Energy and Sampling Constrained Asynchronous Communication

Aslan Tchamkerten, Venkat Chandar, and Giuseppe Caire

***Abstract*—The minimum energy, and, more generally, the minimum input cost, to transmit one bit of information has been recently derived for bursty communication when information is available infrequently at random times at the transmitter. This result assumes that the receiver can sample at no cost all channel outputs.**

Suppose now there is a cost associated to output sampling and that the receiver is constrained to observe only a fraction $\rho \in (0,1]$ of all channel outputs. What is the input cost penalty due to sparse output sampling?

Remarkably, there is no penalty: *regardless* of $\rho > 0$ the asynchronous capacity per unit cost is the same as under full sampling, *i.e.*, when $\rho = 1$. Moreover, there is no penalty in terms of decoding delay with respect to full sampling. This latter result relies on the possibility to sample adaptively; the next sample is a function of past samples. When sampling is non-adaptive it is possible to achieve the full sampling asynchronous capacity per unit cost, but the decoding delay gets multiplied by $1/\rho$. Therefore adaptive sampling strategies are of particular interest in the very sparse sampling regime.

***Index Terms*—Asynchronous communication; bursty communication; capacity per unit cost; energy; sequential decoding; sensor networks; sparse communication; sparse sampling; synchronization**

I. Introduction

In many emerging technologies, communication is sparse and asynchronous, but it is essential that when data is available, it is delivered to the destination as timely and reliably as possible. Examples are sensor networks monitoring rare but critical events, such as earthquakes or forest fires. For such settings, [1] characterized the asynchronous capacity per unit cost based on the following model. There are B bits of information that are made available to the transmitter at some random time ν, and need to be communicated to the receiver. The B bits are coded and transmitted over a memoryless channel using a sequence of symbols that have costs associated with them. The rate $\boldsymbol{R}$ per unit cost is B divided by the cost of the transmitted sequence. Asynchronism is captured here by the fact that the random time ν is not known *a priori* to the receiver. However both transmitter and receiver know that ν is distributed (*e.g.*, uniformly) over a time horizon $[1,\ldots,A]$. At all times before and after the actual transmission, the receiver observes "pure noise." The noise distribution corresponds to a special input "idle symbol" $\star$ being sent across the channel (for example, in the case of a Gaussian channel, this would be the 0, *i.e.*, no transmit signal).

The goal of the receiver is to reliably decode the information bits by sequentially observing the outputs of the channel.

A main result in [1] is a single-letter characterization of the asynchronous capacity per unit cost $\boldsymbol{C}(\beta)$ where

$$\beta \stackrel{\text{def}}{=} \frac{\log A}{B}$$

denotes the *timing uncertainty per information bit*. While this result holds for arbitrary discrete memoryless channels and arbitrary input costs, the underlying model assumes that the receiver is always in the listening mode: every channel output is observed until decoding happens.

In many applications there is a cost associated with the receiver being "on" and sampling, say through A/D conversion. As a proxy for this cost we take the number of channel outputs observed by the receiver. We are thus led to the question "What is the the minimum input cost to transmit one bit of information given that the receiver is constrained to observe only a fraction $\rho \in (0,1]$ of the channel outputs?"

Because communication is asynchronous, transmitter and receiver cannot agree on the information transmission period beforehand. Hence, the transmitter should signal with a sufficiently high energy to allow message detection at the receiver. And, intuitively, the smaller the value of ρ, the larger this energy should be to compensate for partial output observations. As it turns out, this is not the case. In this paper we show that the asynchronous capacity per unit cost satisfies

$$\boldsymbol{C}(\beta,\rho) = \boldsymbol{C}(\beta,1)$$

for any asynchronism level $\beta > 0$ and sampling frequency $0 < \rho \leq 1$. Moreover, not even delay is impacted—by delay we mean the elapsed time between when the message is available and when it is decoded. The same minimal decoding delay as under full sampling can be achieved with an arbitrarily small $\rho > 0$. This result uses the possibility for the receiver to sample adaptively: the next sample can be chosen as a function of past observed samples. In fact, under non-adaptive sampling, it is still possible to achieve the full sampling asynchronous capacity per unit cost, but the decoding delay gets multiplied by a factor $1/\rho$. Therefore, adaptive sampling strategies are of particular interest in the very sparse sampling regime.

