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

In this paper, the trade-off between the number of transmissions (or burstiness) $K_n = e^{n\nu}$ of a user, the asynchronism level $A_n = e^{n\alpha}$ in a slotted strongly asynchronous channel, and the ability to distinguish $M_n = e^{nR}$ messages per transmission with vanishingly error probability is investigated in the asymptotic regime as $n \to \infty$. The receiver must locate and decode, with vanishing error probability in $n$, all of the transmitted messages. The optimal trade-off among $(R, \alpha, \nu)$ is derived. A second model for a bursty user with random access in which the user may access and transmit a message in each block with probability $e^{-n\beta}$ is then considered. The optimal trade-off between $(R, \alpha, \beta)$ is also characterized.

I. Introduction

It is widely believed that Machine-type Communications and Internet of Things are going to be the next dominant paradigm in wireless technology. The traffic pattern imposed by the devices within these networks have unique features different from the ones in human-type communication networks. The communications that take place within these networks is often sporadic and bursty, but must nonetheless 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 asynchronous data bursts of a single user. This extends the authors' work in [1]. In conventional methods the user transmits 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 known to be sub-optimal \[2\]. In this work we do not enforce the usage of pilot symbols, and the codebook serves the dual purpose of synchronization and data transfer. This paper’s central goal is to characterize the trade-off between the reliable transmission rate between one transmitter and one receiver, the burstiness of that transmitter, and the level of asynchronism.

A. Past Work

The problem considered here generalizes the one in \[3\] Remark 3. In \[3\], the authors adopted the ‘per-user’ error criterion and aimed to only recover a large fraction of the transmissions. 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 with 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 \(K_n = 1\), which corresponds to \(\nu = 0\)) in a window of length \(A_n = e^{n\alpha}\) of \(n\) channel uses. They showed that for any \(\alpha\) below the synchronization threshold, \(\alpha_0\), the user can detect the location of the synchronization pattern. In addition they showed that a synchronization pattern consisting of the repetition of a single symbol which induces an output distribution with the maximum divergence from the noise distribution, suffices. 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 a sub-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 \(\alpha\) and \(\nu\).

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.
Recently, the *synchronous* Gaussian massive multiple access channel with random access has been modeled in [7] where the number of users is let to grow linearly in the code blocklength and a random subset of users may try to access the channel. Since then, other versions of “massive number of users” have been proposed in [8], [9]. In [10], we studied a multi-user version of the slotted strongly asynchronous model for a discrete memoryless channel where we assumed that $K_n = e^{n\nu}$ 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. Inner bounds on the trade-off between $(R_i, \nu, \alpha), i \in [1 : K_n]$ were derived, but these were not shown to be tight. What renders the presented version of the problem – a single user transmitting multiple times rather than multiple users transmitting once each – 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. Here, we are able to characterize the trade-offs exactly.

**B. Contributions**

In this paper, we bridge the bursty random access channel model with the asynchronous communication and exactly characterize the trade-off between the number of transmissions (or burstiness) $K_n = e^{n\nu}$ of a user, the asynchronism level $A_n = e^{n\alpha}$ in a slotted bursty and strongly asynchronous channel, and the ability to distinguish $M_n = e^{nR}$ messages per transmission with vanishingly error probability as $n \to \infty$. 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:

1) For synchronization and data transmission ($R > 0$), we find the capacity region $(R, \alpha, \nu)$ and we show in our converse that using the same codebook in all transmissions is optimal.

2) For synchronization only ($R = 0$), our proposed sequential decoder achieves the optimal trade-off. Surprisingly, we show that the optimal synchronization pattern is not fixed and may depend on the asynchronism level $\alpha$.

We also consider a slotted bursty and strongly asynchronous random access channel with asynchronous level $A_n = e^{n\alpha}$ where the number of transmissions of the user is not fixed and the
user may randomly with probability $p_n = e^{-n\beta}$ transmit a message, among $M_n = e^{nR}$ possible ones, within each block of $n$ channel uses. In this case, we show:

3) The exact capacity region $(R, \alpha, \beta)$. As it is clear from the capacity region definition, for the case that $p_n = o(n)$, the capacity region is independent of the value of $\beta$. For the case that $p_n = e^{-n\beta}$ however, the asynchronous window length $A_n = e^{n\alpha}$ increases with the increase of $\beta$ as the number of transmissions to be detected decreases.

C. Paper organization

The rest of this paper is organized as follows. The notation used throughout this paper is defined in section I-D. In section II we introduce the slotted bursty and strongly asynchronous channel model with fixed number of transmissions and derive its capacity region. We also find an equivalent capacity region expression for the special case with zero rate (synchronization only). In section III we introduce a model for slotted bursty and strongly asynchronous channel with random number of transmissions and find the its capacity region. Section IV concludes the paper.

D. 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 distributions on $\mathcal{X}$. As a shorthand notation, we also define $Q_{x^n}(.) := Q(.|x^n)$. We use $y^n_j := [y_{j,1}, ..., y_{j,n}]$, and simply $y^n$ instead of $y^n_1$. The empirical distribution of a sequence $x^n$ is

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

where $\mathbb{1}_{\{A\}}$ is the indicator function of the event $A$ and where $N(a|x^n)$ denotes the number of occurrences of letter $a \in \mathcal{X}$ in the sequence $x^n$; 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 $(X, Y) \sim (P, [PQ])$ coupled via $P_{Y|X}(y|x) = Q(y|x)$, $D(P_1 \parallel P_2)$ for the Kullback Leibler divergence between distribution $P_1$ and $P_2$, and
\[ D(Q_1 \| Q_2 | P) := \sum_{x,y} P(x) Q_1(y|x) \log \frac{Q_1(y|x)}{Q_2(y|x)}. \]

The V-shell of the sequence \( x^n \) is defined as
\[
T_V(x^n) := \left\{ y^n : \frac{N(a,b|x^n, y^n)}{N(a|x^n)} = Q(b|a), \forall (a,b) \in (X,Y) \right\},
\]

where \( N(a,b|x^n, y^n) = \sum_{i=1}^n \mathbb{1}_{\{x_i=a\} \cap \{y_i=b\}} \) is the number of joint occurrences of \((a,b)\) in the sequences \( x^n,y^n \).

