Tighter Bounds on the Capacity of Finite-State Channels Via Markov Set-Chains

Jun Chen, *Member, IEEE*, Haim Permuter, *Member, IEEE*, and Tsachy Weissman, *Senior Member, IEEE*

***Abstract*—The theory of Markov set-chains is applied to derive upper and lower bounds on the capacity of finite-state channels that are tighter than the classic bounds by Gallager. The new bounds coincide and yield single-letter capacity characterizations for a class of channels with the state process known at the receiver, including channels whose long-term marginal state distribution is independent of the input process. Analogous results are established for finite-state multiple access channels.**

***Index Terms*—Channel capacity, finite-state channel, Markov set-chains, multiple access channels.**

I. Introduction

CONSIDER a finite-state channel with transition probability $P_{Y_tS_t|X_tS_{t-1}}$, where $X_t \in \mathcal{X}$, $Y_t \in \mathcal{Y}$, and $S_t \in \mathcal{S}$ are respectively the channel input, the channel output, and the channel state at time t. We assume that the channel is time-invariant, i.e., the transition probability $P_{Y_tS_t|X_tS_{t-1}}$ does not depend on t. Moreover, $|\mathcal{X}|$, $|\mathcal{Y}|$, $|\mathcal{S}|$ are assumed to be finite, where $|\mathcal{A}|$ denotes the cardinality of $\mathcal{A}$ for any set $\mathcal{A}$.

The capacity analysis of this channel model has received considerable attention due to its theoretical significance and practical implications [1]–[11]. A nice review of prior work on this subject, particularly regarding the simulation-based methods and related analytical results, can be found in [10] (see also [11]). In this work, we shall develop a new technique, based on the theory of Markov set-chains, to tackle this long-standing problem.

Denote the channel capacity as C. The capacity of finite-state channels can be characterized using the information spectrum method [12]; however, the resulting capacity formula is in general not computable. Moreover, the exact value of C may depend on the specific assumptions adopted in the definition of channel capacity, which include the realization of the initial state as well as the transmitter and receiver's knowledge of the initial state. To circumvent these subtle technical issues, we shall focus on the computable capacity bounds that are robust to such small variations in the definition of channel capacity. Define

$$\overline{C} = \lim_{k\to\infty} \max_{P_{X_1^k}} \max_{s_0} \frac{1}{k} I\left(X_1^k; Y_1^k | S_0 = s_0\right)$$
$$\underline{C} = \lim_{k\to\infty} \max_{P_{X_1^k}} \min_{s_0} \frac{1}{k} I\left(X_1^k; Y_1^k | S_0 = s_0\right).$$

It was shown in [3] that

$$\underline{C} \leq C \leq \overline{C} \tag{1}$$

and the inequalities in (1) become equalities for indecomposable channels. It is worth noting that (1) is valid for any initial state s_0; moreover, (1) is also valid whether or not the transmitter and the receiver know the initial state. Define

$$\overline{C}_k = \max_{P_{X_1^k}} \max_{s_0} \frac{1}{k} I\left(X_1^k; Y_1^k | S_0 = s_0\right) + \frac{1}{k}\log|\mathcal{S}| \tag{2}$$

$$\underline{C}_k = \max_{P_{X_1^k}} \min_{s_0} \frac{1}{k} I\left(X_1^k; Y_1^k | S_0 = s_0\right) - \frac{1}{k}\log|\mathcal{S}|. \tag{3}$$

It was shown in [3] that

$$\overline{C} = \lim_{k\to\infty} \overline{C}_k = \inf_k \overline{C}_k$$
$$\underline{C} = \lim_{k\to\infty} \underline{C}_k = \sup_k \underline{C}_k.$$

Therefore, we have

$$\underline{C}_k \leq \underline{C} \leq C \leq \overline{C} \leq \overline{C}_k \tag{4}$$

for all $k \geq 1$. Note that (4) provides computable finite-letter upper and lower bounds on the channel capacity, and for indecomposable channels, the bounds are asymptotically tight as $k \to \infty$. However, the complexity of computing $\overline{C}_k$ and $\underline{C}_k$ increases rapidly as k gets larger. Therefore, it is desirable if the bounds given by $\overline{C}_k$ and $\underline{C}_k$ are tight enough even for small k (ideally, $k = 1$). Unfortunately, $\overline{C}_k$ and $\underline{C}_k$ often give loose bounds for small k. First of all, the gap between $\overline{C}_k$ and $\underline{C}_k$ is at least $\frac{2}{k}\log|\mathcal{S}|$, which is not negligible for small k (particularly if the state space $\mathcal{S}$ is large). Furthermore, $\overline{C}_k$ and $\underline{C}_k$ do not coincide even when the second terms in (2) and (3) (i.e., $\pm\frac{1}{k}\log|\mathcal{S}|$) are removed. Indeed, since the behavior of the channel can be dramatically different in different states, the difference between the first terms in (2) and (3) can be as large as $\log|\mathcal{X}|$ when $k = 1$.

To see a possible direction for improving the upper and lower bounds, it is instructive to write (2) and (3) in a slightly different form. Let $\mathcal{P}(\mathcal{A})$ denote the set of probability distributions on $\mathcal{A}$ for any finite set $\mathcal{A}$. Moreover, for any finite sets $\mathcal{A}$ and $\mathcal{B}$,

Manuscript received March 01, 2008; revised May 05, 2009. Date of current version July 14, 2010.

J. Chen is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (e-mail: junchen@ece.mcmaster.ca).

H. Permuter was with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA. He is now with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel (e-mail: haimp@bgu.ac.il).

T. Weissman is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: tsachy@stanford.edu).

Communicated by H. Yamamoto, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2010.2050825

let $\mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{B}) = \{P_{AB} : \text{there exist } P_A \in \mathcal{P}(\mathcal{A}), P_B \in \mathcal{P}(\mathcal{B}) \text{ such that } P_{AB}(a,b) = P_A(a)P_B(b) \text{ for all } a \in \mathcal{A}, b \in \mathcal{B}\}$. It is easy to verify that

$$\overline{C}_k = \max_{P_{X_1^k}} \max_{P_{S_0|X_1^k} \in \mathcal{Q}\left(P_{X_1^k}, \mathcal{P}(\mathcal{S})\right)} \frac{1}{k} I\left(X_1^k; Y_1^k | S_0\right) + \max_{P_{S_0}} \frac{1}{k} H(S_0), \tag{5}$$

$$\underline{C}_k = \max_{P_{X_1^k}} \min_{P_{S_0|X_1^k} \in \mathcal{Q}'\left(P_{X_1^k}, \mathcal{P}(\mathcal{S})\right)} \frac{1}{k} I\left(X_1^k; Y_1^k | S_0\right) - \max_{P_{S_0}} \frac{1}{k} H(S_0) \tag{6}$$

where $\mathcal{Q}(P_{X_1^k}, \mathcal{P}(\mathcal{S})) = \left\{P_{S_0|X_1^k} : P_{X_1^k S_0} \in \mathcal{P}(\mathcal{X}^k \times \mathcal{S})\right\}$ and $\mathcal{Q}'(P_{X_1^k}, \mathcal{P}(\mathcal{S})) = \left\{P_{S_0|X_1^k} : P_{X_1^k S_0} \in \mathcal{P}(\mathcal{X}^k) \times \mathcal{P}(\mathcal{S})\right\}$. Note that any probability distribution on $\mathcal{X}^k \times \mathcal{S}$ can be thought of as a convex combination of probability distributions that are degenerate on (i.e., assign probability one to) a certain element of $\mathcal{X}^k \times \mathcal{S}$; each of these degenerate distributions is trivially in $\mathcal{P}(\mathcal{X}^k) \times \mathcal{P}(\mathcal{S})$. Therefore, we have $\mathcal{P}(\mathcal{X}^k \times \mathcal{S}) = \mathrm{conv}(\mathcal{P}(\mathcal{X}^k) \times \mathcal{P}(\mathcal{S}))$, where $\mathrm{conv}(\mathcal{P}(\mathcal{X}^k) \times \mathcal{P}(\mathcal{S}))$ denotes the convex hull of $\mathcal{P}(\mathcal{X}^k) \times \mathcal{P}(\mathcal{S})$. We shall redefine $\mathcal{Q}\left(P_{X_1^k}, \mathcal{P}(\mathcal{S})\right)$ as $\left\{P_{S_0|X_1^k} : P_{X_1^k S_0} \in \mathrm{conv}(\mathcal{P}(\mathcal{X}^k) \times \mathcal{P}(\mathcal{S}))\right\}$ to make its dependency on $\mathcal{P}(\mathcal{S})$ explicit. More generally, for any nonempty compact set $\mathcal{B} \subseteq \mathcal{P}(\mathcal{S})$, we define

$$\mathcal{Q}\left(P_{X_1^k}, \mathcal{B}\right) = \left\{P_{S_0|X_1^k} : P_{X_1^k S_0} \in \mathrm{conv}(\mathcal{P}(\mathcal{X}^k) \times \mathcal{B})\right\} \tag{7}$$

$$\mathcal{Q}'\left(P_{X_1^k}, \mathcal{B}\right) = \left\{P_{S_0|X_1^k} : P_{X_1^k S_0} \in \mathcal{P}(\mathcal{X}^k) \times \mathrm{conv}(\mathcal{B})\right\}. \tag{8}$$

A simple application of Carathéodory's theorem shows that each point in $\mathrm{conv}(\mathcal{B})$ can be represented as a convex combination of no more than $|\mathcal{S}|$ points in $\mathcal{B}$, and each point in $\mathrm{conv}(\mathcal{P}(\mathcal{X}^k) \times \mathcal{B})$ can be represented as a convex combination of no more than $|\mathcal{X}|^k|\mathcal{S}|$ points in $\mathcal{P}(\mathcal{X}^k) \times \mathcal{B}$. Moreover, since for any $P_{X_1^k S_0} \in \mathcal{P}(\mathcal{X}^k) \times \mathrm{conv}(\mathcal{B})$, there exist $m \in \mathbb{N}$, $\mu_i \in [0,1]$, and $P_{S_0}^{(i)} \in \mathcal{B}, i = 1, \ldots, m$, such that $\sum_{i=1}^m \mu_i = 1$ and

$$\begin{aligned} P_{X_1^k S_0}\left(x_1^k, s_0\right) &= P_{X_1^k}\left(x_1^k\right) \sum_{i=1}^m \mu_i P_{S_0}^{(i)}(s_0) \\ &= \sum_{i=1}^m \mu_i P_{X_1^k}\left(x_1^k\right) P_{S_0}^{(i)}(s_0) \\ & x_1^k \in \mathcal{X}^k, \quad s_0 \in \mathcal{S} \end{aligned}$$

we have $P_{X_1^k S_0} \in \mathrm{conv}(\mathcal{P}(\mathcal{X}^k) \times \mathcal{B})$, which further implies $\mathrm{conv}(\mathcal{P}(\mathcal{X}^k) \times \mathcal{B}) = \mathrm{conv}(\mathcal{P}(\mathcal{X}^k) \times \mathrm{conv}(\mathcal{B}))$.

