_{\{V\in\mathcal{P}^{\mathcal{Y}|\mathcal{X}}: I(PV)=r\}} D(PV\|PQ) \\
\bar{E}_{\mathrm{L}}(r) &= \min_{\lambda\in[0,1]} \min_{\substack{\{V,W\in\mathcal{P}^{\mathcal{Y}|\mathcal{X}}:\\ \lambda I(PV)+(1-\lambda)I(PW)\leq r,\\ I(PV)=I(PW)\}}} (\lambda D(PV\|PQ) \\
&\qquad + (1-\lambda)D(PW\|PQ)) \\
\tilde{E}_{\mathrm{L}}(r) &= \min_{\lambda\in[0,1]} \inf_{\substack{\{V,W\in\mathcal{P}^{\mathcal{Y}|\mathcal{X}}:\\ \lambda I(PV)+(1-\lambda)I(PW)\leq r,\\ I(PV)>I(PW)\}}} (\lambda D(PV\|PQ) \\
&\qquad + (1-\lambda)D(PW\|PQ)).
\end{aligned}$$

Since V and W appear symmetrically in the definition of $\bar{E}_{\mathrm{L}}(r)$, we can without penalty impose the additional constraint that $D(PV\|PQ) \leq D(PW\|PQ)$ in the definition, from which we obtain that $\lambda = 1$ is optimizing, and as a result

$$\bar{E}_{\mathrm{L}}(r) = E_{\mathrm{R}}(r).$$

Hence, it suffices to show that

$$\tilde{E}_{\mathrm{L}}(r) \geq E_{\mathrm{R}}(r) \qquad (163)$$

since $E_{\mathrm{L}}(r) = \min\{\bar{E}_{\mathrm{L}}, \tilde{E}_{\mathrm{L}}\}$.

To verify (163), it is sufficient to establish the following two simple claims:

$$\begin{aligned}
&E'_{\mathrm{R}}(r) \text{ is convex in } r \qquad \text{(Claim 1)} \\
&E_{\mathrm{R}}(r) = E'_{\mathrm{R}}(r). \qquad \text{(Claim 2)}
\end{aligned}$$

Indeed, from these two claims we have

$$\begin{aligned}
E'_{\mathrm{L}}(r') &= \inf_{\{r_1,r_2:\, r_2<r'<r_1\}} \inf_{\substack{\{V,W:\\ I(PV)=r_1,\\ I(PW)=r_2\}}} \frac{r'-r_2}{r_1-r_2} D(PV\|PQ) \\
&\qquad + \frac{r_1-r'}{r_1-r_2} D(PW\|PQ) \\
&= \inf_{\{r_1,r_2:\, r_2<r'<r_1\}} \frac{r'-r_2}{r_1-r_2} E'_{\mathrm{R}}(r_1) + \frac{r_1-r'}{r_1-r_2} E'_{\mathrm{R}}(r_2) \\
&\geq \inf_{\{r_1,r_2:\, r_2<r'<r_1\}} E'_{\mathrm{R}}(r') \qquad (164) \\
&= E_{\mathrm{R}}(r'), \qquad (165)
\end{aligned}$$

where to obtain (164) we have used Claim 1, and where to obtain (165) we have used Claim 2. Finally, using (165) and that $E_{\mathrm{R}}(r)$ is a nonincreasing function of r, we obtain

$$\tilde{E}_{\mathrm{L}}(r) = \min_{r':\, r'\leq r} E'_{\mathrm{L}}(r') \geq \min_{r':\, r'\leq r} E_{\mathrm{R}}(r') \geq E_{\mathrm{R}}(r)$$

and hence (163) follows.

It remains only to establish our two claims.

For Claim 1, let V_1 and V_2 denote the optimizing distributions for $E'_{\mathrm{R}}(r_1)$ and $E'_{\mathrm{R}}(r_2)$, respectively, for some $r_1 \neq r_2$, i.e.,

$$\begin{aligned}
I(PV_1) = r_1 &\quad \text{and} \quad I(PV_2) = r_2 \\
D(PV_1\|PQ) = E'_{\mathrm{R}}(r_1) &\quad \text{and} \quad D(PV_2\|PQ) = E'_{\mathrm{R}}(r_2).
\end{aligned}$$

Hence, for $V = \theta V_1 + (1-\theta)V_2$ with $\theta \in [0,1]$ we have, using the convexity of $I(PV)$ in V

$$\begin{aligned}
I(PV) &\leq \theta I(PV_1) + (1-\theta)I(PV_2) \qquad (166) \\
&= \theta r_1 + (1-\theta) r_2 \qquad (167)
\end{aligned}$$

so we establish Claim 1 by observing that

$$\begin{aligned}
&E'_{\mathrm{R}}(\theta r_1 + (1-\theta) r_2) \\
&\quad\leq D(PV\|PQ) \\
&\quad\leq \theta D(PV_1\|PQ) + (1-\theta) D(PV_2\|PQ) \\
&\quad= \theta E'_{\mathrm{R}}(r_1) + (1-\theta) E'_{\mathrm{R}}(r_2) \qquad (168)
\end{aligned}$$

where to obtain (168) we have used the convexity of $D(PV\|PQ)$ in V.