We also use the notation \( a_n = e^{nb} \) when
\[
\lim_{n \to \infty} \frac{\log a_n}{n} = b.
\]

II. SYSTEM MODEL FOR FIXED NUMBER OF TRANSMISSIONS 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 \((X,Y)\). We also define a noise symbol \( \star \in X \) for which \( Q_\star(y) > 0, \forall y \in Y \).

An \((M,A,K,n,\epsilon)\) code for the slotted bursty and strongly asynchronous discrete memoryless channel with transition probability matrix \( Q(y|x) \) with fixed number of transmissions consists of:

- A message set \( \{1,2,\ldots,M\} \), from which messages are selected uniformly at random.
- Encoding functions \( f_i : [1:M] \to X^n, i \in [1:A], \) where we define \( x^n_i(m) := f_i(m). \)

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 \( x^n_{\nu_i}(m_i) \) in block \( \nu_i \) for \( i \in [1:K] \) and \( \star^n \) in every other block. We denote the chosen blocks and messages as \((\nu_1,m_1),\ldots,(\nu_K,m_K))\).

- A destination decoder function
\[
g(Y^{mA}) = (\widehat{\nu}_1,\widehat{m}_1),\ldots,(\widehat{\nu}_K,\widehat{m}_K)).
\]

The associated average probability of error for the destination decoder function is given by
\[
P_e^{(n)} := \frac{1}{MK\binom{A}{K}} \sum_{(\nu_1,m_1),\ldots,(\nu_K,m_K)} \mathbb{P}[g(y^{mA}) \neq ((\nu_1,m_1),\ldots,(\nu_K,m_K)) | H((\nu_1,m_1),\ldots,(\nu_K,m_K))],
\]

where \( H((\nu_1,m_1),\ldots,(\nu_K,m_K)) \) is the hypothesis that user transmits message \( m_i \) at block \( \nu_i \) with the codebook \( x^n_{\nu_i}(m_i) \), for all \( i \in [1:K] \).
A tuple \((R, \alpha, \nu)\) is said to be achievable if there exists a sequence of codes \((e^n_R, e^n_\alpha, e^n_\nu, n, \epsilon_n)\) with \(P_e^{(n)} \leq \epsilon_n, \epsilon_n \to 0\) as \(n \to \infty\). The capacity region is the set of all possible achievable \((R, \alpha, \nu)\) triplets.

Our main result is as follows.

**Theorem 1.** The capacity region for the slotted bursty and strongly asynchronous discrete memoryless channel with transition probability matrix \(Q(y|x)\) is

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

where

\[
Q_\lambda(\cdot) := \frac{Q_\lambda^1(\cdot)Q_\star^{1-\lambda}(\cdot)}{\sum_{y \in Y} Q_\lambda^1(y)Q_\star^{1-\lambda}(y)}.
\]

**Proof: Achievability.** Codebook generation. The user generates \(A_n\) constant composition – where constant composition means that all codewords have the same empirical distribution – i.i.d. random codebooks, of rate \(R\) and blocklength \(n\), according to the distribution \(P\), one for each available block. **Decoder.** We perform a two-stage decoding. First, the decoder finds the location of the transmitted codewords (first stage, the synchronization stage) and it decodes the messages (second stage, the decoding stage). The probability of error for this two-stage decoder is given by

\[
P_e^{(n)} \leq \mathbb{P}[^{\text{synchronization error}}] + \mathbb{P}[^{\text{decoding error}|^{\text{no synchronization error}}}].
\]

For the first stage, fix a \(T : -D(Q_\star \parallel Q | P) \leq T \leq D(Q \parallel Q_\star | P)\) (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

\[
L(y^n_j, x^n(m)) := \frac{1}{n} \log \frac{Q(y^n_j|x^n_j(m))}{Q_\star^n(y^n_j)} \geq T,
\]

declare a codeword transmission block and a noise block otherwise. Given the hypothesis
the probability of the synchronization error in the first stage is given by

\[ P\left[ \text{synch error} \mid H_{((1,m_1),\ldots,(K_n,m_{K_n}))} \right] \]

\[ \leq P\left[ \bigcup_{j=1}^{K_n} \bigcap_{m=1}^{M_n} L\left(Y_j^n, x_j^n(m)\right) < T \mid H_{((1,m_1),\ldots,(K_n,m_{K_n}))} \right] \]  
(6)

\[ + P\left[ \bigcup_{j=K_n+1}^{A_n} \bigcup_{m=1}^{M_n} L\left(Y_j^n, x_j^n(m)\right) \geq T \mid H_{((1,m_1),\ldots,(K_n,m_{K_n}))} \right] \]  
(7)

\[ \leq \sum_{j=1}^{K_n} \sum_{D(V_1||Q_\lambda||P)-D(V_1||Q||P) < T} Q_{x_j^n(m_j)} \left[ Y^n \in T_{V_1} \left( x_j^n(m_j) \right) \right] \] 
(8)

\[ + e^{nR} \sum_{j=K_n+1}^{A_n} \sum_{D(V_2||Q_\lambda||P)-D(V_2||Q||P) \geq T} Q^n \left[ Y^n \in T_{V_2} (x_j^n(m)) \right] \]  
(9)

\[ = e^{nR} e^{-nD(Q_\lambda||Q_\ast||P)} + e^{n(\alpha+R)} e^{-nD(Q_\lambda||Q_\ast||P)} , \]  
(10)

where \( Q_\lambda \) is defined in (5) and \( \lambda : D(Q_\lambda \parallel Q_\ast||P) - D(Q_\lambda \parallel Q||P) = T \) and where (10) is proved in Appendix A.

