Reliable Communication in the Absence of a Common Clock

Raymond W. Yeung,* Ning Cai,† Siu-Wai Ho,‡ and Aaron B. Wagner§

September 10, 2008

Abstract

We introduce the continuous time *asynchronous channel* as a model for time jitter in a communication system with no common clock between the transmitter and the receiver. We have obtained a simple characterization for an optimal zero-error self-synchronizable code for the asynchronous channel. The capacity of this channel is determined by both a combinatorial approach and a probabilistic approach. Our results unveil the somewhat surprising fact that it is not necessary for the receiver clock to re-synchronize with the transmitter clock within a fixed maximum time in order to achieve reliable communication. This means that no upper limit should be imposed on the runlengths of the self-synchronization code as in the case of runlength limited (RLL) codes which are commonly used in magnetic

recording.

1

1   Introduction

In an information storage system, for example, compact disc, magnetic tape, hard disk, or even the classic vinyl record, information is first stored on a physical medium by a recording system and later on retrieved from the medium by a playback system. In almost all such systems, the accuracy of the information reproduced during playback depends critically on the tracking capability of the playback system. In a digital magnetic recording system, only '0' and '1' can be written on the tape. Suppose there is a long run of the same symbol, say '0', on the tape. During playback, in the presence of timing jitter, which can be due to both fluctuation of the speed of the motor drive as well as the tape being elastic, the total number of '0's of the run may be counted incorrectly. As a result, the information stored on the tape is reproduced incorrectly.

An information storage system can be modeled as a communication system, with the storage medium being the communication channel. The recording process is regarded as transmitting a signal at the input of the channel and the playback process is regarded as receiving a signal at the output of the channel. Motivated by the tracking problem in an information storage system, we will introduce in this paper the *asynchronous channel*.

Consider any continuous time communication system[1] where there is no common clock between the transmitter and the receiver. The fundamental problem in such systems is that it is impossible for the transmitter clock and the receiver clock to advance synchronously with respect to each other at all time, and consequently the receiver may not interpret the received signal correctly. The usual way to tackle this problem is to imbed synchronization cues[2] in the input signal at possibly irregular intervals, hoping that the fluctuation between the two clocks is not excessively large so that the two clocks can always be re-synchronized at the cues.

Time jitter is a serious issue which has been addressed and discussed in several previous works in the context of magnetic recording. In the *discrete memoryless jitter channel*

---

[1] All physical communication systems ultimately are continuous time.

[2] A synchronization marker and the end of a runlength are examples of a synchronization cue.

described in [1], a symbol '1' written into time slot n on the tape may be shifted to time slot $n + i$ with probability p_i, where $s_{min} \leq i \leq s_{max}$, with s_{min} and s_{max} being finite integers. The shifts are assumed to occur in an i.i.d. fashion. In [1], lower and upper bounds on the channel capacity are obtained. By considering special cases of the model, stronger lower and upper bounds have been obtained in [2] (which precedes [1]) and [3].

The discrete-time models studied in [1]-[3] suffer from the drawback that the model itself does not prevent a symbol '1' to "overtake" another symbol '1' previously written on the tape, which is physically impossible. In order to make the model meaningful, a minimum separation between two successive '1's written on the tape is imposed. Such a constraint in the channel input inevitably degrades the performance of the system.

In this paper, instead of a discrete-time model, we capture the effect of time jitter at the physical level by using a continuous-time model. The asynchronous channel we will introduce in the next section is a model not only for the tracking problem arising in information storage systems, but also for any continuous time communication system with no common clock between the transmitter and the receiver. A main contribution of this work is to show that, at least in the absence of noise, it is actually not necessary to impose an upper bound on the time to re-synchronize the transmitter and the receiver clocks in order to achieve reliable communication. This seemingly counter-intuitive result is established in Section 3. Based on an optimal code discussed in Section 3, we determine the capacity of the asynchronous channel in Sections 4 and 5 by a combinatorial approach and a probabilistic approach, respectively. In Section 4, we also obtain an upper bound on the rate loss when we impose the constraint that the transmitter and receiver clocks cannot be *out-of-sync* for a fixed maximum time. Sections 6 to 8 contain proofs of theorems. Concluding remarks are in Section 9.

2   The Asynchronous Channel

In this section, we introduce the *asynchronous channel* as a model for a continuous time asynchronous communication system. The asynchronous channel can be used as a model

for information storage systems such as magnetic recording systems and CD players, or communication systems in which the transmitter and the receiver have no common clock.

Let t be the time index, and let $S(t)$ and $V(t)$ be the input signal and the output signal of the channel, respectively, which are related by $V(t) = S(\tau)$, where

$$\tau(t) = \int_0^t r(u)du \tag{1}$$

and $r(u)$ is a strictly positive function of u, $u \geq 0$. Accordingly, $\tau(0) = 0$, i.e., the transmitter clock and the receiver clock are synchronized to start with. In this case, we say that the asynchronous channel is *noiseless* because it only compresses or stretches the input signal in the time domain without changing its values.

In the context of a magnetic recording system, $r(u)$ may be regarded as the motor speed of the tape drive at time u during playback. Here, the assumption that $r(u)$ is strictly positive means that while the motor speed can vary, the motor is at any moment rotating in the forward direction and can never stop. If $r(u) = 1$ for all $u \geq 0$, we have $V(t) = S(t)$, and we say that the system is synchronous. More generally, we say that the system is synchronous if $r(u)$ is equal to some known constant $c > 0$, for then $S(t)$ can be fully recovered from $V(t)$ as $V(\frac{t}{c})$. When $r(u)$ is not equal to a known constant, we say that the system is asynchronous, and we refer to this phenomenon as *time jitter*.

In the context of a communication system with no common clock, the function $\tau(t)$ can be regarded as the reading of the receiver clock when the reading of the transmitter clock is t. Here, the assumption that $r(u)$ is strictly positive means that the reading of the receiver clock is at any moment advancing with respect to the reading of the transmitter clock.

In our model, we assume that $S(t)$ is a step function which takes values from a finite input alphabet $\mathcal{X} = \{0, 1, 2, \ldots, p-1\}$, where it is assumed that $p \geq 2$ to avoid triviality. An interval on which $S(t)$ takes a constant value is called a *run*, and the lengths of runs are called *runlengths*.

3   The $[\mathcal{A}, \xi]$ Channel and an Optimal Code

In this section, we consider a class of asynchronous channels called the $[\mathcal{A}, \xi]$ channel, where $\mathcal{A}$ is a set of positive real number, and ξ is a real number at least equal to 1. For a technical reason to be explained, we assume that $\mathcal{A}$ is a closed set with the minimum element denoted by d. In a communication session with finite duration, the channel takes as its inputs step functions whose runlengths are elements of the set $\mathcal{A}$, called the *runlength set*. The parameter ξ, called the *jitter ratio*, will be explained in the next paragraph. We will characterize optimal self-synchronizable codes for the $[\mathcal{A}, \xi]$ channel at the end of the section.

Consider an asynchronous channel with the assumption that $r(u)$ can take arbitrary values in the interval $(b^{-1}, a^{-1}]$, where $0 < a < b$. This is a reasonable physical assumption, and as we will see, it also makes the problem mathematically tractable. Then a run of length l in the input is reproduced in the output as a run of length between al and bl. It is a basic observation that a run of length l and a run of length l', where $l < l'$, are always distinguishable in the output if and only if $(b - \epsilon)l < al'$ for any $\epsilon > 0$, or equivalently, $bl \leq al'$. Therefore, we are motivated to define the important parameter $\xi = \frac{b}{a}$, called the *jitter ratio*, so that $bl \leq al'$ becomes $\xi l \leq l'$.

The runlength set $\mathcal{A}$ can be specified according to the physical requirement of the system including the power constraint at the transmitter and the bandwidth constraints at both the transmitter and receiver. For example, $\mathcal{A}$ can be the set of all integers greater than or equal to some integer $d \geq 1$, which corresponds to the constraint that all the runlengths must be integer values at least d.

We have assumed that the runlength set $\mathcal{A}$ is closed. We point out that with this assumption, the set $\mathcal{A}$ cannot contain arbitrarily small real numbers, because otherwise it also contains the value 0 (since $\mathcal{A}$ is closed), a contradiction to the assumption that it only contains positive real numbers.

Let $T > 0$, where T may be infinity, be the duration of a communication session. Let l be the smallest element in $\mathcal{A}$ and consider step functions defined on the interval $[0, T]$

satisfying the following properties:

L1. The duration is at most T.

L2. The runs take values in the input alphabet $\mathcal{X}$.

L3. The lengths of the runs take values in the runlength set $\mathcal{A}$.

A set of step functions satisfying the above conditions is called a *code*, and the step functions are called *codewords* of the code. The class of all such codes is denoted by $\mathcal{C}_L$.

We assume that the receiver can always recognize the end of the communication session from the output signal. As an example, in a magnetic recording system, if the communication session spans the length of the tape, then its end can automatically be detected when the tape is finished during playback.

The assumption L1 implies that the duration of the communication session can be less than T. In the context of a magnetic tape, it means that it is possible to use only part of the tape. We will show in Lemma 1 in the next section that this flexibility in the duration of the communication session (instead of requiring the duration to be exactly equal to T) does not increase the capacity of the system.

Let *Constraint $\mathcal{G}$* be a set of constraints on a code defined as follows. A code $\mathcal{S}$ defined on domain $[0, T]$ is said to satisfy Constraint $\mathcal{G}$ if for all $S(t), S'(t) \in \mathcal{S}$, one of the following holds:

G1. $S(t)$ and $S'(t)$ have different numbers of runs in $\mathcal{A}$;

G2. $S(t)$ and $S'(t)$ have the same number of runs in $\mathcal{A}$ and there exists a run in $\mathcal{A}$ for which they take different values in $\mathcal{X}$;

G3. $S(t)$ and $S'(t)$ have the same number of runs in $\mathcal{A}$ and take the same values in all the runs in $\mathcal{A}$, and there exists an index i such that $\xi l_i \leq l'_i$ or $\xi l'_i \leq l_i$, where $l_i \in \mathcal{A}$ and $l'_i \in \mathcal{A}$ are the lengths of the ith run in $S(t)$ and $S'(t)$, respectively.

Note that a code in $\mathcal{C}_L$ is zero-error if and only if the code satisfies Constraint $\mathcal{G}$.

We refer to an asynchronous channel with runlength set $\mathcal{A}$ and jitter ratio ξ as the $[\mathcal{A}, \xi]$ channel, and refer to a zero-error code in $\mathcal{C}_L$ as an $[\mathcal{A}, \xi]$ self-synchronizable code, or simply an $[\mathcal{A}, \xi]$ code. In this section, we give a characterization of optimal $[\mathcal{A}, \xi]$ codes. In the next two sections, we determine the capacity of the $[\mathcal{A}, \xi]$ channel.

We point out that in the definition of an $[\mathcal{A}, \xi]$ code, though called a self-synchronizable code, has no explicit notion of synchronization except at the end of the communication session. More specifically, when using this code, the transmitter clock and the receiver clock may have no re-synchronization with each other after a long period of time instead of re-synchronizing from time to time with each other as in the usual approach. This novel approach to synchronization is a main contribution of the current work.

