2011 IEEE International Symposium on Information Theory Proceedings

Error Exponents in Asynchronous Communication

Da Wang
EECS Dept., MIT
Cambridge, MA, USA
Email: dawang@mit.edu

Venkat Chandar
Lincoln Laboratory, MIT
Lexington, MA, USA
Email: vchandar@mit.edu

Sae-Young Chung
Dept. of EE, KAIST
Daejeon, Korea
Email: sychung@ee.kaist.ac.kr

Gregory Wornell
EECS Dept., MIT
Cambridge, MA, USA
Email: gww@mit.edu

***Abstract*—Based on recent work on asynchronous communication, this paper proposes a slotted asynchronous channel model and investigates the fundamental limits of asynchronous communication, in terms of *miss* and *false alarm* error exponents. We propose coding schemes that are suitable for various asynchronous communication scenarios, and quantify more precisely the suboptimality of training-based schemes, i.e., communication strategies that separate synchronization from information transmission. In particular, we show that under a broad set of conditions, training-based schemes are suboptimal at all positive rates. Finally, we demonstrate these performance differences by specializing our results to BSCs and AWGN channels.**

***Index Terms*—synchronization, error exponents, training-based schemes**

I. Introduction

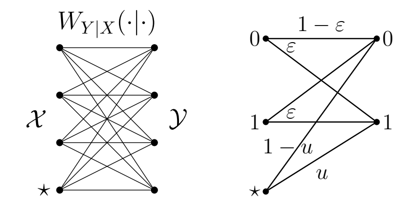
Communication is inherently asynchronous because we need to detect the presence of a codeword correctly before decoding it to the correct message. Traditionally, this asynchronism is handled by separating communication into two sub-problems, *synchronization* and *coding*, where in synchronization, a specific pattern of symbols are used to identify the start of transmitted data/codeword. Therefore, performance improvements for synchronization are in general attained by using better synchronization patterns and/or detection rules (e.g., [1], [2]).

Recently, inspired by emerging applications such as sensor networks, [3] proposes a new framework for asynchronous communication. It extends the classical coding problem to incorporate the requirement of detection and considers synchronization and coding jointly. This implies a question on the "distinguishability" of a channel code, which is the "difference" between channel outputs induced by noise and by codewords in the channel code. [3] investigates this problem from the perspective of minimizing *false alarms*, which is the error of confusing noise as codewords. In this paper, we also introduce *miss* error, a type of error that considers a codeword as noise. In addition, we simplify the asynchronous channel model in [3] by imposing a slotted constraint. This leads to sharper results than [4] on the suboptimality of training-based schemes, and uncovers insights on codes that are more distinguishable and useful for asynchronous communication.

A. Problem Formulation

We consider discrete-time communication over a discrete memoryless channel (DMC), and use the asynchronous channel model proposed in [3], which captures the channel condition when the transmitter is silent.

* This work was supported in part by NSF under Grant No. CCF-1017772, by HP Labs and by MKE under Grant No. NIPA-2011-(C1090-1111-0005).



(a) General model (b) BSC

Fig. 1. (a): an asynchronous DMC with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, "silent" symbol $\star$ and transition probabilities $W_{Y|X}(\cdot|\cdot)$. (b): an asynchronous BSC with crossover probability ε and $\star$ output distribution $\mathsf{Bernoulli}(u)$.

Definition 1 (Asynchronous discrete memoryless channel [3])**.** *An asynchronous discrete memoryless channel (Fig. 1(a)) $(\mathcal{X}, \star, \mathcal{Y}, W)$ is a DMC with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and transition probabilities $W(y|x)$, where the special symbol $\star \in \mathcal{X}$ represents the channel input when the transmitter is silent.*

For example, an asynchronous binary symmetric channel (BSC) has $W(\cdot|\star) = \mathsf{Bernoulli}(u)$ (Fig. 1(b)). Another example is the asynchronous discrete-time additive white Gaussian noise (AWGN) channel, which has average signal power P, average noise power 1, and $W(\cdot|\star) = \mathsf{N}(0,1)$. In other words, when the transmitter is silent, the channel output distribution is the standard Gaussian distribution.

Furthermore, this paper assumes communication is slotted (Fig. 2). In this case, channel outputs in each time slot are induced by either a codeword $c^n(i)$ or the noise sequence $\star^n$.