Conditioning on the ‘no synchronization error’ and having found all \( K_n \) ‘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 analyzing the probability of error in a maximum likelihood decoder. We find a matching (with the achievability) exponentially decaying ‘lower’ bounds on the probability of the missed detection (6) and false alarm (7) error events. By following an argument similar to [3, Eq. 137], we can restrict our attention to constant composition codes. In other words, we assume the use of codes \( x_i^n(\cdot) \) 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

\[
\{ \text{error} \mid H((1, m_1), \ldots, (K, m_K)) \} = \bigcup_{i=1}^{K_n} \left\{ \sum_{i=1}^{K_n} L(Y^n_i, x^n_i(m_i)) \leq \sum_{i=1}^{K_n} L(Y^n_i, x^n_i(\tilde{m}_i)) \right\}
\]

\[
\geq \bigcup_{i \in [1:K_n]} \left\{ L(Y^n_i, x^n_i(m_i)) \leq L(Y^n_j, x^n_j(m)) \right\} \cap \bigcup_{j \in [K_n+1:A_n]} \left\{ L(Y^n_j, x^n_j(m)) \geq T \right\},
\]

(11)

where (11) 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 \( T \in [-D(Q_\lambda \parallel Q \mid P_i^\star), D(Q \parallel Q_\lambda | P_i^\star)] \) where \( i^\star \) is chosen such that

\[
\sup_{i, \lambda: D(Q_\lambda \parallel Q \mid P_i^\star) - D(Q_\lambda \parallel Q \mid P_i^\star) = T} D(Q_\lambda \parallel Q \mid P_i^\star) = D(Q_\lambda^\star \parallel Q \mid P_i^\star).
\]

(13)

The reason for this choice of \( i^\star \) will be become clear later (see (16) and (17)). By (12) we have

\[
P\left[ \text{error} \mid H((1, m_1), \ldots, (K_n, m_{K_n})) \right] \geq P\left[ \bigcup_{i \in [1:K_n]} L(Y^n_i, x^n_i(m)) < T \right] \cdot P\left[ \bigcup_{j \in [K_n+1:A_n]} \bigcup_{m \in [1:M_n]} L(Y^n_j, x^n_j(m)) \geq T \right],
\]

(14)

\[
\geq \left( 1 - e^{-n[\nu - D(Q_\lambda^\star \parallel Q \mid P_i^\star)]]} \right) \cdot \left( 1 - e^{-n[(R + \alpha) - (1 + 6\delta_1) D(Q_\lambda^\star \parallel Q \mid P_i^\star)]]} \right),
\]

(15)

(16)

(17)

where (14) and (15) are due to the independence of \( Y^n_j, j \in [1 : A_n] \) and where (16) and (17) are proved in Appendix B and C respectively. The parameter \( \delta_1 > 0 \) in (17) can be chosen to be arbitrarily small.

With the conventional bound on a synchronous channel \( R < I(P_i^\star, Q) \) and the conditions to drive the lower bound given by (16) and (17) 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 $P_i$, 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. We used the optimal maximal likelihood decoder to do so.

This result shows 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 $K_n = 1$.

Note that adapting the achievability scheme to synchronize only ($R = 0$), we do not need a different synchronization pattern for each block. Using the same synchronization 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.

A. Example

To illustrate the capacity region in Theorem [1], we consider a Binary Symmetric Channel (BSC) $Q$ with cross over probability $\delta$ as it is shown in Figure [1]. We also assume $\star = 0$. For the channel $Q_\lambda$ in (3) we have

$$Q_\lambda(0|0) = 1 - \delta,$$

$$\epsilon_\lambda := Q_\lambda(0|1) = \frac{\delta^\lambda (1 - \delta)^{(1-\lambda)}}{\delta^\lambda (1 - \delta)^{(1-\lambda)} + (1 - \delta)^\lambda \delta^{(1-\delta)}},$$

Fig. 1: Strongly synchronous binary symmetric channel
By changing \( p = \mathbb{P}[x = 0] \in [0, \frac{1}{2}] \) and \( \lambda \in [0, 1] \), we obtain the capacity region shown in Fig 2(a). In addition, the optimal trade off for \( (R, \alpha, \nu = 0) \) can be seen in Fig 2(b) which resembles the one in [5, Fig 1]. The trade off between \( (\alpha, \nu) \) can be seen in Fig 2(c) which has the curvature we expect to see, like the one in Fig 5 in the Appendix.

![Capacity region](image)

Fig. 2: Capacity region of slotted bursty and strongly asynchronous BSC with fixed number of transmissions with cross over probability \( \delta = 0.11 \).

**Theorem 2.** For \( R = 0 \), the capacity region in (2) is equivalent to

\[
\mathcal{R}_{\text{synch}} := \bigcup_{x \in \mathcal{X}, \lambda \in [0,1]} \left\{ \begin{array}{l}
\nu < \alpha \\
\alpha < D(Q_\lambda \parallel Q_*) \\
\nu < D(Q_\lambda \parallel Q_x)
\end{array} \right. \, .
\]  

(18)

**Proof:** \( \mathcal{R}_{\text{synch}} \subseteq \mathcal{R}_{|R=0} \) is trivial since we can restrict the set of distributions \( P \in \mathcal{P}_X \) in \( \mathcal{R}_{|R=0} \) to the distributions with weight one on a single symbol \( x \) and zero weight on all other symbols.