An optimal $[\mathcal{A}, \xi]$ code in $\mathcal{C}_L$ is a zero-error code which contains the largest number of codewords. To obtain an optimal code, we first construct a list $\mathcal{L}(\mathcal{A}, \xi) = \{l_1^*, l_2^*, l_3^*, \ldots, \}$ of runlengths by "Construction $(\mathcal{A}, \xi)$" below recursively.

Construction $(\mathcal{A}, \xi)$:

   Step 1   Let l_1^* be the smallest element of the set $\mathcal{A}$.

   Step j   Having taken $l_1^*, l_2^*, l_3^*, \ldots, l_{j-1}^*$, take l_j^* to be the smallest element of the closed set $\mathcal{A} \cap [\xi l_{j-1}^*, \infty)$.

Note that whether the list $\mathcal{L}(\mathcal{A}, \xi)$ so constructed is finite or countably infinite depends on whether the procedure stops in a finite number of steps.

Theorem 1 *The set of all distinct codewords such that all the runlengths in the codeword are in $\mathcal{L}(\mathcal{A}, \xi)$ is an optimal $[\mathcal{A}, \xi]$ code in $\mathcal{C}_L$.*

Remarks:

i) As mentioned before, when an $[\mathcal{A}, \xi]$ code is used, the transmitter and receiver clocks do not re-synchronize with each other from time to time. In the extreme case, the communication session consists of only one runlength, so that re-synchronization occurs only at the end of the session.

ii) Since an $[\mathcal{A}, \xi]$ code is zero-error, at the end of the communication session, the receiver is able to determine correctly the transmitted codeword, and hence the length of each run in the codeword. In general, this cannot be done before the end of the session. However, for an optimal $[\mathcal{A}, \xi]$ code as constructed, since the runlengths are chosen from a set in which all the runlengths are distinguishable after passing through the $[\mathcal{A}, \xi]$ channel, the receiver in fact can determine the length of each run (as transmitted) as soon as the transition at the end of that run is detected in the received signal. This means that at this point in time the transmitter and receiver clocks can re-synchronize with each other. It is somewhat unexpected that this desirable property, which also is possessed by runlength limited (RLL) codes [5] commonly used in magnetic recording and optical communications, can be retained without sacrificing optimality. We note that an RLL code, which imposes an upper limit on the runlengths, is not optimal in general. See the discussion in Example 2.

Example 1 *For $\mathcal{A} = [a, \infty)$, $\mathcal{L}(\mathcal{A}, \xi) = \{a\xi^t : t = 0, 1, 2, \ldots\}$.*

Example 2 *For $\mathcal{A} = \{a, a+1, \ldots, \}$, where a is a positive integer, the members of the list $\mathcal{L}(\mathcal{A}, \xi)$ are given recursively by*

$$l_1^* = a$$

$$l_j^* = \lceil \xi l_{j-1}^* \rceil, \ j = 2, 3, \ldots.$$

Let $p = 2$, and let the logarithms be in the base 2 so that the capacities are expressed in bits per unit time.

i) Let $a = 1$ and $\xi = 7.1/6$. Consider achieving zero-error communication by an RLL code. Such a code is characterized by two parameters $(d+1)$ and $(k+1)$, which are the minimum and maximum runlengths, respectively. Here, an RLL code attaining the maximum rate is the $(0, 5)$ code, which is explained as follows. First, for $0 \le d \le 5$, the maximum k that can be taken is 5 because $(5 + 1) \cdot \frac{7.1}{6} > 7$, which means that if the runlength 6 is used, then the runlength 7 cannot be used. Thus for $0 \le d \le 5$, we should let $d = 0$ and $k = 5$ in order to maximize the runlength set. For $d > 5$, since $(d + 1) \cdot \frac{7.1}{6} > d + 2$, the runlength $d + 2$ cannot be used and therefore k must be equal to d. In other words, only the runlength d can

7

be used, and such a code is obviously inferior to the $(5,5)$ code, which in turn is inferior to the $(0,5)$ code. Hence, the $(0,5)$ code, with runlength set $\{1, 2, \ldots, 6\}$, is the optimal RLL code that can be employed for zero-error communication.

Alternatively, an $[\mathcal{A}, \xi]$ code with runlength set $\mathcal{L}(\mathcal{A}, \xi) = \{1, 2, ..., 6, 8, 10, 12, ...\}$ can be used. The $(0,5)$ RLL code discussed above cannot be optimal because its runlength set is a proper subset of $\mathcal{L}(\mathcal{A}, \xi)$. We denote the rate of a (d, k) RLL code by $C(d, k)$, and we have $C(0,5) = 0.9881$ and $C(0, \infty) = 1$ [5, P.62]. The rate of the $[\mathcal{A}, \xi]$ code as prescribed should be somewhere in between and it is found to be 0.9914 by the results in the next section. When ξ is large, an optimal $[\mathcal{A}, \xi]$ code can perform much better than an RLL code, as to be shown next.

ii) Consider $a = 2$ and $\xi = 1.51$. Then we see that the $(1, 1)$ RLL code is the best possible RLL code for zero-error communication. The rate of this code is equal to 0 because there are only two codewords. Alternatively, an $[\mathcal{A}, \xi]$ code with $\mathcal{L}(\mathcal{A}, \xi) = \{2, 4, 7, 11, ...\}$ can be used, and the rate is 0.4383. In this example, the gain from not imposing an upper limit on the runlengths is infinite!

Proof of Theorem 1 Consider any $[\mathcal{A}, \xi]$ code $\mathcal{S}$ such that the runlengths of all the codewords are not necessarily in $\mathcal{L}(\mathcal{A}, \xi)$. Then it suffices for us to show that there exists an $[\mathcal{A}, \xi]$ code $\mathcal{S}'$ such that the lengths of all the runs in the codewords are in $\mathcal{L}(\mathcal{A}, \xi)$ and that $\mathcal{S}'$ has at least the same number of codewords as $\mathcal{S}$.

For any positive real number $l \geq l_1^*$ in $\mathcal{L}(\mathcal{A}, \xi)$, let $\tilde{l}$ be the largest element in $\mathcal{L}(\mathcal{A}, \xi)$ which is less than or equal to l. To construct the codewords in $\mathcal{S}'$, we consider each codeword in $\mathcal{S}$. If the lengths of all the runs are in $\mathcal{L}(\mathcal{A}, \xi)$, then we leave the codeword unchanged. Otherwise, we change each runlength l in the codeword to $\tilde{l}$. Obviously, the duration of the resulting step function is shorter than the original codeword and the runlengths of all these step functions are in $\mathcal{L}(\mathcal{A}, \xi)$. This forms the code $\mathcal{S}'$.

We now show that $\mathcal{S}'$ is an $[\mathcal{A}, \xi]$ code. Consider two step functions $S_1(t)$ and $S_2(t)$ in $\mathcal{S}$ which become $S_1'(t)$ and $S_2'(t)$ in $\mathcal{S}'$. If $S_1(t)$ and $S_2(t)$ satisfy G1 or G2, then $S_1'(t)$ and $S_2'(t)$ continue to satisfy G1 or G2 accordingly. If $S_1(t)$ and $S_2(t)$ satisfy G3, then $S_1(t)$

and $S_2(t)$ have the same number of runs and take the same values in all the runs, and there exists an index i such that either $\xi l_{1i} \leq l_{2i}$ or $\xi l_{2i} \leq l_{1i}$, where l_{1i} and l_{2i} are the lengths of the ith run in $S_1(t)$ and $S_2(t)$, respectively. Assume $\xi l_{1i} \leq l_{2i}$. (The case $\xi l_{2i} \leq l_{1i}$ can be proved likewise.) Let $\hat{l}$ be the element in $\mathcal{L}(\mathcal{A}, \xi)$ after $\tilde{l}_{1i}$ in Construction $(\mathcal{A}, \xi)$. Now

$$\xi \tilde{l}_{1i} \leq l_{2i} \tag{2}$$

because $\xi \tilde{l}_{1i} \leq \xi l_{1i} \leq l_{2i}$. It then follows from (2) that

$$\hat{l} \leq l_{2i}, \tag{3}$$

because $\hat{l}$ is the smallest element in $\mathcal{A} \cap [\xi \tilde{l}_{1i}, \infty)$ and $l_{2i} \in \mathcal{A} \cap [\xi \tilde{l}_{1i}, \infty)$. From (3), since $\hat{l} \in \mathcal{L}(\mathcal{A}, \xi)$, we have

$$\hat{l} \leq \tilde{l}_{2i}.$$

On the other hand, since $\hat{l}$ is the smallest element in $\mathcal{A} \cap [\xi \tilde{l}_{1i}, \infty)$, we have

$$\xi \tilde{l}_{1i} \leq \hat{l}.$$

Thus we have $\xi \tilde{l}_{1i} \leq \tilde{l}_{2i}$. Hence we have shown that $S_1'(t)$ and $S_2'(t)$ continue to satisfy G3. This completes the proof. $\qquad \square$

4   The Combinatorial Capacity

The combinatorial capacity of discrete noiseless channels has been studied in Shannon's original paper [4] where only integer cases are considered. This subject is treated extensively in [6] and the proofs for the non-integer cases can be found in [7]-[10]. The combinational capacity of the $[\mathcal{A}, \xi]$ channel will be found via the modification of some well-known results. Let l_1 be the smallest element in $\mathcal{A}$ and consider the set of step functions on the interval $[0, T]$ satisfying the following:

E1. The duration is equal to T.

E2. The runs take values in the input alphabet $\mathcal{X}$.

9

E3. The lengths of the runs, except for possibly the last run, take values in $\mathcal{A}$.

E4. If the length of the last run is not in $\mathcal{A}$, then it is less than l_1.

If the length of the last run is not in $\mathcal{A}$, then it is referred to as the *incomplete run*. Again, we call a set of step functions satisfying the above conditions a code, and call the step functions codewords of the code. The class of all such codes is denoted by $\mathcal{C}_E$. Note that $\mathcal{C}_E$ is different from the class of codes $\mathcal{C}_L$ we have defined in the last section on which an $[\mathcal{A}, \xi]$ self-synchronizable code is defined, specifically because of the requirement for the duration of the step functions (E1) and the possible occurrence of the incomplete run (E3 and E4).

For a class of codes, we refer to an optimal code as a code in the class which satisfies Constraint $\mathcal{G}$ and contains the largest number of codewords. When Constraint $\mathcal{G}$ is applied, the incomplete runs in the codewords (if exist) are ignored. In the following, we will show that an optimal code in $\mathcal{C}_E$ and an optimal code in $\mathcal{C}_L$ have the same asymptotic rate as $T \to \infty$. For this reason, we can analyze the asymptotic rate of an optimal code in $\mathcal{C}_E$ instead of $\mathcal{C}_L$.