For this channel, a length n block code with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and some finite message set $\mathcal{M}_{f_n} = \{1, 2, \ldots, |\mathcal{M}_{f_n}|\}$ is composed of a pair of mappings, encoder mapping $f_n : \mathcal{M}_{f_n} \to \mathcal{X}^n$ and decoder mapping $g_n : \mathcal{A}_n \to \mathcal{M}_{f_n}$, where $\mathcal{A}_n \subset \mathcal{Y}^n$. Given a message m, which is chosen from $\mathcal{M}_{f_n}$ *uniformly*, the encoder maps it to a sequence $x^n(m) \in \mathcal{X}^n$ and transmits this sequence through the channel, where we call $x^n(m)$ the *codeword* for message m and the entire set of codewords $\{x^n(m), m \in \mathcal{M}_{f_n}\}$ a *codebook*. The receiver receives a sequence $y^n \in \mathcal{Y}^n$, where

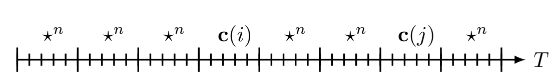


Fig. 2. Slotted channel, with each time slot containing either a codeword or a noise sequence with length n.

978-1-4577-0595-3/11/$26.00 ©2011 IEEE

$W^n(y^n|x^n(m)) \triangleq \prod_{i=1}^n W(y_i|x_i(m))$. If $y^n \in \mathcal{A}_n$, we consider the channel input to be a certain codeword $x^n(m), m \in \mathcal{M}_{f_n}$, otherwise we consider the channel input as $\star^n$. Namely, $\mathcal{A}_n$ is the acceptance region for codewords. In addition, we define $\mathcal{B}_n \triangleq \mathcal{A}_n^c$ to be the rejection region for codewords and denote (f_n, g_n) by $\mathcal{C}^{(n)}$.

The performance of a codebook over a channel W can be characterized by three types of error: 1) *miss*, where we detect a codeword as noise; 2) *false alarm*, where we detect a noise sequence as a codeword; 3) *decoding error*, where after correctly detecting the presence of a codeword, we decode it to an incorrect message. Formally, we define

$$P_{\mathrm{m}}\left(\mathcal{C}^{(n)}\right) \triangleq \max_m P_{\mathrm{m}}(m) \triangleq \max_m W^n\left(\{\mathcal{A}_n\}^c \,|\, x^n(m)\right),$$
$$P_{\mathrm{f}}\left(\mathcal{C}^{(n)}\right) \triangleq W^n\left(\mathcal{A}_n \,|\, \star^n\right),$$
$$P_{\mathrm{d}}\left(\mathcal{C}^{(n)}\right) \triangleq \max_m P_{\mathrm{d}}(m) \triangleq \max_m \sum_{\hat{m} \neq m} W^n\left(g_n^{-1}(\hat{m}) \middle| f_n(m)\right).$$

In addition, we define the rate of a code $\mathcal{C}^{(n)}$ as $R(\mathcal{C}^{(n)}) \triangleq \log|\mathcal{M}_{f_n}|/n$. For a sequence of codebooks $\mathcal{Q} = \{\mathcal{C}^{(n)}, n \in \mathbb{Z}_+\}$, we define its rate as $R_{\mathcal{Q}} = \liminf_{n\to\infty} R(\mathcal{C}^{(n)})$. When the codebook sequence is clear from context, we denote $P_{\mathrm{d}}\left(\mathcal{C}^{(n)}\right)$, $P_{\mathrm{m}}\left(\mathcal{C}^{(n)}\right)$ and $P_{\mathrm{f}}\left(\mathcal{C}^{(n)}\right)$ by $P_{\mathrm{d}}^{(n)}$, $P_{\mathrm{m}}^{(n)}$ and $P_{\mathrm{f}}^{(n)}$, and use R instead of $R_{\mathcal{Q}}$.

Without loss of generality, we assume that for every $y \in \mathcal{Y}$, there exists $x \in \mathcal{X}$ such that $W(x|y) > 0$. Furthermore, we only consider the most interesting case that the support of $W(\cdot|\star)$ is $\mathcal{Y}$. For simplicity, we denote $Q_\star(\cdot) \triangleq W(\cdot|\star)$.