We also prove \( \mathcal{R}_{|R=0} \subseteq \mathcal{R}_{\text{synch}} \) by contradiction and by means of the following Lemma proved in Appendix E.

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

\[
\bigcup_{x \in \mathcal{X}} \{ (D(Q_{\lambda_x} \parallel Q_*|P), D(Q_{\lambda_x} \parallel Q|P)) \},
\]
which are each characterized by $\lambda_x \in [0, 1]$.

We continue the proof by assuming $R|_{R=0} \not\subseteq R^{\text{synch}}$. Then there exists an element

$$r := (r_1, r_2) = (D(Q \parallel Q|P), D(Q_\lambda \parallel Q|P)) \in R|_{R=0},$$

$$(r_1, r_2) \not\in R^{\text{synch}},$$

that is, which lies above all the $\{D(Q_\lambda \parallel Q_\star), D(Q_\lambda \parallel Q_x)\}$ curves for all $x \in \mathcal{X}$. Note that $r$ is in fact in the form of a triplet $(\alpha, \nu, 0)$ but we drop the last element for brevity and because it is clear from context that the last element is zero in this scenario. Hence, for any $x \in \mathcal{X}$, there exists a $\lambda_x$ such that

$$r_1 = D(Q_\lambda \parallel Q|P) > D(Q_{\lambda_x} \parallel Q_\star),$$

$$r_2 = D(Q_\lambda \parallel Q|P) > D(Q_{\lambda_x} \parallel Q_x).$$

As a result

$$D(Q_\lambda \parallel Q|P) > D(Q_{\lambda_x} \parallel Q_\star|P),$$

$$D(Q_\lambda \parallel Q|P) > D(Q_{\lambda_x} \parallel Q|P),$$

which contradicts Lemma\(^{[1]}\) that $(D(Q_\lambda \parallel Q|P), D(Q_\lambda \parallel Q|P))$ is the lower envelope of the set of $\bigcup_{x \in \mathcal{X}} \{(D(Q_{\lambda_x} \parallel Q_\star|P), D(Q_{\lambda_x} \parallel Q|P))\}$ curves and hence the initial assumption that $R|_{R=0} \not\subseteq R^{\text{synch}}$ is not feasible.

Theorem\(^{[2]}\) implies that depending on the value of $\alpha$ and $\nu$, using a repetition synchronization pattern with a single symbol is optimal. This symbol may change depending on the considered value of $\alpha$ and $\nu$. For the ternary channel in Fig\(^{[3(a)]}\) for example, the resulting curves by using symbol $x = 1$ and $x = 2$ are shown in Fig\(^{[3(b)]}\) As it is clear, for the regime $\alpha > 0.356$, symbol $x = 1$ has to be used whereas in the regime $\alpha \leq 0.356$ symbol $x = 2$ has to be used in the synchronization pattern.

III. SYSTEM MODEL FOR RANDOM TRANSMISSIONS AND CAPACITY RESULT

We consider again 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 $Q_\star(y) > 0$, $\forall y \in \mathcal{Y}$.
Fig. 3: Channel with different synchronization pattern symbols for different \((\alpha, \nu)\) regimes.

An \((M, A, p, n, \epsilon)\) code for the *slotted bursty and strongly asynchronous* discrete memoryless channel with transition probability matrix \(Q(y|x)\) with random access is defined as follows.

- A message set \(\{1, 2, \ldots, M\}\), from which messages are selected uniformly at random.
- Encoding functions \(f_i : [1 : M] \rightarrow \mathcal{X}^n, \ i \in [1 : A]\), where we define \(x^n_i(m) := f_i(m)\).
  For each block \(i \in [1 : A]\), the transmitter chooses a message among \(M\) possible ones and transmit \(x^n_i(m_i)\) through the channel with probability \(p\) or remains idle and transmits \(\star^n\) with probability \(1 - p\).
- A destination decoder function
  \[
  g(Y^nA) = ((\nu_1, \hat{m}_1), \ldots, (\nu_k, \hat{m}_k)).
  \]

The associated average probability of error for the destination decoder function is given by

\[
P_e^{(n)} := \sum_{k=1}^{A} \sum_{(\nu_1, m_1), \ldots, (\nu_k, m_k)} \frac{1}{M^k} p^k (1 - p)^{A-k} \mathbb{P}[g(y^nA) \neq ((\nu_1, m_1), \ldots, (m_k, \nu_k))] | H((\nu_1, m_1), \ldots, (\nu_k, m_k))],
\]

where \(H((\nu_1, m_1), \ldots, (\nu_k, m_k))\) is the hypothesis that user transmits message \(m_i\) at block \(\nu_i\) with the codebook \(x^n_{\nu_i}\), for all \(i \in [1 : k]\).

