Training-Based Schemes are Suboptimal for High Rate Asynchronous Communication

Venkat Chandar, Aslan Tchamkerten, and Gregory W. Wornell

***Abstract*—We consider asynchronous point-to-point communication. Building on a recently developed model, we show that training based schemes, i.e., communication strategies that separate synchronization from information transmission, perform suboptimally at high rate.**

***Index Terms*—detection and isolation; sequential decoding; synchronization; training-based schemes**

I. Model and Review of Results

We consider the asynchronous communication setting developed in [1], which provides an extension to Shannon's original point-to-point model for synchronous communication [2].

We recall the setting in [1]. Communication takes place over a discrete memoryless channel characterized by its finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively, and transition probability matrix $Q(y|x)$, for all $y \in \mathcal{Y}$ and $x \in \mathcal{X}$. There are $M \geq 2$ messages $\{1,2,\ldots,M\}$. For each message m there is an associated codeword $c^N(m) \triangleq c_1(m)c_2(m)\ldots c_N(m)$, a string of length N composed of symbols from $\mathcal{X}$.[1] The M codewords form a codebook $\mathcal{C}_N$. The transmitter selects a message m, randomly and uniformly over the message set, and starts sending the corresponding codeword $c^N(m)$ at a random time ν, unknown to the receiver, independent of $c^N(m)$, and uniformly distributed in $\{1,2,\ldots,A\}$. The transmitter and the receiver know the integer $A \geq 1$, which we refer to as the *asynchronism level* between the transmitter and the receiver. If $A = 1$ the channel is said to be synchronized. The capacity of the synchronized channel Q is denoted C, or $C(Q)$ when necessary for clarity.

During information transmission the receiver observes a noisy version of the sent codeword, while before and after the information transmission it observes only noise. Conditioned on the event $\{\nu = k\}$, $k \in \{1,2,\ldots,A\}$, and on the message m to be conveyed, the receiver observes independent symbols $Y_1, Y_2, \ldots$ distributed as follows. If $i \in \{1,2,\ldots,k-1\}$ or $i \in \{k+N, k+N+1,\ldots,A+N-1\}$, the distribution of Y_i is

$$Q_\star(\cdot) \triangleq Q(\cdot|\star)$$

for some fixed $\star \in \mathcal{X}$. At any time $i \in \{k, k+1, \ldots, k+N-1\}$, the distribution of Y_i is

$$Q(\cdot|c_{i-k+1}(m))\ .$$

It should be emphasized that the transition probability matrix $Q(\cdot|\cdot)$, together with the 'no-input' symbol $\star$, characterizes the communication channel. In particular, the $\star$ is not a parameter of the transmitter, i.e., the system designer cannot designate which symbol in the input alphabet is $\star$. This symbol can, however, be used for the codebook design. Throughout the paper, whenever we refer to a certain channel Q, we implicitly assume that the $\star$ symbol is given.

The decoder consists of a sequential test (τ_N, ϕ_N), where τ_N 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_N denotes a decision rule that declares the decoded message. 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 the realizations of $\{Y_i\}_{i=1}^n$, is independent of those of $\{Y_i\}_{i=n+1}^\infty$, for all $n \geq 1$. The function ϕ_N is then defined as any $\mathcal{F}_{\tau_N}$-measurable map taking values in $\{1,2,\ldots,M\}$, where $\mathcal{F}_1, \mathcal{F}_2, \ldots$ is the natural filtration induced by the output process $Y_1, Y_2, \ldots$.

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

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

where $\mathcal{E}$ indicates the event that the decoded message does not correspond to the sent message, and where the

This work was supported in part by the National Science Foundation under Grant No. CCF-0635191, and by a University IR&D Grant from Draper Laboratory.

V. Chandar and G. W. Wornell are with the Dept. EECS, Massachusetts Institute of Technology. (Email: {vchandar,gww}@mit.edu). A. Tchamkerten is with Telecom ParisTech, COMELEC. (Email: aslan.tchamkerten@telecom-paristech.fr).

[1]The symbol '$\triangleq$' stands for 'equal by definition.'

subscripts '$_{m,k}$' indicate the conditioning on the event that message m starts being sent at time k.