We end this section with a brief review of studies related to the above communication model. This model was first introduced in [2], [3]. These works characterize the *synchronization*

This work was supported in part by an Excellence Chair Grant from the French National Research Agency (ACE project).

A. Tchamkerten is with the Department of Communications and Electronics, Telecom ParisTech, 75634 Paris Cedex 13, France. Email: aslan.tchamkerten@telecom-paristech.fr.

V. Chandar is with MIT Lincoln Laboratory, Lexington, MA 02420, USA. Email: vchandar@mit.edu.

G. Caire is with the Viterbi School of Engineering, University of Southern California, Los Angeles, USA. Email: caire@usc.edu.

threshold, *i.e.*, the largest level of asynchronism under which it is still possible to communicate reliably. In [3], [4] capacity is defined as the message length divided by the elapsed time between when information is available and when it is decoded. For this definition, capacity upper and lower bounds are established and shown to be tight for certain channels. In [4] it is also shown that so called training-based schemes, where synchronization and information transmission are performed separately, need not be optimal in particular in the high rate regime.

The finite message regime has been investigated by Polyanskiy in [5] when capacity is defined with respect to the codeword length, *i.e.*, same setting as [1] but with unit cost per transmitted symbol. A main result in [5] is that dispersion—a fundamental quantity that relates rate and error probability in the finite block length regime—is unaffected by the lack of synchronization.

Finally, Shomorony *et al.* [6] generalized the above bursty communication model to a diamond network configuration and provided bounds on the minimum energy needed to convey one bit of information across the network.

In the above models, codeword transmission occurs at once; information transmission lasts the duration of the codeword. Khoshnevisan and Laneman [7] proposed a complementary model where information transmission is intermittent based on a slotted variation of the purely insertion channel model [8].

This paper is organized as follows. Section II contains some background material and extends the model developed in [1] to allow for sparse output sampling. Section III contains the main results, and Section IV provides sketches of the proofs.

II. Model and Performance Criterion

Communication is discrete-time and carried over a discrete memoryless channel characterized by its finite input and output alphabets

$$\mathcal{X} \cup \{\star\} \quad \text{and} \quad \mathcal{Y},$$

respectively, and transition probability matrix $Q(y|x)$ for all $y \in \mathcal{Y}$ and $x \in \mathcal{X} \cup \{\star\}$. The alphabet $\mathcal{X}$ may or may not include $\star$. In any case, we assume that $\star$ is the only symbol with zero cost (see Definition 2). Without loss of generality, we assume that for all $y \in \mathcal{Y}$ there is some $x \in \mathcal{X} \cup \{\star\}$ for which $Q(y|x) > 0$.

Given $B \geq 1$ information bits to be transmitted, a codebook $\mathcal{C}$ consists of

$$M = 2^B$$

codewords of length $n \geq 1$ composed of symbols from $\mathcal{X}$.

A randomly and uniformly chosen message m is made available at the transmitter at a random time ν, independent of m, and uniformly distributed over $[1, \ldots, A]$, where the integer

$$A = 2^{\beta B}$$

characterizes the *asynchronism level* between the transmitter and the receiver, and where the constant $\beta \geq 0$ denotes the *timing uncertainty per information bit*, see Fig. 1.

Only one message arrives over the period $[1, 2, \ldots, A]$. If $A = 1$, the channel is said to be synchronous.

