Real-Time Coding of Markov Sources over Erasure Channels: When is Binning Optimal?

Ashish Khisti and Farrokh Etezadi
University of Toronto
10 King's College Road
Toronto, ON, Canada
Email: {akhisti, fetezadi}@comm.utoronto.ca

Mitchell Trott
HP Labs
1501 Pagemill Road
Palo Alto, California
Email: mitchell.trott@hp.com

Abstract—We study sequential coding of a Markov source process under error propagation constraints. The channel can erase up to B packets in a single burst, but reveals all other packets to the destination. The destination is required to reproduce all the (vector) source sequences sequentially, except those that occur in a window of length $B+W$ following the start of the erasure burst. Our earlier work establishes upper and lower bounds on the compression rate as a function of B and W. In this work we show that for the class of *symmetric* sources, if we restrict to a memoryless encoding function, then a binning based scheme is optimal. Our converse involves a drawing connection between the sequential coding problem and a multiuser source coding problem called Zig-Zag source coding with side information.

I. INTRODUCTION

A tradeoff between compression efficiency and error resilience is fundamental to any video compression system. In live video streaming, an encoder observes a sequence of correlated video frames and produces a compressed bit-stream that is transmitted to the destination. If the underlying channel is an ideal bit-pipe, it is well known that predictive coding [1] achieves the optimum compression rate. Unfortunately in many emerging video distribution networks, such as peer-to-peer systems and mobile systems, packet losses are unavoidable. Predictive coding is highly sensitive to such packet losses and can lead to a significant amount of error propagation. Various techniques are used to practice prevent such losses. Commonly used video coding techniques use a group of picture (GOP) architecture, where intra-frames are periodically inserted to limit the effect of error propagation. Forward error correction codes can also be applied to compressed bit-streams to recover from missing packets [2], [3]. Modifications to predictive coding, such as leaky-DPCM [4], [5], have been proposed in the literature to deal with packet losses. The robustness of distributed video coding techniques in presence of packet losses has been studied in e.g., [6], [7].

Information theoretic analysis of video coding has received significant attention in recent times, see e.g., [8]–[10] and the references therein. These works focus primarily on the source coding aspects of video. The *source process* is a sequence of vectors, each of which is spatially i.i.d. and temporally correlated. Each source vector is sequentially compressed into a bit stream. The destination is required to recover the source vectors in a sequential manner as well. However all of these works assume an ideal channel with no packet losses. To our knowledge even the effect of a single isolated packet loss is not fully understood [11].

In an earlier work [12] we build upon [8], [9] and introduce an information theoretic framework to characterize the tradeoff between error propagation and compression rate. An encoder is revealed source vectors in a sequential manner and compresses them sequentially into channel packets that are then transmitted over a channel. An information theoretic notion of *error propagation* is defined and upper and lower bounds are obtained on the compression rate. The lower bound is based on a careful analysis of information flow during the decoding process whereas the upper bound is based on a binning technique. For a special class of sources, a new technique, *prospicient coding* is proposed, and shown to be optimal. However the optimal compression rate remains an open problem for a large class of sources including the binary symmetric Markov source.

In this paper we consider a class of *symmetric* Markov sources. For this class the minimum compression rate has not yet been characterized. Our main result in this paper is that if we restrict to the class of *memoryless* encoders, then the binning based scheme proposed in [12] is optimal. The converse is established by drawing a connection to a multiuser source coding problem called zig-zag source coding network [13]–[15].

II. PROBLEM STATEMENT

A. Source Model

We consider a semi-infinite stationary vector source process $\{s_t^n\}_{t\geq 0}$ whose symbols (defined over some finite alphabet $\mathcal{S}$) are drawn independently across the spatial dimension and from a first-order Markov chain across the temporal dimension, i.e., for each $t \geq 1$,

$$\Pr(\ s_t^n = s_t^n\ |\ s_{t-1}^n = s_{t-1}^n\ ) = \prod_{j=1}^{n} p_{s_1|s_0}(s_{tj}|s_{t-1,j}), \quad \forall t \geq 1.$$

(1)