Second, we define the average communication rate with respect to the average delay it takes the receiver to react to a sent message, i.e.[2]

$$R \triangleq \frac{\ln M}{\mathbb{E}(\tau_N - \nu)^+} \triangleq \frac{\ln |\mathcal{C}_N|}{\mathbb{E}(\tau_N - \nu)^+}$$

where $\mathbb{E}(\tau_N - \nu)^+$ is defined as

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

where $\mathbb{E}_{m,k}$ denotes the expectation with respect to $\mathbb{P}_{m,k}$, and where x^+ denotes $\max\{0, x\}$. With the above definitions, we now recall the notions of (R, α) coding scheme and capacity function.

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

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

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

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

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

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

It turns out that the exponential scaling of the asynchronism exponent with respect to the codeword length in Definition 1 is natural: asynchronism induces a rate loss with respect to the capacity of the synchronous channel only when it grows at least exponentially with the codeword length [1].

The following theorem, given in [4], provides a non-trivial lower bound to the capacity of asynchronous channels:

Theorem 1. *For a given channel Q, let $\alpha \geq 0$ and let P be a distribution over $\mathcal{X}$ such that*

$$\min_V \max\{D(V\|(PQ)_{\mathcal{Y}}), D(V\|Q_\star)\} > \alpha$$

where the minimization is over all distributions over $\mathcal{Y}$, and where the distribution $(PQ)_{\mathcal{Y}}$ is defined as $(PQ)_{\mathcal{Y}}(y) = \sum_{x\in\mathcal{X}} P(x)Q(y|x)$, $y \in \mathcal{Y}$. Then, the pair $(R = I(PQ), \alpha)$ is achievable.

Corollary 1. *At capacity, it is possible to achieve a strictly positive asynchronism exponent, except for the case when $Q_\star$ corresponds to the capacity-achieving output distribution of the synchronous channel.*[3] *Moreover, the asynchronism exponent achievable at capacity can be arbitrarily large, depending on the channel.*

This is in contrast with training-based schemes. The contribution of this paper, given in the next section, is to show that training-based scheme, in general, achieve a vanishing asynchronism exponent in the limit of the rate going to capacity.

II. TRAINING-BASED SCHEMES

The usual approach to communication is a training-based architecture. In such schemes, each codeword is composed of two parts. The first part, the sync preamble, is a sequence of symbols common to all the codewords, hence carries no information; its only purpose is to help the decoder to locate the sent message. The second part carries information. The decoder operates according to a two-step procedure. First it tries to locate the codeword by seeking the sync preamble. Once the sync preamble is located, it declares a message based on the subsequent symbols. A formal definition of a training-based scheme follows.

Definition 3. *A training-based scheme is a coding scheme $\{(\mathcal{C}_N, (\tau_N, \phi_N))\}_{N\geq 1}$ with the following properties. For some $\varepsilon > 0$, $\eta \in [0, 1]$, and all integers $N \geq 1$*

i. *each codeword in $\mathcal{C}_N$ starts with a string of size ηN that is common to all codewords;*[4]
ii. *the decision time τ_N is such that the event $\{\tau_N = n\}$, conditioned on the ηN observations*

[2] $\ln$ denotes the natural logarithm.

[3] To see this, recall that, given the channel Q, all capacity-achieving input distributions P induce the same output distribution $(PQ)_{\mathcal{Y}}$. Whenever $(PQ)_{\mathcal{Y}}$ differs from $Q_\star$, the min-max expression in Theorem 1 is strictly positive. Therefore capacity is achievable at a strictly positive asynchronism exponent.

[4] To be precise, the string size should be an integer, and instead of having it equal to ηN we should have it equal to $\lfloor \eta N \rfloor$. However, since we are interested in the asymptotic $N \to \infty$, this discrepancy typically vanishes. Similar discrepancies are ignored throughout the paper.

$Y_{n-N+1}^{n-N+\eta N}$,[5] *is independent of all other past observations, i.e.,* Y_1^{n-N} *and* $Y_{n-N+\eta N+1}^{n}$*;*

iii. *the codebook* $\mathcal{C}_N$ *and the decoding time* τ_N *satisfy*

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

for all $k \in \{1, 2, \ldots, A\}$ *.*

Condition i. specifies the size of the sync preamble. Condition ii. indicates that the decoding time should depend only on the sync preamble. Condition iii. imposes that the codeword symbols that follow the sync preamble should not be used to help the decoder locate the codeword. If we remove Condition iii., one could imagine having information symbols with a 'sufficiently biased' distribution to help the decoder locate the codeword position (the 'information symbols' could even start with a second preamble!). In this case the sync preamble is followed by a block of information symbols that also helps the decoder to locate the sent codeword. To avoid this, we impose Condition iii. which says that, once the sync preamble is missed (this is captured by the event $\{\tau_N \geq k+N, \nu = k\}$, the decoder's decision to stop will likely no more depend on the sent codeword since it will occur after $k+2N-1$.