In the rest of the section, we will use ϕ to denote a code and Φ to denote an optimal code in a specified class of codes. Let $\Phi_E(\mathcal{A})$ be an optimal code in $\mathcal{C}_E$. Let $\Phi_E(\mathcal{L}(\mathcal{A}, \xi))$ be an optimal code in the subclass of codes of $\mathcal{C}_E$ such that the codewords satisfy E1 to E4 with $\mathcal{A}$ replaced by $\mathcal{L}(\mathcal{A}, \xi)$ (note that $\mathcal{A}$ and $\mathcal{L}(\mathcal{A}, \xi)$ have the same smallest element). Similarly, let $\Phi_L(\mathcal{L}(\mathcal{A}, \xi))$ be an optimal code in the subclass of codes of $\mathcal{C}_L$ such that the codewords satisfy L1 to L3 with $\mathcal{A}$ replaced by $\mathcal{L}(\mathcal{A}, \xi)$. By Theorem 1, $\Phi_L(\mathcal{L}(\mathcal{A}, \xi))$ is also an optimal code in $\mathcal{C}_L$. We denote the number of codewords in $\Phi_E(\mathcal{A})$ by $|\Phi_E(\mathcal{A})|$, etc. Note that $|\Phi_E(\mathcal{A})|$, $|\Phi_E(\mathcal{L}(\mathcal{A}, \xi))|$, and $|\Phi_L(\mathcal{L}(\mathcal{A}, \xi))|$ are functions of the length T of the communication session. In the discussion that follows, we will temporarily assume the existence of all the limits which will be justified in Theorem 2.

Lemma 1

$$\lim_{T \to \infty} \frac{1}{T} \log |\Phi_E(\mathcal{A})| = \lim_{T \to \infty} \frac{1}{T} \log |\Phi_E(\mathcal{L}(\mathcal{A}, \xi))| = \lim_{T \to \infty} \frac{1}{T} \log |\Phi_L(\mathcal{L}(\mathcal{A}, \xi))|.$$

10

Proof We first construct a set temporarily denoted by Γ through the following procedure. For each codeword in $\Phi_E(\mathcal{A})$, include the codeword in Γ by first removing the incomplete run if it exists. Note that $|\Gamma| = |\Phi_E(\mathcal{A})|$, and if $\Phi_E(\mathcal{A})$ satisfies Constraint $\mathcal{G}$, then Γ continues to satisfies Constraint $\mathcal{G}$. It follows that Γ is a zero-error code in $\mathcal{C}_L$. Let $\mathcal{A}^*$ be the set of runlengths in Γ and henceforth write Γ as $\phi_L(\mathcal{A}^*)$.

Let $\Phi_L(\mathcal{A}^*)$ be an optimal code in the subclass of codes of $\mathcal{C}_L$ such that all the runlengths in a codeword are in $\mathcal{A}^*$. Therefore, $|\phi_L(\mathcal{A}^*)| \le |\Phi_L(\mathcal{A}^*)|$, and

$$\lim_{T\to\infty} \frac{1}{T} \log |\Phi_E(\mathcal{A})| = \lim_{T\to\infty} \frac{1}{T} \log |\phi_L(\mathcal{A}^*)| \le \lim_{T\to\infty} \frac{1}{T} \log |\Phi_L(\mathcal{A}^*)|. \tag{4}$$

Let

$$k = \left\lfloor \frac{T}{l^*} \right\rfloor, \tag{5}$$

where l^* denotes the smallest element in $\mathcal{A}^*$. Then a codeword in $\Phi_L(\mathcal{A}^*)$ contains at least one and at most k runs. Thus we can partition $\Phi_L(\mathcal{A}^*)$ into k subsets $\{\psi_i(\mathcal{A}^*)\}$ for $i = 1, 2, \ldots, k$ such that a codeword in $\Phi_L(\mathcal{A}^*)$ with i runs belongs to $\psi_i(\mathcal{A}^*)$. Assume $\psi_j(\mathcal{A}^*)$ contains the largest number of codewords. Then we can construct a set with size $|\psi_j(\mathcal{A}^*)|$ by appending any valid suffix to each codeword in $\psi_j(\mathcal{A}^*)$ to make the duration of each codeword equal T. Denote this set of codewords by $\phi'_L(A^*)$, and we will show that $\phi'_L(A^*)$ satisfies Constraint $\mathcal{G}$. For any pair of codewords in $\phi'_L(A^*)$, the pair of codewords in $\psi_j(\mathcal{A}^*)$ from which they are derived cannot satisfy G1 because the numbers of runlengths for both are equal to j, and hence they must satisfy either G2 or G3. Then it is readily seen that if the pair of codewords in $\phi'_L(A^*)$ does not satisfy G1, then they must satisfy either G2 or G3 (because they can be distinguished from each other by their first j runs). This shows that the set of codewords $\phi'_L(A^*)$ satisfies Constraint $\mathcal{G}$. Therefore,

$$\frac{|\Phi_L(\mathcal{A}^*)|}{k} \le |\psi_j(\mathcal{A}^*)| \le |\Phi_E(\mathcal{A})|. \tag{6}$$

By (5), we have

$$\lim_{T\to\infty} \frac{\log k}{T} = 0.$$

11

Therefore, (6) becomes

$$\lim_{T\to\infty}\frac{1}{T}\log|\Phi_L(\mathcal{A}^*)| \le \lim_{T\to\infty}\frac{1}{T}(\log k + \log|\Phi_E(\mathcal{A})|) = \lim_{T\to\infty}\frac{1}{T}\log|\Phi_E(\mathcal{A})|. \tag{7}$$

Combining it with (4), we obtain

$$\lim_{T\to\infty}\frac{1}{T}\log|\Phi_E(\mathcal{A})| = \lim_{T\to\infty}\frac{1}{T}\log|\Phi_L(\mathcal{A}^*)|. \tag{8}$$

Similarly, we can show that

$$\lim_{T\to\infty}\frac{1}{T}\log|\Phi_E(\mathcal{L}(\mathcal{A},\xi))| = \lim_{T\to\infty}\frac{1}{T}\log|\Phi_L(\mathcal{L}(\mathcal{A},\xi))|. \tag{9}$$

By Theorem 1 and the definition of an optimal code in $\mathcal{C}_L$, we have

$$\lim_{T\to\infty}\frac{1}{T}\log|\Phi_L(\mathcal{A}^*)| \le \lim_{T\to\infty}\frac{1}{T}\log|\Phi_L(\mathcal{L}(\mathcal{A},\xi))|. \tag{10}$$

From the definitions of $|\Phi_E(\mathcal{A})|$ and $|\Phi_E(\mathcal{L}(\mathcal{A},\xi))|$, we have

$$\lim_{T\to\infty}\frac{1}{T}\log|\Phi_E(\mathcal{A})| \ge \lim_{T\to\infty}\frac{1}{T}\log|\Phi_E(\mathcal{L}(\mathcal{A},\xi))|. \tag{11}$$

Therefore,

$$\begin{aligned}
\lim_{T\to\infty}\frac{1}{T}\log|\Phi_E(\mathcal{L}(\mathcal{A},\xi))| &= \lim_{T\to\infty}\frac{1}{T}\log|\Phi_L(\mathcal{L}(\mathcal{A},\xi))| \\
&\ge \lim_{T\to\infty}\frac{1}{T}\log|\Phi_L(\mathcal{A}^*)| \\
&= \lim_{T\to\infty}\frac{1}{T}\log|\Phi_E(\mathcal{A})| \\
&\ge \lim_{T\to\infty}\frac{1}{T}\log|\Phi_E(\mathcal{L}(\mathcal{A},\xi))|,
\end{aligned}$$

where the equalities and inequalities follow from (9), (10), (8), and (11), respectively. The lemma is proved. □

From the above lemma, we see that the combinatorial capacity of the $[\mathcal{A},\xi]$ channel, given as $\lim_{T\to\infty}\frac{1}{T}\log|\Phi_L(\mathcal{L}(\mathcal{A},\xi))|$ by Theorem 1, is equal to $\lim_{T\to\infty}\frac{1}{T}\log|\Phi_E(\mathcal{L}(\mathcal{A},\xi))|$. In the rest of the section, we will determine this capacity through the determination of the latter limit.

12

Let $\mathcal{L} = \{l_j : j = 1, 2, \ldots, \}$ be a countable set of positive real numbers such that its members are labeled in increasing order, i.e.,

$$l_1 < l_2 < \ldots < l_j < \ldots, \tag{12}$$

and for a positive real number τ, let the τth segment of $\mathcal{L}$ be

$$\mathcal{L}_\tau := \mathcal{L} \cap [0, \tau]. \tag{13}$$

In general, $\mathcal{L}$ is a set of symbol cost which can be time, energy, or some special constraints related to the system. We, however, focus on $\mathcal{L}$ being a list of runlengths. The results we will developed in the rest of the section apply to general runlength sets $\mathcal{L}$ instead of only to specific runlength sets $\mathcal{L}(A, \xi)$.

The "partition function" technique [9] cannot be applied to obtain the capacity of the $[\mathcal{A}, \xi]$ channel because the definition of a codeword therein is different. More fundamentally, this technique is somewhat restrictive because it requires a condition on the set $\mathcal{L}$ to be satisfied before it can be applied. Let

$$\mathcal{D}(t, \mathcal{L}) := \{\tau < t : \text{there exists at least one step function with duration } \tau$$
$$\text{and all the runlengths are in } \mathcal{L} \}.$$

The partition technique can be applied provided that the set $\mathcal{D}(t, \mathcal{L})$ is not too dense, in the sense that for any integer $t \geq 0$,

$$|\mathcal{D}(t, \mathcal{L})| = O(t^K)$$

for some constant $K \geq 0$. This condition, however, is difficult to verify when $\mathcal{L}$ is countably infinite.

Some related results can be found in [7] but the definitions of capacity in [7] are somewhat different from ours. The simplest way to obtain the capacity of the $[\mathcal{A}, \xi]$ channel is to modify the proof of Proposition 1.1 in [8]. Let $N(T)$ denote the number of step functions in $\Phi_E(\mathcal{L}(\mathcal{A}, \xi))$ defined on the interval $[0, T]$, i.e., $N(T) = |\Phi_E(\mathcal{L}(\mathcal{A}, \xi))|$ (for a general $\mathcal{L}$

13

instead of $\mathcal{L}(\mathcal{A}, \xi)$), where $T \geq l_1$. For $T = l_1$, $N(T) = p$. For $T \geq l_1$, $N(T)$ satisfies

$$N(T) = \sum_{\substack{l \in \mathcal{L} \\ l < T}} (p-1)N(T-l) + \sum_{\substack{l \in \mathcal{L} \\ l = T}} p \cdot 1\{T = l\}$$

with $N(T) = \frac{p}{p-1}$ for $0 < T < l_1$ by convention, and

$$1\{T = l\} = \begin{cases} 1 & \text{if } l = T, \\ 0 & \text{if } l \neq T. \end{cases}$$

The above recursive relation for $N(T), T \geq l_1$ can be written more compactly as

$$N(T) = \sum_{l \in \mathcal{L}} (p-1)N(T-l) \tag{14}$$

by further letting $N(0) = \frac{p}{p-1}$ and $N(T) = 0$ for $T < 0$. Note that with our convention for $N(T), 0 \leq T < l_1$, the summand in the above summation is equal to p whenever $0 \leq T - l < l_1$. This is justified because when constraint $\mathcal{G}$ is applied, an incomplete run in the codeword is ignored.

For a list $\mathcal{L}$, let

$$C[p, \mathcal{L}] := \lim_{T \to \infty} \frac{1}{T} \log N(T)$$

if it exists. All the logarithms in this paper are in the same base.

Definition 1 *The capacity of a channel with runlength set $\mathcal{L}$ is defined as $C[p, \mathcal{L}]$ if it exists.*

Now, we modify the proof given in [8] in order to obtain the capacity of the $[\mathcal{A}, \xi]$ channel. Let

$$a(w) = (p-1) \sum_{l \in \mathcal{L}} w^{-l}. \tag{15}$$

If $|\mathcal{L}|$ is finite, $a(w)$ is bounded for all $w > 0$. Consider $|\mathcal{L}| = \infty$ and let w^* denote the infimum of the positive values w for which $a(w)$ is bounded. Note that $w^* \geq 1$ because $a(1)$ is unbounded.