We assume that the prior distribution $p_{s_0}(\cdot)$ and $p_{s_1|s_0}(\cdot|\cdot)$ are selected such that the underlying random variables $\{s_t\}_{t\geq 0}$ constitute a time-invariant and a first-order stationary Markov

$$s_0 \longrightarrow s_1 \longrightarrow s_2 \longrightarrow \cdots \longrightarrow s_{j-1} \rightarrow s_j \rightarrow s_{j+1} \cdots s_{j+B-1} \rightarrow s_{j+B} \rightarrow \cdots \longrightarrow s_{j+B+W-1} \rightarrow s_{j+B+W}$$
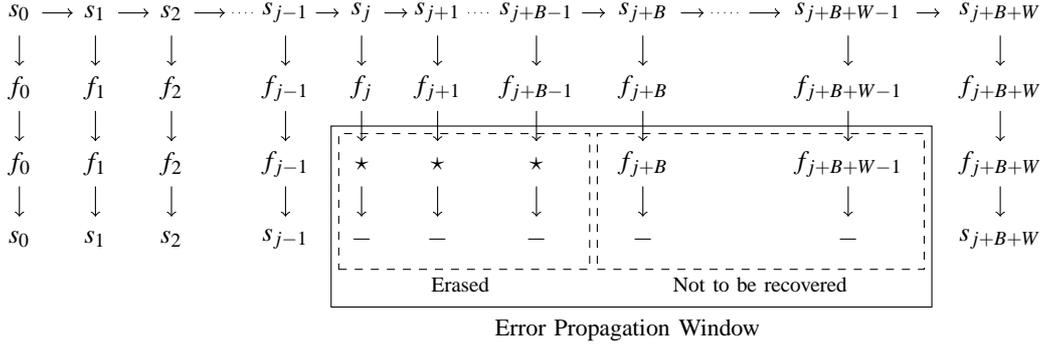
Fig. 1. Problem Setup. As shown in the figure, the channel introduces an erasure burst of length B between the interval $[j, j+B-1]$. The decoder is not required to reproduce the source sequences that occur in the interval $[j, j+B+W-1]$. All other sequences need to be reproduces in an instantaneous and lossless manner.

chain. Of particular interest in this paper is the class of *symmetric* sources where the underlying Markov chain is also reversible i.e., the random variables satisfy $(s_0, \ldots, s_t) \overset{d}{=} (s_t, \ldots, s_0)$, where the equality is in the sense of distribution [16]. Of particular interest to us is the following property satisfied for each $t \geq 1$.

$$p_{s_{t+1},s_t}(s_a, s_b) = p_{s_{t-1},s_t}(s_a, s_b), \quad \forall s_a, s_b \in \mathcal{S} \quad (2)$$

i.e., we can "exchange" the source pair (s_{t+1}^n, s_t^n) with (s_{t-1}^n, s_t^n) without affecting the joint distribution. An important class of sources that are symmetric are the binary sources: $s_t^n = s_{t-1}^n \oplus z_t^n$, where $\{z_t^n\}_{t \geq 1}$ is an i.i.d. binary source process (in both temporal and spatial dimensions) with the marginal distribution $\Pr(z_{t,i} = 0) = p$, the marginal distribution $\Pr(s_{t,i} = 0) = \Pr(s_{t,i} = 1) = \frac{1}{2}$ and $\oplus$ denotes modulo-2 addition.

B. Rate-Recovery Function

A rate-R causal encoder maps the sequence $\{s_i^n\}_{i \geq 0}$ to an index $f_i \in [1, 2^{nR}]$ according to some function

$$f_i = \mathcal{F}_i(s_0^n, \ldots, s_i^n) \quad (3)$$

for each $i \geq 0$. A *memoryless* encoder satisfies $\mathcal{F}_i(s_0^n, \ldots, s_i^n) = \mathcal{F}_i(s_i^n)$ i.e., the encoder does not use the knowledge of the past sequences.

The channel introduces an erasure burst of size B, i.e. for some particular $j \geq 0$, it introduces an erasure burst such that $g_i = \star$ for $i \in \{j, j+1, \ldots, j+B-1\}$ and $g_i = f_i$ otherwise. Upon observing the sequence $\{g_j\}_{j \geq 0}$ the decoder is required to perfectly recover all the source sequences using decoding functions

$$\hat{s}_i^n = \mathcal{G}_i(g_0, g_1, \ldots, g_i), \quad i \notin \{j, \ldots, j+B+W-1\}. \quad (4)$$

It is however not required to produce the source sequences in the window of length $B+W$ following the start of an erasure burst. We call this period the error propagation window. The setup is shown in Fig. 1.