Finally, it can be shown that a large class of training-based schemes considered in practice satisfy the above three conditions.

Theorem 2. *A training-based scheme that achieves a rate* $R \in (0, C(Q)]$ *operates at an asynchronism exponent* α *upper bounded as*

$$\alpha \leq \left(1 - \frac{R}{C}\right) \max_P \min_W \max\{D_1, D_2\},$$

where $D_1 \triangleq D(W\|Q|P)$*, and* $D_2 \triangleq D(W\|Q_\star|P)$*.*[6] *The first maximization is over all distributions over* $\mathcal{X}$ *and the minimization is over all conditional distributions defined over* $\mathcal{X} \times \mathcal{Y}$*.*

The following result is a consequence of Theorem 2.

Corollary 2. *Unless the no-input symbol* $\star$ *does not generate a particular channel output symbol (i.e.,* $Q(y|\star) = 0$ *for some* $y \in \mathcal{Y}$*), training-based schemes achieve a vanishing asynchronism exponent as* $R \to C(Q)$*.*

Proof of Corollary 2: We consider the inequality of Theorem 2 and first upper bound the minimization by choosing $W = Q$. With this choice, the inner maximization becomes $D_2 = D(Q||Q_\star|P)$ (since $D_1 = D(Q||Q|P) = 0$). Maximizing over P yields

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

which is bounded when $Q(y|\star) > 0$ for all $y \in \mathcal{Y}$. Therefore the max-min-max term in the inequality of Theorem 2 is finite and gets multiplied by a term that vanishes as $R \to C(Q)$. ■

Thus, except for degenerate cases, training-based schemes achieve a vanishing asynchronism exponent in the limit of the rate going to capacity. In contrast, from Theorem 1 one deduces that it is possible, in general, to achieve a non-zero asynchronism exponent at capacity, as we saw above.

This suggests that to achieve a high rate under strong asynchronism, separating synchronization from information transmission is suboptimal; the codeword symbols should all play the dual role of information carriers and 'information flags.'

Sketch of Proof of Theorem 2

Consider a training-based scheme $\{(\mathcal{C}_N, (\tau_N, \phi_N))\}_{N \geq 1}$. For simplicity, we assume that the sync preamble distribution of $\mathcal{C}_N$ is the same, equal to P, for all $N \geq 1$. The case of different preamble distributions for different values of N requires a minor extension. The proof consists in showing that if the following two inequalities hold

$$\eta D(W||Q|P) < \alpha \tag{1}$$

$$\eta D(W||Q_\star|P) < \alpha \tag{2}$$

for some conditional distribution W, then the average reaction delay achieved by $\{(\mathcal{C}_N, (\tau_N, \phi_N))\}_{N \geq 1}$ grows exponentially with N. This, in turn, can be shown to imply that the rate is asymptotically equal to zero. Therefore, maximizing over the sync preamble distributions, it is necessary that

$$\alpha \leq \eta \max_P \min_W \max\{D(W||Q|P), D(W||Q_\star|P)\}$$

in order to achieve a strictly positive rate R. The second part of the proof, omitted in this paper, consists in showing that the highest value of η compatible with rate R communication is upper bounded by $(1 - R/C(Q))$. This with the above inequality yields the desired result.

Below we sketch the argument that shows that, if both (1) and (2) hold, the average reaction delay grows exponentially with N.

To keep the presentation simple, in the equations below we omit terms that go to zero in the limit $N \to \infty$. Thus, although the equations may not be valid as written, they become valid in that limit.

[5]We use Y_i^j for $Y_i, Y_{i+1}, \ldots, Y_j$ (for $i \leq j$).

[6]We use the standard notation $D(W\|Q|P)$ for the Kullback-Leibler distance between the joint distributions $P(\cdot)W(\cdot|\cdot)$ and $P(\cdot)Q(\cdot|\cdot)$ (see, e.g., [5, p. 31]).