For Claim 2, if $r = 0$ the claim holds trivially. For the case $r > 0$, we use proof by contradiction. Suppose that the optimizing distribution for $E_{\mathrm{R}}(r)$ is V_* and satisfies $I(PV_*) < r$. Then define $V = \theta V_* + (1-\theta)Q$ with $\theta \in [0,1)$. By the convexity of $D(PV\|PQ)$ in V we have

$$D(PV\|PQ) \leq \theta D(PV_*\|PQ) < D(PV_*\|PQ),$$

and by the convexity of $I(PV)$ in V we have

$$I(PV) \leq \theta I(PV_*) + (1-\theta) I(PQ) \leq r,$$

where the last inequality holds provided that

$$\theta \leq \frac{I(PQ) - r}{I(PQ) - I(PV_*)}.$$

Thus, with such a choice for θ, we see that V is better than V_*, so V_* cannot be the optimizing distribution for $E_{\mathrm{R}}(r)$, which establishes Claim 2. □

Acknowledgment

The authors wish to thank Dr. Ashish Khisti for valuable discussions.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Oct. 1948.

[2] D. Chase, "Communication Over Noisy Channels with no a priori Synchronization Information," Ph.D. dissertation, MIT, Cambridge, MA, Feb. 1968.

[3] R. L. Dobrushin, "Shannon's theorems for channels with synchronization errors," *Probl. Inf. Transm.*, vol. 3, no. 4, pp. 11–26, 1967.

[4] E. Drinea and M. Mitzenmacher, "A simple lower bound for the capacity of the deletion channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4657–4660, Oct. 2006.

[5] S. Diggavi and M. Grossglauser, "On transmission over deletion channels," in *Proc. Allerton Conf. Communication, Control and Computing*, Monticello, IL, Sep. 2001, pp. 573–582.

[6] T. M. Cover, R. J. McEliece, and E. C. Posner, "Asynchronous multiple-access channel capacity," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 4, pp. 409–413, Jul. 1981.

[7] J. Y. N. Hui and P. A. Humblet, "The capacity of the totally asynchronous multiple-access channel," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 2, pp. 207–216, Mar. 1985.

[8] G. S. Poltyrev, "Coding in an asynchronous multiple-access channel," *Probl. Inf. Transm.*, vol. 36, pp. 12–21, Jul.–Sep. 1983.

[9] S. Verdú, "The capacity region of the symbol-asynchronous Gaussian multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 35, no. 4, pp. 733–751, Jul. 1989.

[10] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, vol. 42, no. 2, pp. 1897–1908, 1971.

[11] I. V. Nikiforov, "A generalized change detection problem," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 171–187, Jan. 1995.

[12] T. Lai, "Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 595–607, Mar. 2000.

[13] M. Basseville and I. Nikiforov, Fault Isolation for Diagnosis: Nuisance Rejection and Multiple Hypothesis Testing INRIA, Rapport de Recherche, 4438, 2002.

[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[15] I. Csiszàr and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[16] V. Chandar, A. Tchamkerten, and G. Wornell, "Optimal sequential frame synchronization," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3725–3728, Aug. 2008.

[17] A. Tchamkerten, A. Khisti, and G. W. Wornell, "Information theoretic perspectives on synchronization," in *Proc. Int. Symp. Information Theory*, Seattle, WA, Jul. 2006, pp. 371–375.

[18] A. Tchamkerten and I. E. Telatar, "Variable length coding over an unknown channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2126–2145, May 2006.

[19] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York: McGraw-Hill, 1976.

[20] A. Tchamkerten, V. Chandar, and G. W. Wornell, "On the capacity region of asynchronous channels," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Toronto, ON, Canada, Jul 2008, pp. 1213–1217.

[21] V. Chandar, A. Tchamkerten, and G. W. Wornell, "A novel asynchronous communication paradigm: Detection, isolation, and coding," in *Proc. Int. Workshop on Sequential Methodologies*, Troyes, France.

[22] V. Chandar, A. Tchamkerten, and G. W. Wornell, "Training based schemes are suboptimal for high rate asynchronous communication," in *Proc. IEEE Information Theory Workshop*, Taormina, Italy, Oct. 2009, to be published.

Aslan Tchamkerten received the engineer physicist diploma in 2000, and the Ph.D. degree in communication in 2005, both from the Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.

From 2005 to 2008, he was a Postdoctoral Associate in the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge. Since 2008, he has been with the Communications and Electronics Department, Telecom ParisTech, Paris, France, where he is currently Assistant Professor.

Venkat Chandar received the S.B. and M.Eng. degrees in electrical engineering and computer science, and the S.B. degree in mathematics, from the Massachusetts Institute of Technology (MIT), Cambridge, in 2006.

He is currently working towards the Ph.D. degree at MIT.

Gregory W. Wornell (S'83–M'91–SM'00–F'04) received the B.A.Sc. degree from the University of British Columbia, Vancouver, BC, Canada, and the S.M. and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, all in electrical engineering and computer science, in 1985, 1987, and 1991, respectively.

Since 1991, he has been on the faculty at MIT, where he is Professor of Electrical Engineering and Computer Science, leads the Signals, Information, and Algorithms Laboratory in the Research Laboratory of Electronics, and is Co-Director of the Center for Wireless Networking. He has held visiting appointments at the former AT&T Bell Laboratories, Murray Hill, NJ, the University of California, Berkeley, CA, and Hewlett-Packard Laboratories, Palo Alto, CA. His research interests and publications span the areas of signal processing, digital communication, and information theory, and include algorithms and architectures for wireless and sensor networks, broadband systems, and multimedia environments.

Dr. Wornell has been involved in the Information Theory and Signal Processing Societies of the IEEE in a variety of capacities, and maintains a number of close industrial relationships and activities. He has won a number of awards for both his research and teaching.