On the Capacity of the Slotted Strongly Asynchronous Channel with a Bursty User

Sara Shahi, Daniela Tuninetti and Natasha Devroye
University of Illinois at Chicago, Chicago IL 60607, USA. Email: sshahi7, danielat, devroye @uic.edu

***Abstract*—The slotted strongly asynchronous channel with a bursty user consists of a window of $A_n = e^{n\alpha}$ blocks of length n channel uses. A user transmits a randomly selected message among $M_n = e^{nR}$ different ones in exactly $K_n = e^{n\nu}$ randomly selected but distinct blocks in the window. The receiver must locate and decode, with vanishing error probability in n, each one of the transmitted messages. The optimal tradeoff between (R, α, ν) is derived.**

I. Introduction

Many practical applications, such as sensor networks, deal with tremendous amounts of communication between devices. These transmissions between devices may be bursty, and need to be reliably detected and decoded. For example, each sensor node may want to transmit a signal to the base station only when some incident has taken place.

In this paper, we consider the problem of both detecting and decoding the data bursts. Conventional methods transmit a pilot signal at the beginning of each data burst to notify the decoder of the upcoming data; the decoding phase may be performed using any synchronized decoding method. An alternative approach is to simultaneously detect and decode the codewords. The first approach is optimal when synchronization is done once and the cost of acquiring synchronization is absorbed into the lengthy data stream that follows. For sparse / bursty transmission, as in the problem considered here, the second approach is preferable as the training based schemes are proven to be sub-optimal [1]. In this work we do not enforce the usage of pilot symbols for positive rates, and the codebook serves the purpose of synchronization as well as of data transfer. This paper's central goal is to characterize the tradeoff between the reliable transmission rate between one transmitter and one receiver, the burstiness of that transmitter, and the level of asynchronism.

Past Work. In [2], we studied a multi-user version of the slotted strongly asynchronous channel model where we assumed $K_n = e^{n\nu}$ different users transmit a message among $M_n^{(i)} = e^{nR_i}, i \in [1 : K_n]$ of them only once in an asynchronous window of length $A_n = e^{n\alpha}$ blocks, where each block /slot comprises n channel uses. What renders the presented version of the problem in this paper more tractable is that one is guaranteed that in each block there is at most one transmitted message and we do not need to detect the user's identity. Hence we were able to get to the exact capacity region for this channel model as opposed to upper and lower regions in [2].

The problem considered here also generalizes the one in [3, Remark 3]. In [3], the authors adopted the 'per-user' error criterion and they aimed to only recover a large fraction of the transmissions, while in our work, the error probability is the global / joint probability of error (i.e., an error is declared if *any* of the user's transmissions is in error, and we have an exponential number of transmissions) and we require the exact recovery of the transmission time and codeword in *all* transmissions. The approach in [3] does not extend to the global probability of error criterion for exponential number of transmission ($\nu > 0$) as their achievability relies on the typicality decoder and the derived error bounds do not decay fast enough in blocklength n.

In [4], the authors considered the special case of the problem considered here where a user transmits one synchronization pattern (hence $R = 0$) of length n only once (hence $K_n = 1$) in a window of length $A_n = e^{n\alpha}$ of n channel uses. They showed that for any α below the synchronization threshold, α_0, the user can detect the location of the synchronization pattern. In addition they showed that the synchronization pattern consists of the repetition of a single symbol which induces an output distribution with the maximum divergence from the noise distribution. The typicality decoder introduced in [4] however, even in a slotted channel model, only retrieves one of the trade-off points that we obtain in this paper that corresponds to the sub-exponential number of transmissions. More specifically, the achievability and the converse techniques used in [4] are not applicable for an exponential number of transmissions. We propose new achievability and converse techniques to support an exponential number of transmissions ($K_n = e^{n\nu}$). Interestingly, we show that the symbol used for synchronization may change for different values of α and ν.

The single user strongly asynchronous channel was also considered in [5], where it was shown that the exact transmission time recovery, as opposed to the error criterion in [6] which allows a sub-exponential delay in n, does not change the capacity.

