

On the Information Stability of Channels With Timing Errors

Wei Zeng, Patrick Mitran, Aleksandar Kavčić
DEAS, Harvard University, Cambridge, MA, 02138, USA.

Abstract—In this paper, we consider the class of channels corrupted by timing errors and intersymbol interference (ISI). The timing error is modeled as a discrete-time, discrete-valued random walk process. We prove that this class of channels are information stable for Markov input process of any finite order, and thereby, show achievable transmission rates in terms of asymptotic information rates. The proof we provide is extended from the proof by Dobrushin for the class of memoryless channels with independent timing errors.

I. INTRODUCTION

While classical information theory typically assumes perfect timing information, communication channels as well as data storage channels are normally corrupted by synchronization errors (or timing errors). Most of the earlier literature has been devoted to analyzing the capacity of a special class of channels with timing errors, namely, the insertion/deletion channels. For this class of channels, timing errors are modeled as a complete loss of data symbols (deletion) or repetition of the same data symbols (insertion), and at the output of the channel the indices of the input symbol from which the given output was obtained are unknown. Both analytic bounds ([1–4]) and simulation-based numerical computation of the capacity ([5, 6]) have been proposed.

In [7], Dobrushin proved the *information stability* [8] of the class of *memoryless channels with synchronization errors*, in which each input symbol *independently* of other symbols is transformed into a “word” of random (including zero) length. Thereby, he showed that Shannon’s theorem holds for this class of channels and the capacity is given by

$$C = \lim_{n \rightarrow \infty} \frac{1}{n} \sup_{P(\mathbf{X}^n)} I(\mathbf{X}^n; \bar{\mathbf{Y}}), \quad (1)$$

where \mathbf{X}^n is a vector of n transmitted symbols, and $\bar{\mathbf{Y}}$ is a finite random-length vector of received symbols.

In real communication systems, however, timing errors are generally more complicated and more slowly varying in time than the simple insertion/deletion process, i.e., the timing error can be a fraction of the symbol interval instead of a complete loss or insertion of a symbol. The channels may not be memoryless and the timing error process may also have memory. In this paper, we consider the class of channels corrupted by finite intersymbol interference (ISI) and timing errors. It is well known [9] that channels with finite intersymbol interference belong to the class of indecomposable channels, and the channel capacity is

$$C = \lim_{n \rightarrow \infty} \frac{1}{n} \sup_{P(\mathbf{X}^n)} I(\mathbf{X}^n; \mathbf{Y}^n), \quad (2)$$

where \mathbf{X}^n and \mathbf{Y}^n are both vectors of length n . However, the result in (2) is not directly generalized to the case where timing errors also exist. In this paper, we show a lower bound similar to (1) for the class of ISI channels with Markov timing error processes. This formally justifies the approach taken in [6] to compute capacity bounds that rely on bounding the information rate for Markov inputs.

II. SOURCE AND CHANNEL MODEL

We consider a simple linear-filter channel. Denote by X_k the binary antipodal channel input symbol ($X_k \in \{-1, +1\}$) at time $k \in \mathbb{Z}$. We only consider the case where the random process $\{X_k\}$ is a *stationary* and *ergodic* Markov process of finite order ν . The received waveform $Y(t)$ is assumed to be of the following form

$$Y(t) = \sum_k X_k h(t - kT) + N(t), \quad (3)$$

where T is the symbol interval and $N(t)$ is additive Gaussian noise that is independent of the input.

If no timing errors exist, the receiver will sample $R(t)$ at $t = iT$, for $i \in \mathbb{Z}$. However, if timing uncertainty exists, the real sampling instant will be $t = iT + \mathcal{E}_i$, where \mathcal{E}_i represents the timing uncertainty of the i -th sample. Denote by Y_i the i -th sample at the receiver,

$$Y_i = Y(iT + \mathcal{E}_i) = \sum_{k=-\infty}^{+\infty} X_k \cdot h(iT - kT + \mathcal{E}_i) + N_i. \quad (4)$$

For simplicity, we shall assume that $N_i \sim \mathcal{N}(0, \sigma^2)$ are independent and identically distributed (i.i.d.) Gaussian random variables, and \mathcal{E}_i are independent of the channel input $\{X_k\}$ and noise $\{N_i\}$.

We assume that the baseband channel response function $h(t)$ has *finite support* that satisfies

$$h(t) = 0 \quad \text{for } |t| \geq qT, \quad (5)$$

and q is an integer. Therefore, the i -th sample given by (4) can be rewritten as

$$Y_i = \sum_{k=i-q+\lceil \frac{\mathcal{E}_i}{T} \rceil}^{i+q+\lfloor \frac{\mathcal{E}_i}{T} \rfloor} X_k \cdot h(iT - kT + \mathcal{E}_i) + N_i. \quad (6)$$

We assume \mathcal{E}_i to be discrete-time, discrete-valued random variables that can take one of countably many values $\frac{jT}{Q}$, where j is an arbitrary integer and Q is a fixed positive integer representing the number of quantization levels in each symbol

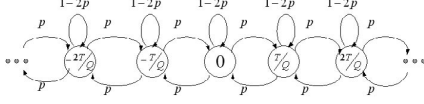


Fig. 1. The state transition diagram of the Markov timing error \mathcal{E}_i .