Although writing $\overline{C}_k$ and $\underline{C}_k$ in the form of (5) and (6) is more cumbersome, once interpreted correctly, it offers an interesting new perspective and suggests possible directions for further improvement. Note that S_0 in (2) and (3) can be naturally interpreted as the channel state at time $t = 0$, and the basic intuition is that upper and lower bounds on the channel capacity can be derived by choosing the best and the worst initial states. On the other hand, S_0 in (5) and (6) is better interpreted as the channel state at time t with $t \to \infty$ (due to a time-shifting argument which will be clear later). Now to derive upper and lower bounds on the channel capacity, one has to optimize over all possible distributions of the channel state as $t \to \infty$. However, since the state process can be affected by the channel input, the limiting marginal distribution of the state process is hard to determine. To circumvent this difficulty, one may simply allow P_{S_0} to be any probability distribution from $\mathcal{P}(\mathcal{S})$. This is exactly the intuition underlying (5) and (6). It will be seen that, to derive tighter capacity bounds, one crucial idea is to find an effective estimate of P_{S_0}.

Note that X_1^k and S_0 are allowed to have an arbitrary joint distribution in (5) while they are assumed to be independent in (6). This difference may seem artificial since $\overline{C}_k$ is unaffected if in (5) we replace $\mathcal{Q}\left(P_{X_1^k}, \mathcal{P}(\mathcal{S})\right)$ with $\mathcal{Q}'\left(P_{X_1^k}, \mathcal{P}(\mathcal{S})\right)$. The purpose of choosing the current form is to motivate the fact that the set of admissible probability distributions $P_{X_1^k S_0}$ for $\overline{C}_k$ is exactly the convex hull of that for $\underline{C}_k$. Indeed, this relation will be preserved in the tightened upper and lower bounds with $\mathcal{P}(\mathcal{S})$ replaced by smaller compact sets.

The main contribution of this paper is a set of new finite-letter upper and lower bounds on the channel capacity. Specifically, we derive new upper and lower bounds C_k^U and C_k^L satisfying

$$\underline{C}_k \le C_k^L \le \underline{C} \le C \le \overline{C} \le C_k^U \le \overline{C}_k$$

where

$$C_k^U = \max_{P_{X_1^k}} \max_s \max_{P_{S_0|X_1^k} \in \mathcal{Q}\left(P_{X_1^k}, \mathcal{A}_s\right)} \frac{1}{k} I\left(X_1^k; Y_1^k | S_0\right) + \frac{1}{k} H(S_0) \tag{9}$$

$$C_k^L = \max_{P_{X_1^k}} \min_s \min_{P_{S_0|X_1^k} \in \mathcal{Q}'\left(P_{X_1^k}, \mathcal{A}_s\right)} \frac{1}{k} I\left(X_1^k; Y_1^k | S_0\right) - \frac{1}{k} H(S_0) \tag{10}$$

and the sets $\mathcal{A}_s$ $(s \in \mathcal{S})$, to be specified later, are used to capture the limiting marginal distribution of the state process. Similar to (5) and (6), S_0 in (9) and (10) should be interpreted as the channel state at time t with $t \to \infty$ rather than the channel state at time $t = 0$.

Due to the channel memory, the channel capacity is intimately related to the long-term behavior of the state process. For the special case in which the state process is unaffected by the channel input, i.e., $P_{S_t|X_t S_{t-1}}(s_t|x_t, s_{t-1})$ does not depend on x_t for all $x_t \in \mathcal{X}$ and all $s_{t-1}, s_t \in \mathcal{S}$, the state process is a homogeneous Markov chain, and its long-term behavior is well understood. However, for the general case, the state process depends on the channel input, which makes the problem more intricate. Fortunately, the theory of Markov set-chains allows us to obtain useful information[1] regarding the long-term behavior of the state process without knowing the channel input. The new capacity bounds are derived by effectively exploiting this information. In contrast, such information is not used in $\overline{C}_k$ and $\underline{C}_k$. By comparing (5) with (9) [as well as (6) with (10)], one can see two improvements. The first improvement, a relative minor one, results from the fact that the two terms in (5) [as well as in (6)] are decoupled while the two terms in (9) [as well as in (10)] are coupled. The second improvement is achieved by replacing $\mathcal{P}(\mathcal{S})$ with $\mathcal{A}_s$. Indeed, the key conceptual difference between the new bounds and the old bounds is succinctly manifested in this second improvement.

The rest of this paper is organized as follows. In Section II, we review some basic definitions and results from the theory of Markov set-chains. Along the way, we also derive a few new results, which will be useful for the later development. New finite-letter upper and lower bounds on the channel capacity are derived in Section III. The capacity bounds are further tightened for the case where the state process is known at the receiver. It is shown that these bounds coincide for a class of channels, yielding a single-letter capacity formula. Analogous results are derived for finite-state multiple access channels in Section IV. Several illustrative examples are given in Section V. We conclude the paper in Section VI. Throughout this paper, the logarithm function is to the base two.

We summarize below a few basic definitions that are used frequently in this paper. For two nonempty sets $\mathcal{A}, \mathcal{B} \subseteq \mathbb{R}^m$ and a scalar $c \in \mathbb{R}$, define

$$\mathcal{A}+\mathcal{B} = \{A+B : A \in \mathcal{A}, B \in \mathcal{B}\}$$
$$c\mathcal{A} = \{cA : A \in \mathcal{A}\}.$$

Let $\mathcal{B}_1, \mathcal{B}_2, \ldots$ be a sequence of nonempty sets in $\mathbb{R}^m$. Define the equation shown at the bottom of the page, where $\|\cdot\|$ is the L^1 norm. If $\limsup_{k\to\infty} \mathcal{B}_k = \liminf_{k\to\infty} \mathcal{B}_k = \mathcal{B}$, we shall write $\lim_{k\to\infty} \mathcal{B}_k = \mathcal{B}$, and refer to $\lim_{k\to\infty} \mathcal{B}_k$ as the sequential limiting set of $\mathcal{B}_1, \mathcal{B}_2, \ldots$. For any two nonempty compact sets $\mathcal{D}_1$ and $\mathcal{D}_2$ in $\mathbb{R}^m$, let

$$\delta(\mathcal{D}_1, \mathcal{D}_2) = \max_{D_1 \in \mathcal{D}_1} \min_{D_2 \in \mathcal{D}_2} \|D_1 - D_2\|$$
$$d(\mathcal{D}_1, \mathcal{D}_2) = \max(\delta(\mathcal{D}_1, \mathcal{D}_2), \delta(\mathcal{D}_2, \mathcal{D}_1)).$$

The function $d(\cdot,\cdot)$ is called the Hausdorff metric. Note that $\delta(\mathcal{D}_1, \mathcal{D}_2) = 0$ whenever $\mathcal{D}_1 \subseteq \mathcal{D}_2$, so while $d(\mathcal{D}_1, \mathcal{D}_2)$ small implies similarity between the sets $\mathcal{D}_1$ and $\mathcal{D}_2$, $\delta(\mathcal{D}_1, \mathcal{D}_2)$ small does not.

[1]This information is contained in $\mathcal{A}_s$.

II. Markov Set-Chains

A finite square matrix is called row-stochastic if all its entries are non-negative and the sum of each row is 1. Without loss of generality, we assume $\mathcal{S} = \{1, 2, \ldots, |\mathcal{S}|\}$. For each $x \in \mathcal{X}$, we shall view $P_{S_t|X_t,S_{t-1}}(\cdot|x,\cdot)$ as an $|\mathcal{S}| \times |\mathcal{S}|$ row-stochastic matrix, the (i,j)-entry of which is equal to $P_{S_t|X_t,S_{t-1}}(j|x,i)$, $i,j \in \mathcal{S}$. Let $\mathcal{T} = \{P_{S_t|X_t,S_{t-1}}(\cdot|x,\cdot)\}_{x\in\mathcal{X}}$. Define $\mathcal{T}^1 = \mathcal{T}$, $\mathcal{T}^2 = \{T_1T_2 : T_1, T_2 \in \mathcal{T}\}$, $\mathcal{T}^3 = \{T_1T_2T_3 : T_1, T_2, T_3 \in \mathcal{T}\}$, and so on. Let $\mathcal{G}_{s,k} = \{G : G \text{ is the } s\text{th row of } T, T \in \mathcal{T}^k\}$ for $s \in \mathcal{S}$ and $k \in \mathbb{N}$. Define $\mathcal{A}_s = \limsup_{k\to\infty} \mathcal{G}_{s,k}$ and $\mathcal{A}'_s = \liminf_{k\to\infty} \mathcal{G}_{s,k}$ for $s \in \mathcal{S}$. The properties of $\mathcal{A}_s$ and $\mathcal{A}'_s$ are summarized in the following lemma. The proof is straightforward and thus omitted.

Lemma 2.1:
1) $\mathcal{A}'_s \subseteq \mathcal{A}_s$;
2) $\mathcal{A}_s$ is a nonempty compact set;
3) $\lim_{k\to\infty} \delta(\mathcal{G}_{s,k}, \mathcal{A}_s) = 0$;
4) For any nonempty compact set $\mathcal{B}$ satisfying $\lim_{k\to\infty} \delta(\mathcal{G}_{s,k}, \mathcal{B}) = 0$, we have $\mathcal{A}_s \subseteq \mathcal{B}$;
5) $\mathcal{A}_s\mathcal{T} \subseteq \mathcal{A}_s$, where $\mathcal{A}_s\mathcal{T} = \{AT : A \in \mathcal{A}_s, T \in \mathcal{T}\}$;
6) If $\mathcal{A}'_s$ is a nonempty compact set, then $\lim_{k\to\infty} \delta(\mathcal{A}'_s, \mathcal{G}_{s,k}) = 0$;
7) For any nonempty compact set $\mathcal{B}'$ satisfying $\lim_{k\to\infty} \delta(\mathcal{B}', \mathcal{G}_{s,k}) = 0$, we have $\mathcal{B}' \subseteq \mathcal{A}'_s$.

Collecting all these properties, we obtain the following theorem.

Theorem 2.2:
1) $\mathcal{A}_s = \mathcal{A}'_s$ if and only if there exists a nonempty compact set $\mathcal{B}$ such that $\lim_{k\to\infty} d(\mathcal{G}_{s,k}, \mathcal{B}) = 0$;
2) $\mathcal{B}\mathcal{T} \subseteq \mathcal{A}_s$ for any $\mathcal{B} \subseteq \mathcal{A}_s$.

Remark: Part 1) of Theorem 2.2 is a special case of [13, Theorem 1].