A rate $R(B, W)$ is feasible if there exists a sequence of encoding and decoding functions and a sequence ε_n that

approaches zero as $n \rightarrow \infty$ such that, $\Pr(s_i^n \neq \hat{s}_i^n) \leq \varepsilon_n$ for all $i \notin \{j, \ldots, j+B+W-1\}$. We seek the minimum feasible rate $R(B, W)$ which we define to be the *rate-recovery* function. The following upper and lower bounds have been established in [12].

Theorem 1 ([12]): For any stationary first-order Markov source process the rate-recovery function satisfies $R^-(B, W) \leq R(B, W) \leq R^+(B, W)$ where

$$R^-(B, W) = H(s_1|s_0) + \frac{1}{W+1} I(s_B; s_{W+B+1}|s_0). \quad (5)$$

$$R^+(B, W) = H(s_1|s_0) + \frac{1}{W+1} I(s_B \; ; \; s_{B+1}|s_0) \quad (6)$$

□

Notice that the upper and lower bound coincide for $W = 0$ and $W \rightarrow \infty$, yielding the rate-recovery function in these cases. The upper bound is obtained via a memoryless binning based scheme. At each time the encoding function f_i in (3) is obtained as the bin-index of an independent Slepian-Wolf codebook [17]. The rate expression for $R^+(B, W)$, which is equivalent to [12]

$$R^+(B, W) = \frac{1}{W+1} H(s_{B+1}, s_{B+2}, \ldots, s_{B+W+1}|s_0) \quad (7)$$

guarantees that the decoder can recover s_{j+B+W}^n following an erasure burst between $[j, j+B-1]$ using the $W+1$ bin indices $f_{j+B}, \ldots, f_{j+B+W}$ and the source sequence s_{j-1}^n before the erasure.

In [12] some counter-examples are provided where the lower bound $R^-(B, W)$ (c.f. (5)) is tight and binning based upper bound $R^+(B, W)$ (c.f. (6)) is not optimal in general. Nevertheless such examples require a special structure and do not include many natural source models such as the binary symmetric sources. The optimal rate-recovery function for the class of symmetric sources remains open. In this paper we establish the optimality of binning based scheme if one restricts to the class of memoryless encoders.

Theorem 2: For the class of symmetric sources that satisfy (2) the rate-recovery function, restricted to the class of

memoryless encoders, is given by

$$R(B,W) = \frac{1}{W+1} H(s_{B+1}, s_{B+2}, \ldots, s_{B+W+1} | s_0). \quad (8)$$

$\square$

Note that the achievability follows immediately from (7). Thus it only remains to show that the lower bound (5) needs to be improved. We have only been able to obtain this improvement for the class of memoryless encoders. For the general encoder structure (3) this remains an open problem. At first glance one may expect that the binning based scheme is always optimal for the class of memoryless encoders. This is however not true. Interestingly the *prospicient encoders* in [12] that improve upon the binning based lower bound are also memoryless. Our proof involves an interesting connection a multi-terminal source coding problem called zig-zag source coding [13]–[15]. In particular we develop a simple approach to lower bound the sum-rate of a zig-zag source coding network with symmetric sources that may be of independent interest.

III. PROOF OF THEOREM 2

The special case when $W = 0$ follows directly from (5). We only need to consider the case when $W \geq 1$. For simplicity in exposition we consider the case when $W = 1$. Then we need to show that

$$R(B, W = 2) \geq \frac{1}{2} H(s_{B+1}, s_{B+2} | s_0) \quad (9)$$

The proof for general $W \geq 1$ follows along similar lines and will be sketched briefly.