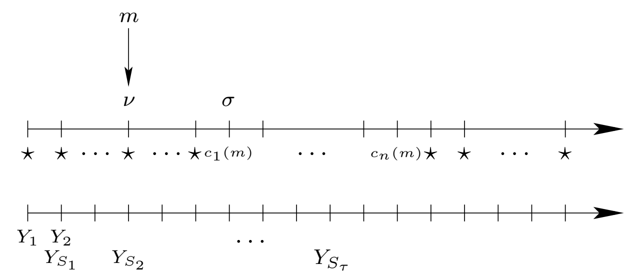

Fig. 1. Time representation of what is sent (upper arrow) and what is received (lower arrow). The "$\star$" represents the "idle" symbol. Message m arrives at time ν and starts being sent at time σ. The receiver samples at the (random) times $S_1, S_2, \ldots$ and decodes at time S_τ based on τ output samples.

Given ν and m, the transmitter chooses a time $\sigma(\nu, m)$ to start sending codeword $c^n(m) \in \mathcal{C}$ assigned to message m. Transmission cannot start before the message arrives or after the end of the uncertainty window, hence $\sigma(\nu, m)$ must satisfy

$$\nu \leq \sigma(\nu, m) \leq A \quad \text{almost surely.}$$

In the rest of the paper, we suppress the arguments ν and m of σ when these arguments are clear from context.

Before and after the codeword transmission, *i.e.*, before time σ and after time $\sigma + n - 1$, the receiver observes "pure noise." Specifically, conditioned on the event $\{\nu = t\}$, $t \in \{1, \ldots, A\}$, and on the message to be conveyed m, the receiver observes independent channel outputs

$$Y_1, Y_2, \ldots, Y_{A+n-1}$$

distributed as follows. For

$$1 \leq i \leq \sigma(t, m) - 1$$

or

$$\sigma(t, m) + n \leq i \leq A + n - 1\,,$$

the Y_i's are "pure noise" symbols, *i.e.*,

$$Y_i \sim Q(\cdot|\star)\,.$$

For $\sigma \leq i \leq \sigma + n - 1$

$$Y_i \sim Q(\cdot|c_{i-\sigma+1}(m))$$

where $c_i(m)$ denotes the ith symbol of the codeword $c^n(m)$.

The receiver operates according to a sampling strategy and a sequential decoder. A sampling strategy consists of "sampling times" which are defined as an ordered collection of random time indices

$$\mathcal{S} = \{(S_1, \ldots, S_\ell) \subseteq \{1, \ldots, A + n - 1\} : S_i < S_j, i < j\}$$

where S_j is interpreted as the jth sampling time.

The sampling strategy is either non-adaptive or adaptive. It is non-adaptive when the sampling times given by $\mathcal{S}$ are all known before communication starts, hence $\mathcal{S}$ is independent of Y_1^{A+n-1}. The strategy is adaptive when the sampling times are functions of past observations. This means that S_1 is an arbitrary value in $\{1, \ldots, A + n - 1\}$, possibly random but independent of Y_1^{A+n-1} and, for $j \geq 2$,

$$S_j = g_j(\{Y_{S_i}\}_{i<j})$$

for some (possibly randomized) function

$$g_j : \mathcal{Y}^{j-1} \to \{S_{j-1}+1, \ldots, A+n-1\}.$$

Notice that ℓ, the total number of output samples, may be random under adaptive sampling, but also under non-adaptive sampling, since the strategy may be randomized (but still independent of the channel outputs Y_1^{A+n-1}).

Once the sampling strategy is fixed, the receiver decodes by means of a sequential test (τ, ϕ), where τ, the decision time, is a stopping time with respect to the sampled sequence

$$Y_{S_1}, Y_{S_2}, \ldots$$

indicating when decoding happens,[1] and where ϕ is the decoding function, *i.e.*, a map

$$\phi : \mathcal{O} \to \{1, 2, \ldots, M\}$$

where

$$\mathcal{O} \stackrel{\text{def}}{=} \{Y_{S_1}, Y_{S_2}, \ldots, Y_{S_\tau}\}$$

is the set of observed samples. Hence, decoding happens at time S_τ on the basis of τ output samples. Since there are at most $A+n-1$ sampling times, τ is bounded by $A+n-1$.