Additional constraints on $\mathcal{T}$ are necessary in order to obtain a finer characterization of $\mathcal{A}_s$. First of all, we need to introduce a few definitions. A square row-stochastic matrix $B = (B_{ij})$ is regular if $\lim_{n\to\infty} B^n$ exists and has rank one, in which case all its rows are the same. Define

$$\lambda(B) = 1 - \min_{i,j} \sum_k \min(B_{ik}, B_{jk}).$$

$$\limsup_{k\to\infty} \mathcal{B}_k = \{B : \text{there exists a subsequence } B_{k_1}, B_{k_2}, \ldots \text{ with } B_{k_n} \in \mathcal{B}_{k_n} \text{ and } k_n \to \infty \text{ as } n \to \infty \text{ such that } \lim_{n\to\infty} \|B_{k_n} - B\| = 0\}$$
$$\liminf_{k\to\infty} \mathcal{B}_k = \{B' : \text{there exists a sequence } B_1, B_2, \ldots \text{ with } B_k \in \mathcal{B}_k \text{ such that } \lim_{k\to\infty} \|B_k - B'\| = 0\}$$

It can be shown [15] that

$$\lambda(B) = \frac{1}{2}\max_{i,j}\sum_k |B_{ik} - B_{jk}| = \max \|\gamma B\| \tag{11}$$

where the maximization is taken over row vectors satisfying $\|\gamma\| = 1$ and $\sum_i \gamma_i = 0$. We call B a scrambling matrix if $\lambda(B) < 1$. It is known [14] that scrambling matrices are regular, but not all regular matrices are scrambling; moreover, if one or more matrices in a product of square row-stochastic matrices is scrambling, so is the product. Let $B_1, B_2, \dots$ be a sequence of square row-stochastic matrices. Define $P^{(m,n)} = B_m B_{m+1} \dots B_{m+n}$. We say that the sequence $B_1, B_2, \dots$ is weakly ergodic if

$$\lim_{n\to\infty} P_{ik}^{(m,n)} - P_{jk}^{(m,n)} = 0$$

for all m, i, j, k. We say the set-chain $\mathcal{T}^1, \mathcal{T}^2, \dots$ is uniformly weakly ergodic if for any $\epsilon > 0$ there is an N such that any $P^{(n)} = T_1 \dots T_n$ with $n \geq N$ and $T_1, \dots, T_n \in \mathcal{T}$ satisfies

$$\max_{i,j,k}\left|P_{ik}^{(n)} - P_{jk}^{(n)}\right| < \epsilon.$$

Theorem 2.3: The following conditions are equivalent:
1) All finite products of matrices from $\mathcal{T}$ are regular;
2) There exists a finite N such that for all $n \geq N$ all products of n matrices from $\mathcal{T}$ are scrambling;
3) Every sequence of matrices from $\mathcal{T}$ is weakly ergodic;
4) There exists a finite N such that for all $n \geq N$ all products of n matrices from $\mathcal{T}$ have a column with all entries nonzero;
5) The set-chain $\mathcal{T}^1, \mathcal{T}^2, \dots$ is uniformly weakly ergodic.

Proof: The equivalence of the first four conditions is known [16]. Wolfowitz [17] proved that Condition 1) implies Condition 5). It is clear that Condition 5) implies Condition 3). Therefore, all these five conditions are equivalent. ■

Remark: Condition 4) is equivalent to Gallager's definition of indecomposable channels [3]. Thomasian [18] proposed an algorithm that can determine, in finite number of steps, whether Condition 1) is satisfied.

For any square row-stochastic matrix T with a stationary distribution π_T, let Π_T be a square matrix with rows equal to π_T.

Theorem 2.4: If all finite products of matrices from $\mathcal{T}$ are regular, then $\mathcal{T}^1, \mathcal{T}^2, \dots$ converges to $\mathcal{T}_R^\infty$ in the Hausdorff metric, where $\mathcal{T}_R^\infty = \mathrm{cl}\left(\{\Pi_T : T \in \mathcal{T}^m, m \geq 1\}\right)$ (i.e., the closure of $\{\Pi_T : T \in \mathcal{T}^m, m \geq 1\}$).

Remark: This theorem is a special case of [13, Theorem 2].

The following corollary is a direct consequence of Theorem 2.2 and Theorem 2.4.

Corollary 2.5: If all finite products of matrices from $\mathcal{T}$ are regular, then

$$\mathcal{A}_s = \mathcal{A}'_s = \{\text{the } s\text{th row of } T : T \in \mathcal{T}_R^\infty\}$$

and $\mathcal{A}_s$ does not depend on s.

Theorem 2.6: If all finite products of matrices from $\mathcal{T}$ are regular, then:
1) $\mathcal{A}_s$ is the unique nonempty compact subset of $\mathcal{P}(\mathcal{S})$ satisfying $\mathcal{A}_s = \mathcal{A}_s\mathcal{T}$;
2) $\mathcal{B}\mathcal{T}^n$ converges to $\mathcal{A}_s$ in the Hausdorff metric with a geometric rate independent of $\mathcal{B}$ for any nonempty compact set $\mathcal{B} \subseteq \mathcal{P}(\mathcal{S})$.

Remark: This result is a special case of [19, Theorem 1]. Part 1) of Theorem 2.6 is particularly useful for obtaining an explicit characterization of $\mathcal{A}_s$ if there exists a natural candidate for $\mathcal{A}_s$ since one just needs to verify whether it is invariant under transformation $\mathcal{T}$. However, in general $\mathcal{A}_s$ does not possess a simple characterization; in this case one may use Part 2) of Theorem 2.6 to compute $\mathcal{A}_s$ numerically.

The following result, which is a direct consequence of Theorem 2.6, provides a way to find inner and outer bounds on $\mathcal{A}_s$.

Corollary 2.7: If all finite products of matrices from $\mathcal{T}$ are regular, then:
1) $\mathcal{A}_s \subseteq \mathcal{B}\mathcal{T}$ if $\mathcal{A}_s \subseteq \mathcal{B}$;
2) $\mathcal{A}_s \subseteq \mathcal{B}$ for any nonempty compact set $\mathcal{B} \subseteq \mathcal{P}(\mathcal{S})$ satisfying $\mathcal{B}\mathcal{T} \subseteq \mathcal{B}$;
3) $\mathcal{B} \subseteq \mathcal{A}_s$ for any compact set $\mathcal{B} \subseteq \mathcal{P}(\mathcal{S})$ satisfying $\mathcal{B} \subseteq \mathcal{B}\mathcal{T}$.

Proof: See Appendix A. ■

Remark: Let $\mathcal{B}_1 = \{\pi_T : T \in \mathcal{T}\}$ and $\mathcal{B}_2 = \mathcal{P}(\mathcal{S})$. If all finite products of matrices from $\mathcal{T}$ are regular, then by Corollary 2.7 we have $\mathcal{B}_1\mathcal{T}^n \subseteq \mathcal{A}_s \subseteq \mathcal{B}_2\mathcal{T}^n$ for any non-negative integer n, where $\mathcal{B}_1\mathcal{T}^0 \triangleq \mathcal{B}_1$, $\mathcal{B}_2\mathcal{T}^0 \triangleq \mathcal{B}_2$. Moreover, it follows from Part 2) of Theorem 2.6 that $\mathcal{B}_1\mathcal{T}^n$ and $\mathcal{B}_2\mathcal{T}^n$ provide asymptotically tight inner and outer bounds on $\mathcal{A}_s$ as n goes to infinity.

We have a complete characterization of $\mathcal{A}_s$ for the following special case.

Corollary 2.8: Let π be a probability distribution in $\mathcal{P}(\mathcal{S})$. We have $\mathcal{A}_s = \{\pi\}$ for all $s \in \mathcal{S}$ if and only if all finite products of matrices from $\mathcal{T}$ are regular, and $\pi_T = \pi$ for all $T \in \mathcal{T}$.

Proof: In view of the fact that $\{\pi\}\mathcal{T} = \{\pi\}$, the "if" part follows directly from Theorem 2.6.

Now we proceed to prove the "only if" part. Let $T = T_1T_2\dots T_m$ be an arbitrary finite product of matrices from $\mathcal{T}$. Since $\mathcal{A}_s = \{\pi\}$ for all $s \in \mathcal{S}$, it follows that $\lim_{n\to\infty} T^n$ exists and all its rows are equal to π. The proof is complete. ■

Intuitively, if the stationary distributions π_T $(T \in \mathcal{T})$ are close to each other, then $\mathcal{A}_s$ should be small. This intuition is formalized in the following corollary.

Corollary 2.9: Assume $\max_{T\in\mathcal{T}} \lambda(T) \triangleq \lambda < 1$. Let $\{r(T)\}_{T\in\mathcal{T}}$ be a set of non-negative numbers satisfying $r(T') \geq \|\pi_T - \pi_{T'}\| + r(T)\lambda(T)$ for all $T, T' \in \mathcal{T}$. Then we have $\mathcal{A}_s \subseteq \mathcal{B}$, where $\mathcal{B} = \{B \in \mathcal{P}(\mathcal{S}) : \|B - \pi_T\| \leq r(T) \text{ for all } T \in \mathcal{T}\}$.

Proof: For any $B \in \mathcal{B}$ and $T, T' \in \mathcal{T}$

$$\begin{aligned}\|BT - \pi_{T'}\| &\leq \|\pi_T - \pi_{T'}\| + \|BT - \pi_T\| \\ &= \|\pi_T - \pi_{T'}\| + \|BT - \pi_T T\| \\ &\leq \|\pi_T - \pi_{T'}\| + \|B - \pi_T\|\lambda(T) \\ &\leq \|\pi_T - \pi_{T'}\| + r(T)\lambda(T) \\ &\leq r(T')\end{aligned}$$

where the second inequality follows from (11). Therefore, we have $BT \in \mathcal{B}$, which further implies $\mathcal{B}\mathcal{T} \subseteq \mathcal{B}$. Now the desired result follows from Corollary 2.7. ■

Remark: Specifically, we can choose $r(T) = r \triangleq \frac{\Delta}{1-\lambda}$ for all $T \in \mathcal{T}$, where $\Delta = \max_{T,T' \in \mathcal{T}} \|\pi_T - \pi_{T'}\|$.

It will be clear that for the purpose of this paper, it suffices to characterize $\mathrm{conv}(\mathcal{A}_s)$ (i.e., the convex hull of $\mathcal{A}_s$). This problem turns out to be simpler.

Theorem 2.10: If all finite products of matrices from $\mathcal{T}$ are regular, then $\mathrm{conv}(\mathcal{A}_s)$ is the unique nonempty compact convex set $\mathcal{B} \subseteq \mathcal{P}(\mathcal{S})$ satisfying $\mathcal{B} = \mathrm{conv}(\mathcal{B}\mathcal{T})$.

Proof: This result can be proved by leveraging Theorem 2.6 and some basic properties of convex sets. The details can be found in Appendix B. ■

Remark: Note that for $|\mathcal{S}| = 2$, any nonempty compact convex set must be a line segment. In this case, one can characterize $\mathcal{A}_s$ explicitly by solving a set of necessary and sufficient algebraic conditions implied by Theorem 2.10. A concrete example is given in Section V (see Example 5.1).

III. Capacity Bounds

Now we proceed to derive new finite-letter bounds on the capacity of finite-state channels. For the ease of reference we reproduce (9) and (10) below

$$C_k^U = \max_{P_{X_1^k}} \max_s \max_{P_{S_0|X_1^k} \in \mathcal{Q}\left(P_{X_1^k}, \mathcal{A}_s\right)} \frac{1}{k} I\left(X_1^k; Y_1^k | S_0\right) + \frac{1}{k} H(S_0)$$

$$C_k^L = \max_{P_{X_1^k}} \min_s \min_{P_{S_0|X_1^k} \in \mathcal{Q}'\left(P_{X_1^k}, \mathcal{A}_s\right)} \frac{1}{k} I\left(X_1^k; Y_1^k | S_0\right) - \frac{1}{k} H(S_0),$$

where $\mathcal{Q}\left(P_{X_1^k}, \mathcal{A}_s\right)$ and $\mathcal{Q}'\left(P_{X_1^k}, \mathcal{A}_s\right)$ are defined in (7) and (8), respectively. It is easy to verify that $C_k^U \le \overline{C}_k$ and $C_k^L \ge \underline{C}_k$ for all $k \ge 1$.