14

We first show that $a(w)$ is continuous on (w^*, ∞) if $w^* < \infty$. For $w \in (w^*, \infty)$, $a(w)$ is obviously monotonically decreasing and $a(w)$ is now shown to be continuous. For $w \in (w^*, \infty)$, let $w' = \frac{w^*+w}{2}$. Then $f_i(w) = w^{-l_i} < (w')^{-l_i}$. Since $w' > w^*$, $a(w')$ is bounded and $\sum_{i=1}^{n}(w')^{-l_i}$ is convergent to $a(w')/(p-1)$. By applying the Weierstrass M-Test [11], $\sum_{i=1}^{n} f_i(w) = \sum_{i=1}^{n} w^{-l_i}$ is uniformly convergent to $\sum_{i=1}^{\infty} w^{-l_i} = a(w)/(p-1)$ on (w^*, ∞). Together with $f_i(w)$ being continuous on (w^*, ∞) for each integer i, $a(w)/(p-1)$ and also $a(w)$ are continuous on (w^*, ∞).

Now, suppose $w^* < \infty$. We are going to show that there is a unique w_0 such that $a(w_0) = 1$ if $a(w^*) > 1$. If $a(w^*) = \infty$, let $\gamma = 2$; otherwise, let $\gamma = \frac{a(w^*)+1}{2} > 1$. Then there exists $m > 0$ such that

$$(p-1) \sum_{l \in \mathcal{L}_m} (w^*)^{-l} > \gamma.$$

Let $\epsilon > 0$ be sufficiently small such that

$$
\begin{aligned}
a(w^* + \epsilon) &= (p-1) \sum_{l \in \mathcal{L}} (w^* + \epsilon)^{-l} \\
&\geq (p-1) \sum_{l \in \mathcal{L}_m} (w^* + \epsilon)^{-l} \\
&> \gamma \\
&> 1,
\end{aligned}
$$

and let $s^* = w^* + \epsilon$. On the other hand, $a(w)$ converges for any $w \in (w^*, \infty)$, and hence there exists a real number $k > 0$ such that

$$(p-1) \sum_{l \in \mathcal{L} \backslash \mathcal{L}_k} (s^*)^{-l} < 0.5.$$

Then for all $w > s^*$,

$$(p-1) \sum_{l \in \mathcal{L} \backslash \mathcal{L}_k} w^{-l} < 0.5. \tag{16}$$

Therefore, we can find $s > s^*$ such that

$$(p-1) \sum_{l \in \mathcal{L}_k} s^{-l} < 0.5.$$

15

Together with (16), we have $a(s) < 1$. Since $a(w)$ is continuous on (w^*, ∞), $a(w)$ is continuous on $[s^*, s]$. Together with the intermediate value theorem and $a(w)$ being strictly decreasing on $[s^*, s]$, there is a unique w_0 such that $a(w_0) = 1$. Similar to [8], we will show in Section 6 (Proof of Theorem 2) that

$$\lim_{T \to \infty} \frac{1}{T} \log N(T) = \log w_0 \qquad \text{for } a(w^*) > 1 \qquad (17)$$

and

$$\lim_{T \to \infty} \frac{1}{T} \log N(t) = \log w^* \qquad \text{for } a(w^*) \leq 1. \qquad (18)$$

The function $a(w)$ has played an important role in our discussion so far. Let us further introduce the characteristic function

$$\begin{aligned} Q(\mathcal{L}, z) &= 1 - a(z^{-1}) \\ &= 1 - (p-1) \sum_{l \in \mathcal{L}} z^l. \end{aligned}$$

It is then obvious that

$$w_0 \text{ is a root of } a(\cdot) = 1 \iff w_0^{-1} \text{ is a root of } Q(\mathcal{L}, \cdot) = 0. \qquad (19)$$

Note that if $|\mathcal{L}| = \infty$, $Q(\mathcal{L}, z)$ converges for $z = 0$ and diverges for all $z > 1$, but it may or may not converge for $0 < z < 1$. We now define the *radius of convergence*, denoted by r, of $Q(\mathcal{L}, z)$. If $Q(\mathcal{L}, z)$ converges only at $z = 0$, then $r = 0$. Otherwise, r is the unique positive real number such that $Q(\mathcal{L}, z)$ converges for all $z < r$ and diverges for all $z > r$. In fact, we have $r = (w^*)^{-1}$ and $Q(\mathcal{L}, z)$ can be shown to be a decreasing and continuous function on $(0, r)$. If $r > 0$ and $Q(\mathcal{L}, r) < 0$, there is a unique positive real root of $Q(\mathcal{L}, z)$.

From (17) and the discussion that follows, the capacity $C[p, \mathcal{L}]$ can be given in terms of the characteristic function $Q(\mathcal{L}, z)$. This is stated in the corollary of the next theorem which will be proved in Section 6. When $p = 2$, this theorem reduces to Proposition 1.1 in [8].

Theorem 2 *The capacity of a channel with runlength set $\mathcal{L}$ is as follows:*

16

i) If $r = 0$, i.e. $Q(\mathcal{L}, z) = -\infty$ for all $z > 0$, then $C[p, \mathcal{L}] = \infty$.

ii) Suppose $r > 0$. If $Q(\mathcal{L}, r) < 0$, then

$$C[p, \mathcal{L}] = -\log \mu(\mathcal{L}), \tag{20}$$

where $\mu(\mathcal{L})$ is the unique positive real root of $Q(\mathcal{L}, z)$. If $Q(\mathcal{L}, r) \geq 0$,

$$C[p, \mathcal{L}] = -\log r. \tag{21}$$

Corollary 1 *The capacity of the $[\mathcal{A}, \xi]$ channel is*

$$C[p, \mathcal{L}(\mathcal{A}, \xi)] = -\log \mu(\mathcal{L}(\mathcal{A}, \xi)),$$

where $\mu(\mathcal{L}(\mathcal{A}, \xi))$ is the unique positive real root of $Q(\mathcal{L}(\mathcal{A}, \xi), z)$.

Proof We put $\mathcal{L} = \mathcal{L}(\mathcal{A}, \xi)$. In light of Theorem 2 Part ii), it suffices to show that $a(w)$ always converges for $w > 1$. Recall that $\mathcal{L}(\mathcal{A}, \xi) = \{l_1^*, l_2^*, l_3^*, \ldots\}$. From Construction ($\mathcal{A}$, ξ), it is readily checked that $l_i^* \geq l_1^* \xi^{i-1}$ for all i. Therefore, we have for $w > 1$,

$$
\begin{aligned}
a(w) &= (p-1) \sum_{l \in \mathcal{L}(\mathcal{A}, \xi)} w^{-l} \\
&\leq (p-1) \sum_{k=0}^{\infty} w^{-l_1^* \xi^k}.
\end{aligned}
$$

Since $\xi > 1$, there exists an integer K_0 such that $\xi^k > k$ for $k \geq K_0$. For $w > 1$, we have

$$
\begin{aligned}
a(w) &\leq (p-1) \sum_{k=0}^{\infty} w^{-l_1^* \xi^k} \\
&\leq (p-1) \sum_{k=0}^{K_0-1} w^{-l_1^* \xi^k} + (p-1) \sum_{k=K_0}^{\infty} w^{-l_1^* k} \\
&= (p-1) \sum_{k=0}^{K_0-1} w^{-l_1^* \xi^k} + (p-1) \frac{w^{-l_1^* K_0}}{1 - w^{-l_1^*}},
\end{aligned}
$$

which proves that $a(w)$ is bounded from above. Together with $a_\tau(w) = (p-1) \sum_{l \in \mathcal{L}_\tau} w^{-l}$ being an increasing function in τ, $a(w) = \lim_{\tau \to \infty} a_\tau(w)$ converges to a finite value. Therefore, we have shown that $w^* \leq 1$, $r \geq 1$. Since $Q(\mathcal{L}(\mathcal{A}, \xi), r) < 0$, we can apply (20) to obtain the capacity of the $[\mathcal{A}, \xi]$ channel. $\square$

Corollary 2 *If $\mathcal{L}$ is a subset of integers, $Q(\mathcal{L}, z)$ converges on $[0, \theta]$ for all $\theta \in (0, 1)$ and $-\log \mu(\mathcal{L}) \leq \log p$.*

Theorem 2 also provides an easy way to determine $C[p, \mathcal{L}]$ when $Q(\mathcal{L}, z)$ has a closed form.

Example 3 *Consider binary codes whose runlengths are multiples of k. Then $\mathcal{L} = \{k, 2k, 3k, \ldots\}$ and*

$$Q(\mathcal{L}, z) = 1 - \sum_{j=1}^{\infty} z^{kj} = 1 - \frac{z^k}{1 - z^k}$$

for $z \in [0, 1)$. Therefore, $\mu(\mathcal{L}) = 2^{-\frac{1}{k}}$, and $C[p, \mathcal{L}] = 1/k$ bit per unit time.

The following example illustrates the scenario in Theorem 2 Part i).

Example 4 *Consider a channel with runlength set*

$$\mathcal{L} = \{l_1, l_2, \ldots\} = \left\{ 1, \frac{3}{2}, \frac{5}{3}, \frac{7}{4}, \ldots \right\},$$

where $l_i = \frac{2i-1}{i}$ for $i \geq 1$. Then for $0 < z < 1$,

$$
\begin{aligned}
Q(\mathcal{L}, z) &= 1 - (p-1) \sum_{l \in \mathcal{L}} z^l \\
&\leq 1 - (p-1) \sum_{l \in \mathcal{L}} z^2 \\
&= -\infty.
\end{aligned}
$$

By Theorem 2 Part i), the capacity of this channel is infinite.

In general, $Q(\mathcal{L}, z)$ may not have a closed form so that $\mu(\mathcal{L})$ is difficult to compute. In this case, we may need to approximate $\mu(\mathcal{L})$ by $\mu_m(\mathcal{L})$ for a sufficiently large m, where $\mu_m(\mathcal{L})$ is the unique positive root of

$$Q_m(\mathcal{L}, z) = 1 - (p-1) \sum_{l \in \mathcal{L}_m} z^l,$$

and $\mathcal{L}_m$ is defined in (13). Thus the convergence rate of $\{\mu_m(\mathcal{L})\}$ is of practical interest.

18

Theorem 3 (The Convergence Rate) *For any $0 < \theta_1 < \theta_2$ with $Q(\mathcal{L}, \theta_1) \geq 0$ and $Q(\mathcal{L}, \theta_2) < 0$,*

$$0 \leq \mu_m(\mathcal{L}) - \mu(\mathcal{L}) \leq \frac{\sum_{l \in \mathcal{L} \setminus \mathcal{L}_m} {\theta_2}^l}{\sum_{l \in \mathcal{L}'_m} l\mu_m(\mathcal{L})^{l-1} + \sum_{l \in \mathcal{L}''_m} l{\theta_1}^{l-1}}, \tag{22}$$

where $\mathcal{L}'_m = \{l \in \mathcal{L}_m : l \leq 1\}$ and $\mathcal{L}''_m = \{l \in \mathcal{L}_m : l > 1\}$.