Contributions. In this paper, we investigate the trade-off between the number of transmissions (or burstiness) $K_n = e^{n\nu}$ of the user and the asynchronism level $A_n = e^{n\alpha}$ in a slotted strongly asynchronous channel while distinguishing $M_n = e^{nR}$ messages with vanishingly error probability in n. The slotted assumption restricts the transmission times to be integer multiples of the blocklength n; this assumption simplifies the error analysis yet captures the essence of the problem. We show:

978-1-5090-3097-2/17/$31.00 ©2017 IEEE

1) For synchronization and data transmission ($R > 0$), we find the capacity region (R, α, ν) and we show in our converse that using the same codebook in all transmissions is optimal. In addition, we show that performing a two stage decoding where the receiver first finds the locations of transmitted codewords and then decodes the messages is optimal.
2) For synchronization only ($R = 0$), our proposed sequential decoder achieves the optimal tradeoff. This implies that almost real-time synchronization (due to sequential decoding) can be achieved. Surprisingly, we show that the optimal synchronization pattern is not fixed and it may change depending on the considered value of asynchronism level α and burstiness level ν.

Notation. Capital letters represent random variables that take on lower case letter values in calligraphic letter alphabets. A stochastic kernel / transition probability from $\mathcal{X}$ to $\mathcal{Y}$ is denoted by $Q(y|x), \forall(x,y) \in \mathcal{X}\times\mathcal{Y}$, and the output marginal distribution induced by $P \in \mathcal{P}_{\mathcal{X}}$ through the channel Q as $[PQ](y) := \sum_x P(x)Q(y|x), \forall y \in \mathcal{Y}$ where $\mathcal{P}_{\mathcal{X}}$ is the space of all input distributions on $\mathcal{X}$. As a shorthand notation, we also define $Q_{x^n}(.) := Q(.|x^n)$. We use $y_j^n := [y_{j,1}, ..., y_{j,n}]$, and simply y^n instead of y_1^n. The empirical distribution of a sequence x^n is

$$\widehat{P}_{x^n}(a) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{x_i=a\}}, \forall a \in \mathcal{X}, \tag{1}$$

where $\mathbb{1}_{\{A\}}$ is the indicator function of the event A; when using (1) the target sequence x^n is usually clear from the context so we may drop the subscript x^n. We also use $I(P,Q)$ to denote the mutual information between random variable $(X,Y) \sim PQ$, $D(P_1 \parallel P_2)$ for the Kullback Leibler divergence between distribution P_1 and P_2, and $D(Q_1 \parallel Q_2|P) := \sum_{x,y\in\mathcal{X}\times\mathcal{Y}} P(x)Q_1(y|x)\log\frac{Q_1(y|x)}{Q_2(y|x)}$. The Q-shell of the sequence x^n is defined as

$$T_Q(x^n) := \left\{ y^n : \frac{\sum_{i=1}^{n} \mathbb{1}_{\left\{\substack{x_i=a,\\ y_i=b}\right\}}}{\sum_{i=1}^{n}\mathbb{1}_{\{x_i=a\}}} = Q(b|a), \forall(a,b) \in (\mathcal{X},\mathcal{Y}) \right\}.$$

II. System model and capacity result

We consider a discrete memoryless channel with transition probability matrix $Q(y|x)$ defined over all (x,y) in the finite input and output alphabets $(\mathcal{X},\mathcal{Y})$. We also define a noise symbol $\star \in \mathcal{X}$ for which $Q_\star(y) > 0$, $\forall y \in \mathcal{Y}$. An (M, A, K, n, ϵ) code for the slotted bursty and strongly asynchronous discrete memoryless channel with transition probability matrix $Q(y|x)$, $(x,y) \in (\mathcal{X},\mathcal{Y})$, is defined as follows. Encoding functions $f_i : [1 : M] \to \mathcal{X}^n$, $i \in [1 : A]$, where we define $x_i^n(m) := f_i(m)$. The transmitter chooses uniformly at random one set of K blocks for transmission out of the $\binom{A}{K}$ possible ones, and a set of K messages from M^K possible ones, also uniformly at random, and sends $x_{\nu_i}^n(m_i)$ in block ν_i for $i \in [1 : K]$ and $\star^n$ in every other block. We denote the chosen blocks and messages as $((\nu_1, m_1), \ldots, (\nu_K, m_K))$. The associated average probability of error for the destination decoder function

$$g(\mathcal{Y}^{nA}) = ((\widehat{\nu}_1, \widehat{m_1}), \ldots, (\widehat{\nu}_K, \widehat{m_K})),$$

is given by

$$P_e^{(n)} := \frac{1}{M^K\binom{A}{K}} \sum_{(\nu_1,m_1),\ldots,(\nu_K,m_K)} \mathbb{P}[g(y^{nA}) \neq ((\nu_1,m_1),\ldots,(\nu_K,m_K))\,|H_{((\nu_1,m_1),\ldots,(\nu_K,m_K))}],$$

where $H_{((m_1,\nu_1),\ldots,(m_K,\nu_K))}$ is the hypothesis that user transmits message m_i at block ν_i with the codebook $x_{\nu_i}^n(m_i)$, for all $i \in [1 : K]$.