Theorem 3.1: $C_k^U \ge \overline{C}$ for all $k \ge 1$.

Proof: See Appendix C. ■

Corollary 3.2: $\overline{C} = \inf_k C_k^U = \lim_{k\to\infty} C_k^U$.

Proof: This result follows directly from the fact that $\overline{C} = \inf_k \overline{C}_k = \lim_{k\to\infty} \overline{C}_k$ and $\overline{C}_k \ge C_k^U \ge \overline{C}$ for all $k \ge 1$. ■

Lemma 3.3: For any positive integers k_1, k_2, and $k = k_1 + k_2$, if $P_{X_1^k S_0} \in \mathrm{conv}\left(\mathcal{P}(\mathcal{X}^k) \times \mathcal{A}_s\right)$ for some $s \in \mathcal{S}$, then $P_{X_1^{k_1} S_0} \in \mathrm{conv}\left(\mathcal{P}(\mathcal{X}^{k_1}) \times \mathcal{A}_s\right)$ and $P_{X_{k_1+1}^k S_{k_1}} \in \mathrm{conv}\left(\mathcal{P}(\mathcal{X}^{k_2}) \times \mathcal{A}_s\right)$, where S_{k_1} is the channel state at time k_1 induced by the initial state S_0 and channel input $X_1^{k_1}$.

Proof: See Appendix D. ■

Theorem 3.4: $kC_k^U \le k_1 C_{k_1}^U + k_2 C_{k_2}^U$ for any positive integers k_1, k_2, and $k = k_1 + k_2$.

Proof: See Appendix E. ■

Remark: Theorem 3.4 implies that $\lim_{k\to\infty} C_k^U = \inf_k C_k^U$, which is consistent with Corollary 3.2.

Now we proceed to derive lower bounds on the channel capacity.

Theorem 3.5: $C_k^L \le \underline{C}$ for all $k \ge 1$.

Proof: See Appendix F. ■

Corollary 3.6: $\underline{C} = \sup_k C_k^L = \lim_{k\to\infty} C_k^L$.

Proof: This result follows directly from the fact that $\underline{C} = \sup_k \underline{C}_k = \lim_{k\to\infty} \underline{C}_k$ and $\underline{C}_k \le C_k^L \le \underline{C}$ for all $k \ge 1$. ■

Lemma 3.7: For any positive integer k, if $P_{X_1^k S_0} \in \mathcal{P}(\mathcal{X}^k) \times \mathrm{conv}(\mathcal{A}_s)$ for some $s \in \mathcal{S}$, then $P_{S_k} \in \mathrm{conv}(\mathcal{A}_s)$, where S_k is the channel state at time k induced by the initial state S_0 and channel input X_1^k.

Proof: See Appendix G. ■

Theorem 3.8: $kC_k^L \ge k_1 C_{k_1}^L + k_2 C_{k_2}^L$ for any positive integers k_1, k_2, and $k = k_1 + k_2$.

Proof: See Appendix H. ■

It is obvious that we can get other upper and lower capacity bounds by replacing $\mathcal{A}_s$ $(s \in \mathcal{S})$ with larger compact sets. This fact is summarized in the following corollary.

Corollary 3.9: For any collection of nonempty compact sets $\{\mathcal{B}_s\}_{s\in\mathcal{S}}$ satisfying $\mathcal{A}_s \subseteq \mathcal{B}_s \subseteq \mathcal{P}(\mathcal{S})$, $s \in \mathcal{S}$, see the equation shown at the bottom of the next page, where $\mathcal{B} = \bigcup_s \mathcal{B}_s$. We have $C_k^U \le C_k^U(\{\mathcal{B}_s\}_{s\in\mathcal{S}}) \le C_k^U(\mathcal{B}) \le \overline{C}_k$ and $C_k^L \ge C_k^L(\{\mathcal{B}_s\}_{s\in\mathcal{S}}) \ge C_k^L(\mathcal{B}) \ge \underline{C}_k$ for all $k \ge 1$.

The capacity bounds take a particularly simple form in the following case.

Corollary 3.10: If there exists a probability distribution $\pi \in \mathcal{P}(\mathcal{S})$ such that $\mathcal{A}_s = \{\pi\}$ for all $s \in \mathcal{S}$, then

$$C_k^U = \max_{P_{X_1^k}} \frac{1}{k} \sum_{s_0} \pi(s_0) I\left(X_1^k; Y_1^k | S_0 = s_0\right) + \frac{1}{k} H(\pi)$$

$$C_k^L = \max_{P_{X_1^k}} \frac{1}{k} \sum_{s_0} \pi(s_0) I\left(X_1^k; Y_1^k | S_0 = s_0\right) - \frac{1}{k} H(\pi)$$

for all $k \ge 1$, where $H(\pi) \triangleq -\sum_s \pi(s) \log \pi(s)$. Specifically, this condition is satisfied if and only if all finite products of matrices from $\mathcal{T}$ are regular, and $\pi_T = \pi$ for all $T \in \mathcal{T}$.

Remark: In this case $C_k^U - C_k^L = \frac{2}{k} H(\pi) \le \frac{2}{k} \log |\mathcal{S}|$ for all $k \ge 1$. Moreover, we have

$$C = \lim_{k\to\infty} \max_{P_{X_1^k}} \frac{1}{k} \sum_{s_0} \pi(s_0) I\left(X_1^k; Y_1^k | S_0 = s_0\right).$$

To demonstrate the usefulness of the new capacity bounds, two illustrative examples are given in Section V (see Examples 5.2 and 5.3).

Now we proceed to study the case where the state process $\{S_t\}_{t=0}^{\infty}$ is known at the receiver. Although C_k^U and C_k^L are directly applicable here with Y_1^k replaced by $\left(Y_1^k, S_1^k\right)$, it turns out that one can derive better upper and lower bounds for this scenario. Let C^S denote the channel capacity in this setting. Define

$$\overline{C}^S = \lim_{k\to\infty} \frac{1}{k} \max_{P_{X_1^k}} \max_{s_0} I\left(X_1^k; Y_1^k, S_1^k | S_0 = s_0\right)$$

$$\underline{C}^S = \lim_{k\to\infty} \frac{1}{k} \max_{P_{X_1^k}} \min_{s_0} I\left(X_1^k; Y_1^k, S_1^k | S_0 = s_0\right)$$

$$\overline{C}_k^S = \max_{P_{X_1^k}} \max_{s_0} \frac{1}{k} I\left(X_1^k; Y_1^k, S_1^k | S_0 = s_0\right)$$

$$\underline{C}_k^S = \max_{P_{X_1^k}} \min_{s_0} \frac{1}{k} I\left(X_1^k; Y_1^k, S_1^k | S_0 = s_0\right)$$

$$C_k^{S,U} = \max_{P_{X_1^k}} \max_{s_0} \max_{P_{S_0|X_1^k} \in \mathcal{Q}\left(P_{X_1^k}, \mathcal{A}_{s_0}\right)} \frac{1}{k} I\left(X_1^k; Y_1^k, S_1^k | S_0\right)$$

$$C_k^{S,L} = \max_{P_{X_1^k}} \min_{s_0} \min_{P_{S_0|X_1^k} \in \mathcal{Q}'\left(P_{X_1^k}, \mathcal{A}_{s_0}\right)} \frac{1}{k} I\left(X_1^k; Y_1^k, S_1^k | S_0\right).$$

In view of (1), it is clear that $\underline{C}^S \le C^S \le \overline{C}^S$, where the inequalities become equalities for indecomposable channels.

Theorem 3.11:
1) $\overline{C}_k^S \ge C_k^{S,U} \ge \overline{C}^S$ for all $k \ge 1$;
2) $\underline{C}_k^S \le C_k^{S,L} \le \underline{C}^S$ for all $k \ge 1$.

Proof: See Appendix I. ■

Remark: For the case where the state process is not available at the receiver, we have an additional term $\frac{1}{k}H(S_0)$ in the capacity bounds. This term can be interpreted as the information regarding $\{S_{mk}\}_{m=1}^{\infty}$ that the genie provides to the receiver. For the case where the state process is known at the receiver, the role of genie becomes superfluous, and consequently the term $\frac{1}{k}H(S_0)$ can be dropped.

The results collected in the following theorem are easy to verify. The proof is omitted.

Theorem 3.12:
1) $kC_k^{S,U} \le k_1 C_{k_1}^{S,U} + k_2 C_{k_2}^{S,U}$, $kC_k^{S,L} \ge k_1 C_{k_1}^{S,L} + k_2 C_{k_2}^{S,L}$, $k\overline{C}_k^S \le k_1 \overline{C}_{k_1}^S + k_2 \overline{C}_{k_2}^S$, $k\underline{C}_k^S \ge k_1 \underline{C}_{k_1}^S + k_2 \underline{C}_{k_2}^S$ for any positive integers k_1, k_2, and $k = k_1 + k_2$;
2) $\overline{C}^S = \inf_k C_k^{S,U} = \inf_k \overline{C}_k^S = \lim_{k\to\infty} C_k^{S,U} = \lim_{k\to\infty} \overline{C}_k^S$;
3) $\underline{C}^S = \sup_k C_k^{S,L} = \sup_k \underline{C}_k^S = \lim_{k\to\infty} C_k^{S,L} = \lim_{k\to\infty} \underline{C}_k^S$.

The new bounds yield a single-letter capacity formula for the following case.

Theorem 3.13: If there exists a probability distribution $\pi \in \mathcal{P}(\mathcal{S})$ such that $\mathcal{A}_s = \{\pi\}$ for all $s \in \mathcal{S}$, then

$$C_k^{S,U} = C_k^{S,L} = C^S$$

for all $k \ge 1$.

Corollary 3.14: If all finite products of matrices from $\mathcal{T}$ are regular, and $\pi_T = \pi$ for all $T \in \mathcal{T}$, then

$$C^S = \max_{P_{X_1}} \sum_{s_0} \pi(s_0) I(X_1; Y_1, S_1 | S_0 = s_0). \tag{12}$$

Proof: It is a direct consequence of Theorem 3.13 and Corollary 2.8. ■

Remark: An illustrative example is given in Section V (see Example 5.4).