A tuple \((R, \alpha, \beta)\) is said to be achievable if there exists a sequence of codes \((e^{nR} , e^{n\alpha} , e^{-n\beta} , n, \epsilon_n)\) with \(P_e^{(n)} \leq \epsilon_n, \epsilon_n \rightarrow 0\) as \(n \rightarrow \infty\). The capacity region is the set of all possible achievable \((R, \alpha, \beta)\) triplets.
Theorem 3. The capacity region for slotted strongly asynchronous random access channel $Q$ is

$$
R := \bigcup_{\lambda \in [0,1], P \in \mathcal{P}_X} \left\{ \begin{array}{ll}
\alpha + R < D(Q_\lambda \parallel Q, P) \\
\alpha - \beta < D(Q_\lambda \parallel Q|P) \\
R < I(P, Q)
\end{array} \right\}.
$$

Proof: Achievability. The encoder and decoder are the same as the one given for the achievability proof of Theorem 1 except that the number of active blocks is not fixed. We denote $p_n := e^{-n\beta}$ and $H_k$ to be the hypothesis that the user is active in $k$ blocks. By the symmetry of the probability of error among hypotheses with the same number of occupied blocks, we can write

$$
P_e^{(n)} = \sum_{k=0}^{A_n} \binom{A_n}{k} p_n^k (1 - p_n)^{A_n - k} \mathbb{P}[\text{error}|H_k]
$$

$$
\leq \sum_{k=0}^{A_n} \binom{A_n}{k} p_n^k (1 - p_n)^{A_n - k} \mathbb{P}[\text{synchronization error}|H_k] + \sum_{k=0}^{A_n} \binom{A_n}{k} p_n^k (1 - p_n)^{A_n - k} \mathbb{P}[\text{decoding error}|H_k, \text{no synchronization error}].
$$

With similar steps as those in the proof of Theorem 1 we obtain

$$
\mathbb{P}[\text{synchronization error}|H_k] \leq k e^{-nD(Q_\lambda \parallel Q|P)} + e^{nR} (e^{n\alpha} - k) e^{-nD(Q_\lambda \parallel Q, P)},
$$

where $\lambda : D(Q_\lambda \parallel Q, P) - D(Q_\lambda \parallel Q|P) = T$. By (20) and (22), we can write

$$
\sum_{k=0}^{A_n} \binom{A_n}{k} p_n^k (1 - p_n)^{(A_n - k)} \mathbb{P}[\text{synchronization error}|H_k]
$$

$$
\leq e^{n\alpha} e^{-n\beta} e^{-nD(Q_\lambda \parallel Q|P)} + e^{n(R - \beta)} e^{-nD(Q_\lambda \parallel Q, P)},
$$

which goes to zero for

$$
\alpha - \beta < D(Q_\lambda \parallel Q|P),
$$

$$
\alpha + R < D(Q_\lambda \parallel Q, P).
$$

For the decoding stage, with the same strategy as the one in Theorem 1 we obtain the third bound in (19).
**Converse.** The converse argument is also similar to the converse proof of Theorem 1. It can be shown that

\[
P \left[ \text{error} \mid H_k \right] \geq \left( 1 - \frac{e^D(Q_{\lambda^*} \mid Q \mid P_{\lambda^*})}{k} \right) \cdot \left( 1 - \frac{e^{-n[D(1 + 6\delta_1)D(Q_{\lambda^*} \mid Q_{\lambda^*} \mid P_{\lambda^*})]}}{A_n - k} \right).
\]

Hence

\[
P[\text{error}] \geq \sum_{k=1}^{A_n-1} \binom{A_n}{k} \left( e^{-n\beta} \right)^k \left( 1 - e^{-n\beta} \right)^{A_n-k} \left( 1 - \frac{e^{nD_1}}{k} \right) \left( 1 - \frac{e^{n(D_2-R)}}{A_n - k} \right)
\]

\[
\geq 1 - (1 - e^{-n\beta})A_n - e^{-n\beta}A_n - \frac{2e^{nD_1}}{e^{-n\beta}e^{n\alpha}} - \frac{2e^{n(D_2-R)}}{(1 - e^{-n\beta})e^{n\alpha}},
\]

(23)

where

\[
D_1 := D(Q_{\lambda^*} \mid Q \mid P_{\lambda^*}),
\]

\[
D_2 := D(Q_{\lambda^*} \mid Q_{\lambda^*} \mid P_{\lambda^*}),
\]

and where (23) is proved in Appendix F. This retrieves the first two bounds in (19). The third bound in (19) is by the usual bound on the reliable rate of a synchronous channel.

A. Example

We consider the same BSC channel defined in Example II-A and illustrate its capacity region for the slotted bursty and strongly asynchronous channel with random access in Fig 4(a). For values of \( \beta > D(Q \mid Q_{\lambda^*} \mid P) = 2.3527 \), the capacity region is similar to the to the case \( \beta = 2.3527 \) and the surface remains unchanged. This is also apparent in Fig 4(b) where the trade-off between \((\alpha, \beta)\) is depicted. This is in fact obvious from the capacity region in Theorem 3 since for values of \( \beta > D(Q \mid Q_{\lambda^*} \mid P) \) the second bound becomes redundant and the capacity region for \((\alpha, R)\) becomes the same as the one for only one transmission as the one in [5, Fig 1].

IV. CONCLUSION

In this paper we study a slotted bursty and strongly asynchronous discrete memoryless 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. We derive the exact trade-off among \((R, \alpha, \nu)\) by finding matching achievability and converse bounds where we use a sequential decoder in the achievability and an optimal maximum a posteriori decoder in the converse. For the case that the number of transmissions of the user is not fixed and the user may access the channel with probability \( e^{-n\beta} \), we characterize the trade of between \((R, \alpha, \beta)\).
Fig. 4: Capacity region of slotted bursty and strongly asynchronous BSC with random access with cross over probability $\delta = 0.11$.