Definition 2 (Miss, false alarm, and decoding error exponents)**.** *Given an asynchronous DMC $(\mathcal{X}, \star, \mathcal{Y}, W)$ and a codebook sequence $\mathcal{Q} = \{\mathcal{C}^{(n)}, n \in \mathbb{Z}_+\}$ with rate R and all three error probabilities vanishing asymptotically, define its* miss error exponent *as* $e_{\mathrm{m}}(\mathcal{Q}) \triangleq \liminf_{n\to\infty} -\frac{1}{n}\log P_{\mathrm{m}}^{(n)}$. *Similarly, we define its* false alarm error exponent e_{f} *and* decoding error exponent e_{d} *in terms of* $P_{\mathrm{f}}^{(n)}$ *and* $P_{\mathrm{d}}^{(n)}$.

A triplet of numbers $(e_{\mathrm{m}}(\mathcal{Q}), e_{\mathrm{f}}(\mathcal{Q}), e_{\mathrm{d}}(\mathcal{Q}))$ is called achievable *if they can be achieved simultaneously, and is denoted by $\mathcal{E}(\mathcal{Q}, R)$. In addition, we let $\mathcal{E}(R)$ be the closure of the region $\{\mathcal{E}(\mathcal{Q}, R) : R_{\mathcal{Q}} \geq R\}$.*

The most general problem—characterization of the achievable error exponent region $\mathcal{E}(R)$—is open and this paper focuses on the false alarm and miss errors by setting $e_{\mathrm{d}} = 0$. In particular, this paper first investigates the reliability functions for false alarm and miss errors, which are important for various communication scenarios.

Definition 3 (Reliability functions)**.** *For an asynchronous DMC $(\mathcal{X}, \star, \mathcal{Y}, W)$, given a rate R, we define the* false alarm reliability function *and the* miss reliability function *as*

$$E_{\mathrm{f}}(R) \triangleq \sup_{(e_{\mathrm{m}}=0, e_{\mathrm{f}}, e_{\mathrm{d}}=0)\in\mathcal{E}(R)} e_{\mathrm{f}}$$
$$E_{\mathrm{m}}(R) \triangleq \sup_{(e_{\mathrm{m}}, e_{\mathrm{f}}=0, e_{\mathrm{d}}=0)\in\mathcal{E}(R)} e_{\mathrm{m}}.$$

These two reliability functions are characterized in Section II and Section III respectively. Then the tradeoff between miss and false alarm exponents is investigated in Section IV. With these characterizations, we compare the detection performance of training-based schemes to the optimal performance in Section V. This problem has been investigated in [4], where training is shown to be suboptimal at high rate. In our work, we show training is suboptimal at almost all rates and give a more precise quantification on its performance loss.

B. Notation

Most notations in this paper follow [5]. Specifically, a *constant composition code* is a code where all its codewords have the same type (empirical distribution), and for distributions $P(\cdot), Q(\cdot) \in \mathcal{P}(\mathcal{X})$ and conditional distributions $W(\cdot|\cdot) : \mathcal{X} \to \mathcal{Y}$, $V(\cdot|\cdot) : \mathcal{X} \to \mathcal{Y}$, define

$$[P \cdot W](x, y) \triangleq W(y|x) P(x)$$
$$[P \cdot W]_Y(y) \triangleq \sum_{x\in\mathcal{X}} W(y|x) P(x)$$
$$I(P, W) \triangleq \sum_{x\in\mathcal{X}, y\in\mathcal{Y}} P(x)W(y|x)\log\frac{W(y|x)}{\sum_{x\in\mathcal{X}} W(y|x)P(x)}.$$
$$D(V \| W|P) \triangleq \mathbb{E}_P\left[D\left(V(\cdot|P) \| W(\cdot|P)\right)\right] = \sum_{x\in\mathcal{X}} P(x) D\left(V(\cdot|x) \| W(\cdot|x)\right)$$

where $D(V \| W|P)$ is the expectation of the conditional information divergence between $V(\cdot|\cdot)$ and $W(\cdot|\cdot)$ under $P(\cdot)$.

II. False alarm reliability function

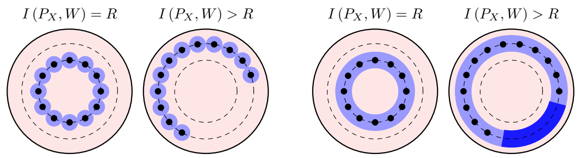
This section provides a complete characterization of the false alarm reliability function (Theorem 1). We show that an i.i.d. codebook is sufficient to achieve optimal performance, and different codebook designs have different implications for decoding procedures.