Let $\{(\mathcal{C}_N, (\tau_N, \phi_N))\}_{N\geq 1}$ be a training-based scheme with preamble empirical distribution equal to P. By property ii., the stopping time τ_N is such that the event $\{\tau_N = n\}$ depends only on the realizations of $Y_{n-N+1}^{n-N+\eta N}$. For simplicity, instead of τ_N, we are going to consider the shifted stopping time $\tau'_N \triangleq \tau_N - (1-\eta)N$ whose decision to stop at a certain moment depends on immediate ηN previously observed symbols. Clearly, τ'_N can be written as

$$\tau'_N = \inf\{i \geq 1 : S_i = 1\},$$

where each S_i is some (decision) function defined over $Y_{i-\eta N+1}^{i}$ and that take on the values 0 or 1.

The condition iii. in terms of τ'_N becomes

$$\mathbb{P}(\tau'_N \geq k + N + \eta N - 1 | \tau'_N \geq k + \eta N, \nu = k) \geq \varepsilon \tag{3}$$

for all $k \in \{1, 2, \ldots, A\}$.

Let us define the events

$$\begin{aligned}
&\mathcal{E}_1 = \{\tau'_N \geq \nu + \eta N\}\\
&\mathcal{E}_2 = \{S_i = 0 \text{ for } i \in \{\nu + N + \eta N - 1, \ldots, 3A/4\}\}\\
&\mathcal{E}_3 = \{\tau'_N \geq \nu + N + \eta N - 1\}\\
&\mathcal{E}_4 = \{\nu \leq A/4\}\ .
\end{aligned}$$

We lower bound the reaction delay as

$$\mathbb{E}((\tau'_N - \nu)^+) \geq \mathbb{E}((\tau'_N - \nu)^+ | \mathcal{E}_1, \mathcal{E}_4)\mathbb{P}(\mathcal{E}_1, \mathcal{E}_4), \tag{4}$$

and consider the two terms on the right-side separately.

We first show that $\mathbb{E}((\tau'_N - \nu)^+ | \mathcal{E}_1, \mathcal{E}_4) = \Omega(A)$.[7] We have

$$\begin{aligned}
&\mathbb{E}((\tau'_N - \nu)^+ | \mathcal{E}_1, \mathcal{E}_4)\\
&\quad \geq \mathbb{E}((\tau'_N - \nu)^+ | \mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_4)\mathbb{P}(\mathcal{E}_2, \mathcal{E}_3 | \mathcal{E}_1, \mathcal{E}_4)\\
&\quad = \mathbb{E}((\tau'_N - \nu)^+ | \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_4)\mathbb{P}(\mathcal{E}_2, \mathcal{E}_3 | \mathcal{E}_1, \mathcal{E}_4)\\
&\quad = \mathbb{E}((\tau'_N - \nu)^+ | \tau'_N \geq 3A/4, \nu \leq A/4)\mathbb{P}(\mathcal{E}_2, \mathcal{E}_3 | \mathcal{E}_1, \mathcal{E}_4)\\
&\quad \geq \frac{A}{2}\mathbb{P}(\mathcal{E}_2, \mathcal{E}_3 | \mathcal{E}_1, \mathcal{E}_4)
\end{aligned} \tag{5}$$

where the first equality holds since $\mathcal{E}_3 \subset \mathcal{E}_1$, and where the second equality holds since $\mathcal{E}_2 \cap \mathcal{E}_3 = \{\tau'_N > 3A/4\}$. We now prove that $\mathbb{P}(\mathcal{E}_2 | \mathcal{E}_1, \mathcal{E}_4)$ and $\mathbb{P}(\mathcal{E}_3 | \mathcal{E}_1, \mathcal{E}_4)$ have large probabilities for large N. This implies that $\mathbb{P}(\mathcal{E}_2, \mathcal{E}_3 | \mathcal{E}_1, \mathcal{E}_4)$ has a probability bounded away from zero for N large enough. This together with (5) implies that $\mathbb{E}((\tau'_N - \nu)^+ | \mathcal{E}_1, \mathcal{E}_4) = \Omega(A)$ as claimed above.

[7] $\Omega(\cdot)$ refers to the standard Landau order notation.