The proof of Theorem 3 is deferred to Section 7. In addition to the convergence rate, this theorem has the following physical interpretation. Recall from Remark ii) following Theorem 1 that for a $[\mathcal{A}, \xi]$ code, the transmitter and receiver clocks can always re-synchronize with each other at the end of a runlength. This means that when the set of runlengths is $\mathcal{L}_m$, the transmitter and receiver clocks can always re-synchronize with each other within a horizon of length m. Thus if we impose the constraint that the transmitter and receiver clocks cannot be out-of-sync for more than m time units, the rate loss incurred is at most

$$- \log \mu(\mathcal{L}) + \log \mu(\mathcal{L}_m),$$

which by (22) is upper bounded by

$$- \log(\mu(\mathcal{L}_m) - \delta_m) + \log \mu(\mathcal{L}_m),$$

where δ_m denotes the RHS of (22) and we have invoked the monotonicity of the logarithmic function.

Corollary 3 *If $\mathcal{L}$ is a subset of positive integers and $Q(\mathcal{L}, \theta) < 0$, then*

$$0 \leq \mu_m(\mathcal{L}) - \mu(\mathcal{L}) \leq \frac{\theta^{m+1}}{(1 - \theta)\left(\sum_{l \in \mathcal{L}'_m} l\mu_m(\mathcal{L})^{l-1} + \tilde{d}p^{-(\tilde{d}-1)}\right)}, \tag{23}$$

where $\tilde{d}$ is the smallest element of $\mathcal{L}''_m$.

Proof Observe that

$$Q(\mathcal{L}, p^{-1}) = 1 - (p-1)\sum_{l \in \mathcal{L}} p^{-l} \geq 1 - (p-1)\sum_{l=1}^{\infty} p^{-l} = 0$$

19

since $\mathcal{L}$ is a subset of positive integers. Then the corollary can be proved by applying Theorem 3 with $\theta_1 = p^{-1}$ and $\theta_2 = \theta$ together with the inequalities

$$\sum_{l \in \mathcal{L}''_m} l p^{-(l-1)} \geq \tilde{d} p^{-(\tilde{d}-1)}$$

and

$$\sum_{l \in \mathcal{L} \backslash \mathcal{L}_m} \theta^l \leq \sum_{l=m+1}^{\infty} \theta^l = \frac{\theta^{m+1}}{1-\theta}.$$

$\square$

5   The Probabilistic Capacity

The Maxentropic Theorem for Markov information sources, which is an important result in information theory, can go back to Shannon's original work [4], and the Maxentropic Theorem for (dk) sequences can be obtained as its special case (cf. [5] for a detailed discussion and further references). For our problem, when the list $\mathcal{L}$ of runlengths is an infinite set, the information source has infinite memory and therefore is not Markovian. So the approach for proving the Maxentropic Theorem via transition matrices does not work. In this section, the theorem will be obtained by way of the characteristic function $Q(\cdot, \cdot)$.

The Maxentropic Theorem not only asserts that the probabilistic capacity and the combinatorial capacity are the same for $\mathcal{L}$ being an infinite set of runlengths, but also provides insight into the asymptotic behavior of an optimal code. In fact, this insight has led to the construction of an optimal code recently reported in [12].

In the following, we are only interested in those $\mathcal{L}$ with a positive radius of convergence, i.e. $r > 0$, and $Q(\mathcal{L}, r) < 0$. This means $w^* = r^{-1}$ is finite and $a(w^*) > 1$, from which the existence of $\mu(\mathcal{L})$, the unique positive real root of $Q(\mathcal{L}, z)$, is guaranteed. Note $\mathcal{L}(\mathcal{A}, \xi)$ always satisfy the above requirements which has been shown in the proof of Corollary 1.

Definition 2 *For a given list $\mathcal{L}$ of runlengths and an alphabet $\mathcal{X}$ with cardinality p, let $\mathcal{S}(\mathcal{L})$ be the set of step functions on $[0, \infty)$ taking values in $\mathcal{X}$ whose runlengths are in $\mathcal{L}$, $\mathcal{K}(\mathcal{L})$ be the set of stochastic processes $X(t)$ on $[0, \infty)$ taking values almost surely (a.s.) in*

$\mathcal{S}(\mathcal{L})$, $X^T(t) := \{X(t) : 0 \leq t \leq T\}$ be the Tth segment of stochastic process $X(t)$, and $P_{X(t)}$ and $P_{X^T(t)}$ be the distributions of $X(t)$ and $X^T(t)$, respectively. Then the entropy rate of $X(t) \in \mathcal{K}(\mathcal{L})$ is defined by

$$r(X(t)) := \lim_{T \to \infty} -\frac{1}{T} \mathbf{E} \log P_{X^T(t)}(X^T(t)) \tag{24}$$

and the maxentropy of $\mathcal{K}(\mathcal{L})$ is defined by

$$r(\mathcal{L}) := \max_{X(t) \in \mathcal{K}(\mathcal{L})} r(X(t)). \tag{25}$$

In order for the maxentropy to be well defined, we have to show that the maximum in (25) is achievable. Under a reasonable condition, this will be done in the proof of the next theorem. Specifically, we will show that the stochastic process $X^*(t) \in \mathcal{K}(\mathcal{L})$ defined below achieves the maxentropy.

Any step function $S(t) \in \mathcal{S}(\mathcal{L})$ can be uniquely determined by a pair of sequences, $l^\infty = (l_1, l_2, l_3, \ldots)$ and $y^\infty = (y_1, y_2, y_3, \ldots)$, where $l_i \in \mathcal{L}$ and $y_j \in \mathcal{X}$ are respectively the ith runlength and the value taken by the jth run of $S(t)$. We will write $\Im(S(t)) = (l^\infty, y^\infty)$ and $\Im^{-1}(l^\infty, y^\infty) = S(t)$. Similarly, a stochastic process $X(t) \in \mathcal{K}(\mathcal{L})$ is uniquely determined by a pair of stochastic sequences $L(i)$ and $Y(j)$. Likewise, write $\Im(X(t)) = (L(i), Y(j))$ and $\Im^{-1}(L(i), Y(j)) = X(t)$. This allows us to define $X^*(t)$ by defining $\Im(X^*(t))$. For the list $\mathcal{L}$, $\mu(\mathcal{L})$ satisfies $1 - (p-1)\sum_{l \in \mathcal{L}}(\mu(\mathcal{L}))^l = Q(\mathcal{L}, \mu(\mathcal{L})) = 0$, and so $\sum_{l \in \mathcal{L}}(p-1)(\mu(\mathcal{L}))^l = 1$. Thus $P_{\mu(\mathcal{L})} := \{P_{\mu(\mathcal{L})}(l) = (p-1)(\mu(\mathcal{L}))^l : l \in \mathcal{L}\}$ is a probability distribution over $\mathcal{L}$. Let $\{L^*(i)\}_{i=1}^\infty$ be an i.i.d. stochastic sequence with distribution $P_{\mu(\mathcal{L})}$ and $\{Y^*(j)\}_{j=1}^\infty$ be a Markov chain which is independent of $\{L^*(i)\}_{i=1}^\infty$ and has the following distribution, where a is any fixed letter in $\mathcal{X}$.

$$Pr(Y^*(1) = x) = \begin{cases} 0 & \text{if } x = a \\ \frac{1}{p-1} & \text{otherwise,} \end{cases} \tag{26}$$

and for $j = 2, 3, \ldots,$

$$Pr(Y^*(j) = x_j | Y^*(j-1) = x_{j-1}) = \begin{cases} 0 & \text{if } x_j = x_{j-1} \\ \frac{1}{p-1} & \text{otherwise.} \end{cases} \tag{27}$$

21

Define $X^*(t) = \Im^{-1}(L^*(i), Y^*(j))$. Then $X^*(t) \in \mathcal{K}(\mathcal{L})$. Notice that if instead of (26) we let $Y^*(1)$ be uniformly distributed over $\mathcal{X}$, then the entropy would be increased, but for the entropy rate in (24) which we are interested in, it does not make any difference.

Theorem 4 (Maxentropic Theorem) *For any* $X(t) \in \mathcal{K}(\mathcal{L})$,

$$r(X(t)) \leq -\log \mu(\mathcal{L}). \tag{28}$$

In the above, equality is achieved by $X^*(t)$, *and hence*

$$r(\mathcal{L}) = -\log \mu(\mathcal{L}). \tag{29}$$

6 Proof of Theorem 2

We first prove Part ii) and we let $w^* = r^{-1}$ and w^* is finite. Suppose $Q(\mathcal{L}, r) < 0$ which means $a(w^*) > 1$. We have already shown the existence of w_0 such that $a(w_0) = 1$. Note that $\mu(\mathcal{L}) = (w_0)^{-1}$. In order to show (20), we only need to prove (17). A sequence of positive numbers b_k is defined recursively by

$$b_1 = (p-1)^{-1} w_0^{-l_1}$$

and

$$b_{k+1} = \left[(p-1) \sum_{l \in \mathcal{L}: l \leq k l_1} w_0^{-l} \right] b_k$$

for positive integer k. Then

$$b_k = (p-1)^{-1} w_0^{-l_1} \prod_{j=1}^{k-1} \left[\sum_{l \in \mathcal{L}: l \leq j l_1} (p-1) w_0^{-l} \right]$$

for $k = 1, 2, \ldots$ Similar to [8], we will prove that for $0 \leq T \leq k l_1$,

$$(p-1) b_k w_0^T \leq N(T) \leq p w_0^T \tag{30}$$

by mathematical induction. For $k = 1$ and $0 \leq T \leq l_1$,

$$
\begin{aligned}
(p-1)b_1 w_0^T &\leq (p-1)b_1 w_0^{l_1} \\
&= 1 \\
&< \frac{p}{p-1} \qquad (= N(T) \text{ if } 0 \leq T < l_1) \\
&\leq p \qquad\quad (= N(T) \text{ if } T = l_1) \\
&\leq pw_0^T,
\end{aligned}
$$

since $p \geq 2$ and $w_0 \geq 1$. Therefore, (30) holds. Assume that (30) holds for some positive integer k. Then for $kl_1 < T \leq (k+1)l_1$, we have

$$
\begin{aligned}
(p-1)b_{k+1} w_0^T &= (p-1)^2 b_k \sum_{l \in \mathcal{L}: l \leq kl_1} w_0^{T-l} \\
&\leq (p-1)^2 b_k \sum_{l \in \mathcal{L}} w_0^{T-l} \\
&\leq (p-1) \sum_{l \in \mathcal{L}} N(T-l) \qquad (= N(T)) \\
&\leq (p-1) \sum_{l \in \mathcal{L}} pw_0^{T-l} \\
&= w_0^T p \left[(p-1) \sum_{l \in \mathcal{L}} w_0^{-l} \right] \\
&= pw_0^T a(w_0) \\
&= pw_0^T.
\end{aligned}
$$

Therefore, we conclude that (30) holds for all k and $0 < T \leq kl_1$ (regardless of whether $a(w^*) > 1$ or not). Since

$$
\begin{aligned}
b_{k+1} &= \left[(p-1) \sum_{l \in \mathcal{L}: l \leq kl_1} w_0^{-l} \right] b_k \\
&\leq \left[(p-1) \sum_{l \in \mathcal{L}} w_0^{-l} \right] b_k \\
&= a(w_0)b_k \\
&= b_k,
\end{aligned}
$$