Assume that an erasure-burst spans time indices $j - B, \ldots, j - 1$. The decoder must recover

$$\hat{s}_{j+1}^n = G_{j+1} \left(f_0^{j-B-1}, f_j, f_{j+1} \right). \quad (10)$$

From Fano's inequality, we have,

$$H \left(s_{j+1}^n \mid f_0^{j-B-1}, f_j, f_{j+1} \right) \leq n\varepsilon_n. \quad (11)$$

Furthermore if there is no erasure until time j then

$$\hat{s}_j^n = G_j \left(f_0^j \right) \quad (12)$$

must hold. Hence from Fano's Inequalty,

$$H \left(s_j^n \mid f_0^j \right) \leq n\varepsilon_n. \quad (13)$$

Our aim is to combine (11) and (13) to establish the following lower bound on the sum-rate

$$R_j + R_{j+1} \geq H(s_{j+1} | s_j) + H(s_j | s_{j-B-1}). \quad (14)$$

The lower bound then follows since

$$R \geq \max(R_j, R_{j+1}) \quad (15)$$

$$\geq \frac{1}{2}(R_j + R_{j+1}) \quad (16)$$

$$\geq \frac{1}{2}(H(s_{j+1} | s_j) + H(s_j | s_{j-B-1})) \quad (17)$$

$$= \frac{1}{2}(H(s_{j+1} | s_j, s_{j-B-1}) + H(s_j | s_{j-B-1})) \quad (18)$$

$$= \frac{1}{2} H(s_{j+1}, s_j | s_{j-B-1}) = \frac{1}{2} H(s_{B+1}, s_{B+2} | s_0) \quad (19)$$

thus establishing (9).

To establish (14) we make a connection to a multi-terminal source coding problem in Fig. 2.

A. Zig-Zag Source Coding

Consider the source coding problem with side information illustrated in Fig. 2(a). In this setup there are four source sequences drawn i.i.d. from a joint distribution $p(s_{j+1}, s_j, s_{j-1}, s_{j-B-1})$. The two encoders j and $j+1$ are revealed source sequences s_j^n and s_{j+1}^n and the two decoders j and $j+1$ are revealed sources s_{j-1}^n and s_{j-B-1}^n. The encoders operate independently and compress the source sequences to f_j and f_{j+1} at rates R_j and R_{j+1} respectively. Decoder j has access to (f_j, s_{j-1}^n) while decoder $j+1$ has access to $(f_j, f_{j+1}, s_{j-B-1}^n)$ and are interested in reproducing,

$$\hat{s}_j^n = \hat{G}_j(f_j, s_{j-1}^n) \quad (20)$$

$$\hat{s}_{j+1}^n = \hat{G}_{j+1}(f_j^{j+1}, s_{j-B-1}^n) \quad (21)$$

respectively such that $\Pr(s_i^n \neq \hat{s}_i^n) \leq \varepsilon_n$ for $i = j, j+1$.

When s_{j-B-1}^n is a constant sequence, the problem has been studied in [13], [15]. A complete single letter characterization involving an auxiliary random variable is obtained. Fortunately in the present case of symmetric sources a simple lower bond can be obtained using the following observation.

Lemma 1: The set of all achievable rate-pairs (R_j, R_{j+1}) for the problem in Fig. 2(a) is identical to the set of all achievable rate-pairs for the problem in Fig. 2(b) where the side information sequence s_{j-1}^n at decoder 1 is replaced by the side information sequence s_{j+1}^n.

The proof of Lemma 1 follows by observing that the capacity region for the problem in Fig. 2(a) depends on the joint distribution $p(s_j, s_{j+1}, s_{j-1}, s_{j-B-1})$ only via the *marginal* distributions $p(s_j, s_{j-1})$ and $p(s_{j+1}, s_j, s_{j-B-1})$. When the source is symmetric the distributions $p(s_j, s_{j-1})$ and $p(s_j, s_{j+1})$ are identical. The formal proof will be omitted.

Thus it suffices to lower bound the achievable sum rate for the problem in Fig. 2(b). First upon applying the Slepian-Wolf lower bound to encoder $j+1$

$$nR_{j+1} \geq H(s_{j+1}^n | s_{j-B-1}^n, f_j) - n\varepsilon_n \quad (22)$$

and to bound R_j

$$nR_j \geq H(f_j) = I(f_j; s_j^n | s_{j-B-1}^n)$$

$$\geq H(s_j^n | s_{j-B-1}^n) - H(s_j^n | s_{j-B-1}^n, f_j)$$

$$\geq nH(s_{B+1} | s_0) - H(s_j^n | s_{j-B-1}^n, f_j)$$

$$\qquad + H(s_j^n | s_{j-B-1}^n, s_{j+1}^n, f_j) - n\varepsilon_n \quad (23)$$

$$= nH(s_j | s_{j-B-1}) - I(s_j^n; s_{j+1}^n | s_{j-B-1}^n, f_j) - n\varepsilon_n$$