A code $(\mathcal{C}, \mathcal{S}, (\tau, \phi))$ is defined as a codebook, a receiver sampling strategy, and a decoder (decision time and decoding function). Throughout the paper, whenever clear from context, we often refer to a code using the codebook symbol $\mathcal{C}$ only, leaving out an explicit reference to the sampling strategy and to the decoder.

Definition 1 (Error probability)**.** The maximum (over messages) decoding error probability of a code $\mathcal{C}$ is defined as

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

where the subscripts "m, t" denote conditioning on the event that message m is available at time $\nu = t$, and where $\mathcal{E}_m$ denotes the error event that the decoded message does not correspond to m, *i.e.*,

$$\mathcal{E}_m \stackrel{\text{def}}{=} \{\phi(\mathcal{O}) \neq m\}.$$

Definition 2 (Cost of a Code)**.** The (maximum) cost of a code $\mathcal{C}$ with respect to a cost function $k : \mathcal{X} \to [0, \infty]$ is defined as

$$K(\mathcal{C}) \stackrel{\text{def}}{=} \max_m \sum_{i=1}^{n} k(c_i(m)).$$

Notice that when $\star \in \mathcal{X}$ the transmitter can multiplex idle channel uses in between other code symbols at no cost. When $\star \notin \mathcal{X}$ then $k(x) > 0$ for any $x \in \mathcal{X}$.

Below, $\mathbb{P}_m$ denotes the output distribution conditioned on the sending of message m. Hence, by definition we have

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

[1] Recall that a (deterministic or randomized) stopping time τ with respect to a sequence of random variables $Y_1, Y_2, \ldots$ is a positive, integer-valued, random variable such that the event $\{\tau = t\}$, conditioned on the realization of $Y_1, Y_2, \ldots, Y_t$, is independent of the realization of $Y_{t+1}, Y_{t+2}, \ldots$, for all $t \geq 1$.

Definition 3 (Sampling Frequency of a Code)**.** Given $\varepsilon > 0$, the sampling frequency of a code $\mathcal{C}$, denoted by $\rho(\mathcal{C}, \varepsilon)$, is the relative number of channel outputs that are observed until a message is declared. Specifically, it is defined as the smallest $r \geq 0$ such that

$$\min_m \mathbb{P}_m(\tau / S_\tau \leq r) \geq 1 - \varepsilon.$$

(Recall that S_τ refers to the last sampling time.)

Definition 4 (Delay of a Code)**.** Given $\varepsilon > 0$, the (maximum) delay of a code $\mathcal{C}$, denoted by $d(\mathcal{C}, \varepsilon)$, is defined as the smallest integer l such that

$$\min_m \mathbb{P}_m(S_\tau - \nu \leq l - 1) \geq 1 - \varepsilon.$$

We now define capacity per unit cost when the receiver is constraint to observe only a limited number of channel outputs.

Definition 5 (Asynchronous Capacity per Unit Cost under Sampling Constraint)**.** $\boldsymbol{R}$ is an achievable rate per unit cost at timing uncertainty per information bit β, and sampling frequency ρ if there exists a sequence of codes $\{\mathcal{C}_B\}$ and a sequence of positive numbers ε_B with $\varepsilon_B \stackrel{B\to\infty}{\longrightarrow} 0$ such that for all B large enough

1) $\mathcal{C}_B$ operates at timing uncertainty per information bit β;
2) the maximum error probability $\mathbb{P}(\mathcal{E}|\mathcal{C}_B)$ is at most ε_B;
3) the rate per unit cost $B/K(\mathcal{C}_B)$ is at least $\boldsymbol{R} - \varepsilon_B$;
4) the sampling frequency satisfies $\rho(\mathcal{C}, \varepsilon_B) \leq \rho + \varepsilon_B$;
5) the delay satisfies[2]

$$\frac{1}{B} \log(d(\mathcal{C}_B, \varepsilon_B)) \leq \varepsilon_B.$$

The asynchronous capacity per unit cost, denoted by $\boldsymbol{C}(\beta, \rho)$, is the supremum of achievable rates per unit cost.