A tuple (R, α, ν) is said to be achievable if there exists a sequence of codes $(e^{nR}, e^{n\alpha}, e^{n\nu}, n, \epsilon_n)$ with $P_e^{(n)} \le \epsilon_n, \epsilon_n \to 0$ as $n \to \infty$. The capacity region is the set of all possible achievable (R, α, ν) triplets.

The main result is as follows.

Theorem 1. *The capacity region for the slotted bursty and strongly asynchronous discrete memoryless channel is*

$$\mathcal{R} := \bigcup_{\lambda\in[0,1], P\in\mathcal{P}_{\mathcal{X}}} \left\{ \begin{array}{c} \nu < \alpha \\ \alpha + R < D(Q_\lambda \parallel Q_\star|P) \\ \nu < D(Q_\lambda \parallel Q|P) \\ R < I(P,Q) \end{array} \right\}, \tag{2}$$

where

$$Q_\lambda(.) := \xi_x Q_x^\lambda(.)Q_\star^{1-\lambda}(.), \tag{3}$$

for $\xi_x := \sum_{y\in\mathcal{Y}} Q_x^\lambda(y)Q_\star^{1-\lambda}(y)$ a normalization factor.

Proof: **Achievability.** *Codebook generation.* The user generates A_n constant composition i.i.d. random codebooks, of rate R and blocklength n, according to the distribution P, one for each available block. Since the code is the same for all blocks, we drop the subscript of the encoder function of individual blocks and simply use $x^n(.)$ for all of them. *Decoder.* We perform a two-stage decoding. First, the decoder finds the location of the transmitted codewords (synchronization stage) and it decodes the messages (decoding stage). The probability of error for this two-stage decoder is given by

$$P_e^{(n)} \le \mathbb{P}[\text{synchronization error}] \tag{4}$$
$$+ \mathbb{P}[\text{decoding error}|\text{no synchronization error}]. \tag{5}$$

For the first stage, fix a $T : -D(Q_\star \parallel Q|P) \le T \le D(Q \parallel Q_\star|P)$ (which can be changed for different trade-off points). At each block $j, j \in [1 : A_n]$, if there exists *any* message $m, m \in [1 : e^{nR}]$ such that

$$L\left(y_j^n, x^n(m)\right) := \frac{1}{n}\log\frac{Q\left(y_j^n|x_j^n(m)\right)}{Q_{\star^n}(y_j^n)} \ge T,$$

declare a codeword transmission block and a noise block otherwise. Given the hypothesis $H_{((m_1,1),\ldots,(m_{K_n},K_n))}$ (we drop the conditioning on this hypothesis in the following equations so as to simplify the notation), the probability of synchronization error in the first stage is given by

$$
\begin{aligned}
P_{e_{\text{sync}}}^{(n)} &\le \sum_{j=1}^{K_n} \mathbb{P}\left[\bigcap_{m=1}^{M_n} L\left(Y_j^n, x^n(m)\right) < T\right] \\
&\quad + \sum_{j=K_n+1}^{A_n} \mathbb{P}\left[\bigcup_{m=1}^{M_n} L\left(Y_j^n, x^n(m)\right) \ge T\right] \\
&\le e^{n\nu} \sum_{\substack{V_1:\\ D(V_1\|Q_\star|P)-D(V_1\|Q|P)<T}} Q_{x^n(m)}\left[Y^n \in T_{V_1}\left(x^n(m)\right)\right] \\
&\quad + e^{n(\alpha+R)} \sum_{\substack{V_2:\\ D(V_2\|Q_\star|P)-D(V_2\|Q|P)\ge T}} Q_\star^n\left[Y^n \in T_{V_2}(x^n(m))\right] \\
&\doteq e^{n\nu} e^{-nD(Q_\lambda\|Q|P)} + e^{n(\alpha+R)} e^{-nD(Q_\lambda\|Q_\star|P)}, \qquad (6)
\end{aligned}
$$

where Q_λ is defined in (3) and $\lambda : D(Q_\lambda \parallel Q_\star|P) - D(Q_\lambda \parallel Q|P) = T$ and where (6) is proved in Appendix A. The notation $\doteq$ is also defined in [7]. Having found the K_n 'not noisy' blocks, we use a typicality decoder on the super-block of length nK_n to distinguish among e^{nK_nR} different message combinations and hence we get the bound $R < I(P, Q)$.