23

we see that b_k is non-increasing with respect to k. Since

$$1 = a(w_0) = (p-1) \sum_{l \in \mathcal{L}: l \leq kl_1} w_0^{-l} + (p-1) \sum_{l \in \mathcal{L}: l > kl_1} w_0^{-l},$$

we have

$$
\begin{aligned}
\sum_{k=1}^{\infty} \left[1 - (p-1) \sum_{l \in \mathcal{L}: l \leq kl_1} w_0^{-l} \right] &= \sum_{k=1}^{\infty} (p-1) \sum_{l \in \mathcal{L}: l > kl_1} w_0^{-l} \qquad (31) \\
&= \sum_{l \in \mathcal{L}} \sum_{1 \leq k < \frac{l}{l_1}} (p-1) w_0^{-l} \\
&\leq \sum_{l \in \mathcal{L}} \frac{l}{l_1} (p-1) w_0^{-l} \\
&= -\frac{w_0}{l_1} a'(w_0),
\end{aligned}
$$

where $a'(w_0)$ denotes the derivative of $a(w)$ at w_0. Since we have assumed that $a(w^*) > 1$ and $a(w)$ is continuous on (w^*, ∞), we have $w_0 > w^*$. Therefore, $a(w)$ converges in the neighborhood of w_0, which implies $-a'(w_0)$ exists and the summation on the LHS of (31) converges. It is easy to verify that for any $0 < c_k < 1$ where $k = 1, 2, \ldots$, if $\sum_{k=1}^{\infty} c_k < \infty$, then $\prod_{k=1}^{\infty} (1 - c_k) > 0$. By letting c_k be the kth term in the summation on the LHS of (31), we have $\sum_{k=1}^{\infty} c_k < \infty$, so that

$$
\begin{aligned}
0 &< \prod_{k=1}^{\infty} (1 - c_k) \\
&= \prod_{k=1}^{\infty} \left((p-1) \sum_{l \in \mathcal{L}: l \leq kl_1} w_0^{-l} \right) \\
&= (p-1) w_0^{l_1} \lim_{k \to \infty} b_k.
\end{aligned}
$$

Since (30) holds for $0 < T \leq kl_1$, for any $t > 0$, it holds for $k = \lceil \frac{T}{l_1} \rceil$. For this choice of k, upon taking logarithm and dividing by t in (30), we conclude that

$$\lim_{T \to \infty} \frac{1}{T} \log N(T) = \log w_0 \qquad \text{for } a(w^*) > 1,$$

i.e., (17).

Now, we are going to show (21) by considering $Q(\mathcal{L}, r) \geq 0$. It is sufficient to prove (18). Consider $a(w^*) \leq 1$. Since $a(w^* - \epsilon)$ is unbounded for any $0 < \epsilon < w^*$, there exists a real

24

number m such that

$$(p-1) \sum_{l \in \mathcal{L}_m} (w^* - \epsilon)^{-l} > 1.$$

As

$$(p-1) \sum_{l \in \mathcal{L}_m} (w^*)^{-l} < a(w^*) \leq 1,$$

we can find $w_0' \in (w^* - \epsilon, w^*)$ such that

$$(p-1) \sum_{l \in \mathcal{L}_m} (w_0')^{-l} = 1.$$

If we consider a channel with runlengths given in $\mathcal{L}_m$, the previous results show that the capacity of the channel is $\log w_0'$. Since we are considering a channel supporting more runlengths as $\mathcal{L}_m \subset \mathcal{L}$, we have

$$\lim_{t \to \infty} \frac{1}{t} \log N(t) \geq \log w_0' \geq \log(w^* - \epsilon),$$

so that upon letting $\epsilon \to 0$, we have

$$\lim_{t \to \infty} \frac{1}{t} \log N(t) \geq \log w^*.$$

On the other hand, we recall from the foregoing that the upper bound on $N(t)$ in (30) also holds for $a(w^*) \leq 1$. Since $w_0 = w^*$ in this case, from the upper bound in (30), we have

$$N(t) \leq p(w^*)^t \tag{32}$$

for all $t > 0$, which implies

$$\lim_{t \to \infty} \frac{1}{t} \log N(t) \leq \log w^*,$$

hence proving (18). The proof of Part ii) is also completed.

For Part i), i.e. $Q(\mathcal{L}, z) = -\infty$ for any $z > 0$, we are going to show that $C[p, \mathcal{L}] = \infty$. For any $s > 0$, let $w_1 = 2^s$. Since $Q(\mathcal{L}, (w_1)^{-1}) = -\infty$, there exists an integer m such that

$$1 - (p-1) \sum_{l \in \mathcal{L}_m} (w_1)^{-l} < 0.$$

25

The previous part of the proof has already shown that the capacity of $C[p, \mathcal{L}_m] > \log w_1 = s \log 2$. Together with $C[p, \mathcal{L}] \geq C[p, \mathcal{L}_m]$, $C[p, \mathcal{L}] > s \log 2$. Since $s > 0$ is arbitrary, we have

$$C[p, \mathcal{L}] = \infty.$$

7 Proof of Theorem 3

Since $Q_m(\mathcal{L}, \cdot)$ is monotonically decreasing (for $0 \leq z < r$), its inverse function $Q_m^{-1}(\mathcal{L}, \cdot)$ exists and has derivative

$$
\begin{aligned}
&\frac{dQ_m^{-1}(\mathcal{L}, u)}{du} \\
&= \left[\frac{dQ_m(\mathcal{L}, z)}{dz} \Big|_{z=Q_m^{-1}(\mathcal{L}, u)} \right]^{-1} \\
&= -\frac{1}{(p-1) \sum_{l \in \mathcal{L}_m} l(Q_m^{-1}(\mathcal{L}, u))^{l-1}}.
\end{aligned}
\tag{33}
$$

Moreover, $\mu(\mathcal{L})$ exists because we have $\theta_2 > 0$ and $Q(\mathcal{L}, \theta_2) < 0$. From

$$0 = Q(\mathcal{L}, \mu(\mathcal{L})) = Q_m(\mathcal{L}, \mu(\mathcal{L})) - (p-1) \sum_{l \in \mathcal{L} \setminus \mathcal{L}_m} \mu(\mathcal{L})^l,$$

we have

$$Q_m(\mathcal{L}, \mu(\mathcal{L})) = (p-1) \sum_{l \in \mathcal{L} \setminus \mathcal{L}_m} \mu(\mathcal{L})^l,
\tag{34}$$

or

$$Q_m^{-1}\left(\mathcal{L}, (p-1) \sum_{l \in \mathcal{L} \setminus \mathcal{L}_m} \mu(\mathcal{L})^l \right) = \mu(\mathcal{L}).$$

On the other hand, since $\mu_m(\mathcal{L})$ is a root of $Q_m(\mathcal{L}, z)$, we have

$$Q_m^{-1}(\mathcal{L}, 0) = \mu_m(\mathcal{L}).$$

Then by Lagrange's mean value theorem and (33), there exists an

$$\eta \in (0, (p-1) \sum_{l \in \mathcal{L} \setminus \mathcal{L}_m} \mu(\mathcal{L})^l),$$

26

such that

$$
\begin{aligned}
0 \ & \le \ \mu_m(\mathcal{L}) - \mu(\mathcal{L}) \\
& = \ Q_m^{-1}(\mathcal{L}, 0) - Q_m^{-1}\left(\mathcal{L}, (p-1) \sum_{l \in \mathcal{L} \setminus \mathcal{L}_m} \mu(\mathcal{L})^l\right) \\
& = \ \frac{1}{(p-1) \sum_{l \in \mathcal{L}_m} l(Q_m^{-1}(\mathcal{L}, \eta))^{l-1}} \left[(p-1) \sum_{l \in \mathcal{L} \setminus \mathcal{L}_m} \mu(\mathcal{L})^l\right] \\
& = \ \frac{\sum_{l \in \mathcal{L} \setminus \mathcal{L}_m} \mu(\mathcal{L})^l}{\sum_{l \in \mathcal{L}_m} l(Q_m^{-1}(\mathcal{L}, \eta))^{l-1}}.
\end{aligned}
\tag{35}
$$

By the monotonicity of $Q_m^{-1}(\mathcal{L}, \cdot)$, we have

$$
\mu(\mathcal{L}) \le Q_m^{-1}(\mathcal{L}, \eta) \le \mu_m(\mathcal{L}).
$$

In the course of lower bounding the denominator in (35), we have to distinguish two cases. From the inequality above, for $l \in \mathcal{L}'_m$, i.e., $l \le 1$ (so that $l - 1 \le 0$), we have

$$
(Q_m^{-1}(\mathcal{L}, \eta))^{l-1} \ge \mu_m(\mathcal{L})^{l-1},
$$

and for $l \in \mathcal{L}''_m$, i.e., $l > 1$, we have

$$
(Q_m^{-1}(\mathcal{L}, \eta))^{l-1} \ge \mu(\mathcal{L})^{l-1}.
$$

Thus

$$
\sum_{l \in \mathcal{L}_m} l(Q_m^{-1}(\mathcal{L}, \eta))^{l-1} \ge \sum_{l \in \mathcal{L}'_m} l\mu_m(\mathcal{L})^{l-1} + \sum_{l \in \mathcal{L}''_m} l\mu(\mathcal{L})^{l-1}.
$$

It then follows that

$$
\begin{aligned}
0 \ & \le \ \mu_m(\mathcal{L}) - \mu(\mathcal{L}) \\
& \le \ \frac{\sum_{l \in \mathcal{L} \setminus \mathcal{L}_m} \mu(\mathcal{L})^l}{\sum_{l \in \mathcal{L}'_m} l\mu_m(\mathcal{L})^{l-1} + \sum_{l \in \mathcal{L}''_m} l\mu(\mathcal{L})^{l-1}} \\
& \le \ \frac{\sum_{l \in \mathcal{L} \setminus \mathcal{L}_m} \theta_2{}^l}{\sum_{l \in \mathcal{L}'_m} l\mu_m(\mathcal{L})^{l-1} + \sum_{l \in \mathcal{L}''_m} l\theta_1{}^{l-1}};
\end{aligned}
$$

the last inequality is justified because $Q(\mathcal{L}, \theta_1) \ge 0$ and $Q(\mathcal{L}, \theta_2) < 0$ imply

$$
\theta_1 \le \mu(\mathcal{L}) < \theta_2.
$$

This proves the theorem.

8 Proof of Theorem 4

In this section we will prove Theorem 4, the Maxentropic Theorem. The upper bound (28) is actually a direct consequence of Theorem 2 as we now explain. Let

$$N'(T) = |\Phi_L(\mathcal{L}(\mathcal{A}, \xi) \cap [0, T])|,$$

the size of an optimal code in the subclass of the code $\mathcal{C}_L$ such that all the runlengths of a codeword are in $\mathcal{L}(\mathcal{A}, \xi) \cap [0, T]$, as previously defined in Section 4. For all $X(t) \in \mathcal{K}(\mathcal{L})$ and all $T > l_1$, upon truncating the last run if its length is not in $\mathcal{L}$, $X^T(t)$ takes values in the set of step functions on the interval $[0, T]$ satisfying L1 to L3 with $\mathcal{A}$ replaced by $\mathcal{L}$ a.s. Therefore, we have

$$
\begin{aligned}
r(X(t)) &= \lim_{T \to \infty} -\frac{1}{T} \mathbf{E} \log P_{X^T(t)}(X^T(t)) \\
&= \lim_{T \to \infty} \frac{1}{T} H(X^T(t)) \\
&\leq \lim_{T \to \infty} \frac{1}{T} \log[(p-1)N'(T)] \\
&= \lim_{T \to \infty} \frac{1}{T} \log N(T) \\
&= -\log \mu(\mathcal{L}),
\end{aligned}
$$

where the last equality follows from Theorem 2.