V. 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 $Q_{x^n}[L(Y^n, x^n) \geq T]$ in (8) and (9)

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

$$Q_{x^n} \left[ \frac{1}{n} \log \frac{Q(Y^n|x^n)}{Q_{x^n}(Y^n)} < T \right] = \sum_{\hat{Q}} \sum_{Y^n \in T_{\hat{Q}}(x^n)} Q_{x^n} \left[ Y^n \right]$$

$$= \sum_{\hat{Q}} \sum_{D(\hat{Q}||Q_1|P) - D(\hat{Q}||Q_1|P) < T} e^{-nD(\hat{Q}||Q_1|P)}$$

$$= \exp \left[ -n \min_{\hat{Q}} \frac{D(\hat{Q} \parallel Q_1|P)}{D(\hat{Q}||Q_1|P) - D(\hat{Q}||Q_1|P) < T} \right],$$

October 20, 2017 DRAFT
where (24) is by [11, Lemma 2.6] and (25) is due to the fact that \[ \hat{Q} : D(\hat{Q} \parallel Q_\ast | P) - D(\hat{Q} \parallel Q | P) < T \] is only polynomial in \( n \) [11, Lemma 2.2]. We find the solution of the optimization in (25), by solving the equivalent Lagrangian function which is defined as

\[
J(\hat{Q}) := \sum_{x,y} P(x) \hat{Q}(y|x) \log \frac{\hat{Q}(y|x)}{Q(y|x)} - \nu(\sum_{x,y} P(x) \hat{Q}(y|x) - 1)
- \lambda \left( \sum_{x,y} P(x) \hat{Q}(y|x) \log \frac{Q(y|x)}{Q_\ast(y)} - T \right).
\]

By differentiating \( J \) with respect to \( \hat{Q} \) and setting it equal to zero, we get

\[
Q_\text{opt}(y|x) = \frac{Q(y|x)^\lambda Q_\ast(y)^{1-\lambda}}{\sum_{y \in \mathcal{Y}} Q(y|x)^\lambda Q_\ast(y)^{1-\lambda}},
\]

where \( \lambda : D(\hat{Q} \parallel Q_\ast | P) - D(\hat{Q} \parallel Q | P) = T \).

By (25) and (26) and with a similar argument we get

\[
Q_x^n \left[ \frac{1}{n} \log \frac{Q(Y^n|x^n)}{Q_\ast(Y^n)} < T \right] \geq e^{-nD(Q_\lambda \parallel Q_\ast | P)},
\]

\[
Q_\ast^n \left[ \frac{1}{n} \log \frac{Q(Y^n|x^n)}{Q_\ast(Y^n)} \geq T \right] \geq e^{-nD(Q_\lambda \parallel Q_\ast | P)}.
\]

**B. Proof of (16)**

The main trick in the proof of (16) is to find an equivalent event and lower bound the probability of that event instead. In this regard, we can write

\[
\mathbb{P} \left[ \bigcup_{i \in [1:K_n]} \frac{1}{n} \log \frac{Q(Y^n_i|x^n_i(m_i))}{Q_\ast(Y^n_i)} < T \right]
= \mathbb{P}[Z_1 \geq 1]
\geq 1 - \frac{\text{Var}[Z_1]}{\mathbb{E}^2[Z_1]}
\geq 1 - e^{-n(\nu - D(Q_{\lambda_i} \parallel Q_\ast | P_\ast))},
\]

where we define

\[
Z_1 := \sum_{i=1}^{K_n} \xi_i, \quad \xi_i \sim \text{Bernoulli} \left( e^{-nD(Q_{\lambda_i} \parallel Q_\ast | P_\ast)} \right).
\]

Equation (30) is by equivalence of the events to the one in (29) and where the inequality in (31) is by [12, Appendix 8A] and (32) is by the choice of \( i^* \) in (13).
C. Proof of (17)

To find a lower bound on the term in (15), we proceed as before by writing

\[
\mathbb{P} \left[ \bigcup_{j \in [K_n+1:A_n]} \bigcup_{m \in [1:M_n]} \frac{1}{n} \log \frac{Q(Y_j^n|x_j^n(m))}{Q_n(Y_j^n)} \geq T \right]
\]

\[
\geq \mathbb{P} [Z_2 \geq 1]
\]

\[
\geq 1 - \frac{\text{Var}[Z_2]}{\mathbb{E}^2[Z_2]}
\]

\[
\geq e^{-n\left[ (R+\alpha) - (1+6\delta_1)D(Q_{\lambda\epsilon} || Q_{\lambda\delta} | P_{\lambda}) \right]},
\]

where we have defined

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

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

\[
e^{-nD(Q_{\lambda\epsilon} || Q_{\lambda\delta} | P_{\lambda})} \geq q_j \geq e^{-nD(Q_{\lambda\epsilon} || Q_{\lambda\delta} | P_{\lambda})(1+3\delta_1)}.
\]

The equality in (34) is true because the two events in the probabilities are the same, inequality (35) is again by [12, Appendix 8A]. The first inequality in (37) is by the simple union bound and (28), and the second inequality in (37) 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 \(Q_{n\epsilon} \left[ \bigcup_{m \in [1:M_n]} \frac{1}{n} \log \frac{Q(Y_j^n|x_j^n(m))}{Q_n(Y_j^n)} \geq T \right]\) in (36)

We first define a new typical set \(T^{\delta}_{Q_{\lambda\epsilon}}\) as follows.