Notice that here we require that the delay is sub-exponential in B. By contrast, a more general case was treated in [1] where exponential delay may be used to reduce the timing uncertainty per information bit. Two basic observations are in order:

- $\boldsymbol{C}(\beta, \rho)$ is a non-increasing function of β for fixed ρ;
- $\boldsymbol{C}(\beta, \rho)$ is an non-decreasing function of ρ for fixed β.

In particular, we have for any fixed $\beta \geq 0$

$$\max_{\rho \geq 0} \boldsymbol{C}(\beta, \rho) = \boldsymbol{C}(\beta, 1).$$

The capacity under full sampling $\boldsymbol{C}(\beta, 1)$ is characterized in the following theorem:

Theorem 1 ([1] Theorem 1)**.** *For any* $\beta \geq 0$

$$\boldsymbol{C}(\beta, 1) = \max_X \min \left\{ \frac{I(X;Y)}{\mathbb{E}[k(X)]}, \frac{I(X;Y) + D(Y||Y_\star)}{\mathbb{E}[k(X)](1+\beta)} \right\}, \tag{2}$$

where $\max_X$ *denotes maximization with respect to the channel input distribution* P_X*, where* $(X, Y) \sim P_X(\cdot)Q(\cdot|\cdot)$*, where* $Y_\star$ *denotes the random output of the channel when the idle symbol* $\star$ *is transmitted (i.e.,* $Y_\star \sim Q(\cdot|\star)$*), where* $I(X;Y)$ *denotes the mutual information between* X *and* Y*, and where*

[2] Throughout the paper log is always to the base 2.

$D(Y||Y_\star)$ denotes the divergence (Kullback-Leibler distance) between the distributions of Y and $Y_\star$.[3] ■

Let P_{X^*} be a capacity per unit cost achieving input distribution, *i.e.*, X^* achieves the maximum in (2). As shown in the converse proof of [1, Theorem 1], codes that achieve the capacity per unit cost can be restricted to codes of (asymptotically) constant composition P_{X^*}. Specifically, we have

$$\frac{B}{n_B(P_{X^*})\mathbb{E}[k(X^*)]} = \boldsymbol{C}(\beta,1)(1-o(1)) \quad (B\to\infty)$$

where $n_B(P_{X^*})$ denotes the length of the P_{X^*}-constant composition codes achieving $\boldsymbol{C}(\beta,1)$. Now define

$$n_B^* \stackrel{\text{def}}{=} \min_{P_{X^*}} n_B(P_{X^*}) = \min_{X\in\mathcal{P}} \frac{B}{\boldsymbol{C}(\beta,1)\mathbb{E}[k(X)]}$$

where

$$\mathcal{P} \stackrel{\text{def}}{=} \{X : X \text{ achieves the maximum in (2)}\}.$$

From the achievability and converse of [1, Theorem 1] $\{n_B^*\}$ represent the smallest achievable delays for codes $\{\boldsymbol{C}_B\}$ achieving the asynchronous capacity per unit cost under *full sampling* $\boldsymbol{C}(\beta,1)$ in the sense that

$$d(\mathcal{C}_B,\varepsilon_B) \geq n_B^*(1+o(1)) \quad (B\to\infty)$$

for any $\varepsilon_B \to 0$ as $B\to\infty$.