Note that the condition in Theorem 3.13 (as well as Corollary 3.14) is fulfilled if the state process is a regular homogeneous Markov chain independent of the channel input. In this case, we have

$$C^S = \max_{P_{X_1}} \sum_{s_0} \pi(s_0) I(X_1; Y_1 | S_1, S_0 = s_0). \tag{13}$$

since $I(X_1; S_1 | S_0 = s_0) = 0$ for all $s_0 \in \mathcal{S}$. However, the reverse is not true, i.e., $\mathcal{A}_s = \{\pi\}$ for all $s \in \mathcal{S}$ does not imply that the state process is independent of the channel input. Indeed, it is easy to construct finite-state channels for which the channel input can affect the transition probability matrix of the state process but not its limiting marginal distribution. For such kind of channels, one cannot reduce (12) to (13) in general since the state process can carry some information from the channel input.

$$C_k^U(\{\mathcal{B}_s\}_{s\in\mathcal{S}}) = \max_{P_{X_1^k}} \max_{s} \max_{P_{S_0|X_1^k} \in \mathcal{Q}\left(P_{X_1^k}, \mathcal{B}_s\right)} \frac{1}{k} I\left(X_1^k; Y_1^k | S_0\right) + \frac{1}{k} H(S_0)$$

$$C_k^L(\{\mathcal{B}_s\}_{s\in\mathcal{S}}) = \max_{P_{X_1^k}} \min_{s} \min_{P_{S_0|X_1^k} \in \mathcal{Q}'\left(P_{X_1^k}, \mathcal{B}_s\right)} \frac{1}{k} I\left(X_1^k; Y_1^k | S_0\right) - \frac{1}{k} H(S_0)$$

$$C_k^U(\mathcal{B}) = \max_{P_{X_1^k}} \max_{P_{S_0|X_1^k} \in \mathcal{Q}\left(P_{X_1^k}, \mathcal{B}\right)} \frac{1}{k} I\left(X_1^k; Y_1^k | S_0\right) + \frac{1}{k} H(S_0)$$

$$C_k^L(\mathcal{B}) = \max_{P_{X_1^k}} \min_{P_{S_0|X_1^k} \in \mathcal{Q}'\left(P_{X_1^k}, \mathcal{B}\right)} \frac{1}{k} I\left(X_1^k; Y_1^k | S_0\right) - \frac{1}{k} H(S_0)$$

IV. Finite-State Multiple Access Channels

Results analogous to those in the previous section can be established for finite-state multiple access channels. Although the derivations are conceptually similar, a few new technical issues arise in the context of finite-state multiple access channels. Furthermore, it is instructive to re-examine the concepts developed for finite-state channels in a more general setting. A close comparison with the results in the previous section will be made, and the subtle differences will be pointed out when they appear.

To simplify the notations, we shall only consider finite-state multiple access channels with two transmitters and one receiver. All the results can be extended in a straightforward manner to the case with an arbitrary number of transmitters. Let $P_{Y_tS_t|X_{1,t}X_{2,t}S_{t-1}}$ be a finite-state multiple access channel, where $X_{i,t} \in \mathcal{X}_i$ $(i = 1,2)$, $Y_t \in \mathcal{Y}$, and $S_t \in \mathcal{S}$ are, respectively, the channel input from transmitter i $(i = 1,2)$, the channel output, and the channel state at time t. We assume that the channel is time-invariant, i.e., that the transition probability $P_{Y_tS_t|X_{1,t}X_{2,t}S_{t-1}}$ does not depend on t. Moreover, $|\mathcal{X}_1|$, $|\mathcal{X}_2|$, $|\mathcal{Y}|$, $|\mathcal{S}|$ are assumed to be finite; in particular, we let $\mathcal{S} = \{1, 2, \dots, |\mathcal{S}|\}$. We shall reuse the notations $\mathcal{T}$, $\mathcal{G}_{s,k}$, and $\mathcal{A}_s$. It should be noted that $\mathcal{T} = \{P_{S_t|X_{1,t}X_{2,t}S_{t-1}}(\cdot|x_1, x_2, \cdot)\}_{x_1\in\mathcal{X}_1, x_2\in\mathcal{X}_2}$ in the current setting. The capacity region of finite-state multiple access channel $P_{Y_tS_t|X_{1,t}X_{2,t}S_{t-1}}$ is denoted as $\mathcal{R}$.

For any nonempty compact set $\mathcal{B} \subseteq \mathcal{P}(\mathcal{S})$, see the equation shown at the bottom of the page. Moreover, given any $P_{X_{1,1}^kX_{2,1}^k} \in \mathcal{P}(\mathcal{X}_1^k) \times \mathcal{P}(\mathcal{X}_2^k)$, define

$$\mathcal{V}\left(P_{X_{1,1}^kX_{2,1}^k}, \mathcal{B}\right) = \left\{P_{S_0|X_{1,1}^kX_{2,1}^k} : P_{X_{1,1}^kX_{2,1}^kS_0} \in \Xi_k(\mathcal{B})\right\}$$
$$\mathcal{V}'\left(P_{X_{1,1}^kX_{2,1}^k}, \mathcal{B}\right) = \left\{P_{S_0|X_{1,1}^kX_{2,1}^k} : P_{X_{1,1}^kX_{2,1}^kS_0} \in \Xi'_k(\mathcal{B})\right\}.$$

Lemma 4.1:

1) For any $P_{X_{1,1}^kX_{2,1}^kS_0} \in \Xi_k(\mathcal{B})$, we have $P_{X_{1,1}^kX_{2,1}^k} \in \mathcal{P}(\mathcal{X}_1^k) \times \mathcal{P}(\mathcal{X}_2^k)$.
2) $\Xi_k(\mathcal{P}(\mathcal{S})) = \Big\{P_{X_{1,1}^kX_{2,1}^kS_0} \in \mathcal{P}(\mathcal{X}_1^k \times \mathcal{X}_2^k \times \mathcal{S}) : P_{X_{1,1}^kX_{2,1}^k} \in \mathcal{P}(\mathcal{X}_1^k) \times \mathcal{P}(\mathcal{X}_2^k)\Big\}$.
3) There is no loss of generality to assume $m_1 = m_2$ and $\max(m_1, m_2) \le |\mathcal{X}_1|^k|\mathcal{X}_2|^k|\mathcal{S}|$ in the definition of $\Xi_k(\mathcal{B})$.
4) $\Xi_k(\mathcal{B}) = \Xi_k(\mathrm{conv}(\mathcal{B}))$.
5) For any positive integers k_1, k_2, and $k = k_1 + k_2$, if $P_{X_{1,1}^kX_{1,2}^kS_0} \in \Xi_k(\mathcal{A}_s)$, then $P_{X_{1,1}^{k_1}X_{2,1}^{k_1}S_0} \in \Xi_{k_1}(\mathcal{A}_s)$ and $P_{X_{1,k_1+1}^kX_{2,k_1+1}^kS_{k_1}} \in \Xi_{k_2}(\mathcal{A}_s)$, where S_{k_1} is the channel state at time k_1 induced by the initial state S_0 and channel inputs $X_{1,1}^k$ and $X_{2,1}^k$.
6) For any positive integer k, if $P_{X_{1,1}^kX_{1,2}^kS_0} \in \Xi'_k(\mathcal{A}_s)$, then $P_{S_k} \in \mathrm{conv}(\mathcal{A}_s)$, where S_k is the channel state at time k induced by the initial state S_0 and channel inputs $X_{1,1}^k$ and $X_{2,1}^k$.

Proof: See Appendix J. ■

Remark: It is easy to show that $\Xi_k(\mathcal{B})$ is compact (for any nonempty compact set $\mathcal{B} \subseteq \mathcal{P}(\mathcal{S})$) by leveraging Part 3) of Lemma 4.1. It is worth noting that for $P_{X_{1,1}^kX_{2,1}^kS_0} \in \Xi_k(\mathcal{B})$, although $X_{1,1}^k$ and $X_{2,1}^k$ are independent, $X_{1,1}^k$ and $X_{2,1}^k$ given S_0 are not independent in general. A detailed discussion of this point will be given later in this section.

See the first equation at the bottom of the next page, where $P_{X_{1,1}^kX_{2,1}^k}$ in the definition of $\mathcal{R}_k\left(P_{X_{1,1}^kX_{2,1}^k}, s_0\right)$ should be interpreted as $P_{X_{1,1}^kX_{2,1}^k|S_0=s_0}$. See the second equation at the bottom of the next page. It is easy to verify that for any $P_{X_{1,1}^kX_{2,1}^k} \in \mathcal{P}(\mathcal{X}_1^k) \times \mathcal{P}(\mathcal{X}_2^k)$

$$\bigcap_{P_{S_0|X_{1,1}^kX_{2,1}^k} \in \mathcal{V}'\left(P_{X_{1,1}^kX_{2,1}^k}, \mathcal{P}(\mathcal{S})\right)} \mathcal{R}_k\left(P_{X_{1,1}^kX_{2,1}^kS_0}\right) = \bigcap_{s_0} \mathcal{R}_k\left(P_{X_{1,1}^kX_{2,1}^k}, s_0\right).$$

Therefore, we can write $\underline{\mathcal{R}}$ alternatively as the third equation at the bottom of the next page. It follows from [20, Lemmas 26 and 27] that the limits in the definition of $\overline{\mathcal{R}}$ and $\underline{\mathcal{R}}$ exist; furthermore [20, Theorems 9 and 11] imply that

$$\underline{\mathcal{R}} \subseteq \mathcal{R} \subseteq \overline{\mathcal{R}}. \tag{14}$$

$$\Xi_k(\mathcal{B}) = \Bigg\{P_{X_{1,1}^kX_{2,1}^kS_0} : \text{there exist } m_1, m_2 \in \mathbb{N}, P_{X_{1,1}^k}^{(i)} \in \mathcal{P}(\mathcal{X}_1^k), P_{X_{2,1}^k}^{(j)} \in \mathcal{P}(\mathcal{X}_2^k), P_{S_0}^{(i,j)} \in \mathcal{B}$$
$$\mu_{1,i}, \mu_{2,j} \in [0,1], i = 1, \dots, m_1, j = 1, \dots, m_2, \text{ such that } \sum_{i=1}^{m_1} \mu_{1,i} = \sum_{j=1}^{m_2} \mu_{2,j} = 1$$
$$\text{and } P_{X_{1,1}^kX_{2,1}^kS_0}\left(x_{1,1}^k, x_{2,1}^k, s_0\right) = \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} \mu_{1,i}\mu_{2,j} P_{X_{1,1}^k}^{(i)}\left(x_{1,1}^k\right) P_{X_{2,1}^k}^{(j)}\left(x_{2,1}^k\right) P_{S_0}^{(i,j)}(s_0)$$
$$\text{for } x_{1,1}^k \in \mathcal{X}_1^k, x_{2,1}^k \in \mathcal{X}_2^k, s_0 \in \mathcal{S}\Bigg\}$$
$$\Xi'_k(\mathcal{B}) = \mathcal{P}(\mathcal{X}_1^k) \times \mathcal{P}(\mathcal{X}_2^k) \times \mathrm{conv}(\mathcal{B})$$

Note that (14) is valid for any initial state s_0; it is also valid whether or not the transmitters and the receiver know the initial state. Moreover, for indecomposable finite-state multiple access channels, we have $\underline{\mathcal{R}} = \mathcal{R} = \overline{\mathcal{R}}$ [20, Theorem 12]. See the fourth equation at the bottom of the page, where $[t]^+ \triangleq \max(t, 0)$. See the first equation shown at the bottom of the next page. It is known [20] that

$$\overline{\mathcal{R}} = \lim_{k\to\infty} \overline{\mathcal{R}}_k = \bigcap_k \overline{\mathcal{R}}_k$$

$$\underline{\mathcal{R}} = \lim_{k\to\infty} \underline{\mathcal{R}}_k = \mathrm{cl}\left(\bigcup_k \underline{\mathcal{R}}_k\right).$$

Therefore, we have

$$\underline{\mathcal{R}}_k \subseteq \underline{\mathcal{R}} \subseteq \mathcal{R} \subseteq \overline{\mathcal{R}} \subseteq \overline{\mathcal{R}}_k$$

for all $k \geq 1$.