Theorem 1 (False alarm reliability function)**.** *An asynchronous DMC $(\mathcal{X}, \star, \mathcal{Y}, W)$ has false alarm reliability function*

$$E_{\mathrm{f}}(R) = \max_{P_X : I(P_X, W) \geq R} D(P_Y \| Q_\star) + I(P_X, W) - R \quad (1)$$
$$= \max_{P_X : I(P_X, W) = R} D(P_Y \| Q_\star), \quad (2)$$

where $P_Y(\cdot) = [P_X \cdot W]_Y$.

We omit the proof (including the converse, cf. [6] for details) and present two strategies to achieve the above reliability function. The first strategy corresponds to (1), which indicates a more flexible codebook design. In this case, an i.i.d. codebook with *any* distribution P_X such that $I(P_X, W) \geq R$ can be used. However, this flexibility requires a typicality decoder, which declares a message m if there exist only one m such that $(x^n(m), y^n) \in \mathcal{T}^n_{[P_{XY}]_\delta}$, or noise sequence $\star^n$ if there is no such m, otherwise decoding error. The second coding strategy corresponds to (2) and imposes a stronger constraint on the input distribution—an i.i.d. codebook with distribution



(a) Typicality decoding (b) Detection based on type

Fig. 3. The geometry of acceptance regions for codewords (shaded regions) shows that detection based on typicality is more flexible than detection based on type. Each concentric circle corresponds to a different input distribution and a larger circle corresponds to the input distribution with higher $I(P_X, W)$. The darken portion in Fig. 3(b) is the unnecessary part that leads to suboptimal performance when $I(P_X, W) > R$.

P_X such that $I(P_X, W) = R$ is required. This allows a *two-stage* decoding strategy. In the first stage, we detect based on the type of the channel output and the receiver simply conducts a binary hypothesis test between distributions P_Y and $Q_\star$. In the second stage, the decoder follows the regular channel code decoding procedure if the test result is P_Y, or declare a noise sequence $\star^n$ otherwise. When no codeword is sent, this detection process is conceptually simpler, because it only conducts one hypothesis testing, while in principle, the one-stage decoder needs to check every codeword in the codebook.

Fig. 3 illustrates the difference between these two strategies. To maximize the false alarm error exponent, we can employ a regular channel code and set the acceptance region $\mathcal{A}_n$ just large enough to keep P_{d} and P_{m} small, which approximately corresponds to the union of typical shells of codewords. Therefore, typicality decoding achieves optimal performance as long as the channel code is reliable and satisfies our rate requirement (Fig. 3(a)). By contrast, the type detection strategy based on (2) does not take the detailed codebook structure into account, and hence requires a stricter codebook design. If the codebook is generated by a distribution P_X such that $I(P_X, W) > R$, then it would not be optimal as $\mathcal{A}_n$ is set to larger than necessary, as illustrated in Fig. 3(b).

Finally, we note that (1) and (2) are equivalent because expression in (1) is linear and hence convex in P_X and the set $\{P_X : I(P_X, W) \geq R\}$ is compact.

Examples of the false alarm reliability functions for BSC and AWGN channel are shown later in Fig. 6 of Section V.

III. Miss reliability function

This section provides the lower and upper bounds of the miss reliability function (Theorem 2) and shows a constant composition codebook with type decoding can achieve the lower bound.

Theorem 2 (Miss reliability function)**.** *The miss reliability function of an asynchronous DMC $(\mathcal{X}, \star, \mathcal{Y}, W)$ satisfies*

$$\underline{E_{\mathrm{m}}}(R) \leq E_{\mathrm{m}}(R) \leq \overline{E_{\mathrm{m}}}(R),$$

where we define $Q_V \triangleq [P_X \cdot V]_Y$, $P_Y(\cdot) \triangleq [P_X \cdot W]_Y$, and

$$\underline{E_{\mathrm{m}}}(R) = \max_{P_X : I(P_X, W) \geq R} \min_{V : Q_V = Q_\star} D(V \,\|\, W | P_X),$$

$$\overline{E_{\mathrm{m}}}(R) = \max_{P_X : I(P_X, W) \geq R} D(Q_\star \,\|\, W | P_X),$$

The proof for lower bound is included in [6] and the upper bound is based on an upper bound result for single-message unequal error protection in [7]. Below we provide a sketch on the coding strategy to achieve the lower bound as well as some intuition on the upper bound.