Our main results, stated in the next section, say that the capacity per unit cost under sampling frequency $0\rho \in (0,1)$ is the same as under full sampling, *i.e.*, $\rho = 1$. To achieve this, non-adaptive sampling is sufficient. However, if we also want to achieve minimum delay, then adaptive sampling is necessary. In fact, non-adaptive sampling strategies that achieve the capacity per unit cost have a delay that grows at least as n_B^*/ρ.

III. Results

In the sequel we denote by $\boldsymbol{C}_{\text{a}}(\beta,\rho)$ and $\boldsymbol{C}_{\text{na}}(\beta,\rho)$ the capacity per unit cost when restricted to adaptive and non-adaptive sampling, respectively. Our first result characterizes the capacity per unit cost under non-adaptive sampling.

Theorem 2 (Non-adaptive sampling)**.** *Under non-adaptive sampling it is possible to achieve the full-sampling capacity per unit cost, i.e.,*

$$\boldsymbol{C}_{na}(\beta,\rho) = \boldsymbol{C}(\beta,1) \quad \text{for any } \beta,\rho\in(0,1].$$

Furthermore, when $\star \in \mathcal{X}$, codes $\{\mathcal{C}_B\}$ that achieve rate $\gamma\boldsymbol{C}(\beta,1)$, $0\leq\gamma\leq 1$, are such that

$$\lim_{\gamma\to 1}\liminf_{B\to\infty}\frac{d(\mathcal{C}_B,\varepsilon_B)}{n_B^*} \geq \frac{1}{\rho},$$

and when $\star\notin\mathcal{X}$ we have

$$\lim_{\gamma\to 1}\liminf_{B\to\infty}\frac{d(\mathcal{C}_B,\varepsilon_B)}{n_B^*} \geq \frac{1+\rho}{\rho}.$$

Hence, even with a negligible fraction of the channel outputs it is possible to achieve the full-sampling capacity per unit cost. However, this comes at the expense of delay, which gets multiplied by a factor $1/\rho$ or $(1+\rho)/\rho$ depending on whether or not $\star$ can be used for code design.

Under adaptive sampling we have:

Theorem 3 (Adaptive sampling)**.** *Under adaptive sampling it is possible to achieve the full-sampling capacity per unit cost, i.e.,*

$$\boldsymbol{C}_{na}(\beta,\rho) = \boldsymbol{C}(\beta,1) \quad \text{for any } \beta,\rho\in(0,1].$$

Furthermore, the minimum delay of codes that achieve the capacity per unit cost is given by

$$d(\mathcal{C}_B,\varepsilon_B) = n_B^*(1+o(1)).$$

The first part of Theorem 3 immediately follows from the first part of Theorem 2, since the set of adaptive sampling strategies include the set of non-adaptive sampling strategies. The interesting part of Theorem 3 is that adaptive sampling strategies guarantee minimal delay *regardless* of the sampling rate ρ, as long as it is non-zero.

We end this section by considering the specific case when $\beta = 0$, *i.e.*, when the channel is synchronous. For a given sampling frequency ρ, the receiver gets to see only a fraction ρ of the transmitted codeword (whether sampling is adaptive or non-adaptive) and hence $\boldsymbol{C}(0,\rho) = \rho\,\boldsymbol{C}(0,1)$ for any $\rho\geq 0$.

How is it possible that a sparse output sampling induces a rate per unit cost loss for synchronous communication ($\beta = 0$), but not for asynchronous communication ($\beta > 0$) as we saw in Theorems 2 and 3? The reason for this apparent contradiction is that when $\beta > 0$, the level of asynchronism is exponential in B. Therefore, even if the receiver is constrained to sample only a fraction ρ of the channel outputs, it may still occasionally sample fully over, say, $\Theta(B)$ channel outputs, and still satisfy the overall constraint that the fraction shouldn't exceed ρ.[4]

IV. Sketch of the Proofs

In this section we describe the achievability schemes yielding Theorems 2 and 3 and provide the main ideas behind the delay converses. The detailed analysis (error probability, rate per unit cost, and sampling rate) of the achievability and converse arguments can be found in [9].