For $\mathbb{P}(\mathcal{E}_2 | \mathcal{E}_1, \mathcal{E}_4)$ we have

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_2 | \mathcal{E}_1, \mathcal{E}_4) &= \mathbb{P}(\mathcal{E}_2 | \mathcal{E}_4)\\
&= \mathbb{P}(S_{\nu+N+\eta N-1}^{A/4} = 0 | \nu \leq A/4)\\
&= \frac{1}{A/4}\sum_{k=1}^{A/4} \mathbb{P}(S_{k+N+\eta N-1}^{A/4} = 0 | \nu = k)\\
&= \frac{1}{A/4}\sum_{k=1}^{A/4} \mathbb{P}_\star(S_{k+N+\eta N-1}^{A/4} = 0)\\
&\geq \frac{1}{A/4}\sum_{k=1}^{A/4} \mathbb{P}_\star(S_1^{A/4} = 0)\\
&= \mathbb{P}_\star(S_1^{A/4} = 0)\\
&= \mathbb{P}_\star(\tau'_N > 3A/4)
\end{aligned} \tag{6}$$

where $\mathbb{P}_\star$ denotes the output distribution under pure noise, i.e., when the Y_i's are i.i.d. according to $Q_\star$. For the first equality we used the independence between $\mathcal{E}_2$ and $\mathcal{E}_1$ conditioned on $\mathcal{E}_4$. For the fourth equality we noted that, conditioned on $\{\nu = k\}$, the event $S_{k+N+\eta N-1}^{3A/4}$ is independent of the sent codeword (prefix and information sequence), hence its probability is $\mathbb{P}_\star$.

Now, the event $\{\tau'_N > 3A/4\}$ only depends on the output symbols up to time $3A/4$. The probability of this event under $\mathbb{P}_\star$ is thus the same as under the probability distribution induced by the sending of a message *after* time $3A/4$. Therefore, since the probability of error vanishes for large N, and that a message starts being sent after time $3A/4$ with (large) probability $1/4$, we must have $\mathbb{P}_\star(\tau'_N > 3A/4) \approx 1$ for large N. Hence from (6) we have

$$\mathbb{P}(\mathcal{E}_2 | \mathcal{E}_1, \mathcal{E}_4) \approx 1 \tag{7}$$

for large N. Now consider $\mathbb{P}(\mathcal{E}_3 | \mathcal{E}_1, \mathcal{E}_4)$. Using (3), we have

$$\mathbb{P}(\mathcal{E}_3 | \mathcal{E}_1, \mathcal{E}_4) \geq \varepsilon. \tag{8}$$

From (7) and (8) we deduce that $\mathbb{P}(\mathcal{E}_2, \mathcal{E}_3 | \mathcal{E}_1, \mathcal{E}_4)$ is the (conditional) probability of the intersection of two large probability events. Therefore $\mathbb{P}(\mathcal{E}_2, \mathcal{E}_3 | \mathcal{E}_1, \mathcal{E}_4)$ has a probability bounded away from zero as $N \to \infty$. Hence, we have shown that

$$\mathbb{E}((\tau'_N - \nu)^+ | \mathcal{E}_1, \mathcal{E}_4) = \Omega(A) \tag{9}$$

as claimed earlier.

Second, we prove that

$$\mathbb{P}(\mathcal{E}_1, \mathcal{E}_4) = \Omega(e^{-\eta N D_1}\text{poly}(N)), \tag{10}$$

where $D_1 = D(W\|Q|P)$, P denotes the type of the preamble, and $\text{poly}(N)$ denotes a quantity that goes to 0 at most polynomially quickly as a function of N.

We expand $\mathbb{P}(\mathcal{E}_1, \mathcal{E}_4)$ as

$$\mathbb{P}(\mathcal{E}_1, \mathcal{E}_4) = \frac{1}{A}\sum_{k=1}^{A/4} \mathbb{P}_k(\tau'_N \geq k + \eta N), \tag{11}$$

where $\mathbb{P}_k$ represents the probability distribution of the output conditioned on the event $\{\nu = k\}$. Further, by picking a conditional distribution W defined over $\mathcal{X} \times \mathcal{Y}$ such that $\mathbb{P}_k(Y_k^{k+\eta N-1} \in \mathcal{T}_W^{\eta N}(P)) > 0$,[8] we lower the term in the above sum as

$$\begin{aligned}\mathbb{P}_k(\tau'_N \geq k + \eta N) \geq &\mathbb{P}_k(\tau'_N \geq k + \eta N | Y_k^{k+\eta N-1} \in \mathcal{T}_W^{\eta N}(P)) \\ &\times \mathbb{P}_k(Y_k^{k+\eta N-1} \in \mathcal{T}_W^{\eta N}(P)) \,.\end{aligned} \tag{12}$$

We lower bound each of the two terms on the right-side of (12).