$$= nH(s_j | s_{j-B-1}) - H(s_{j+1}^n | s_{j-B-1}^n, f_j) + H(s_{j+1}^n | s_{j-B-1}^n, s_j^n, f_j) - n\varepsilon_n$$

$$= nH(s_j | s_{j-B-1}) - H(s_{j+1}^n | s_{j-B-1}^n, f_j) + nH(s_{j+1} | s_j) - n\varepsilon_n \quad (24)$$

where (23) follows by applying Fano's inequality to since s_j^n can be recovered from (s_{j+1}^n, f_j) and hence $H(s_j^n | s_{j-B-1}^n, s_{j+1}^n, f_j) \leq n\varepsilon_n$ holds and (24) follows form the Markov relation $s_{j+1}^n \to s_j^n \to (f_j, s_{j-B-1}^n)$. Observe that (14) follows by summing (22) and (24).
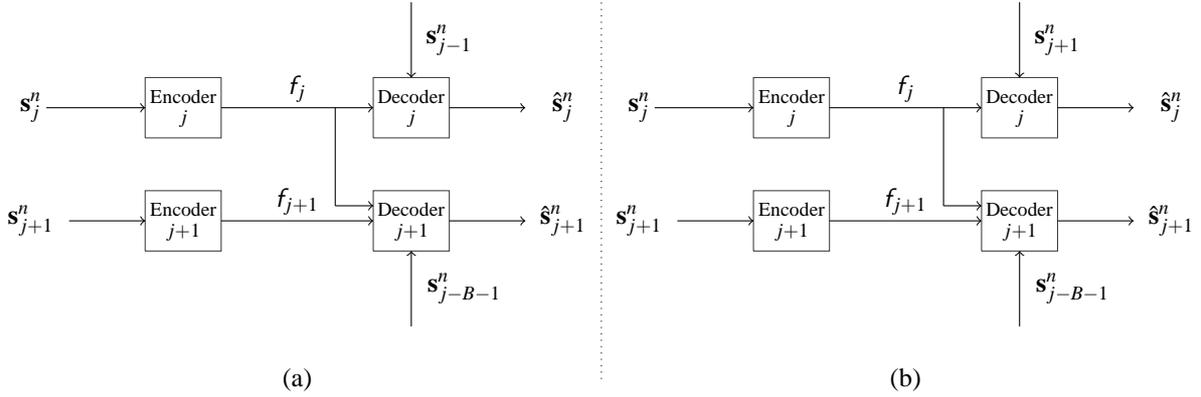
Fig. 2. Connection between the streaming problem and Zig-Zag source coding problem. The setup on the right is identical to the setup on the left, except with the side information sequence s_{j-1}^n replaced with s_{j+1}^n. However the rate region for both problems turns out to be identical for symmetric Markov sources.

B. Connection between Streaming and Zig-Zag Coding Problems

It remains to show that the lower bound on the Zig-Zag coding problem also constitutes a lower bound on the original problem.

Lemma 2: Suppose that the encoding function $f_j = \mathcal{F}_j(s_j^n)$ is memoryless. Suppose that there exist decoding functions $\hat{s}_j^n = \mathcal{G}_j(f_0^j)$ and $\hat{s}_{j+1}^n = \mathcal{G}_{j+1}(f_0^{j-B-1}, f_j, f_{j+1})$ such that $\Pr(\hat{s}_j^n \neq s_j^n)$ and $\Pr(\hat{s}_{j+1}^n \neq s_{j+1}^n)$ both vanish to zero as $n \to \infty$. Then

$$H(s_j^n | s_{j-1}^n, f_j) \leq n\varepsilon_n \qquad (25)$$

$$H(s_{j+1}^n | s_{j-B-1}^n, f_j, f_{j+1}) \leq n\varepsilon_n \qquad (26)$$

also hold.