Sketch of Achievability of Theorem 2: Given B bits of information to be transmitted, the codebook $\mathcal{C}$ is a random almost constant composition code as in [1, Proof of Theorem 1], *i.e.*, each symbol of each codeword appears roughly the same number of times and the cost of each codeword is

$$n\mathbb{E}[k(X)](1+o(1))$$

as $n\to\infty$ where X denotes the random input and where $k(\cdot)$ is the input cost function of the channel.

Case $\star\in\mathcal{X}$: Information transmission is as follows. Codeword symbols can be transmitted only at integer multiples of $1/\rho$ (for simplicity we assume that $1/\rho$ is an integer). Multiples of $1/\rho$ from now on are referred to as transmission times. Given a message m available at time ν, the transmitter sends the corresponding codeword $c^n(m)$ during the first n information

[3] $Y_\star$ can be interpreted as "pure noise."

[4] If over a long trip we have a high-mileage drive, we can still push the car a few times without impacting the overall mileage.

transmission times coming at time $\geq \nu$. In between transmission times the transmitter sends $\star$. Hence, the transmitter sends the "expanded" codeword

$$c_1(m) \star \ldots \star c_2(m) \star \ldots \star c_3(m)\{\ldots\ldots\}c_n(m)$$

starting at time $\sigma = \sigma(\nu) = \min\{t : t/\rho \geq \nu\}$.

Sampling is performed only at the transmission times. At transmission time t, the decoder performs a typicality decoding on the basis of the last output samples y^n and all the codewords $\{c^n(m)\}$. If there is a unique message $c^n(m)$ that is jointly typical with y^n, the decoder stops and declares that message m was sent. If two (or more) codewords $c^n(m)$ and $c^n(m')$ are typical with y^n, the decoder stops and declares one of the corresponding messages at random. If no codeword is typical with y^n, the decoder repeats the procedure at the next transmission time, *i.e.*, at time $t + 1/\rho$.

Case $\star \notin \mathcal{X}$: By contrast with the previous case where codewords are expanded through the filling of $\star$'s, here codewords keep the same length as under full sampling—since $\star \notin \mathcal{X}$ codeword expansion is costly.

Parse the entire sequence $\{1, 2, \ldots, A + n - 1\}$ into consecutive superperiods of size n/ρ. The periods of duration n occurring at the end of each superperiod are referred to as transmission periods. Given ν, the codeword starts being sent over the first transmission period occurring after ν. In particular, if ν happens over a transmission period, then the transmitter delays the codeword transmission to the next superperiod.

The receiver samples only the transmission periods. At the end of a transmission period, the decoder performs typicality decoding as in the previous case. In particular, if no codeword is typical with y^n, the decoder waits for the next transmission period to occur, samples it, and repeats the decoding procedure. ■

Sketch of Delay Converse of Theorem 2:

Case $\star \in \mathcal{X}$: To achieve the capacity per unit cost, the decoder needs to acquire a number of codeword samples at least equal to a large fraction ≈ 1 of n_B^*. This follows from the converse of [1, Theorem 1] and from the definition of n_B^*. Now, given the sampling constraint and that the sampling strategy is non-adaptive, over most consecutive periods of size N the decoder will be able to sample no more than $\approx \rho N$ symbols—recall that ν is uniformly distributed over $[1, 2, \ldots, A]$. As a consequence, a decoder that achieves capacity per unit cost experiences a delay at least equal to n_B^*/ρ.

Case $\star \notin \mathcal{X}$: The argument for the case $\star \in \mathcal{X}$ holds irrespectively of whether or not $\star \in \mathcal{X}$. To acquire n codeword symbols the receiver must spend a time $\approx n/\rho$ channel uses. However, when $\star \in \mathcal{X}$ we can construct an "expanded" code of length n/ρ whose cost is the same as under full sampling as described above.