Converse. The main technical difficulty and innovation in the proof relies on finding a matching (with the achievability) exponentially decaying 'lower' bounds on the false alarm and missed detection error events. By following an argument similar to [5, Eq. 137], we can restrict our attention to constant composition codes. In other words, we assume the use of codes $x_i^n(.)$ with constant compositions $P_i, i \in [1 : A_n]$ in each block but we will see later that using a single composition in all blocks is optimal. Given the hypothesis $H_{((1,m_1)\ldots(K_n,m_{K_n}))}$, with a maximum likelihood decoder (which achieves the minimum average probability of error) and for any $T \in \mathbb{R}$, the error events are given by

$$
\begin{aligned}
&\left\{\text{error}|H_{((1,m_1),\ldots,(K,m_K))}\right\} = \bigcup_{\substack{((l_1,\tilde{m}_1)\ldots,(l_{K_n},\tilde{m}_{K_n}))\\ \neq((1,m_1),\ldots,(K_n,m_{K_n}))}} \\
&\left\{\sum_{i=1}^{K_n} L\left(Y_i^n, x_i^n(m_i)\right) \le \sum_{i=1}^{K_n} L\left(Y_{l_i}^n, x_{l_i}^n(\tilde{m}_i)\right)\right\} \\
&\supseteq \bigcup_{\substack{i\in[1:K_n]\\ j\in[K_n+1:A_n]\\ m\in[1:M_n]}} \left\{L\left(Y_i^n, x_i^n(m_i)\right) \le L\left(Y_j^n, x_i^n(m)\right)\right\} \qquad (7) \\
&\supseteq \left\{\bigcup_{i\in[1:K_n]} \left\{L\left(Y_i^n, x_i^n(m_i)\right) < T\right\}\right\} \bigcap \\
&\left\{\bigcup_{\substack{j\in[K_n+1:A_n]\\ m\in[1:M_n]}} \left\{L\left(Y_j^n, x_j^n(m)\right) \ge T\right\}\right\}, \qquad (8)
\end{aligned}
$$

where (7) is the union over the events that *(any message, noisy block, correct codebook)* is selected instead of one of the *(correct message, correct block, correct codebook)*'s. We also further restrict $T \in [-D(Q_\star \parallel Q|P_{i^\star}), D(Q \parallel Q_\star|P_{i^\star})]$ and define $i^\star$ as

$$
\sup_{\substack{i,\lambda_i:\\ D(Q_{\lambda_i}\|Q_\star|P_i)-D(Q_{\lambda_i}\|Q|P_i)=T}} D(Q_{\lambda_i} \parallel Q|P_i) = D(Q_{\lambda_{i^\star}} \parallel Q|P_{i^\star}). \qquad (9)
$$

The reason for this choice of $i^\star$ will become clear later (see (12) and (13)). By (8) we have

$$
\begin{aligned}
&\mathbb{P}\left[\text{error}\middle|H_{((1,m_1),\ldots,(K_n,m_{K_n}))}\right] \\
&\ge \mathbb{P}\left[\bigcup_{i\in[1:K_n]} L\left(Y_i^n, x_i^n(m_i)\right) < T\right] \qquad (10) \\
&\cdot \mathbb{P}\left[\bigcup_{\substack{j\in[K_n+1:A_n]\\ m\in[1:M_n]}} L\left(Y_j^n, x_j^n(m)\right) \ge T\right] \qquad (11) \\
&\ge \left(1 - e^{-n\left[\nu - D\left(Q_{\lambda_{i^\star}}\|Q|P_{i^\star}\right)\right]}\right) \cdot \qquad (12) \\
&\left(1 - e^{-n\left[(R+\alpha)-(1+6\delta_1)D\left(Q_{\lambda_{i^\star}}\|Q_\star|P_{i^\star}\right)\right]}\right). \qquad (13)
\end{aligned}
$$

where (10) and (11) are due to the independence of $Y_j^n, j \in [1 : A_n]$ and where (12) and (13) are proved in Appendix B and C, respectively.