interval T . We further assume that the process $\{\mathcal{E}_i\}$ is *slowly* varying with time, and can be represented by the following random walk process

$$\mathcal{E}_{i+1} = \mathcal{E}_i + \Delta_{i+1}, \quad (7)$$

$$P(\Delta_i = \delta_i) = \begin{cases} p & \text{if } \delta_i = \pm \frac{T}{Q}, \\ 1 - 2p & \text{if } \delta_i = 0. \end{cases} \quad (8)$$

The slowly time-varying assumption is satisfied if $p \ll 1$. The initial value of this random process is $\mathcal{E}_0 = 0$. In practical systems, this is generally achieved by using preambles ahead of each block of data symbols. The timing error increments Δ_i are assumed to be i.i.d. and independent of all previous samples Y_j and previous timing errors \mathcal{E}_j , where $j < i$. Fig. 1 shows the state transition diagram of \mathcal{E}_i . More complicated Markov models than this three-state first-order Markov chain can be adopted without changing the nature of the problem and the proof methodology. The simple model in (7)-(8), however, simplifies the notation and the proofs.

Notational Conventions: We denote a sequence of variables X_i, X_{i+1}, \dots, X_j by X_j^i . For a given block of input bits X_1^k , there are L_k readback samples $Y_1^{L_k}$ out of the first k input symbol intervals (notice that the vector length L_k is also random because of timing errors). Throughout the paper, uppercase letters denote random variables, lowercase letters denote their realizations, and boldface letters denote vectors or sequences. We denote by \mathbb{B}^k the set of length- k binary vectors, and \mathbb{R} the set of finite length vectors with real values.

III. THE FINITE-STATE MACHINE CHANNEL

In this section, we formulate the channel model described in Sec. II into a finite-state machine channel. Assume that the i -th sample falls inside the k -th input symbol interval, for some $k \in \mathbb{Z}^+$, i.e.,

$$iT + \mathcal{E}_i = (k-1)T + \frac{M_i + 1}{Q}T, \quad (9)$$

where $M_i \in \{0, 1, \dots, Q-1\}$ is the mod Q reduced timing error. Note that the sequence $\{M_i\}$ itself is a first order Markov chain. From (7) and (8), we have

$$\begin{aligned} P_{M_i|M_{i-1}}(m_i|m_{i-1}) \\ = \begin{cases} p & \text{if } m_i = (m_{i-1} \pm 1) \bmod Q, \\ 1 - 2p & \text{if } m_i = m_{i-1}. \end{cases} \end{aligned} \quad (10)$$

We know that the distribution of each sample Y_i depends only on the values of M_i and at most $2q$ consecutive binary input symbols X_{k-v}^{k+v-1} , for some integer $k \in \mathbb{Z}^+$. Therefore, for the class of channels given by (6), (7) and (8), we can define a channel state by M_i and $\lambda = \max(2q, \nu)$ binary symbols.

Definition 1: [channel state] Define the *channel state* for the i -th sample by $S_i \in \mathbb{S} = \{(m, a_1^\lambda)\}$, where $m \in$

$\{0, 1, \dots, Q-1\}$ and a_1^λ represent λ consecutive input symbols. The last $2q$ symbols of a_1^λ correspond to the $2q$ consecutive input symbols that determine the value of Y_i . \square Obviously $|\mathbb{S}| = 2^\lambda \cdot Q$. Note that for a specific sample, it is not important to know the time indices of the input symbols corresponding to a_1^λ , the value of the binary vector is sufficient.

When transmitting a finite-length vector X_1^k , we assume $X_{-\lambda+1}^0$ are generated according to the steady state distribution of input process, and λ extra channel input symbols $X_{k+1}^{k+\lambda}$ are generated before the transmission terminates.

Let L_k denote the number of output samples within the first k input symbol intervals $(0, kT]$, i.e.,

$$L_k T + \mathcal{E}_{L_k} \leq kT < (L_k + 1)T + \mathcal{E}_{L_k+1}. \quad (11)$$

Obviously, L_k is a random number, the distribution of which depends on k . We denote the realization of L_k by ℓ_k . From the random walk assumption in (7) and (8), the interval T' between any two adjacent samples satisfies $\frac{Q-1}{Q}T \leq T' \leq \frac{Q+1}{Q}T$. Therefore

$$\frac{Q}{Q+1}k \leq \ell_k \leq \frac{Q}{Q-1}k. \quad (12)$$

IV. INFORMATION STABILITY

In this section we review some definitions of information stability from [7, 8].