When $\star \notin \mathcal{X}$ codewords cannot be expanded without incurring a cost loss. Hence, to achieve capacity per unit cost it is necessary that the decoder samples in blocks of n_B^* symbols. Now, the sampling constraint means that, on average, the gap between sampled blocks grows like n_B^*/ρ. However, if the message arrives at a time ν close to the beginning of a block, then in addition to waiting until the next block, the message must wait for most of the current block before being transmitted. The reason for this is that, close to capacity, we cannot afford to miss any portion other than negligible of the codeword—for otherwise the error probability is large. Therefore, the delay must grow at least as $n_B^*/\rho + n_B^*$. ■

Proof Sketch of Theorem 3: We use a random codebook construction as in the proof sketch of Theorem 2. Now, intuitively, an optimal sampling strategy should sample sparsely, with a sampling frequency of no more than ρ, under pure noise—for otherwise the sampling constraint would be violated. It should also sample the entire sent codeword, and so densely sample during message transmission—for otherwise a rate per unit cost penalty is incurred.

In order to illustrate the proposed asymptotically optimal sampling/decoding strategy, we introduce the following notation. Let $\tilde{Y}_a^b$ denote the random vector obtained by extracting the components of the output process Y_t at $t \in [a, b]$ of the form $t = \lceil j/\rho \rceil$ for non-negative integer j. Notice that, for any $t \geq \ell$ and $\ell \gg 1$, $\tilde{Y}_{t-\ell+1}^t$ contains $\approx \rho\ell$ samples. The strategy starts in the sparse mode, taking samples at times $S_j = \lceil j/\rho \rceil$, $j = 1, 2, \ldots$. At each j, the receiver computes the empirical distribution (or type)

$$\hat{P}_j = \hat{P}_{\tilde{Y}_{S_j - \log(n)+1}^{S_j}}$$

of the sampled output in the most recent window of length $\log(n)$. If the probability under pure noise of observing this empirical distribution is $> \frac{1}{n^2}$, the mode is kept unchanged and we repeat this test at the next round $j + 1$. Instead, if it is $\leq \frac{1}{n^2}$, then we switch to the dense sampling mode, taking samples continuously for at most n time steps. At each of these steps the receiver applies the same sequential typicality decoder as in the proof of Theorem 2, based on the past n output samples. If no codeword is typical with the channel outputs after these n times steps, sampling is switched back to the sparse mode. ■

References

[1] V. Chandar, A. Tchamkerten, and D. Tse, "Asynchronous capacity per unit cost," *to appear in the IEEE Trans. Inform. Theory*, 2012.

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

[3] A. Tchamkerten, V. Chandar, and G. W. Wornell, "Communication under strong asynchronism," *IEEE Trans. Inform. Theory*, vol. 55, no. 10, pp. 4508–4528, Oct. 2009.

[4] ——, "Asynchronous communication: capacity bounds and suboptimality of training," *to appear in the IEEE Trans. Inform. Theory*, 2012.

[5] Y. Polyanksiy, "Asynchronous communicaton: exact synchronization, universality, and dispersion," *to appear in the IEEE Trans. Inform. Theory*, 2012.

[6] I. Shomorony, R. H. Etkin, F. Parvaresh, and A. Avestimehr, "Bounds on the minimum energy-per-bit for bursty traffic in diamond networks," in *IEEE Int. Symp. Information Theory (ISIT)*, July 2012.

[7] M. Khoshnevisan and J. N. Laneman, "Achievable rates for intermittent communication," in *IEEE Int. Symp. Information Theory (ISIT)*, July 2012.

[8] R. L. Dobrushin, "Asymptotic bounds on the probability of error for the transmission of messages over a memoryless channel using feedback," *Probl. Kibern.*, vol. 8, pp. 161–168, 1963.

[9] A. Tchamkerten, V. Chandar, and G. Caire, "Energy and sampling constrained asynchronous communication," *submitted to IEEE Trans. Inform. Th.*