For the first term, a change of measure argument reveals that

$$\begin{aligned}&\mathbb{P}_k(\tau'_N \geq k + \eta N | Y_k^{k+\eta N-1} \in \mathcal{T}_W^{\eta N}(P)) \\ &\quad = \mathbb{P}_\star(\tau'_N \geq k + \eta N | Y_k^{k+\eta N-1} \in \mathcal{T}_W^{\eta N}(P)) \,.\end{aligned} \tag{13}$$

To see this, one expands

$$\mathbb{P}_k(\tau'_N \geq k + \eta N | Y_k^{i+\eta N-1} \in \mathcal{T}_W^{\eta N}(P))$$

by further conditioning on individual sequences in $\mathcal{T}_W^{\eta N}(P)$. Then, one uses the fact that, conditioned on a particular such sequence, the channel outputs outside the time window $\{k, k+1, \ldots, k+\eta N - 1\}$ are distributed according to noise, i.e., i.i.d. according to $Q_\star$.

For the second term we have

$$\mathbb{P}_k(Y_k^{k+\eta N-1} \in \mathcal{T}_W^{\eta N}(P)) \geq \text{poly}(N) e^{-\eta N D_1} \tag{14}$$

using [5, Lemma 2.6, p. 32], where $D_1 \triangleq D(W\|Q|P)$. Combining (11), (12), (13), and (14) we get

$$\begin{aligned}&\mathbb{P}(\mathcal{E}_1, \mathcal{E}_4) \\ &\geq \text{poly}(N)\frac{e^{-\eta N D_1}}{A} \times \\ &\quad \times \sum_{k=1}^{A/4} \mathbb{P}_\star(\tau'_N \geq i + \eta N | Y_k^{k+\eta N-1} \in \mathcal{T}_W^{\eta N}(P)) \\ &\geq \text{poly}(N)\frac{e^{-\eta N(D_1 - D_2)}}{A} \times \\ &\quad \times \sum_{k=1}^{A/4} \mathbb{P}_\star(\tau'_N \geq i + \eta N, Y_k^{i+\eta N-1} \in \mathcal{T}_W^{\eta N}(P)) \,,\end{aligned} \tag{15}$$

where $D_2 \triangleq D(W\|Q_\star|P)$, and where for the second inequality we again used [5, Lemma 2.6, p. 32].

[8] The set $\mathcal{T}_W^{\eta N}(P)$ corresponds to all output sequences $y^{\eta N}$ that, together with the preamble, have joint type equal to $P(\cdot)W(\cdot|\cdot)$.

Now, assuming that $\alpha > \eta D_2$, one can show that

$$\sum_{k=1}^{A/4} \mathbb{P}_\star(\tau'_N \geq k + \eta N, Y_k^{k+\eta N-1} \in \mathcal{T}_W^{\eta N}(P)) = \Omega(A e^{-\eta D_2})$$

using the union bound. Therefore, under the above assumption we get from (15) the desired claim that

$$\mathbb{P}(\mathcal{E}_1, \mathcal{E}_4) = \Omega(e^{-\eta N D_1}\text{poly}(N)) \,. \tag{16}$$

From (4), (9), and (16), we conclude that if $\alpha > \eta D_2$ then

$$\mathbb{E}((\tau'_N - \nu)^+) \geq \Omega(A e^{-\eta N D_1}\text{poly}(N)) \,.$$

Therefore, letting $A = e^{N\alpha}$, we deduce that, if, in addition to the inequality $\alpha > \eta D_2$, we also have $\alpha > \eta D_1$, the average reaction delay $\mathbb{E}((\tau'_N - \nu)^+)$ grows exponentially with N. ■

Concluding Remarks

Synchronization and information transmission of virtually all practical communication systems are performed separately, on the basis of different communication bits. Moreover, in general, the rate of these strategies is computed with respect to the information transmission time period, ignoring the delay overhead caused by various hand-shake protocols used to guarantee synchronization. In these cases, the notions of 'high rate' or 'capacity-achieving' communication strategies clearly raises questions.

Building on an extension of Shannon's original point-to-point synchronous communication channel model to assess the overall rate performance of asynchronous communication systems, we showed that training-based schemes perform suboptimally at high rates. In this regime, it is necessary to envision communication strategies that integrate synchronization into information transmission.

Acknowledgments

We thank the reviewer for valuable comments.

References

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

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

[3] T. Cover and J. Thomas, *Elements of information theory*. New York: Wiley, 2006.

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

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