Definition 2: [information density] Define the information density of the random variables X and Y as the function

$$i(X; Y) = \log \frac{P_{XY}(X, Y)}{P_X(X)P_Y(Y)}. \quad (13)$$

\square

The mutual information is the mean of information density

$$I(X; Y) = E[i(X; Y)]. \quad (14)$$

Definition 3: [information-stable sequence] A sequence of pairs of random variables (or vectors) $(\mathbf{X}^k, \mathbf{Y}^k)$ having information density $i(\mathbf{X}^k; \mathbf{Y}^k)$ will be called *information-stable* if for all sufficiently large k , we have $0 \leq I(\mathbf{X}^k; \mathbf{Y}^k) < \infty$, and $\forall \epsilon > 0$ we have

$$\lim_{k \rightarrow \infty} P\left(\left|\frac{i(\mathbf{X}^k; \mathbf{Y}^k)}{I(\mathbf{X}^k; \mathbf{Y}^k)} - 1\right| > \epsilon\right) = 0. \quad (15)$$

\square

Definition 4: [information-stable channels] Let a length- k channel be specified by transition probability function $P^k(\cdot|\cdot)$ and a set \mathbb{V}^k of probability distributions on the space of transmitted symbols. We say that the sequence of length- k channels is *information-stable* if there exists an information-stable sequence of pairs of variables $(\mathbf{X}^k, \mathbf{Y}^k)$ such that the k -th pair is connected by the channel $(\mathbb{V}^k, P^k(\cdot|\cdot))$ and

$$\lim_{k \rightarrow \infty} \frac{I(\mathbf{X}^k; \mathbf{Y}^k)}{C_k} = 1, \quad (16)$$

where $C_k = \sup_{P(\mathbf{X}^k) \in \mathbb{V}^k} I(\mathbf{X}^k; \mathbf{Y}^k)$. \square

It was shown by Dobrushin [7,8] that if the channel is information stable, then Shannon's theorem holds, i.e., there exists a channel capacity

$$C = \lim_{k \rightarrow \infty} \frac{1}{k} \sup_{P(\mathbf{X}^k) \in \mathbb{V}^k} I(\mathbf{X}^k; \mathbf{Y}^k), \quad (17)$$

such that for all rate $R < C$ and every $\epsilon > 0$, we can choose an integer k^* large enough, such that for any $k \geq k^*$, there exist coding schemes for the length- k channel that can achieve probability of decoding errors less than ϵ . In the next section, we will prove information stability for the class of channels given by (6) (7), and (8) when the class of input distributions \mathbb{V}^k are stationary and ergodic Markov distributions of any finite order. This will then demonstrate that any rate below the information rate achieved by a Markov process is itself achievable.

V. INFORMATION STABILITY FOR CHANNELS WITH TIMING ERRORS

For a fixed input sequence $\mathbf{x} = x_1^\infty$, the state sequence (S_0, S_1, \dots) can be regarded as a *nonhomogeneous* Markov chain. In order to measure the dependence of $P(s_n|s_0, x_1^\infty)$ on the initial state s_0 , we introduce the following definition of *distance* that is similar (but different) to [9] (p.106):

Definition 5: $[d_n(s_0, s'_0|\mathbf{x})]$ Define the *state distance* $d_n(s_0, s'_0|\mathbf{x})$ related to a given input sequence \mathbf{x} as

$$d_n(s_0, s'_0|\mathbf{x}) \triangleq \sum_{S_{L_n} \in \mathbb{S}} |P(S_{L_n}|s_0, \mathbf{x}, n) - P(S_{L_n}|s'_0, \mathbf{x}, n)|. \quad (18)$$

Since L_n represents the number of samples within $(0, nT]$, the variable S_{L_n} represents the channel state after the transmission of n input symbols. Notice that the random variable L_n also depends on the initial state. \square

We now introduce a result extended from [9].

Lemma 1: Suppose that for some fixed $m > 0$ and $\mu > 0$, for all input sequences \mathbf{x} there exists a choice for the state $s_{L_m} \in \mathbb{S}$ (s_{L_m} can depend upon \mathbf{x}) such that

$$p(s_{L_m}|\mathbf{x}, s_0, m) > \mu, \quad \text{for all } s_0 \quad (19)$$

then $d_n(s_0, s'_0|\mathbf{x})$ is non-increasing and approaches 0 exponentially with n , *uniformly* in all \mathbf{x} , s_0 and s'_0 . Furthermore

$$d_n(s_0, s'_0|\mathbf{x}) < 2(1 - \mu)^{n/m-1}. \quad (20)$$

Proof: The proof is extended from the similar argument as lemma 4.6.1 and 4.6.2 in [9], by observing that for $\forall a \in \mathbb{S}, b \in \mathbb{S}$, we have $P(S_{L_n} = b|S_{L_{n-1}} = a, s_0, \mathbf{x}, n-1) = P(S_{L_n} = b|S_{L_{n-1}} = a, s'_0, \mathbf{x}, n-1)$. \blacksquare

Lemma 2 (4A.1 in [9]): Let X, Y, Z, S be a joint ensemble and let S be drawn from a set of cardinality K . Then

$$|I(X; Y|Z, S) - I(X; Y|Z)| \leq \log K. \quad (21)$$

We now use lemma 1 and 2 to prove the following theorem:

Theorem 1: Consider the class of channels with timing errors given by (6), (7) and (8). Denote by $\mathbf{X}^n = X_1^n$ the order ν Markov input sequence of length n , and $\mathbf{Y}^n = Y_1^{L_n}$

the corresponding output symbols. Denote by $S_{L_n} \in \mathbb{S}$ the channel state at the last sample. The mutual information

$$I(\mathbf{X}^n; \mathbf{Y}^n, S_{L_n}|s_0) = \sum_{\mathbf{x}^n \in \mathbb{B}^n, s_{\ell_n} \in \mathbb{S}} \int_{\mathbf{y}^n \in \mathbb{R}} P(\mathbf{y}^n, s_{\ell_n}|\mathbf{x}^n, s_0) \cdot P(\mathbf{x}^n) \log \frac{P(\mathbf{y}^n, s_{\ell_n}|\mathbf{x}^n, s_0)}{P(\mathbf{y}^n, s_{\ell_n}|s_0)} d\mathbf{y}^n$$

is asymptotically independent of the initial state. Further,

$$\lim_{n \rightarrow \infty} \frac{1}{n} |I(\mathbf{X}^n; \mathbf{Y}^n, S_{L_n}|s_0) - I(\mathbf{X}^n; \mathbf{Y}^n, S_{L_n}|s'_0)| = 0 \quad (22)$$

uniformly on all input distributions $P(\mathbf{X}^n)$ and $\forall s_0 \in \mathbb{S}$ and $\forall s'_0 \in \mathbb{S}$.

Proof: Let m be any fixed positive integer ($0 < m < n$). We first consider a *genie-aided* channel, where the decoder knows L_m , the number of samples within the first m data symbols. For simplicity, we denote $\mathbf{X}_1 = X_1^m$, $\mathbf{X}_2 = X_{m+1}^n$, $\mathbf{Y}_1 = Y_1^{L_m}$, $\mathbf{Y}_2 = Y_{L_m+1}^{L_n}$. The mutual information for this genie-aided channel can be expanded by using the chain rule:

$$\begin{aligned} I(\mathbf{X}^n; \mathbf{Y}^n, S_{L_n}, L_m|s_0) &= I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}^n, S_{L_n}, L_m|s_0) \\ &= I(\mathbf{X}_1; \mathbf{Y}^n, S_{L_n}, L_m|s_0) + I(\mathbf{X}_2; \mathbf{Y}_1, L_m|s_0, \mathbf{X}_1) \\ &\quad + I(\mathbf{X}_2; \mathbf{Y}_2, S_{L_n}|s_0, \mathbf{X}_1, \mathbf{Y}_1, L_m) \\ &= I(\mathbf{X}_1; \mathbf{Y}^n, S_{L_n}, L_m|s_0) + I(\mathbf{X}_{m+1}^{m+q}; \mathbf{Y}_1, L_m|s_0, \mathbf{X}_1) \\ &\quad + I(\mathbf{X}_2; \mathbf{Y}_2, S_{L_n}|s_0, \mathbf{X}_1, \mathbf{Y}_1, L_m). \end{aligned} \quad (23)$$

The last equality is based on the fact that given \mathbf{X}_1 , the output $\mathbf{Y}_1^{L_m}$ depend on at most q symbols in \mathbf{X}_2 , due to the finite-support assumption in (5). Notice that

$$\begin{aligned} 0 &\leq I(\mathbf{X}_1; \mathbf{Y}^n, S_{L_n}, L_m|s_0) \leq m \log 2, \\ 0 &\leq I(\mathbf{X}_{m+1}^{m+q}; \mathbf{Y}_1, L_m|s_0, \mathbf{X}_1) \leq q \log 2, \end{aligned} \quad (24)$$

we obtain

$$\begin{aligned} &|I(\mathbf{X}^n; \mathbf{Y}^n, S_{L_n}, L_m|s_0) - I(\mathbf{X}^n; \mathbf{Y}^n, S_{L_n}, L_m|s'_0)| \\ &\leq (m+q) \log 2 + |I(\mathbf{X}_2; \mathbf{Y}_2, S_{L_n}|s_0, \mathbf{X}_1, \mathbf{Y}_1, L_m) \\ &\quad - I(\mathbf{X}_2; \mathbf{Y}_2, S_{L_n}|s'_0, \mathbf{X}_1, \mathbf{Y}_1, L_m)|. \end{aligned} \quad (25)$$

From lemma 2, conditioning the mutual information on $S_{L_m} \in \mathbb{S}$ will change its value by at most $\log(2^\lambda Q)$, therefore

$$\begin{aligned} &|I(\mathbf{X}_2; \mathbf{Y}_2, S_{L_n}|s_0, \mathbf{X}_1, \mathbf{Y}_1, L_m) - I(\mathbf{X}_2; \mathbf{Y}_2, S_{L_n}|s'_0, \mathbf{X}_1, \mathbf{Y}_1, L_m)| \\ &\leq 2 \log(2^\lambda Q) + |I(\mathbf{X}_2; \mathbf{Y}_2, S_{L_n}|s_0, \mathbf{X}_1, \mathbf{Y}_1, L_m, S_{L_m}) \\ &\quad - I(\mathbf{X}_2; \mathbf{Y}_2, S_{L_n}|s'_0, \mathbf{X}_1, \mathbf{Y}_1, L_m, S_{L_m})| \\ &= 2 \log(2^\lambda Q) + |I(\mathbf{X}_2; \mathbf{Y}_2, S_{L_n}|s_0, S_{L_m}) \\ &\quad - I(\mathbf{X}_2; \mathbf{Y}_2, S_{L_n}|s'_0, S_{L_m})| \\ &= \left| \sum_{s_{\ell_m} \in \mathbb{S}} [P(s_{\ell_m}|s_0) - P(s_{\ell_m}|s'_0)] \cdot I(\mathbf{X}_2; \mathbf{Y}_2, S_{L_n}|s_{\ell_m}) \right| \\ &\quad + 2 \log(2^\lambda Q) \\ &\leq \sum_{s_{\ell_m} \in \mathbb{S}} |P(s_{\ell_m}|s_0) - P(s_{\ell_m}|s'_0)| \cdot I(\mathbf{X}_2; \mathbf{Y}_2, S_{L_n}|s_{\ell_m}) \\ &\quad + 2 \log(2^\lambda Q) \\ &\stackrel{(a)}{\leq} d_m^*(s_0, s'_0) \log(2^{n-m}) + 2 \log(2^\lambda Q), \end{aligned} \quad (26)$$