With the conventional bound on a synchronous channel $R < I(P_{i^\star}, Q)$ and the conditions to drive the lower bound given by (12) and (13) to zero (the same lower bound holds for the average probability of error over all hypothesis), the proof is complete. Moreover, since the converse bounds matches our achievability bounds, using only one code distribution $P_{i^\star}$ in all blocks is optimal. ■

What makes this proof more challenging than that in prior work [6] is that we have to find the *exact decay rate* and the *exact trade off* for missed detections and false alarms and here we used the optimal maximal likelihood decoder to do so.

This result show that an exponential number of transmissions is possible at the expense of a reduced rate and/or reduced asynchronous window length compared to the case of only one transmission.

We should note that if we customize the achievability for the case of synchronization only, $R = 0$, we do not need to have a different sync pattern for each block. Using the same sync pattern in every block suffices to drive the probability of error in the synchronization stage to zero and since it matches the converse, it is optimal. In addition for $R = 0$, the capacity region in (2) is equivalent to

$$
\mathcal{R}^{\text{synch}} := \bigcup_{x\in\mathcal{X},\lambda\in[0,1]} \begin{Bmatrix} \nu < \alpha \\ \alpha < D(Q_\lambda \parallel Q_\star) \\ \nu < D(Q_\lambda \parallel Q_x) \end{Bmatrix}, \qquad (14)
$$

the proof of which can be found in [8] and is roughly as follows. $\mathcal{R}^{\text{synch}} \subseteq \mathcal{R}|_{R=0}$ is trivial. If $\mathcal{R}|_{R=0} \not\subseteq \mathcal{R}^{\text{synch}}$, then there exists an element $r^\star \in \mathcal{R}|_{R=0}, r^\star \notin \mathcal{R}^{\text{synch}}$, that is, which lies above all the $(D(Q_\lambda \parallel Q_\star), D(Q_\lambda \parallel Q_x))$ curves for all $x \in \mathcal{X}$. In addition, it can be proved as in [8] that the

$(D(Q_\lambda \parallel Q_\star|P), D(Q_\lambda \parallel Q|P))$ curve characterized by $\lambda \in [0,1]$ is the lower envelope of the set of curves

$$\{(D(Q_{\lambda_x} \parallel Q_\star|P), D(Q_{\lambda_x} \parallel Q|P))\}_{x\in\mathcal{X}},$$

which are each characterized by $\{\lambda_x \in [0,1]\}_{x\in\mathcal{X}}$, which is a contradiction. This implies that depending on the value of α and ν, using a repetition sync pattern with a single symbol is optimal. This symbol may change depending on the considered value of α and ν.

III. Conclusion

In this paper we find the exact trade-off between (R, α, ν) in a strongly asynchronous channel where a user transmits a randomly selected message among $M_n = e^{nR}$ messages in each one of the $K_n = e^{n\nu}$ randomly selected blocks of the available $A_n = e^{n\alpha}$ blocks.

Acknowledgement The work of the authors was partially funded by NSF under award 1422511. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the NSF.

Appendix

A. Calculation of $Q_{x^n}[L(Y^n, x^n) < T]$ and $Q_{\star^n}[L(Y^n, x^n) \geq T]$

For every sequence x^n with composition P and every distribution Q_{x^n} on $\mathcal{Y}^n$ we have

$$\begin{aligned}
&Q_{x^n}\left[\frac{1}{n}\log\frac{Q(Y^n|x^n)}{Q_{\star^n}(Y^n)} < T\right] \\
&= \sum_{\substack{\widehat{Q}:\\ D(\widehat{Q}\|Q_\star|P)-D(\widehat{Q}\|Q|P)<T}} Q_{x^n}\left[Y^n \in T_{\widehat{Q}}(x^n)\right] \\
&= \sum_{\substack{\widehat{Q}:\\ D(\widehat{Q}\|Q_\star|P)-D(\widehat{Q}\|Q|P)<T}} e^{-nD(\widehat{Q}\|Q|P)} \qquad (15)\\
&\doteq \exp\left[-n \min_{\substack{\widehat{Q}:\\ D(\widehat{Q}\|Q_\star|P)-D(\widehat{Q}\|Q|P)<T}} D(\widehat{Q} \parallel Q|P)\right], \qquad (16)
\end{aligned}$$