Define[2]

$$\widetilde{\mathcal{R}} = \limsup_{k\to\infty} \widetilde{\mathcal{R}}_k \tag{15}$$

[2]Under certain definitions of capacity region (see, e.g., [20]), one can replace lim sup in (15) by lim inf and the resulting $\widetilde{\mathcal{R}}$ is still an outer bound on $\mathcal{R}$. We choose this more conservative definition of $\widetilde{\mathcal{R}}$ so that $\mathcal{R} \subseteq \widetilde{\mathcal{R}}$ holds regardless how $\mathcal{R}$ is defined. Moreover, it can be shown that $\widetilde{\mathcal{R}}$ is a compact set under the current definition.

$$\mathcal{R}_k\left(P_{X_{1,1}^k X_{2,1}^k S_0}\right) = \left\{(R_1,R_2)\in\mathbb{R}_+^2 : \begin{array}{l} R_1 \leq \frac{1}{k} I\left(X_{1,1}^k; Y_1^k | X_{2,1}^k, S_0\right) \\ R_2 \leq \frac{1}{k} I\left(X_{2,1}^k; Y_1^k | X_{1,1}^k, S_0\right) \\ R_1 + R_2 \leq \frac{1}{k} I\left(X_{1,1}^k, X_{2,1}^k; Y_1^k | S_0\right) \end{array}\right\}$$

$$\mathcal{R}_k\left(P_{X_{1,1}^k X_{2,1}^k}, s_0\right) = \left\{(R_1,R_2)\in\mathbb{R}_+^2 : \begin{array}{l} R_1 \leq \frac{1}{k} I\left(X_{1,1}^k; Y_1^k | X_{2,1}^k, S_0 = s_0\right) \\ R_2 \leq \frac{1}{k} I\left(X_{2,1}^k; Y_1^k | X_{1,1}^k, S_0 = s_0\right) \\ R_1 + R_2 \leq \frac{1}{k} I\left(X_{1,1}^k, X_{2,1}^k; Y_1^k | S_0 = s_0\right) \end{array}\right\}$$

$$\overline{\mathcal{R}} = \lim_{k\to\infty} \mathrm{conv}\left(\bigcup_{P_{X_{1,1}^k X_{2,1}^k} \in \mathcal{P}(\mathcal{X}_1^k)\times\mathcal{P}(\mathcal{X}_2^k)} \; \bigcup_{P_{S_0|X_{1,1}^k X_{2,1}^k} \in \mathcal{V}\left(P_{X_{1,1}^k X_{2,1}^k}, \mathcal{P}(\mathcal{S})\right)} \mathcal{R}_k\left(P_{X_{1,1}^k X_{2,1}^k S_0}\right)\right)$$

$$\underline{\mathcal{R}} = \lim_{k\to\infty} \mathrm{conv}\left(\bigcup_{P_{X_{1,1}^k X_{2,1}^k} \in \mathcal{P}(\mathcal{X}_1^k)\times\mathcal{P}(\mathcal{X}_2^k)} \; \bigcap_{P_{S_0|X_{1,1}^k X_{2,1}^k} \in \mathcal{V}'\left(P_{X_{1,1}^k X_{2,1}^k}, \mathcal{P}(\mathcal{S})\right)} \mathcal{R}_k\left(P_{X_{1,1}^k X_{2,1}^k S_0}\right)\right)$$

$$\underline{\mathcal{R}} = \lim_{k\to\infty} \mathrm{conv}\left(\bigcup_{P_{X_{1,1}^k X_{2,1}^k} \in \mathcal{P}(\mathcal{X}_1^k)\times\mathcal{P}(\mathcal{X}_2^k)} \; \bigcap_{s_0} \mathcal{R}_k\left(P_{X_{1,1}^k X_{2,1}^k}, s_0\right)\right)$$

$$\overline{\mathcal{R}}_k\left(P_{X_{1,1}^k X_{2,1}^k S_0}\right) = \left\{(R_1,R_2)\in\mathbb{R}_+^2 : \begin{array}{l} R_1 \leq \frac{1}{k} I\left(X_{1,1}^k; Y_1^k | X_{2,1}^k, S_0\right) + \frac{1}{k}\log|\mathcal{S}| \\ R_2 \leq \frac{1}{k} I\left(X_{2,1}^k; Y_1^k | X_{1,1}^k, S_0\right) + \frac{1}{k}\log|\mathcal{S}| \\ R_1 + R_2 \leq \frac{1}{k} I\left(X_{1,1}^k, X_{2,1}^k; Y_1^k | S_0\right) + \frac{1}{k}\log|\mathcal{S}| \end{array}\right\}$$

$$\underline{\mathcal{R}}_k\left(P_{X_{1,1}^k X_{2,1}^k S_0}\right) = \left\{(R_1,R_2)\in\mathbb{R}_+^2 : \begin{array}{l} R_1 \leq \left[\frac{1}{k} I\left(X_{1,1}^k; Y_1^k | X_{2,1}^k, S_0\right) - \frac{1}{k}\log|\mathcal{S}|\right]^+ \\ R_2 \leq \left[\frac{1}{k} I\left(X_{2,1}^k; Y_1^k | X_{1,1}^k, S_0\right) - \frac{1}{k}\log|\mathcal{S}|\right]^+ \\ R_1 + R_2 \leq \left[\frac{1}{k} I\left(X_{1,1}^k, X_{2,1}^k; Y_1^k | S_0\right) - \frac{1}{k}\log|\mathcal{S}|\right]^+ \end{array}\right\}$$

where

$$\widetilde{\mathcal{R}}_k = \bigcup_{P_{X_{1,1}^k X_{2,1}^k} \in \mathcal{P}(\mathcal{X}_1^k)\times\mathcal{P}(\mathcal{X}_2^k)} \bigcup_{s_0} \mathcal{R}_k\left(P_{X_{1,1}^k X_{2,1}^k}, s_0\right).$$

It was shown in [20] that

$$\mathcal{R} \subseteq \widetilde{\mathcal{R}} \subseteq \overline{\mathcal{R}}. \tag{16}$$

Note that $\widetilde{\mathcal{R}}$ is in general not computable. However, we shall show that it can be leveraged to derive computable outer bounds on the capacity region.

See the second equation shown at the bottom of the page. Define the third equation shown at the bottom of the page.

Theorem 4.2:
1) $\overline{\mathcal{R}}_k \supseteq \mathcal{R}_k^O \supseteq \widetilde{\mathcal{R}}$ for all $k \geq 1$;
2) $\underline{\mathcal{R}}_k \subseteq \mathcal{R}_k^I \subseteq \underline{\mathcal{R}}$ for all $k \geq 1$.

Proof: See Appendix K. ∎

Theorem 4.3:
1) $k\mathcal{R}_k^O \subseteq k_1\mathcal{R}_{k_1}^O + k_2\mathcal{R}_{k_2}^O$ for any positive integers k_1, k_2, and $k = k_1 + k_2$;
2) $k\mathcal{R}_k^I \supseteq k_1\mathcal{R}_{k_1}^I + k_2\mathcal{R}_{k_2}^I$ for any positive integers k_1, k_2, and $k = k_1 + k_2$.

$$\overline{\mathcal{R}}_k = \operatorname{conv}\left(\bigcup_{P_{X_{1,1}^k X_{2,1}^k} \in \mathcal{P}(\mathcal{X}_1^k)\times\mathcal{P}(\mathcal{X}_2^k)} \bigcup_{P_{S_0|X_{1,1}^k X_{2,1}^k} \in \mathcal{V}\left(P_{X_{1,1}^k X_{2,1}^k}, \mathcal{P}(\mathcal{S})\right)} \overline{\mathcal{R}}_k\left(P_{X_{1,1}^k X_{2,1}^k S_0}\right)\right)$$

$$\underline{\mathcal{R}}_k = \operatorname{conv}\left(\bigcup_{P_{X_{1,1}^k X_{2,1}^k} \in \mathcal{P}(\mathcal{X}_1^k)\times\mathcal{P}(\mathcal{X}_2^k)} \bigcap_{P_{S_0|X_{1,1}^k X_{2,1}^k} \in \mathcal{V}'\left(P_{X_{1,1}^k X_{2,1}^k}, \mathcal{P}(\mathcal{S})\right)} \underline{\mathcal{R}}_k\left(P_{X_{1,1}^k X_{2,1}^k S_0}\right)\right)$$

$$\mathcal{R}_k^O\left(P_{X_{1,1}^k X_{2,1}^k S_0}\right) = \left\{(R_1,R_2)\in\mathbb{R}_+^2 : \begin{array}{l} R_1 \leq \frac{1}{k}I\left(X_{1,1}^k; Y_1^k | X_{2,1}^k, S_0\right) + \frac{1}{k}H(S_0) \\ R_2 \leq \frac{1}{k}I\left(X_{2,1}^k; Y_1^k | X_{1,1}^k, S_0\right) + \frac{1}{k}H(S_0) \\ R_1 + R_2 \leq \frac{1}{k}I\left(X_{1,1}^k, X_{2,1}^k; Y_1^k | S_0\right) + \frac{1}{k}H(S_0) \end{array}\right\}$$

$$\mathcal{R}_k^I\left(P_{X_{1,1}^k X_{2,1}^k S_0}\right) = \left\{(R_1,R_2)\in\mathbb{R}_+^2 : \begin{array}{l} R_1 \leq \left[\frac{1}{k}I\left(X_{1,1}^k; Y_1^k | X_{2,1}^k, S_0\right) - \frac{1}{k}H(S_0)\right]^+ \\ R_2 \leq \left[\frac{1}{k}I\left(X_{2,1}^k; Y_1^k | X_{1,1}^k, S_0\right) - \frac{1}{k}H(S_0)\right]^+ \\ R_1 + R_2 \leq \left[\frac{1}{k}I\left(X_{1,1}^k, X_{2,1}^k; Y_1^k | S_0\right) - \frac{1}{k}H(S_0)\right]^+ \end{array}\right\}$$

$$\mathcal{R}_k^O = \operatorname{conv}\left(\bigcup_{P_{X_{1,1}^k X_{2,1}^k} \in \mathcal{P}(\mathcal{X}_1^k)\times\mathcal{P}(\mathcal{X}_2^k)} \mathcal{R}_k^O\left(P_{X_{1,1}^k X_{2,1}^k}\right)\right)$$

$$\mathcal{R}_k^I = \operatorname{conv}\left(\bigcup_{P_{X_{1,1}^k X_{2,1}^k} \in \mathcal{P}(\mathcal{X}_1^k)\times\mathcal{P}(\mathcal{X}_2^k)} \mathcal{R}_k^I\left(P_{X_{1,1}^k X_{2,1}^k}\right)\right)$$

where

$$\mathcal{R}_k^O\left(P_{X_{1,1}^k X_{2,1}^k}\right) = \bigcup_{s} \bigcup_{P_{S_0|X_{1,1}^k X_{2,1}^k} \in \mathcal{V}\left(P_{X_{1,1}^k X_{2,1}^k}, \mathcal{A}_s\right)} \mathcal{R}_k^O\left(P_{X_{1,1}^k X_{2,1}^k S_0}\right)$$

$$\mathcal{R}_k^I\left(P_{X_{1,1}^k X_{2,1}^k}\right) = \bigcap_{s} \bigcap_{P_{S_0|X_{1,1}^k X_{2,1}^k} \in \mathcal{V}'\left(P_{X_{1,1}^k X_{2,1}^k}, \mathcal{A}_s\right)} \mathcal{R}_k^I\left(P_{X_{1,1}^k X_{2,1}^k S_0}\right)$$

Proof: See Appendix L. ∎

In view of Theorems 4.2 and 4.3, the following corollary is a direct consequence of [20, Lemmas 5 and 6].