where $d_m^*(s_0, s'_0) = \sup_{\mathbf{x}} d_m(s_0, s'_0|\mathbf{x})$. Equation (a) is based on the fact that $I(\mathbf{X}_2; \mathbf{Y}_2, S_{L_n}|s_{\ell_m}) < \log(2^{n-m})$. Substituting (26) into (25), for any fixed m we obtain

$$|I(\mathbf{X}^n; \mathbf{Y}^n, S_{L_n}, L_m|s_0) - I(\mathbf{X}^n; \mathbf{Y}^n, S_{L_n}, L_m|s'_0)|$$

$$\leq d_m^*(s_0, s'_0) \log(2^{n-m}) + \log(2^{2\lambda+q+m}Q^2). \quad (27)$$

Now, we consider the mutual information in (22) for the original (non genie-aided) channel. By using the chain rule and (12) we have

$$\begin{aligned} & |I(\mathbf{X}^n; \mathbf{Y}^n, S_{L_n}, L_m | s_0) - I(\mathbf{X}^n; \mathbf{Y}^n, S_{L_n} | s_0)| \\ &= |I(\mathbf{X}^n; L_m | s_0, \mathbf{Y}^n, S_{L_n})| \leq \log\left(\frac{2Qm}{Q^2 - 1}\right). \end{aligned} \quad (28)$$

Using (27) and (28), we see that for any fixed m

$$\begin{aligned} & |I(\mathbf{X}^n; \mathbf{Y}^n, S_{L_n} | s_0) - I(\mathbf{X}^n; \mathbf{Y}^n, S_{L_n} | s'_0)| \\ &\leq d_m^*(s_0, s'_0) \log(2^{n-m}) + \log(2^{2\lambda+q+m}Q^2) + \log\left(\frac{2Qm}{Q^2 - 1}\right)^2. \end{aligned}$$

From Lemma 1 (we can easily prove that the condition (19) is satisfied here), for any $\varepsilon > 0$, we can first choose m sufficiently large, such that $d_m^*(s_0, s'_0) \leq \varepsilon$, then by letting n be sufficiently large we get

$$\frac{1}{n} |I(\mathbf{X}^n; \mathbf{Y}^n, S_{L_n} | s_0) - I(\mathbf{X}^n; \mathbf{Y}^n, S_{L_n} | s'_0)| \leq \varepsilon \log 2. \quad (29)$$

Therefore, we have proved (22). ■

Definition 6: The Markov channel capacity C^ν for binary Markov sources of order ν is defined as

$$C^\nu \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} C_n^\nu, \quad (30)$$

where

$$\begin{aligned} C_n^\nu &\triangleq \sup_{P^\nu(\mathbf{X}^n)} I(\mathbf{X}^n; \mathbf{Y}^n) \\ &= \sum_{\mathbf{x}^n \in \mathbb{B}^n} \int_{\mathbf{y}^n \in \mathbb{R}} P^\nu(\mathbf{x}^n) P(\mathbf{y}^n | \mathbf{x}^n) \log \frac{P(\mathbf{y}^n | \mathbf{x}^n)}{P(\mathbf{y}^n)} \quad (31) \end{aligned}$$

and $P^\nu(\cdot)$ is the set of stationary and ergodic Markov distributions of order ν . □

The following lemma provides a sufficient condition for the existence of the limit given in (30).

Lemma 3: Given a sequence $\{C_n\}$, and constants $A > 0$ and $B > 0$, such that for all $m \in \mathbb{Z}^+$ and $n \in \mathbb{Z}^+$

$$0 \leq C_n \leq nA \quad (32)$$

$$C_{m+n} \leq C_m + C_n + B, \quad (33)$$

then the limit $C = \lim_{n \rightarrow \infty} \frac{1}{n} C_n$ exists.