To achieve $\underline{E_{\mathrm{m}}}(R)$, we use a constant composition codebook with type P_X such that $I(P_X, W) \geq R$. Note that an i.i.d. codebook is suboptimal here because atypical codewords produced during the i.i.d. codebook generation process would be harmful for the miss error exponent. Then to ensure P_{f} is small, the typicality shell of the noise sequence $\star^n$, which has type $Q_\star$, should be roughly included in the rejection region $\mathcal{B}_n$. Therefore, if a channel realization V makes the output type Q_V "similar" to $Q_\star$, a miss error occurs. Based on this, we partition the channel realization V by the divergence between Q_V and $Q_\star$, and define

$\mathcal{V}_A \triangleq \{V : D(Q_V \,\|\, Q_\star) > \lambda_n\}$, where $\lambda_n \to 0$ as $n \to \infty$. Then we can assign the following acceptance and rejection region, $\mathcal{A}_n = \bigcup_i \bigcup_{V \in \mathcal{V}_A} \mathcal{T}_V^n(x^n(i))$ and $\mathcal{B}_n = \{\mathcal{A}_n\}^c$, which are shown to achieve $\underline{E_{\mathrm{m}}}(R)$.

The upper bound $\overline{E_{\mathrm{m}}}(R)$ can easily be derived by interchanging the noise and codebook in Theorem 1. Pick a codeword with type P_X such that $I(P_X, W) \geq R$ to be the noise, and consider a codebook consisting of the single codeword $\star^n$. The false alarm reliability function for the new problem is an upper bound on $E_{\mathrm{m}}(R)$, but we must modify (2) to handle noise sequences with an arbitrary type. This requires that we average over the exponent associated with each symbol of the noise sequence, so (2) becomes $D(Q_\star \,\|\, W | P_X)$.

IV. Tradeoffs between false alarm and miss error exponents

In addition to maximizing either false alarm or miss error exponent, sometimes we may require positive error exponents for both false alarm and miss. In this case, we are interested in the tradeoff between these two exponents, or the "capacity region" of the $(e_{\mathrm{m}}, e_{\mathrm{f}})$ pair at a given rate R. Characterizing this tradeoff is more involved than characterizing the reliability functions and only achievability results on DMC and AWGN channels are presented (cf. [6] for achievability proofs, a multi-letter outer bound for the error exponents capacity region of DMC and a single-letter outer bound for BSC).

A. DMC

Theorem 3 (Achievability via constant composition codebook)**.** *For an asynchronous DMC $(\mathcal{X}, \star, \mathcal{Y}, W)$, given a rate R and a miss error exponent constraint e_{m}, the following lower bound of the false alarm reliability function is achievable via a sequence of constant composition codebooks*

$$\underline{E_{\mathrm{f}}}(R, e_{\mathrm{m}}) = \max_{P_X : I(P_X, W) \geq R} \min_{V : D(V \| W | P_X) \leq e_{\mathrm{m}}} \left[D(Q_V \,\|\, Q_\star) + |I(P_X, V) - R|^+ \right],$$

where Q_V is defined in Theorem 2.

Proof for this theorem is in [6] and here we provide some intuition of the achievability strategy.

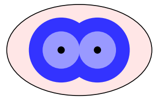


Fig. 4. Acceptance region $\mathcal{A}_n$ (shaded region) for two codewords. The region with lighter shade indicates the typical shell of the codeword, and the darker ring-like region indicates additional acceptance region needed to satisfy the miss error exponent requirement.

Given a miss error exponent requirement e_{m}, we need to make the acceptance region $\mathcal{A}_n$ large enough to make the miss error exponentially small. Meanwhile, we only need to keep the decoding error probability small (but not exponentially small). This scenario is illustrated for two codewords in Fig. 4. Intuitively, if $P_{\mathrm{m}}^{(n)} \dot{\leq} e^{-ne_{\mathrm{m}}}$, for any V such that $D\left(V \,\|\, W|P_X\right) \leq e_{\mathrm{m}}$, we have $\min_m V^n\left(\mathcal{A}_n | x^n(m)\right) \geq 1-\varepsilon$. In this way, we can represent $\mathcal{A}_n$ as the union of all possible typical V-shell of $x^n(m)$ and derive the achievable false alarm error exponent, which eventually leads to Theorem 3 (cf. [6] for details). Note that when $e_{\mathrm{m}} = 0$, $V = W$. In this case, Theorem 3 reduces to Theorem 1.