where (15) is by [9, Lemma 2.6] and (16) is due to the fact that $\left|\widehat{Q} : D(\widehat{Q} \parallel Q_\star|P) - D(\widehat{Q} \parallel Q|P) < T\right|$ is only polynomial in n [9, Lemma 2.2] and hence the overall sum is quantified by the term with minimum exponent. We find the solution of the optimization in (16), by solving the equivalent Lagrangian function which is defined as

$$\mathcal{J}(\widehat{Q}) := \sum_{x,y} P(x)\widehat{Q}(y|x)\log\frac{\widehat{Q}(y|x)}{Q(y|x)} - \nu(\sum_{x,y} P(x)\widehat{Q}(y|x) - 1) \\ -\lambda\left(\sum_{x,y} P(x)\widehat{Q}(y|x)\log\frac{Q(y|x)}{Q_\star(y)} - T\right).$$

By differentiating $\mathcal{J}$ with respect to $\widehat{Q}$ and setting it equal to zero, we get

$$Q^{\text{opt}}(y|x) = Q_\lambda(y|x) = \frac{Q(y|x)^\lambda Q_\star(y)^{1-\lambda}}{\sum_{y\in\mathcal{Y}} Q(y|x)^\lambda Q_\star(y)^{1-\lambda}}, \qquad (17)$$

where $\lambda : D(\widehat{Q} \parallel Q_\star|P) - D(\widehat{Q} \parallel Q|P) = T$.

By (16) and (17) and with a similar argument we get

$$Q_{x^n}\left[\frac{1}{n}\log\frac{Q(Y^n|x^n)}{Q_\star(Y^n)} < T\right] \doteq e^{-nD(Q_\lambda\|Q|P)}, \qquad (18)$$

$$Q_{\star^n}\left[\frac{1}{n}\log\frac{Q(Y^n|x^n)}{Q_{\star^n}(Y^n)} \geq T\right] \doteq e^{-nD(Q_\lambda\|Q_\star|P)}. \qquad (19)$$

B. Proof of (12)

$$\begin{aligned}
&\mathbb{P}\left[\bigcup_{i\in[1:K_n]} \frac{1}{n}\log\frac{Q(Y_i^n|x_i^n(m_i))}{Q_{\star^n}(Y_i^n)} < T\right] \qquad (20)\\
&= \mathbb{P}[Z_1 \geq 1] \qquad (21)\\
&\geq 1 - \frac{\text{Var}[Z_1]}{\mathbb{E}^2[Z_1]} \qquad (22)\\
&\geq 1 - e^{-n\left(\nu - D\left(Q_{\lambda_{i}^\star}\|Q|P_{i^\star}\right)\right)}, \qquad (23)
\end{aligned}$$

where we have defined

$$Z_1 := \sum_{i=1}^{K_n} \xi_i, \qquad \xi_i \sim \text{Bernoulli}\left(e^{-nD(Q_{\lambda_i}\|Q|P_i)}\right).$$

Equation (21) is by equivalence of the events to the one in (20) and where the inequality in (22) is by [7, Appendix 8A] and (23) is by the choice of $i^\star$ in (9).

C. Proof of (13)

To find a lower bound on the term in (11), we proceed as before by writing

$$\begin{aligned}
&\mathbb{P}\left[\bigcup_{j\in[K_n+1:A_n]}\bigcup_{m\in[1:M_n]} \frac{1}{n}\log\frac{Q\left(Y_j^n|x_j^n(m)\right)}{Q_{\star^n}(Y_j^n)} \geq T\right] \qquad (24)\\
&= \mathbb{P}[Z_2 \geq 1] \qquad (25)\\
&\geq 1 - \frac{\text{Var}[Z_2]}{\mathbb{E}^2[Z_2]} \qquad (26)\\
&\geq e^{-n\left[(R+\alpha)-(1+6\delta_1)D\left(Q_{\lambda_{i^\star}}\|Q_\star|P_{i^\star}\right)\right]},
\end{aligned}$$

where we have defined

$$Z_2 := \sum_{j\in[K_n+1:A_n]} \zeta_j, \qquad \zeta_j \sim \text{Bernoulli}(q_j),$$

$$q_j := Q_{\star^n}\left[\bigcup_{m\in[1:M_n]} \frac{1}{n}\log\frac{Q(Y_j^n|x_j^n(m))}{Q_{\star^n}(Y_j^n)} \geq T\right],$$

$$e^{nR}e^{-nD(Q_{\lambda_j}\|Q_\star|P_j)} \geq q_j \geq e^{nR}e^{-nD(Q_{\lambda_j}\|Q_\star|P_j)(1+3\delta_1)}. \qquad (27)$$