Proof: Without loss of generality, let $n = km + q$, where $k \in \mathbb{Z}^+$ is any fixed integer and $q \in \{0, 1, \dots, k-1\}$. From (33), we have

$$\frac{C_n}{n} = \frac{C_{km+q}}{km+q} \leq \frac{mC_k + C_q + (m+1)B}{km+q} \quad (34)$$

Since both sides of (34) are bounded sequences as easily verified from (32), the lim sup exists and we have

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{C_n}{n} &\leq \limsup_{m \rightarrow \infty} \frac{mC_k + C_q + (m+1)B}{km+q} \\ &= \frac{C_k + B}{k}. \end{aligned} \quad (35)$$

Since (35) holds for any fixed $k \in \mathbb{Z}^+$, therefore we have

$$\limsup_{n \rightarrow \infty} \frac{C_n}{n} \leq \liminf_{k \rightarrow \infty} \frac{C_k}{k}. \quad (36)$$

This proves the existence of $C = \lim_{n \rightarrow \infty} \frac{C_n}{n}$. ■

Next, we prove that the limit in (30) exists.

Theorem 2: Let $\mathbf{X}^n \in \mathbb{B}^n$ be a Markov source of order ν with distribution $P^\nu(\mathbf{X}^n)$, and let $\mathbf{Y}^n \in \mathbb{R}$ be the corresponding output of the channel given by (6), (7) and (8). The limit $C^\nu \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} \sup_{P^\nu(\mathbf{X}^n)} I(\mathbf{X}^n; \mathbf{Y}^n)$ in (30) exists.

Proof: Since the input \mathbf{X}^n is a binary vector, we have

$$0 \leq I(\mathbf{X}^n; \mathbf{Y}^n) \leq n \log 2. \quad (37)$$

Next, consider the channel input $\mathbf{X}^{m+n} = X_1^{m+n}$ and the corresponding output $\mathbf{Y}^{m+n} = Y_1^{L_{m+n}}$. Let $\mathbf{X}_1 = X_1^m$, and $\mathbf{X}_2 = X_{m+1}^{m+n}$ be two input sub-blocks, and let $\mathbf{Y}_1 = Y_1^{L_m}$, $\mathbf{Y}_2 = Y_{L_m+1}^{L_{m+n}}$ be the corresponding output sub-blocks, with L_m defined in (11). Obviously, \mathbf{Y}^{m+n} is functionally dependent on the two variables \mathbf{Y}_1 and \mathbf{Y}_2 , therefore

$$\begin{aligned} I(\mathbf{X}^{m+n}; \mathbf{Y}^{m+n}) &\leq I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_1, \mathbf{Y}_2) \\ &= I(\mathbf{X}_1; \mathbf{Y}_1) + I(\mathbf{X}_1; \mathbf{Y}_2 | \mathbf{Y}_1) + I(\mathbf{X}_2; \mathbf{Y}_1 | \mathbf{X}_1) \\ &\quad + I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{X}_1, \mathbf{Y}_1). \end{aligned} \quad (38)$$

Since the noise N_i are i.i.d. and the source is Markov of order ν , we have

$$\begin{aligned} I(\mathbf{X}_1; \mathbf{Y}_2 | \mathbf{Y}_1) &\leq I(\mathbf{X}_1; \mathbf{Y}_2 | \mathbf{Y}_1, S_{L_m}) + \log(2^\lambda Q) \\ &= H(\mathbf{Y}_2 | \mathbf{Y}_1, S_{L_m}) - H(\mathbf{Y}_2 | \mathbf{X}_1, \mathbf{Y}_1, S_{L_m}) + \log(2^\lambda Q) \\ &\leq \log(2^\lambda Q). \end{aligned} \quad (39)$$

Similarly, we can prove the following inequalities

$$\begin{aligned} I(\mathbf{X}_2; \mathbf{Y}_1 | \mathbf{X}_1) &= H(\mathbf{Y}_1 | \mathbf{X}_1) - H(\mathbf{Y}_1 | \mathbf{X}_1, X_{m+1}^{m+n}) \\ &\leq \log 2^q \end{aligned} \quad (40)$$

$$\begin{aligned} I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{X}_1, \mathbf{Y}_1) &\leq I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{X}_1, \mathbf{Y}_1, S_{L_m}) + \log(2^\lambda Q) \\ &= I(\mathbf{X}_2; \mathbf{Y}_2 | S_{L_m}) + \log(2^\lambda Q) \\ &\leq I(\mathbf{X}_2; \mathbf{Y}_2) + 2 \log(2^\lambda Q). \end{aligned} \quad (41)$$

By substituting (39), (40) and (41) into (38), and then taking the supremum of both sides over all Markov distributions of order ν , we have

$$C_{m+n}^\nu \leq C_m^\nu + C_n^\nu + \log(2^{3\lambda+q}Q^3), \quad (42)$$

where the last term does not depend on m or n . Thus by using Lemma 3, we prove the existence of the limit C^ν . ■

Lemma 4 (4.3 in [7]): Let $f(a)$, where $a \in \mathcal{A}$ be a function with values in the set \mathcal{G} . Denote by $|f^{-1}(g)|$ the number of solutions of the equation $f(a) = g$, where $g \in \mathcal{G}$. Then

$$E(|i(X; Y) - i(f(X); Y)|) \leq \max_{g \in \mathcal{G}} \log |f^{-1}(g)|. \quad (43)$$

Since the limit in (30) exists, in order to prove the information stability of the channel defined in (6), (7) and (8) for binary Markov sources of any order ν , it is sufficient ([5, 8]) to prove the following:

Theorem 3: For any $\nu \in \mathbb{Z}^+$, there exists a sequence of binary processes $\mathbf{X}^n = X_1^n$, and the corresponding output $\mathbf{Y}^n = Y_1^{L_n}$ through the channel defined by (6), (7) and (8), such that for any $\rho > 0$

$$\lim_{n \rightarrow \infty} P(|i(\mathbf{X}^n; \mathbf{Y}^n) - nC^\nu| > n\rho) = 0, \quad (44)$$

where C^ν is given in (30).