B. AWGN channel

Theorem 4 (Achievability for the AWGN channel)**.** *Given a rate R, for an asynchronous AWGN channel with average power constraint P, the following error exponent pairs $(e_{\mathrm{f}}(\eta), e_{\mathrm{m}}(\eta))$ are achievable:*

$$e_{\mathrm{f}}(\eta) \leq \max_{(a,b)\in[0,1]^2} \min_{0\leq r\leq \eta-b} \left[\frac{r^2}{2a^2} + I_{\chi_1^2}\left(\frac{(\eta-r)^2}{b^2}\right)\right]$$

$$e_{\mathrm{m}}(\eta) \leq \max_{(a,b)\in[0,1]^2} \min_{\eta-b\sqrt{P_c+1}\leq r\leq\eta} \left[\frac{(r-a\sqrt{P_s})^2}{2a^2} + I_S\left(P_c, \frac{(\eta-r)^2}{b^2}\right)\right],$$

where P_c and P_s satisfy $R = \log\left(1+P_c\right)/2$, $P_s = P - P_c$, and $b < \eta < a\sqrt{P_s} + b\sqrt{P_c+1}$,

$$I_{\chi_1^2}(x) \triangleq \frac{1}{2}(x - \ln x - 1)$$

$$I_S(P,\eta) \triangleq \frac{1}{2}\left(P + \eta - \sqrt{1+4P\eta} - \log\left[\frac{\sqrt{1+4P\eta}-1}{2P}\right]\right).$$

The above error exponents can be achieved via the following codebook and heuristic decoding rule.

Given a rate R, we generate a codebook as follows: choose e^{nR} points uniformly from the surface of a $(n-1)$-dimensional sphere with radius $\sqrt{nP_c}$, and let these points be $\hat{X}^{n-1}(1), \hat{X}^{n-1}(2), \cdots$. Then let $X^n(i) = \left(\sqrt{nP_s}, \hat{X}_1(i), \cdots, \hat{X}_{n-1}(i)\right), i = 1, 2, \cdots, e^{nR}$, and use $\left\{X^n(i),\, i = 1, 2, \cdots, e^{nR}\right\}$ as a codebook, which is named "clustered spherical codebook" due to its geometric structure, as shown in Fig. 5. Note that here we use nP_c amount of power for communication, which is just enough to support reliable communication at rate R, and allocate the rest of the power nP_s for synchronization.

Inspired by high dimensional geometry, we develop a heuristic detection rule that allows asymptotic performance

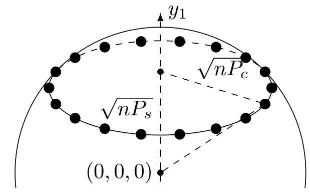


Fig. 5. A clustered spherical codebook ($n = 3$). All the codewords are clustered into a $(n-1)$-sphere (circle when $n = 3$) on a n-sphere.

analysis:

$$\mathcal{A}_n = \left\{y^n : ay_1 + b\|y_2^n\| \geq \sqrt{n}\eta\right\} \qquad \text{(declare a codeword)}$$

$$\mathcal{B}_n = \left\{y^n : ay_1 + b\|y_2^n\| < \sqrt{n}\eta\right\}, \qquad \text{(declare noise)}$$

where $a \in [0,1]$ and $b \in [0,1]$ are weights to be selected. Intuitively, when a codeword is transmitted, y_1 should be large as $X^n(0) = \sqrt{nP_s}$, and $\|y_2^n\|$ should also be large as all codewords reside on the $(n-1)$-sphere. We take a linear combination of these two to take both factors into account, and by optimizing a and b at each rate R, good performance can be obtained. Theorem 4 then follows the analysis in [6].

V. Training-based schemes is suboptimal almost everywhere

Under the unslotted model, [4] defines training-based schemes precisely and shows that training-based schemes achieve vanishing false alarm error exponent at capacity except for degenerate cases. Using the slotted model, this paper simplifies the definition of training-based schemes and is able to quantify the suboptimality of these schemes more precisely *at any rate* $R \in [0, C)$.