Corollary 4.4:
1) $\overline{\mathcal{R}} \supseteq \lim_{k\to\infty} \mathcal{R}_k^O = \bigcap_k \mathcal{R}_k^O \supseteq \widetilde{\mathcal{R}}$;
2) $\lim_{k\to\infty} \mathcal{R}_k^I = \mathrm{cl}\Big(\bigcup_k \mathcal{R}_k^I\Big) = \underline{\mathcal{R}}$.

A few comparisons with the results in the previous section are now in place. To emphasize their analogous roles, we can form the following pairs: $(\underline{C}, \underline{\mathcal{R}})$, $(\underline{C}_k, \underline{\mathcal{R}}_k)$, and $(C_k^L, \mathcal{R}_k^I)$. In contrast, the situation for the outer bounds is more complicated. It might be tempting to form the following pairs: $(\overline{C}, \overline{\mathcal{R}})$, $(\overline{C}_k, \overline{\mathcal{R}}_k)$, and $(C_k^U, \mathcal{R}_k^O)$. However, such pairings are natural but not exact. A close look reveals that although $\overline{C} = \lim_{k\to\infty} \overline{C}_k = \lim_{k\to\infty} C_k^U$, it is unclear whether the second equality in $\overline{\mathcal{R}} = \lim_{k\to\infty} \overline{\mathcal{R}}_k = \lim_{k\to\infty} \mathcal{R}_k^O$ holds in general. Actually it is also reasonable to relate $\widetilde{\mathcal{R}}$ with $\overline{C}$. Moreover, to complement $\widetilde{\mathcal{R}}$, one would naturally expect a finite-letter outer bound on $\mathcal{R}$ in a form analogous to $\overline{C}_k$ in (2). However, a direct generalization of $\overline{C}_k$ in the form of (2) does not seem to yield a valid outer bound. In contrast, $\overline{C}_k$ in the form of (5) does have a counterpart in the setting of multiple access channels, which is $\overline{\mathcal{R}}_k$. This leads to a puzzling phenomenon: $\overline{C}_k$ in the form of (2) does not permit a direct generalization while its equivalent form in (5) does. The reason is somewhat subtle. In order to obtain an outer bound on $\mathcal{R}$ in a form analogous to $\overline{C}_k$ in (2), one needs the following assumption: the inputs of the two transmitters from time $t+1$ on are independent conditioned on S_t. Although this assumption holds when $t = 0$, it is in general not true. The crucial idea underlying the derivation of $\overline{\mathcal{R}}_k$ and $\mathcal{R}_k^O$ is to go beyond conditional independence. Indeed, it can be verified from the definition of $\overline{\mathcal{R}}_k$ and $\mathcal{R}_k^O$ that the inputs of the two transmitters from time $t+1$ on are not necessarily independent conditioned on S_t although they are mutually independent. In contrast, the requirement of conditional independence is void in the point-to-point case since there is only one transmitter. In this sense, $\overline{C}_k$ in the form of (2) is less fundamental than its equivalent form in (5) since the latter one is extendable to more general scenarios.

Now consider the case where the state process is available at the receiver. Let $\mathcal{R}^S$ denote the capacity region in this setting. See the first equation shown at the bottom of the page, where $P_{X_{1,1}^k X_{2,1}^k}$ in the definition of $\mathcal{R}_k^S\left(P_{X_{1,1}^k X_{2,1}^k}, s_0\right)$ should be interpreted as $P_{X_{1,1}^k X_{2,1}^k | S_0 = s_0}$. See the second equation shown at the bottom of the page. In view of (14) and (16), it is clear that $\underline{\mathcal{R}}^S \subseteq \mathcal{R}^S \subseteq \widetilde{\mathcal{R}}^S \subseteq \overline{\mathcal{R}}^S$; moreover, for indecomposable finite-state multiple access channels, we have $\underline{\mathcal{R}}^S = \mathcal{R}^S = \widetilde{\mathcal{R}}^S = \overline{\mathcal{R}}^S$.

See the equation shown at the bottom of the next page. The following theorem is easy to verify. The proof is omitted.

$$\mathcal{R}_k^S\left(P_{X_{1,1}^k X_{2,1}^k S_0}\right) = \left\{ (R_1, R_2) \in \mathbb{R}_+^2 : \begin{array}{l} R_1 \le \frac{1}{k} I\left(X_{1,1}^k; Y_1^k, S_1^k | X_{2,1}^k, S_0\right) \\ R_2 \le \frac{1}{k} I\left(X_{2,1}^k; Y_1^k, S_1^k | X_{1,1}^k, S_0\right) \\ R_1 + R_2 \le \frac{1}{k} I\left(X_{1,1}^k, X_{2,1}^k; Y_1^k, S_1^k | S_0\right) \end{array} \right\}$$

$$\mathcal{R}_k^S\left(P_{X_{1,1}^k X_{2,1}^k}, s_0\right) = \left\{ (R_1, R_2) \in \mathbb{R}_+^2 : \begin{array}{l} R_1 \le \frac{1}{k} I\left(X_{1,1}^k; Y_1^k, S_1^k | X_{2,1}^k, S_0 = s_0\right) \\ R_2 \le \frac{1}{k} I\left(X_{2,1}^k; Y_1^k, S_1^k | X_{1,1}^k, S_0 = s_0\right) \\ R_1 + R_2 \le \frac{1}{k} I\left(X_{1,1}^k, X_{2,1}^k; Y_1^k, S_1^k | S_0 = s_0\right) \end{array} \right\}$$

$$\overline{\mathcal{R}}^S = \lim_{k\to\infty} \mathrm{conv}\left(\bigcup_{P_{X_{1,1}^k X_{2,1}^k} \in \mathcal{P}(\mathcal{X}_1^k) \times \mathcal{P}(\mathcal{X}_2^k)} \; \bigcup_{P_{S_0 | X_{1,1}^k X_{2,1}^k} \in \mathcal{V}\left(P_{X_{1,1}^k X_{2,1}^k}, \mathcal{P}(\mathcal{S})\right)} \mathcal{R}_k^S\left(P_{X_{1,1}^k X_{2,1}^k S_0}\right) \right)$$

$$\underline{\mathcal{R}}^S = \lim_{k\to\infty} \mathrm{conv}\left(\bigcup_{P_{X_{1,1}^k X_{2,1}^k} \in \mathcal{P}(\mathcal{X}_1^k) \times \mathcal{P}(\mathcal{X}_2^k)} \; \bigcap_{P_{S_0 | X_{1,1}^k X_{2,1}^k} \in \mathcal{V}'\left(P_{X_{1,1}^k X_{2,1}^k}, \mathcal{P}(\mathcal{S})\right)} \mathcal{R}_k^S\left(P_{X_{1,1}^k X_{2,1}^k S_0}\right) \right)$$

$$\widetilde{\mathcal{R}}^S = \limsup_{k\to\infty} \left(\bigcup_{P_{X_{1,1}^k X_{2,1}^k} \in \mathcal{P}(\mathcal{X}_1^k) \times \mathcal{P}(\mathcal{X}_2^k)} \; \bigcup_{s_0} \mathcal{R}_k^S\left(P_{X_{1,1}^k X_{2,1}^k}, s_0\right) \right)$$

Theorem 4.5:
1) $\overline{\mathcal{R}}_k^S \supseteq \mathcal{R}_k^{S,O} \supseteq \overline{\mathcal{R}}^S$, $\underline{\mathcal{R}}_k^S \subseteq \mathcal{R}_k^{S,I} \subseteq \underline{\mathcal{R}}^S$ for all $k \geq 1$;
2) $k\mathcal{R}_k^{S,O} \subseteq k_1\mathcal{R}_{k_1}^{S,O} + k_2\mathcal{R}_{k_2}^{S,O}$, $k\mathcal{R}_k^{S,I} \supseteq k_1\mathcal{R}_{k_1}^{S,I} + k_2\mathcal{R}_{k_2}^{S,I}$, $k\overline{\mathcal{R}}_k^S \subseteq k_1\overline{\mathcal{R}}_{k_1}^S + k_2\overline{\mathcal{R}}_{k_2}^S$, $k\underline{\mathcal{R}}_k^S \supseteq k_1\underline{\mathcal{R}}_{k_1}^S + k_2\underline{\mathcal{R}}_{k_2}^S$ for any positive integers k_1, k_2, and $k = k_1 + k_2$;
3) $\widetilde{\mathcal{R}}^S \subseteq \lim_{k\to\infty} \mathcal{R}_k^{S,O} = \bigcap_k \mathcal{R}_k^{S,O} \subseteq \lim_{k\to\infty} \overline{\mathcal{R}}_k^S = \bigcap_k \overline{\mathcal{R}}_k^S = \overline{\mathcal{R}}^S$;
4) $\lim_{k\to\infty} \mathcal{R}_k^{S,I} = \lim_{k\to\infty} \underline{\mathcal{R}}_k^S = \mathrm{cl}\Big(\bigcup_k \mathcal{R}_k^{S,I}\Big) = \mathrm{cl}\Big(\bigcup_k \underline{\mathcal{R}}_k^S\Big) = \underline{\mathcal{R}}^S$.

The finite-letter inner and outer bounds coincide for the following case, yielding a single-letter characterization of the capacity region.

Theorem 4.6: If there exists a probability distribution $\pi \in \mathcal{P}(\mathcal{S})$ such that $\mathcal{A}_s = \{\pi\}$ for all $s \in \mathcal{S}$, then

$$\mathcal{R}_k^{S,O} = \mathcal{R}_k^{S,I} = \mathcal{R}^S$$

for all $k \geq 1$. Specifically, the condition is satisfied if and only if all finite products of matrices from $\mathcal{T}$ are regular, and $\pi_T = \pi$ for all $T \in \mathcal{T}$.

Remark: An illustrative example is given in Section V (see Example 5.5).

V. Examples

Example 5.1: Let $\mathcal{T} = \{T_1, T_2\}$, where

$$T_1 = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}, \quad T_2 = \begin{pmatrix} 1-c & c \\ d & 1-d \end{pmatrix}.$$

It is easy to verify that all finite products of matrices from $\mathcal{T}$ are regular if and only if $0 \leq \min(a,b,c,d) \leq \max(a,b,c,d) \leq 1$, $0 < a+b < 2$, and $0 < c+d < 2$. We assume these conditions are satisfied.

Assume $\mathrm{conv}(\mathcal{A}_s)$ is a line segment joining $(1-e, e)$ and $(1-f, f)$ with $e \geq f$. Note that $\mathrm{conv}(\mathcal{A}_s)T_1$ is a line segment with two endpoints $(1-\alpha_1, \alpha_1)$ and $(1-\alpha_2, \alpha_2)$, where

$$\alpha_1 = a(1-e) + (1-b)e$$
$$\alpha_2 = a(1-f) + (1-b)f.$$

Similarly, $\mathrm{conv}(\mathcal{A}_{s_0})T_2$ is a line segment with two endpoints $(1-\beta_1, \beta_1)$ and $(1-\beta_2, \beta_2)$, where

$$\beta_1 = c(1-e) + (1-d)e$$
$$\beta_2 = c(1-f) + (1-d)f.$$

We have the following 8 cases.