**Definition 1.** For \(\epsilon\) and \(\delta\) define

\[
T^{\delta}_{Q_{\lambda\epsilon}}(x^n) := \left\{ y^n : \sum_{a,b} \frac{1}{n} \mathcal{N}(a,b|x^n,y^n) \log \frac{Q(b|a)}{Q_{\epsilon}(b)} \geq T, \right. \]

\[
\left. \left| \frac{1}{n} \mathcal{N}(a,b|x^n,y^n) - P(a)Q_{\lambda\epsilon}(b|a) \right| < \delta, \forall (a,b) \in \mathcal{X} \times \mathcal{Y} \right\}.
\]

The new constraint \(\sum_{a,b} \frac{1}{n} \mathcal{N}(a,b|x^n,y^n) \log \frac{Q(b|a)}{Q_{\epsilon}(b)} \geq T\) that we included in the typical set definition ensures that all the sequences \(y^n\) that belong to \(T^{\delta}_{Q_{\lambda\epsilon}}\) will also satisfy \(\frac{1}{n} \log \frac{Q(y^n|x^n)}{Q_{n\epsilon}(y^n)} \geq T\).
In addition, define
\[
\Delta := \sum_{a,b} P(a)Q_{\lambda+\epsilon}(b|a) \log \frac{Q(b|a)}{Q_*(b)} - T,
\]
where \( \Delta > 0 \) since \( T = \sum_{a,b} P(a)Q_{\lambda}(b|a) \log \frac{Q(b|a)}{Q_*(b)} \) is decreasing in \( \lambda \) \[13\]. By the Law of Large Numbers
\[
Q_{\lambda+\epsilon}^n \left[ \frac{1}{n} N(a,b|x^n,Y^n) - P(a)Q_{\lambda+\epsilon}(b|a) \right] > \delta|x^n| \rightarrow 0
\]
and
\[
Q_{\lambda+\epsilon}^n \left[ \sum_{a,b} \frac{1}{n} N(a,b|x^n,Y^n) \log \frac{Q(b|a)}{Q_*(b)} \geq T|x^n| \right] \rightarrow 0
\]
and hence for any \( \delta > 0 \) there exists \( n_1 \) such that for all \( n \geq n_1 \) we have
\[
Q_{\lambda+\epsilon}^n \left[ T_{\epsilon,\delta}^\delta x^n(x^n)|x^n\right] > 1 - \delta_1.
\]
(38)

We now state a relation between the 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 \(D_{Q_{\lambda+\epsilon}}^n(m)\), the optimal (and disjoint) decoding region of \(x^n(m)\) for channel \(Q_{\lambda+\epsilon}^n\) 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.,
\[
P_{e}^{(n),\text{opt}} = \sum_{m=1}^{M_n} \frac{1}{M_n} Q_{\lambda+\epsilon}^n \left[ Y^n \not\in D_{Q_{\lambda+\epsilon}}^n(m) | x^n(m) \right]
\leq \sum_{m=1}^{M_n} \frac{1}{M_n} Q_{\lambda+\epsilon}^n \left[ Y^n \not\in T_{\epsilon,\delta}^{\delta} (x^n(m)) | x^n(m) \right] \leq \delta_1
\]
(39)
where the inequality in (39) is by (38). 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
\[
Q_{\lambda+\epsilon}^n \left[ Y^n \not\in D_{Q_{\lambda+\epsilon}}^n(m) | x^n(m) \right] < 2\delta_1;
\]
(40)
otherwise, the average probability of error for the decoding regions \(D_{Q_{\lambda+\epsilon}}^n(m)\) will be larger than \(\delta_1\) 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 (38) and (40) for any \(x^n(m) \in\)
we have

\[ Q_{\lambda+\epsilon}^n \left[ T_{Q_{\lambda+\epsilon}^n}^\delta \left( x^n(m) \right) \cap D_{Q_{\lambda+\epsilon}^n}^\epsilon (m) \right] \geq 1 - 3\delta_1. \]  

(41)

Therefore, we can write

\[
Q_{\star}^n \left[ \bigcup_{m \in [1:MN_2]} \frac{1}{n} \log \frac{Q(Y^n_i|x^n(m))}{Q_{\star}^n(Y^n_i)} \geq T \right] 
\geq Q_{\star}^n \left[ \bigcup_{m \in [1:MN_2]} T_{Q_{\lambda+\epsilon}^n}^\delta \left( x^n(m) \right) \right] 
\geq Q_{\star}^n \left[ \bigcup_{m \in [1:MN_2]} T_{Q_{\lambda+\epsilon}^n}^\delta \left( x^n(m) \right) \cap D_{Q_{\lambda+\epsilon}^n}^\epsilon (m) \right] 
= \sum_{m=1}^{MN_2} Q_{\star}^n \left[ T_{Q_{\lambda+\epsilon}^n}^\delta \left( x^n(m) \right) \cap D_{Q_{\lambda+\epsilon}^n}^\epsilon (m) \right] 
= \frac{1}{2} e^{nR} e^{-nD(Q_{\lambda+\epsilon}^n \parallel Q_{\star}^n)(1+3\delta_1)} 
\]  

(42)

where (42) is by [11, Eq. 5.21] and (41).