We omit the proof due to space constraint. The conditions in (25) and (26) show that any rate that is achievable in the original problem is also achieved in the zig-zag source network. Hence a lower bound to the source network also constitutes a lower bound to the original problem.

C. Extension to Arbitrary $W > 1$

Finally we comment of the extension of the above approach to $W = 2$. We now consider three encoders $t \in \{j, j+1, j+2\}$. Encoder t observes a source sources s_t^n and compresses it into an index $f_j \in [1, 2^{nR_j}]$. The corresponding decoders are revealed s_{t-1}^n for $t \in \{j, j+1\}$ and the decoder $j+2$ is revealed s_{j-B-1}^n. By an argument analogous to Lemma 1 the rate region is equivalent to the case when decoders j and $j+1$ are instead revealed s_{j+1}^n and s_{j+2}^n respectively. For this new setup it is easy to show that decoder $j+2$ must reconstruct $(s_j^n, s_{j+1}^n, s_{j+2}^n)$ given $(s_{j-B-1}^n, f_j, f_{j+1}, f_{j+2})$. The sum rate must therefore satisfy $R_j + R_{j+1} + R_{j+2} \geq \frac{1}{3} H(s_j, s_{j+1}, s_{j+2} | s_{j-B-1})$. Using an extension of Lemma 2 we can show that the proposed lower bound also continues to hold for the original streaming problem. This completes the proof. The extension to any arbitrary $W > 1$ is completely analogous.

REFERENCES

[1] T. Berger, *Rate Distortion Theory, A Mathematical Basis for Data Compression.* NJ: Prentice-Hall Englewood Cliffs, 1971.

[2] W. Tan and A. Zakhor, "Video multicast using layered fec and scalable compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 373–386, 2001.

[3] Z. Li, A. Khisti, and B. Girod, "Forward error correction for low-delay packet video," in *Packet Video Workshop*, 2010.

[4] Y. Huang, Y. Kochman, and G. Wornell, "Causal transmission of colored source frames over a packet erasure channel," in *DCC*, 2010, pp. 129–138.

[5] D. Connor, "Techniques for reducing the visibility of transmission errors in digitally encoded video signals," *IEEE Trans. Comm.*, vol. 21, pp. 695–706, 1973.

[6] A. Sehgal, A. Jagmohan, and N. Ahuja, "Wyner-ziv coding of video: An error-resilient compression framework," *IEEE Trans. on Multimedia*, vol. 6, pp. 249–258, 2004.

[7] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Inform. Theory*, vol. 49, pp. 626–643, Mar. 2003.

[8] H. Viswanathan and T. Berger, "Sequential coding of correlated sources," *Information Theory, IEEE Transactions on*, vol. 46, no. 1, pp. 236 –246, jan 2000.

[9] N. Ma and P. Ishwar, "On delayed sequential coding of correlated sources," *IEEE Trans. Inform. Theory*, vol. 57, pp. 3763–3782, 2011.

[10] E. Yang, L. Zheng, D. He, and Z. Zhang, "Rate distortion theory for causal video coding: Characterization, computation algorithm, and comparison," *IEEE Trans. Inform. Theory*, vol. 57, pp. 5258–5280, 2011.

[11] J. Wang, V. Prabhakaran, and K. Ramchandran, "Syndrome-based robust video transmission over networks with bursty losses," in *ICIP*, Atlanta, GA, 2006.

[12] F. Etezadi, A. Khisti, and M. Trott, "Prospicient real-time coding of Markov sources over burst erasure channels: Lossless case," in *Data Compression Conference (DCC)*, 2012 (To Appear).

[13] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, G. Longo, Ed. Springer-Verlag, 1977, ch. 4.

[14] Y. Oohama, "Gaussian zig-zag source coding problem." *IEICE Transactions*, pp. 1889–1903, 2010.

[15] I. Csiszar and J. Korner, "Towards a general theory of source networks," *Information Theory, IEEE Transactions on*, vol. 26, no. 2, pp. 155 – 165, mar 1980.

[16] D. Aldous, "Reversible Markov Chains and Random Walks on Graphs (Chapter 3)," unpublished notes, available at http://www.stat.berkeley.edu/ aldous/RWG/book.html, Sep. 2002.

[17] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, pp. 471–480, July 1973.