The definition of training-based schemes under the slotted model is straightforward. To transmit nR bits of information in n channel uses, the best training-based scheme uses a capacity-achieving code with block length nR/C for information transmission, and the rest $k = (1 - R/C)n$ symbols for synchronization (Fig. 6). In addition to this code design constraint, training also limits the detection algorithm to operate on the k synchronization symbols only.

For the case of maximizing false alarm error exponents, it is not difficult to see that the best synchronization word is $s^* s^* \cdots s^*$, where $s^* = \arg\max_{s\in\mathcal{X}} D\left(W\left(\cdot\,|\,s\right) \| Q_\star\right)$. Then standard large deviation arguments show that for training-based schemes, the maximum achievable false alarm error exponent is $E_{\mathrm{t}}(R) = (1 - R/C)\, D\left(W\left(\cdot\,|\,s^*\right) \| Q_\star\right)$. Therefore, $E_{\mathrm{f}}(R) - E_{\mathrm{t}}(R)$ is the gap between the maximum false alarm error exponent attained by training-based schemes and by the optimal scheme. Furthermore, Theorem 5 shows this gap is strictly positive under a broad set of conditions.

Theorem 5 (The suboptimality of training-based scheme)**.** *For an asynchronous DMC $(\mathcal{X}, \star, \mathcal{Y}, W)$ and $0 < R \leq C$, in*

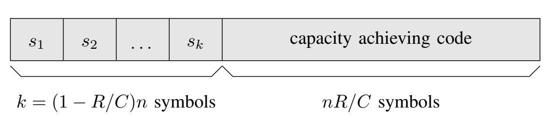
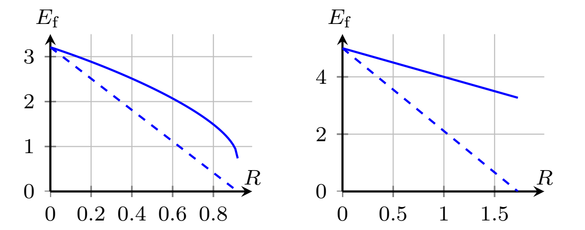


Fig. 6. Training-based scheme, where the first $(1-R/C)n$ symbols are used for synchronization, and the next nR/C symbols are coded as a capacity-achieving code for information transmission.



(a) BSC with $\varepsilon = 0.01, u = 0.1$ (b) AWGN channel with SNR=10

Fig. 7. Performance gaps (in terms of false alarm error exponent) between training-based scheme and joint synchronization and coding scheme on BSC and AWGN channel. The gap is larger at higher rates.

general, $E_t(0) = E_f(0)$ *and* $E_t(R) \le E_f(R)$*. Furthermore, if the capacity achieving output distribution* P_Y^* *satisfies* $D\left(P_Y^* \| Q_\star\right) > 0$*, then for all* $R > 0$*,* $E_t(R) < E_f(R)$*.*

This theorem is based on the fact that $E_f(R)$ is concave [6].

In Fig. 7, we demonstrate that for both BSC and AWGN channels, the performance gap is more significant in the high rate regime, because here training-based approaches uses most of the degrees of freedom for information transmission, leaving few for synchronization, resulting poor performance.

For the case of positive error exponents for both miss and false alarms, performance analysis becomes more complicated, because converses for both training and joint synchronization and coding schemes are unknown. However, for two special yet important asynchronous channels, BSC with $u = 0.5$ and AWGN channel, we can find the best training schemes, and show that there exist joint synchronization and coding schemes that achieve better tradeoffs, hence demonstrating the suboptimality of training.

BSC with $u = 0.5$**:** For the asynchronous BSC with $u = 0.5$, it is not difficult to see that by symmetry, the synchronization sequence with identical symbols attains the best performance. Hence by standard large deviation arguments, the optimal tradeoff between false alarm and miss error exponents satisfies $e_m \le (1 - R/C)\, D\left(q_\lambda \| \varepsilon\right)$ and $e_f \le (1 - R/C)\, D\left(q_\lambda \| u\right)$, where $q_\lambda = \varepsilon^\lambda u^{1-\lambda}/(\varepsilon^\lambda u^{1-\lambda} + (1-\varepsilon)^\lambda (1-u)^{1-\lambda})$, $0 \le \lambda \le 1$.