The equality in (25) is true because the two events in the probabilities are the same, inequality (26) is again by [7, Appendix 8A]. The first inequality in (27) is by the simple union bound and (19), and the second inequality in (27) is proved in Appendix D. We should again note that $\zeta_j, j \in [K_n+1 : A_n]$, are independent since $Y_j^n, j \in [K_n+1 : A_n]$ are independent.

D. Lower bound on $Q_{\star^n}\left[\bigcup_{m\in[1:M_n]} \frac{1}{n}\log\frac{Q(Y_j^n|x^n(m))}{Q_{\star^n}(Y_j^n)} \geq T\right]$

We first define a new typical set $T^{\delta}_{Q^n_{\lambda+\epsilon}}$ as follows.

Definition 1. *For ϵ and δ define*

$$T^{\delta}_{Q^n_{\lambda+\epsilon}}(x^n) := \Bigg\{ y^n : \left|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\left\{\substack{x_i=a\\ y_i=b}\right\}} - P(a)Q_{\lambda+\epsilon}(b|a)\right| < \delta, \quad \forall (a,b)\in\mathcal{X}\times\mathcal{Y}, \sum_{a,b}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\left\{\substack{x_i=a\\ y_i=b}\right\}}\log\frac{Q(b|a)}{Q_\star(b)} \geq T \Bigg\}.$$

The new constraint $\sum_{a,b}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\left\{\substack{x_i=a\\ y_i=b}\right\}}\log\frac{Q(b|a)}{Q_\star(b)} \geq T$ that we included in the typical set definition ensures that all the sequences y^n that belong to $T^{\delta}_{Q^n_{\lambda+\epsilon}}$ will also satisfy $\log\frac{Q(y^n|x^n)}{Q_{\star^n}(y^n)} \geq T$.

In addition, define

$$\Delta := \sum_{a,b} P(a)Q_{\lambda+\epsilon}(b|a)\log\frac{Q(b|a)}{Q_\star(b)} - T,$$

where $\Delta > 0$ since $T = \sum_{a,b} P(a)Q_{\lambda}(b|a)\log\frac{Q(b|a)}{Q_\star(b)}$ is decreasing in λ [10]. By the Law of Large Numbers

$$Q^n_{\lambda+\epsilon}\left[\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\left\{\substack{x_i=a\\ Y_i=b}\right\}} - P(a)Q_{\lambda+\epsilon}(b|a)\right| > \delta | x^n\right] \to 0,$$

$$Q^n_{\lambda+\epsilon}\left[\sum_{a,b}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\left\{\substack{x_i=a\\ Y_i=b}\right\}}\log\frac{Q(b|a)}{Q_\star(b)} \geq T | x^n\right] \to 0$$

and hence for any $\delta_1 > 0$ there exists n_1 such that for all $n \geq n_1$ we have

$$Q^n_{\lambda+\epsilon}\left[T^{\delta}_{Q_{\lambda+\epsilon}}(x^n)|x^n\right] > 1-\delta_1. \tag{28}$$

We now state a relation between optimal decoding regions and a suboptimal decoder. Given a set of codewords $(x^n(1),\ldots,x^n(M_n))$ and the set of output sequences y^n, we denote by $D_{Q^n_{\lambda+\epsilon}}(m)$, the optimal (and disjoint) decoding region of $x^n(m)$ for channel $Q^n_{\lambda+\epsilon}$ which corresponds to the Maximum A Posteriori (MAP) decoder. The MAP decoder minimizes the average probability of error among different messages and any other suboptimal decoder will have (on average) a larger probability of error, i.e.,