Proof: (The proof we provide is extended from the proof of (4.8) in [7].) From theorem 2 and the definition in (30) and (31), we know that there exists a sequence of order ν Markov process $\mathbf{X}^k = X_1^k$ with pmf $P_k^\nu(\mathbf{X})$, and the corresponding channel output $\mathbf{Y}^k = Y_1^{L_k}$ such that

$$\lim_{k \rightarrow \infty} I_k = \lim_{k \rightarrow \infty} \frac{1}{k} I(\mathbf{X}^k; \mathbf{Y}^k) = \lim_{k \rightarrow \infty} \frac{1}{k} I(\mathbf{X}^k; \mathbf{Y}^k, S_{L_k}) = C^\nu. \quad (45)$$

We now construct a new *super-sequence* $\mathbf{X}^n = X_1^n = (X_1^k X_{k+1}^{2k} \cdots X_{(g-1)k+1}^{gk} X_{gk+1}^{gk+r})$, according to the distribution $P_k^\nu(\mathbf{X})$, as the input to the channel, where $n = gk + r$ is an arbitrary integer and $0 \leq r < k$. We denote the *sub-sequence* $X_{(t-1)k+1}^{tk}$ by \mathbf{X}_t , for $1 \leq t \leq g$, and denote X_{gk+1}^{gk+r} by \mathbf{X}_{g+1} .

We first consider a *genie-aided* auxiliary channel, where \mathbf{Y}^n is the channel output when \mathbf{X}^n is the channel input, and a genie provides the receiver the following side information:

- The number of samples corresponding to each sub-sequence $X_{(t-1)k+1}^{tk}$, i.e., the values ℓ_{tk} for $1 \leq t \leq g$ are known to the receiver. For simplicity, we denote $Y_{\ell_{(t-1)k+1}}^{\ell_{tk}}$ by \mathbf{Y}_t , for $1 \leq t \leq g$, and $Y_{\ell_{gk+1}}^{\ell_n}$ by \mathbf{Y}_{g+1} .
- The channel states at the beginning of each sub-sequence, i.e., $\xi_t = S_{\ell_{tk}} \in \mathbb{S}$ for $1 \leq t \leq g$ are known to the receiver.

Denote the output of the genie-aided channel by $\tilde{\mathbf{Y}}^n = (\mathbf{Y}_1 \mathbf{Y}_2 \cdots \mathbf{Y}_{g+1}, \xi_1 \xi_2 \cdots \xi_g)$. The actual transition law of the genie-aided channel is

$$q_{n,k}(\tilde{\mathbf{Y}}^n | \mathbf{X}^n) = q_{n,k}(\tilde{\mathbf{Y}}^n | \mathbf{X}^n, \xi_0) \\ = P(\mathbf{Y}_{g+1} | \mathbf{X}_{g+1}, \xi_g) \prod_{t=1}^g P(\mathbf{Y}_t, \xi_t | \mathbf{X}_t, \xi_{t-1}) \quad (46)$$

where $P(\bullet | \bullet)$ is the transmission law of the original length- k channel. From the finite-order Markov property, the *information density* [7] between the input and output is

$$i(\mathbf{X}^n; \tilde{\mathbf{Y}}^n) = \sum_{t=1}^g i(\mathbf{X}_t; \mathbf{Y}_t, \xi_t | \xi_{t-1}) + i(\mathbf{X}_{g+1}; \mathbf{Y}_{g+1} | \xi_g). \quad (47)$$

Therefore, for any $\rho > 0$ we have

$$P(|i(\mathbf{X}^n; \tilde{\mathbf{Y}}^n) - nC^\nu| > n\rho) \\ \leq P\left(\left|\frac{i(\mathbf{X}^n; \tilde{\mathbf{Y}}^n)}{n} - I_k\right| > \frac{\rho}{2}\right) + P(|I_k - C^\nu| > \frac{\rho}{2}) \\ \leq P\left(\left|\frac{1}{g} \left[\sum_{t=1}^g \frac{1}{k} \cdot i(\mathbf{X}_t; \mathbf{Y}_t, \xi_t | \xi_{t-1})\right] + \frac{1}{n} i(\mathbf{X}_{g+1}; \mathbf{Y}_{g+1} | \xi_g) - I_k\right| > \frac{\rho}{2}\right) \\ + P(|I_k - C^\nu| > \frac{\rho}{2}). \quad (48)$$