To complete the proof we have to prove $X^*(t), t \in [0, \infty)$ defined by $\{L^*(i)\}_{i=1}^{\infty}$ and $\{Y^*(j)\}_{j=1}^{\infty}$ in the last part of Section 5 achieves the upper bound $-\log \mu(\mathcal{L})$. By the weak law of large numbers, $I^{-1} \sum_{i=1}^{I} L^*(i)$ is convergent to $\mathbf{E}L^*$ in probability as $I \to \infty$. This means that for any given positive real number T and an arbitrarily small but positive δ, by letting $i(T) = \lfloor \frac{T}{\mathbf{E}L^* + \delta} \rfloor$ and $U(I) = \sum_{i=1}^{I} L^*(i)$, for any $\epsilon > 0$ and sufficiently large T,

$$Pr(U(i(T)) < T) \geq Pr\left[\sum_{i=1}^{i(T)} L^*(i) < i(T)(\mathbf{E}L^* + \delta)\right] > 1 - \epsilon. \tag{36}$$

Now we define

$$L^{*i(T)} = (L^*(1), L^*(2), L^*(3), \dots, L^*(i(T)))$$

28

and

$$Y^{*i(T)} = (Y^*(1), Y^*(2), Y^{*T}(3), \dots, Y^*(i(T))).$$

Under the condition $U(i(T)) < T$, $(L^{*i(T)}, Y^{*i(T)})$ is a function of $X^{*T}(t)$. Therefore,

$$
\begin{aligned}
\mathbf{E}[-\log P_{X^{*T}(t)}(X^{*T}(t)) | U(i(T)) < T] \\
&= H(X^{*T}(t) | U(i(T)) < T) \\
&\geq H(L^{*i(T)}, Y^{*i(T)} | U(i(T)) < T) \\
&= H(L^{*i(T)} | U(i(T)) < T) + H(Y^{*i(T)}), \quad (37)
\end{aligned}
$$

where the last inequality follows from the fact that $\{Y^*(j)\}_{j=1}^{\infty}$ is independent of $\{L^*(i)\}_{i=1}^{\infty}$ as well as $U(I)$ for any I. To bound the first term on the RHS of (37), we have to estimate $H(L^{*i(T)} | U(i(T)) \geq T)$. To simplify notation, we introduce two sets, $\mathcal{L}^I := \{l^I = (l_1, l_2, l_3, \dots, l_I) : l_i \in \mathcal{L}, \text{ for } i = 1, 2, 3, \dots, I\}$ and $\mathcal{U}(I) = \{u = \sum_{i=1}^{I} l_i : l^I \in \mathcal{L}^I\}$. Then

$$
\begin{aligned}
H(L^{*i(T)} | U(i(T)) \geq T) \\
&= \mathbf{E}\left[-\log P_{L^{*i(T)}}(L^{*i(T)}) | U(i(T)) \geq T\right] \\
&= \mathbf{E}\left[-\log[(p-1)^{i(T)}(\mu(\mathcal{L}))^{\sum_{i=1}^{i(T)} L^*(i)}] | U(i(T)) \geq T\right] \\
&= -i(T)\log(p-1) - \log\mu(\mathcal{L})\mathbf{E}\left[\sum_{i=1}^{i(T)} L^*(i) | U(i(T)) \geq T\right] \\
&= -i(T)\log(p-1) - \log\mu(\mathcal{L})\mathbf{E}\left[U(i(T)) | U(i(T)) \geq T\right] \\
&= -i(T)\log(p-1) \\
&\quad - \log\mu(\mathcal{L})\frac{1}{Pr[U(i(T)) \geq T]} \sum_{u \in \mathcal{U}(i(T)) : u \geq T} u Pr[U(i(T)) = u]. \quad (38)
\end{aligned}
$$

Here the second equality follows since $(L^*(1), L^*(2), L^*(3), \dots, L^*(i(T)))$ is an i.i.d. sequence of random variables with distribution $P_{\mu(\mathcal{L})}$. For all $l^{i(T)} := (l_1, l_2, l_3, \dots, l_{i(T)}) \in \mathcal{L}^{i(T)}$, we have

$$Pr[L^{*i(T)} = l^{i(T)}] = \prod_{i=1}^{i(T)} (p-1)\mu^{l_i}(\mathcal{L}) = (p-1)^{i(T)}(\mu(\mathcal{L}))^{\sum_{i=1}^{i(T)} l_i}. \quad (39)$$

29

Among the terms on the RHS of (38), $\sum_{u \in \mathcal{U}(i(T)):u \geq T} uPr[U(i(T)) = u]$ is difficult to evaluate due to the general form of $\mathcal{L}$. So we instead have to find an upper bound. To this end, we introduce a random variable $V(I) = \sum_{i=1}^{I}[L^*(i) - \mathbf{E}L^*] = U(I) - I\mathbf{E}L^*$ and a set $\mathcal{V}(I) = \{v = u - I\mathbf{E}L^* : u \in \mathcal{U}(I)\}$, and write

$$\sum_{u \in \mathcal{U}(i(T)):u \geq T} uPr[U(i(T)) = u]$$

$$= \sum_{v \in \mathcal{V}(i(T)):v \geq T - i(T)\mathbf{E}L^*} [v + i(T)\mathbf{E}L^*]Pr[V(i(T)) = v]$$

$$= \sum_{v \in \mathcal{V}(i(T)):v > T - i(T)\mathbf{E}L^*} vPr[V(i(T)) = v] + Pr[V(i(T)) \geq T - i(T)\mathbf{E}L^*]i(T)\mathbf{E}L^*$$

$$= \sum_{v \in \mathcal{V}(i(T)):v \geq T - i(T)\mathbf{E}L^*} vPr[V(i(T)) = v] + Pr[U(i(T)) \geq T]i(T)\mathbf{E}L^*. \tag{40}$$

Recalling that $i(T) = \lfloor \frac{T}{\mathbf{E}L^* + \delta} \rfloor$, we obtain $T \geq i(T)[\mathbf{E}L^* + \delta]$ and therefore $\{v : v \geq T - i(T)\mathbf{E}L^*\} \subset \{v : v \geq i(T)\delta\}$. This observation allows us to bound the first term on the RHS of (40) as follows.

$$\sum_{v \in \mathcal{V}(i(T)):v \geq T - i(T)\mathbf{E}L^*} vPr[V(i(T)) = v] \leq \sum_{v \in \mathcal{V}(i(T)):v \geq i(T)\delta} vPr[V(i(T)) = v]$$

$$\leq \sum_{v \in \mathcal{V}(i(T)):v \geq i(T)\delta} \frac{v^2}{i(T)\delta}Pr[V(i(T)) = v]$$

$$\leq \frac{1}{i(T)\delta}\mathbf{E}V^2(i(T))$$

$$= \frac{1}{i(T)\delta}\mathbf{E}[U(i(T)) - i(T)EL^*]^2,$$

where $\mathbf{E}[U(i(T)) - i(T)EL^*]^2$ can be seen as the variance of $U(i(T))$. Since $U(I) = \sum_{i=1}^{I} L^*(i)$ where $\{L^*(i)\}_{i=1}^{I}$ are i.i.d., we have

$$\sum_{v \in \mathcal{V}(i(T)):v \geq T - i(T)\mathbf{E}L^*} vPr[V(i(T)) = v] \leq \frac{1}{\delta}\mathbf{E}\left[L^* - \mathbf{E}L^*\right]^2. \tag{41}$$

By combining (38), (40) and the above inequality, we obtain the bound

$$H(L^{*i(T)}|U(i(T)) \geq T)$$

$$\leq -i(T)\log(p-1) - \frac{\log \mu(\mathcal{L})}{Pr[U(i(T)) \geq T]}\left(\frac{1}{\delta}\mathbf{E}[L^* - \mathbf{E}L^*]^2 + Pr[U(i(T)) \geq T]i(T)\mathbf{E}L^*\right)$$

$$= i(T)(-\log \mu(\mathcal{L})\mathbf{E}L^* - \log(p-1)) - \frac{\log \mu(\mathcal{L})}{\delta Pr[U(i(T)) \geq T]}\mathbf{E}[L^* - \mathbf{E}L^*]^2 \tag{42}$$

30

Define a binary random variable

$$
B = \begin{cases} 0 & \text{if } U(i(T)) < T \\ 1 & \text{otherwise.} \end{cases}
$$

Then B is a function of $L^{*i(T)}$, and so

$$
\begin{aligned}
H(L^{*i(T)}) &= H(L^{*i(T)}B) \\
&= H(B) + Pr(B=0)H(L^{*i(T)}|B=0) + Pr(B=1)H(L^{*i(T)}|B=1). \quad (43)
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&Pr[U(i(T)) < T]H(L^{*i(T)}|U(i(T)) < T) \\
&= H(L^{*i(T)}) - Pr[U(i(T)) \geq T]H(L^{*i(T)}|U(i(T)) \geq T) - H(B). \quad (44)
\end{aligned}
$$

Then by the definition of $L^{*i(T)}$, it is not hard for us to calculate

$$
\begin{aligned}
H(L^{*i(T)}) &= \sum_{i=1}^{i(T)} H(L^*(i)) \\
&= i(T)H(L^*) \\
&= -i(T)\mathbf{E}[\log(p-1)(\mu(\mathcal{L}))^{L^*}] \\
&= -i(T)[\log(p-1) + \log \mu(\mathcal{L})\mathbf{E}L^*]. \quad (45)
\end{aligned}
$$

We next substitute (42) and (45) into (44) and divide both sides of the resulting inequality by $Pr[U(i(T)) < T]$. By using (36) and the fact that B is a binary random variable with $H(B) \leq 1$, we can obtain a lower bound on $H(L^{*i(T)}|U(i(T)) < T)$, namely

$$
H(L^{*i(T)}|U(i(T)) < T) \quad (46)
$$
$$
\geq i(T)[-\log(p-1) - \log \mu(\mathcal{L})\mathbf{E}L^*] + \frac{\log \mu(\mathcal{L})}{\delta(1-\epsilon)}\mathbf{E}[L^* - \mathbf{E}L^*]^2 - \frac{1}{1-\epsilon}. \quad (47)
$$

Let us now return to (37) and consider the second term on the RHS. From the definition of $\{Y^*(j)\}_{j=1}^{\infty}$, it is obvious that $H(Y^{*i(T)}) = \sum_{j=1}^{i(T)} H(Y^*(j)) = i(T) \log(p-1)$. We then substitute this and (46) into (37) to obtain a final bound on $\mathbf{E}[-\log P_{X^{*T}(t)}(X^{*T}(t))]$,