Specializing the results in Section IV-A, given $\delta \in (u, s)$, where $s = (1-\varepsilon)p + \varepsilon(1-p)$, we can achieve any $(e_f(\delta), e_m(\delta))$ such that $e_f(\delta) \le D\left(\delta \| u\right)$ and $e_m(\delta) \le \min_{\kappa \in [\delta - \bar{p}\varepsilon, \kappa^*]} \left[\bar{p} D\left((\delta - \kappa)/\bar{p} \,\|\, \varepsilon\right) + p D\left(\kappa/p \| \bar{\varepsilon}\right)\right]$, where $\bar{x} \triangleq 1 - x$ and $\kappa^* = \min\{\delta, p(1-\varepsilon)\}$.

We compare the performance of constant composition codebook and training in Fig. 8, and show that the former achieves a much better tradeoff than the latter, especially when we have a strong requirement for e_m.

AWGN channel: Unlike the DMC, where the allocation between synchronization and communication channel uses matters, for the AWGN channel, it is the allocation between synchronization power and communication power that matters. Therefore, the best training scheme has the same codebook structure as the clustered spherical codebook in Section IV-B, but the detection is based on the synchronization power $\sqrt{nP_s}$ only. It can be shown that we can achieve any $(e_f(\eta), e_m(\eta))$

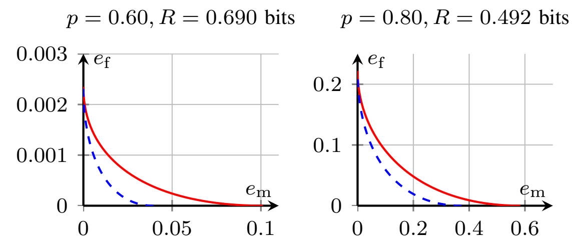


Fig. 8. Performance comparison between constant composition codebook (solid line) and training (dashed line) for a BSC with $\varepsilon = 0.05$ and $u = 0.5$.

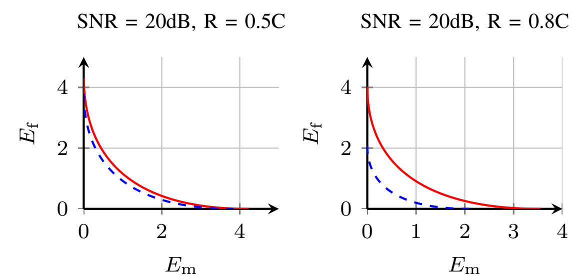


Fig. 9. Clustered spherical codebook with heuristic detection (dashed line) is better than training (solid line) for AWGN channels, especially at high rate and/or low SNR.

such that $e_m(\eta) \le (\sqrt{P_s} - \eta)^2/2$ and $e_f(\eta) \le \eta^2/2$. Then applying results in Section IV-B, we get the performance comparisons in Fig. 9. At low rates, the clustered spherical codebook and training perform almost equally well, but at high rates, the clustered spherical codebook achieves a much better e_m–e_f tradeoff.

These examples on BSC and AWGN channel demonstrate that with better codebook designs and detection strategies, joint synchronization and coding can achieve significant performance improvements over training-based schemes, especially at high rates or when we have strict requirements on the miss error probability. On the other hand, if we communicate at low rate, we may use training without much performance penalty, gaining the benefit of faster detection.

REFERENCES

[1] R. Scholtz, "Frame synchronization techniques," *IEEE Trans. Commun.*, vol. 28, no. 8, pp. 1204–1213, 1980.
[2] J. Massey, "Optimum frame synchronization," *IEEE Trans. Commun.*, vol. 20, no. 2, pp. 115–119, 1972.
[3] A. Tchamkerten, V. Chandar, and G. W. Wornell, "Communication under strong asynchronism," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4508–4528, 2009.
[4] V. Chandar, A. Tchamkerten, and G. W. Wornell, "Training-based schemes are suboptimal for high rate asynchronous communication," in *Proc. IEEE Inf. Th. Workshop (ITW)*, Taormina, Italy, 2009, pp. 389–393.
[5] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. New York: Academic Press, 1981.
[6] D. Wang, "Distinguishing codes from noise : fundamental limits and applications to sparse communication," Master thesis, Massachusetts Institute of Technology, Dept. of EECS, 2010.
[7] S. Borade, Barış Nakiboğlu, and L. Zheng, "Unequal error protection: An Information-Theoretic perspective," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5511–5539, 2009.