From (45), $\forall \epsilon > 0$ we can always choose k sufficiently large such that the second term on the right hand side of (48) is less than $\frac{\epsilon}{2}$. Then, by using Theorem 1 and the law of large numbers, for any k we can choose g sufficiently large such that the first term on the right hand side of (48) is less than $\frac{\epsilon}{2}$. Therefore, there exists $n_0 \in \mathbb{Z}^+$, such that $\forall n > n_0$

$$P(|i(\mathbf{X}^n; \tilde{\mathbf{Y}}^n) - nC^\nu| > n\rho) < \epsilon. \quad (49)$$

Notice that the above result is only for the *genie-aided* channel. To change back to the original channel, we resort to Lemma 4 and introduce the following function

$$f(\tilde{\mathbf{Y}}^n) = \mathbf{Y}^n. \quad (50)$$

We need to determine the maximal number of inverse images of the above function for a given received sequence. According to (12), the number of possible values of ℓ_{tk} , given $\ell_{(t-1)k}$, is at most $\frac{2Q}{Q^2-1}$. Therefore,

$$|f^{-1}(\mathbf{Y}^n)| \leq |\mathbb{S}|^g \cdot \left(\frac{2Q}{Q^2-1}\right)^g = \left(\frac{2^{2q+1}Q^2}{Q^2-1}\right)^g. \quad (51)$$

By using Chebyshev's inequality, we can prove

$$P(|i(\mathbf{X}^n; \mathbf{Y}^n) - nC^\nu| > n\rho) \\ \leq P(|i(\mathbf{X}^n; \mathbf{Y}^n) - i(\mathbf{X}^n; \tilde{\mathbf{Y}}^n)| + |i(\mathbf{X}^n; \tilde{\mathbf{Y}}^n) - nC^\nu| > n\rho) \\ \leq P(|i(\mathbf{X}^n; \mathbf{Y}^n) - i(\mathbf{X}^n; \tilde{\mathbf{Y}}^n)| > \frac{n\rho}{2}) + P(|i(\mathbf{X}^n; \tilde{\mathbf{Y}}^n) - nC^\nu| > \frac{n\rho}{2}) \\ \leq \frac{E(|i(\mathbf{X}^n; \mathbf{Y}^n) - i(\mathbf{X}^n; \tilde{\mathbf{Y}}^n)|)}{\frac{n\rho}{2}} + P(|i(\mathbf{X}^n; \tilde{\mathbf{Y}}^n) - nC^\nu| > \frac{n\rho}{2}) \\ \leq \frac{2g \cdot \log\left(\frac{2^{2q+1}Q^2}{Q^2-1}\right)}{n\rho} + P(|i(\mathbf{X}^n; \tilde{\mathbf{Y}}^n) - nC^\nu| > \frac{n\rho}{2}). \quad (52)$$

The first term on the right hand side of (52) can be made arbitrarily small by letting $k \rightarrow \infty$. The second term can be made arbitrarily small for sufficiently large k and g according to (49). Therefore, $\forall \rho > 0$ and $\forall \epsilon > 0$, we can choose sufficiently large k and g , such that

$$P(|i(\mathbf{X}^n; \mathbf{Y}^n) - nC^\nu| > n\rho) < \epsilon. \quad (53)$$

From (45) and (53), using a standard diagonalization argument we have proven (44). ■

Hence we have shown the information stability of the channel for any binary Markov input of finite order. This indicates that for any rate less than C^ν given by (30), there exists a coding scheme that can achieve arbitrarily small probability of decoding error for sufficiently large n .

VI. CONCLUSION

We have proven the information stability for the class of channels corrupted by timing errors, finite ISI and additive Gaussian noise, when the input is a stationary and ergodic Markov process of finite order. Therefore, we have derived a set of lower bounds on the achievable transmission rates for such a channel.

REFERENCES

- [1] R. G. Gallager, "Sequential decoding for binary channels with noise and synchronization errors," 1961, unpublished Lincoln Lab. report 25 G-2.
- [2] J. D. Ullman, "On the capabilities of codes to correct synchronization errors," *IEEE Trans. Information Theory*, vol. 13, no. 5, pp. 95–105, January 1967.
- [3] S. N. Diggavi and M. Grossglauser, "On transmission over deletion channels," *Allerton Conference 2001, Monticello, Illinois*, October 2001.
- [4] E. Drinea and M. Mitzenmacher, "On lower bounds for the capacity of deletion channels," in *Proc. of the IEEE International Symposium on Information Theory*, Chicago, Illinois, June 2004, p. 227.
- [5] R. L. Dobrushin, "The computation on a computer of the channel capacity of a line with symbols drop-out," *Problems of Information Transmission*, vol. 4, no. 3, pp. 92–95, 1968.
- [6] A. Kavcic and R. Motwani, "Insertion/deletion channels: Reduced-state lower bounds on channel capacities," in *Proc. IEEE International Symposium on Information Theory 2004*, Chicago, Illinois, June 2004, p. 229.
- [7] R. L. Dobrushin, "Shannon's theorems for channels with synchronization errors," *Problems of Information Transmission*, vol. 3, no. 4, pp. 18–36, 1967.
- [8] —, "General formulation of Shannon's main theorem in information theory," *American Math. Soc. Trans.*, vol. 33, pp. 323–438, 1963.
- [9] R. G. Gallager, *Information Theory and Reliable Communication*. USA: John Wiley & Sons, 1968.