$$\begin{aligned} P_e^{(n),\text{opt}} &= \sum_{m=1}^{M_n}\frac{1}{M_n}Q^n_{\lambda+\epsilon}\left[Y^n \notin D_{Q^n_{\lambda+\epsilon}}(m)|x^n(m)\right] \\ &\leq \sum_{m=1}^{M_n}\frac{1}{M_n}Q^n_{\lambda+\epsilon}\left[Y^n \notin T^{\delta}_{Q^n_{\lambda+\epsilon}}(x^n(m))\,|x^n(m)\right] \leq \delta_1 \end{aligned} \tag{29}$$

where the inequality in (29) is by (28). Now, if we drop half of the codewords in $(x^n(1),\ldots,x^n(M_n))$ with the largest probability of the error, the remaining half must must all satisfy

$$Q^n_{\lambda+\epsilon}\left[Y^n \notin D_{Q^n_{\lambda+\epsilon}}(m)|x^n(m)\right] < 2\delta_1; \tag{30}$$

otherwise, the average probability of error for the decoding regions $D_{Q^n_{\lambda+\epsilon}}(m)$ will be larger than δ_1 and we reach a contradiction.

We now restrict our attention to this half of the codebook (which without loss of generality we assume is the first $\frac{M_n}{2}$ codewords) and hence by (28) and (30) for any $x^n(m) \in \left(x^n(1),\ldots,x^n(\frac{M_n}{2})\right)$ we have

$$Q^n_{\lambda+\epsilon}\left[T^{\delta}_{Q^n_{\lambda+\epsilon}}(x^n(m)) \cap D_{Q^n_{\lambda+\epsilon}}(m)|x^n(m)\right] \geq 1-3\delta_1. \tag{31}$$

Therefore, we can write

$$\begin{aligned} &Q_{\star^n}\left[\bigcup_{m\in[1:M_n]}\frac{1}{n}\log\frac{Q(Y_i^n|x^n(m))}{Q^n_\star(Y_i^n)} \geq T\right] \\ &\geq Q_{\star^n}\left[\bigcup_{m\in[1:\frac{M_n}{2}]} T^{\delta}_{Q^n_{\lambda+\epsilon}}(x^n(m))\right] \\ &\geq Q_{\star^n}\left[\bigcup_{m\in[1:\frac{M_n}{2}]} T^{\delta}_{Q^n_{\lambda+\epsilon}}(x^n(m)) \cap D_{Q^n_{\lambda+\epsilon}}(m)\right] \\ &= \sum_{m=1}^{\frac{M_n}{2}} Q_{\star^n}\left[T^{\delta}_{Q^n_{\lambda+\epsilon}}(x^n(m)) \cap D_{Q^n_{\lambda+\epsilon}}(m)\right] \\ &= \frac{1}{2}e^{nR}e^{-nD(Q_{\lambda+\epsilon}\|Q_\star|P)(1+3\delta_1)} \end{aligned} \tag{32}$$

where (32) is by [9, Eq. 5.21] and (31).

In addition, due to continuity of the divergence, as $\epsilon \to 0$, we have $D(Q_{\lambda+\epsilon} \parallel Q_\star|P) \to D(Q_\lambda \parallel Q_\star|P)$.

REFERENCES

[1] A. Tchamkerten, V. Chandar, and G. W. Wornell, "Asynchronous communication: Capacity bounds and suboptimality of training," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1227–1255, March 2013.

[2] S. Shahi, D. Tuninetti, and N. Devroye, "On the capacity of strong asynchronous multiple access channels with a large number of users," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 1486–1490.

[3] A. Tchamkerten, V. Chandar, and G. Caire, "Energy and sampling constrained asynchronous communication," *IEEE Transactions on Information Theory*, vol. 60, no. 12, pp. 7686–7697, Dec 2014.

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

[5] Y. Polyanskiy, "Asynchronous communication: Exact synchronization, universality, and dispersion," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1256–1270, March 2013.

[6] V. Chandar, A. Tchamkerten, and D. Tse, "Asynchronous capacity per unit cost," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1213–1226, March 2013.

[7] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Camrbidge University Press, 2012.

[8] S. Shahi, D. Tuninetti, and N. Devroye, "On the Capacity of the Slotted Strongly Asynchronous Channel with a Bursty User," *arXiv:1701.08783v2*, May 2017.

[9] I. Csiszar and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[10] R. Blahut, "Hypothesis testing and information theory," *IEEE Transactions on Information Theory*, vol. 20, no. 4, pp. 405–417, Jul 1974.

978-1-5090-3097-2/17/$31.00 ©2017 IEEE