In addition, due to continuity of the divergence, as \( \epsilon \to 0 \), we have

\[ D(Q_{\lambda+\epsilon} \parallel Q_{\star}^n | P) \to D(Q_{\lambda} \parallel Q_{\star} | P). \]

E. Proof of Lemma 1

We provide the proof for a binary alphabet \( \mathcal{X} = \{a, b\} \) in a proof by contradiction. The proof for the general \( |\mathcal{X}| > 2 \) is a straightforward generalization. For \( x = a, b \) define

\[ E_0^{(x)}(\lambda_x) := D(Q_{\lambda_x} \parallel Q_{\star}), \]
\[ E_1^{(x)}(\lambda_x) := D(Q_{\lambda_x} \parallel Q_x). \]

Assume that the claim of the Lemma [1] is not valid and hence there exists \( (\lambda_a, \lambda_b, \tilde{\lambda}) \in [0, 1]^3 \) such that

\[ D(Q_{\lambda_a} \parallel Q | P) < D(Q_{\tilde{\lambda}} \parallel Q | P), \]
\[ D(Q_{\lambda_b} \parallel Q_{\star} | P) < D(Q_{\tilde{\lambda}} \parallel Q_{\star} | P), \]
or equivalently
\[
\begin{align}
\rho E_1^{(a)}(\lambda_a) + \tilde{\rho} E_1^{(b)}(\lambda_b) &< \rho E_1^{(a)}(\tilde{\lambda}) + \tilde{\rho} E_1^{(b)}(\tilde{\lambda}), \quad (43a) \\
\rho E_0^{(a)}(\lambda_a) + \tilde{\rho} E_0^{(b)}(\lambda_b) &< \rho E_0^{(a)}(\tilde{\lambda}) + \tilde{\rho} E_0^{(b)}(\tilde{\lambda}), \quad (43b)
\end{align}
\]
where \( \rho := \mathbb{P}(x = a) \) and \( \tilde{\rho} = 1 - \lambda = \mathbb{P}(x = b) \). By [13, Theorem 2] we can exclude the cases where \( \lambda_a, \lambda_b < \tilde{\lambda} \) and \( \lambda_a, \lambda_b > \tilde{\lambda} \) and assume \( \lambda_a < \tilde{\lambda} < \lambda_b \), which implies
\[
E_1^{(x)}(\lambda_a) > E_1^{(x)}(\tilde{\lambda}) > E_1^{(x)}(\lambda_b),
\]
\[
E_0^{(x)}(\lambda_a) < E_0^{(x)}(\tilde{\lambda}) < E_0^{(x)}(\lambda_b),
\]
for \( x \in \{a, b\} \). Hence, by rearranging (43) and by dividing the two equations, we get
\[
\frac{E_1^{(a)}(\lambda_a) - E_1^{(a)}(\tilde{\lambda})}{E_0^{(a)}(\lambda_a) - E_0^{(a)}(\tilde{\lambda})} > \frac{E_1^{(b)}(\tilde{\lambda}) - E_1^{(b)}(\lambda_b)}{E_0^{(b)}(\tilde{\lambda}) - E_0^{(b)}(\lambda_b)}. \quad (44)
\]
Note since the \( (E_0^{(x)}(\lambda), E_1^{(x)}(\lambda)) \) curve is convex and strictly decreasing, we have
\[
\frac{\partial E_1^{(a)}(E_0^{(a)}(\lambda))}{\partial \lambda} \bigg|_{\lambda = \tilde{\lambda}} \geq \frac{E_1^{(a)}(\lambda_a) - E_1^{(a)}(\lambda_b)}{E_0^{(a)}(\lambda_a) - E_0^{(a)}(\lambda_b)}, \quad (45)
\]
\[
\frac{E_1^{(b)}(\lambda) - E_1^{(b)}(\lambda_b)}{E_0^{(b)}(\lambda) - E_0^{(b)}(\lambda_b)} \geq \frac{\partial E_1^{(b)}(E_0^{(b)}(\lambda))}{\partial \lambda} \bigg|_{\lambda = \tilde{\lambda}}, \quad (46)
\]
where \( \frac{\partial E_1^{(x)}(E_0^{(x)}(\lambda))}{\partial \lambda} \) is the slope of the \( (E_0^{(x)}(\lambda), E_1^{(x)}(\lambda)) \), which can be visually seen in Fig 5. However, according to [13, Theorem 6], the slope of the \( (E_0^{(x)}(\lambda), E_1^{(x)}(\lambda)) \) curve at \( \lambda = \tilde{\lambda} \) is equal to \( \frac{\tilde{\lambda} - 1}{\lambda} \) and is independent of \( x \).

Putting (44), (45) and (46) together, we reach a contradiction and the proof is complete.
Fig. 5: Slope at $\lambda = \tilde{\lambda}$ is larger than the slope of the line between $\lambda_a$ and $\tilde{\lambda}$.

F. Proof of (23)

Note that

$$\sum_{k=1}^{A_n-1} \binom{A_n}{k} p^k (1-p)^{A_n-k} \frac{1}{k} = \frac{1}{A_n + 1} \sum_{k=1}^{A_n-1} \binom{A_n + 1}{k + 1} p^k (1-p)^{A_n-k} \frac{k + 1}{k}$$

$$\leq \frac{2}{A_n + 1} \sum_{k=1}^{A_n-1} \binom{A_n + 1}{k + 1} p^k (1-p)^{A_n-k}$$

$$\leq \frac{2}{p(A_n + 1)} \sum_{j=0}^{A_n+1} \binom{A_n + 1}{j} p^j (1-p)^{A_n+1-j}$$

$$= \frac{2}{p(A_n + 1)} \leq \frac{2}{pA_n},$$

and similarly

$$\sum_{k=1}^{A_n-1} \binom{A_n}{k} p^k (1-p)^{A_n-k} \frac{1}{A_n - k} \leq \frac{2}{(1-p)A_n}.$$
[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] X. Chen, T. Y. Chen, and D. Guo, “Capacity of gaussian many-access channels,” *IEEE Transactions on Information Theory*, vol. 63, no. 6, pp. 3516–3539, June 2017.

[8] Y. Polyanskiy, “A perspective on massive random-access,” in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 2523–2527.

[9] L. Liu and W. Yu, “Massive connectivity with massive mimo-part i: Device activity detection and channel estimation,” *arXiv preprint arXiv:1706.06438*, 2017.

[10] 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.

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

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

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