31

namely

$$\begin{aligned}
&\mathbf{E}[-\log P_{X^{*T}(t)}(X^{*T}(t))|U(i(T)) < T] \\
&\geq \quad -i(T)\log \mu(\mathcal{L})\mathbf{E}L^* + \frac{\log \mu(\mathcal{L})}{\delta(1-\epsilon)}\mathbf{E}(L^* - \mathbf{E}L^*)^2 - \frac{1}{1-\epsilon}.
\end{aligned} \tag{48}$$

So by (36) and

$$\begin{aligned}
\mathbf{E}[-\log P_{X^{*T}(t)}(X^{*T}(t))] &= H(X^{*T}(t)) \\
&\geq H(X^{*T}(t)|U(i(T))) \\
&= Pr[U(i(T)) < T]\mathbf{E}[-\log P_{X^{*T}(t)}(X^{*T}(t))|U(i(T)) < T] + \\
&\quad Pr[U(i(T)) \geq T]\mathbf{E}[-\log P_{X^{*T}(t)}(X^{*T}(t))|U(i(T)) \geq T],
\end{aligned}$$

we have

$$\begin{aligned}
&r(X^*(t)) \\
&= \lim_{T\to\infty} \frac{1}{T}\mathbf{E}[-\log P_{X^{*T}(t)}(X^{*T}(t))] \\
&\geq \lim_{T\to\infty} \frac{1}{T}Pr[U(i(T)) < T]\mathbf{E}[-\log P_{X^{*T}(t)}(X^{*T}(t))|U(i(T)) < T] \\
&\geq (1-\epsilon)\lim_{T\to\infty} \frac{1}{T}\mathbf{E}[-\log P_{X^{*T}(t)}(X^{*T}(t))|U(i(T)) < T] \\
&\geq \frac{1-\epsilon}{\mathbf{E}L^* + \delta}\lim_{i(T)\to\infty} \frac{1}{i(T)}\mathbf{E}[-\log P_{X^{*T}(t)}(X^{*T}(t))|U(i(T)) < T] \\
&\geq \frac{1-\epsilon}{\mathbf{E}L^* + \delta}\left\{[-\log \mu(\mathcal{L})\mathbf{E}L^*] + \lim_{i(T)\to\infty} \frac{1}{i(T)}\left[\frac{\log \mu(\mathcal{L})}{\delta(1-\epsilon)}\mathbf{E}(L^* - \mathbf{E}L^*)^2 - \frac{1}{1-\epsilon}\right]\right\} \\
&= -\frac{\mathbf{E}L^*(1-\epsilon)}{\mathbf{E}L^* + \delta}\log \mu(\mathcal{L}).
\end{aligned} \tag{49}$$

Finally, by letting $\delta \to 0$ and $\epsilon \to 0$ in (49), we obtain $r(X^*(t)) \geq -\log \mu(\mathcal{L})$, completing the proof.

9 Conclusion

In this paper, we have explored a new approach to tackle time jitter in a communication system with no common clock between the transmitter and the receiver. To start with,

we only make the assumption the transmitter and receiver clocks can synchronize with each other explicitly at the beginning and the end of a communication session. Unlike the traditional approach, we do not impose the constraint that the transmitter and receiver clocks have to re-synchronize with each other from time to time, in particular, within a fixed maximum time.

To capture the effect of time jitter at the physical level of a communication system, we have introduced the asynchronous channel. Unlike previous models in the literature, our channel model is in continuous time. We have obtained a simple characterization of an optimal zero-error self-synchronizable code for the asynchronous channel. We also have proved the Maxentropic Theorem which implies the equivalence between the combinatorial capacity and the probabilistic capacity of the channel. With the insight provided by this theorem, a construction of an optimal code has recently been given in [12].

Our characterization of an optimal code for the noiseless asynchronous channel reveals that in designing efficient codes for the purpose of synchronization, no upper limit should be imposed on the runlengths as in runlength limited (RLL) codes which are commonly used in magnetic recording. We have shown this somewhat surprising fact when the channel is noiseless; whether this remains to be true when the channel is noisy is unknown. After all, how to incorporate both time jitter and noise into the same model while keeping the model analytically tractable is perhaps the biggest challenge for future research in this direction.

Acknowledgment

References

[1] P. M. J. Baggen, "An Information-Theoretic Approach to Timing Jitter", Ph.D. thesis, University of California at San Diego (UCSD), 1991.

[2] S. Shamai and E. Zehavi, "Bounds on the Capacity of the Bit-Shift Magnetic Recording Channel", in *IEEE Trans. Inform. Theory*, IT-37: 863-872, 1991.

[3] D. Arnold, A. Kavčić, R. Kötter, H.-A. Loeliger and P.O. Vontobel, "The Binary Jitter Channel: A New Model for Magnetic Recording", in *Proc. 2000 IEEE Int. Symposium Inform. Theory (ISIT 2000)*, Sorrento, Italy, June 25-30, 2000.

[4] C. E. Shannon, A Mathematical Theory of Communiction, *Bell Tech. J.*, V. 27, pp. 379-423, July 1948.

[5] K. A. S. Immink, *Codes for Mass Data Storage Systems*, Shannon Foundation Publishers, The Netherlands, 1999.

[6] R. E. Blahut, *Principles and Practice of Information Theory*, Reading, Mass.: Addison-Wesley Publishing Company, 1987.

[7] R. M. Krause, "Channels Which Transmit Letters of Unequal Duration", in *Information and Control*, V. 5, pp. 13-24, 1962.

[8] I. Csiszár, "Simple Proofs of Some Theorems on Noiseless Channels", in *Information and Control*, V. 14, pp. 285-298, 1969.

[9] A. Khandekar, R. McEliece and E. Rodemich, "The Discrete Noiseless Channel Revisited", in *Proceedings of ISCTA'99*, 1999.

[10] C. Pimentel and B. F. Uchôa-Filho, "A Combinatorial Approach to Finding the Capacity of the Discrete Noiseless Channel", in *IEEE Trans. Inform. Theory*, IT-49: 2024-2028, 2003.

[11] R. G. Bartle and D. R. Sherbert, *Introduction to Real Analysis*, 2nd Ed., New York : Wiley, 1992.

[12] N. Cai, S.-W. Ho, and R. W. Yeung, "Probabilistic Capacity and Optimal Coding for Asynchronous Channel", IEEE Information Theory Workshop (ITW 2007), Lake Tahoe, USA, Sept. 2-6, 2007.

Raymond W. Yeung (S'85-M'88-SM'92-F'03) was born in Hong Kong on June 3, 1962. He received the B.S., M.Eng., and Ph.D. degrees in electrical engineering from Cornell University, Ithaca, NY, in 1984, 1985, and 1988, respectively.

He was on leave at Ecole Nationale Supérieure des Télécommunications, Paris, France, during fall 1986. He was a Member of Technical Staff of AT&T Bell Laboratories from 1988 to 1991. Since 1991, he has been with the Department of Information Engineering, The Chinese University of Hong Kong, where he is now a chair professor. He is the author of the textbooks *A First Course in Information Theory* (Kluwer Academic/Plenum, 2002) and its revision *Information Theory and Network Coding* (Springer 2008). He has held visiting positions at Cornell University, Nankai University, the University of Bielefeld, the University of Copenhagen, Tokyo Institute of Technology, and Munich University of Technology. His research interests include information theory and network coding. He has been a Consultant in a project of Jet Propulsion Laboratory, Pasadena, CA, for salvaging the malfunctioning Galileo Spacecraft and a Consultant for NEC, USA.

Dr. Yeung was a member of the Board of Governors of the IEEE Information Theory Society from 1999 to 2001. He has served on the committees of a number of information theory symposiums and workshops. He was General Chair of the First and the Fourth Workshop on Network, Coding, and Applications (NetCod 2005 and 2008), a Technical Co-Chair for the 2006 IEEE International Symposium on Information Theory, and a Technical Co-Chair for the 2006 IEEE Information Theory Workshop, Chengdu, China. He currently serves as an Editor-at-Large of *Communications in Information and Systems*, an Editor of *Foundation and Trends in Communications and Information Theory* and of *Foundation and Trends in Networking*, and was an Associate Editor for Shannon Theory of this Transactions from 2003 to 2005. He was a recipient of the Croucher Foundation Senior Research Fellowship for 2000/2001, the Best Paper Award (Communication Theory) of the 2004 International Conference on Communications, Circuits and System (with C. K. Ngai), the 2005 IEEE Information Theory Society Paper Award (for his paper "Linear network coding" co-authored with S.-Y. R. Li and N. Cai), and the Friedrich Wilhelm Bessel Research

Award of the Alexander von Humboldt Foundation in 2007. He is a Fellow of the IEEE and the Hong Kong Institution of Engineers.

Ning Cai received the B.S. degree in mathematics from the Normal College of Beijing, Beijing, China in 1982, the M.S. degree in mathematics from Academia Sinica, Beijing, China, in 1984, and the Dr. degree in mathematics from the University of Bielefeld, Bielefeld, Germany, in 1988.

During 1984-1986, he worked in the Institute of Systems Sciences, Academia Sinica, Beijing, China. During 1988-1989, he was with the Department of Mathematics, Statistics and Computer Science, the University of Illinois, Chicago, USA. From 1989 to 1998, he was a Miss. Mitarbeiter in the Department of Mathematics, the University of Bielefeld, Germany and from 1998 to 1999, he was with the School of Computing, the National University of Singapore. From 2000 to 2001, he was with the Department of Information Engineering, The Chinese University of Hong Kong. From 2002 to 2004 he was with the Department of Mathematics, the University of Bielefeld, Germany. In 2005 he visited the Department of Information Engineering, The Chinese University of Hong Kong. Since 2006, he is a Distinguished Professor in the State Key Lab. of Integrated Services Networks (ISN), the Xidian University, China.

Dr. Cai is a recipient of the 2005 IEEE Information Theory Society Paper Award (for his paper "Linear network coding" co-authored with S.-Y. R. Li and R. W. Yeung). He has served on the committees of a number of information theory symposiums and workshops. His research interests include network coding, Shannon and quantum information theory, and combinatorics and its applications in communication and computer science.

S.-W. Ho (S'05-M'07) was born in Hong Kong on Nov 16, 1978. He received the B.S., M.Phil., and Ph.D. degrees in information engineering from The Chinese University of Hong Kong, Hong Kong, in 2000, 2003, and 2006, respectively. Since 2006, he has been a post-doctoral research fellow in the Department of Electrical Engineering, Princeton University, Princeton, NJ.

Dr. Ho was a recipient of the Croucher Foundation Fellowship for 2006/2008. His research interests are in Shannon Theory, Data Communications and Recording Systems and Quantum Information Theory.

Aaron B. Wagner(S'00-M'05) received the B.S. degree in Electrical Engineering from the University of Michigan, Ann Arbor, in 1999 and the M.S. and Ph.D. degrees in Electrical Engineering and Computer Sciences from the University of California, Berkeley, in 2002 and 2005, respectively. During the 2005-2006 academic year, he was a Postdoctoral Research Associate in the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign and a Visiting Assistant Professor in the School of Electrical and Computer Engineering at Cornell University. Since 2006, he has been an Assistant Professor in the School of Electrical and Computer Engineering at Cornell.

He has received the NSF CAREER award, the David J. Sakrison Memorial Prize from the U.C. Berkeley EECS Dept., the Bernard Friedman Memorial Prize in Applied Mathematics from the U.C. Berkeley Dept. of Mathematics, and the Distinguished Achievement Award in Electrical Engineering from the University of Michigan College of Engineering.