1) $(1-e, e) = (1-\alpha_1, \alpha_1)$ and $(1-f, f) = (1-\beta_2, \beta_2)$. This implies

$$e = \frac{a}{a+b}, \quad f = \frac{c}{c+d}.$$

Moreover, we must have $e \geq \max(\alpha_2, \beta_1) \geq \min(\alpha_2, \beta_1) \geq f$, i.e.,

$$\frac{a}{a+b} \geq \max\left\{\frac{ad+(1-b)c}{c+d}, \frac{a(1-d)+bc}{a+b}\right\}$$
$$\frac{c}{c+d} \leq \min\left\{\frac{ad+(1-b)c}{c+d}, \frac{a(1-d)+bc}{a+b}\right\}.$$

Note that if we further have $\beta_1 \geq \alpha_2$, then $\mathcal{A}_s = \mathrm{conv}(\mathcal{A}_s)$. For example, the above inequalities are satisfied when $a = 0.38$, $b = 0.1$, $c = 0.1$, $d = 0.2$.

2) $(1-e, e) = (1-\beta_1, \beta_1)$ and $(1-f, f) = (1-\alpha_2, \alpha_2)$. This follows from Case 1) by exchanging a with c and b with d.

$$\overline{\mathcal{R}}_k^S = \mathrm{conv}\left(\bigcup_{P_{X_{1,1}^k X_{2,1}^k} \in \mathcal{P}(\mathcal{X}_1^k)\times\mathcal{P}(\mathcal{X}_2^k)} \ \bigcup_{P_{S_0|X_{1,1}^k X_{2,1}^k} \in \mathcal{V}\left(P_{X_{1,1}^k X_{2,1}^k}, \mathcal{P}(\mathcal{S})\right)} \mathcal{R}_k^S\left(P_{X_{1,1}^k X_{2,1}^k S_0}\right)\right)$$

$$\underline{\mathcal{R}}_k^S = \mathrm{conv}\left(\bigcup_{P_{X_{1,1}^k X_{2,1}^k} \in \mathcal{P}(\mathcal{X}_1^k)\times\mathcal{P}(\mathcal{X}_2^k)} \ \bigcap_{P_{S_0|X_{1,1}^k X_{2,1}^k} \in \mathcal{V}'\left(P_{X_{1,1}^k X_{2,1}^k}, \mathcal{P}(\mathcal{S})\right)} \mathcal{R}_k^S\left(P_{X_{1,1}^k X_{2,1}^k S_0}\right)\right)$$

$$\mathcal{R}_k^{S,O} = \mathrm{conv}\left(\bigcup_{P_{X_{1,1}^k X_{2,1}^k} \in \mathcal{P}(\mathcal{X}_1^k)\times\mathcal{P}(\mathcal{X}_2^k)} \ \bigcup_s \ \bigcup_{P_{S_0|X_{1,1}^k X_{2,1}^k} \in \mathcal{V}\left(P_{X_{1,1}^k X_{2,1}^k}, \mathcal{A}_s\right)} \mathcal{R}_k^S\left(P_{X_{1,1}^k X_{2,1}^k S_0}\right)\right)$$

$$\mathcal{R}_k^{S,I} = \mathrm{conv}\left(\bigcup_{P_{X_{1,1}^k X_{2,1}^k} \in \mathcal{P}(\mathcal{X}_1^k)\times\mathcal{P}(\mathcal{X}_2^k)} \ \bigcap_s \ \bigcap_{P_{S_0|X_{1,1}^k X_{2,1}^k} \in \mathcal{V}'\left(P_{X_{1,1}^k X_{2,1}^k}, \mathcal{A}_s\right)} \mathcal{R}_k^S\left(P_{X_{1,1}^k X_{2,1}^k S_0}\right)\right)$$

3) $(1-e,e)=(1-\alpha_2,\alpha_2)$ and $(1-f,f)=(1-\beta_1,\beta_1)$. This implies

$$e=\frac{a+c-ac-bc}{a+b+c+d-ac-ad-bc-bd}$$
$$f=\frac{a+c-ac-ad}{a+b+c+d-ac-ad-bc-bd}.$$

Moreover, we must have $e \geq \max(\alpha_1,\beta_2) \geq \min(\alpha_1,\beta_2) \geq f$, i.e., see the first equation shown at the bottom of the page. Note that if we further have $\beta_2 \geq \alpha_1$, then $\mathcal{A}_s = \mathrm{conv}(\mathcal{A}_s)$. For example, the above inequalities are satisfied when $a=0.89$, $b=0.89$, $c=0.74$, $d=0.8$.

4) $(1-e,e)=(1-\beta_2,\beta_2)$ and $(1-f,f)=(1-\alpha_1,\alpha_1)$. This follows from Case 3) by exchanging a with c and b with d.

5) $(1-e,e)=(1-\alpha_1,\alpha_1)$ and $(1-f,f)=(1-\beta_1,\beta_1)$. This implies

$$e=\frac{a}{a+b},\quad f=\frac{a(1-d)+bc}{a+b}.$$

Moreover, we must have $e \geq \max(\alpha_2,\beta_2) \geq \min(\alpha_2,\beta_2) \geq f$, i.e., see the second equation shown at the bottom of the page. Note that if we further have $\beta_2 \geq \alpha_2$, then $\mathcal{A}_s = \mathrm{conv}(\mathcal{A}_s)$. For example, the above inequalities are satisfied when $a=0.55$, $b=0.1$, $c=0.89$, $d=0.8$.

6) $(1-e,e)=(1-\beta_1,\beta_1)$ and $(1-f,f)=(1-\alpha_1,\alpha_1)$. This follows from Case 5) by exchanging a with c and b with d.

7) $(1-e,e)=(1-\alpha_2,\alpha_2)$ and $(1-f,f)=(1-\beta_2,\beta_2)$. This implies

$$e=\frac{ad+(1-b)c}{c+d},\quad f=\frac{c}{c+d}.$$

Moreover, we must have $e \geq \max(\alpha_1,\beta_1) \geq \min(\alpha_1,\beta_1) \geq f$, i.e., see the third equation shown at the bottom of the page. Note that if we further have $\beta_1 \geq \alpha_1$, then $\mathcal{A}_s = \mathrm{conv}(\mathcal{A}_s)$. For example, the above inequalities are satisfied when $a=0.89$, $b=0.89$, $c=0.26$, $d=0.4$.

8) $(1-e,e)=(1-\beta_2,\beta_2)$ and $(1-f,f)=(1-\alpha_2,\alpha_2)$. This follows from Case 7) by exchanging a with c and b with d.

It is worth noting that $\mathcal{A}_s$ itself might not be a line segment. For example, assume $a \neq b$ and $c=d=\frac{1}{2}$. It is easy to verify that $\mathcal{A}_s = \left\{(\frac{1}{2},\frac{1}{2})T_1^m, 0\leq m\leq\infty\right\}$, where $\left(\frac{1}{2},\frac{1}{2}\right)T_1^0 \triangleq \left(\frac{1}{2},\frac{1}{2}\right)$ and $\left(\frac{1}{2},\frac{1}{2}\right)T_1^\infty \triangleq \left(\frac{b}{a+b},\frac{a}{a+b}\right)$. Clearly, $\mathcal{A}_s$ is not a line segment.

$$\frac{a+c-ac-bc}{a+b+c+d-ac-ad-bc-bd}$$
$$\geq \max\left\{\frac{a(b+d-ad-bd)+(1-b)(a+c-ac-bc)}{a+b+c+d-ac-ad-bc-bd},\frac{c(b+d-bc-bd)+(1-d)(a+c-ac-ad)}{a+b+c+d-ac-ad-bc-bd}\right\}$$
$$\frac{a+c-ac-ad}{a+b+c+d-ac-ad-bc-bd}$$
$$\leq \min\left\{\frac{a(b+d-ad-bd)+(1-b)(a+c-ac-bc)}{a+b+c+d-ac-ad-bc-bd},\frac{c(b+d-bc-bd)+(1-d)(a+c-ad-ad)}{a+b+c+d-ac-ad-bc-bd}\right\}$$

$$\frac{a}{a+b}\geq \max\left\{\frac{a[ad+b(1-c)]+(1-b)[a(1-d)+bc]}{a+b},\frac{c[ad+b(1-c)]+(1-d)[a(1-d)+bc]}{a+b}\right\}$$
$$\frac{a(1-d)+bc}{a+b}\leq \min\left\{\frac{a[ad+b(1-c)]+(1-c)[a(1-d)+bc]}{a+b},\frac{c[ad+b(1-c)]+(1-d)[a(1-d)+bc]}{a+b}\right\}$$

$$\frac{ad+(1-b)c}{c+d}\geq \max\left\{\frac{a[(1-a)d+bc]+(1-b)[ad+(1-b)c]}{c+d},\frac{c[(1-a)d+bc]+(1-d)[ad+(1-b)c]}{c+d}\right\}$$
$$\frac{c}{c+d}\leq \min\left\{\frac{a[(1-a)d+bc]+(1-b)[ad+(1-b)c]}{c+d},\frac{c[(1-a)d+bc]+(1-d)[ad+(1-b)c]}{c+d}\right\}$$

Example 5.2: Let $\mathcal{X} = \mathcal{Y} = \mathcal{S} = \{1,2\}$. Denote the transition probability matrices $P_{S_t|X_tS_{t-1}}(\cdot|1,\cdot)$ and $P_{S_t|X_tS_{t-1}}(\cdot|2,\cdot)$ by T_1 and T_2, respectively, where

$$T_1 = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}, \quad T_2 = \begin{pmatrix} 1-c & c \\ d & 1-d \end{pmatrix}.$$

We assume $0 \leq \min(a,b,c,d) \leq \max(a,b,c,d) \leq 1$, $0 < a+b < 2$, $0 < c+d < 2$, and

$$\frac{a}{b} = \frac{c}{d} = \gamma.$$

Note that we allow $\gamma = 0$ or $\gamma = \infty$. By Corollary 2.8, we have $\mathcal{A}_s = \{\pi\}$ for all $s \in \mathcal{S}$, where

$$\pi(1) = \frac{1}{\gamma+1}.$$

Now it follows from Corollary 3.10 that

$$C_1^U = \max_{P_{X_1}} \sum_{s_0} \pi(s_0) I(X_1;Y_1|S_0 = s_0) + H(\pi)$$
$$C_1^L = \max_{P_{X_1}} \sum_{s_0} \pi(s_0) I(X_1;Y_1|S_0 = s_0) - H(\pi).$$

Note that the gap between C_1^U and C_1^L is $2H(\pi)$, which converges to 0 as $\gamma \to 0$ or $\gamma \to \infty$.

Consider the following two cases:

1) The channel transition probability is of the form

$$P_{Y_tS_t|X_tS_{t-1}}(y_t,s_t|x_t,s_{t-1}) = P_{Y_t|X_tS_{t-1}}(y_t|x_t,s_{t-1}) P_{S_t|X_tS_{t-1}}(s_t|x_t,s_{t-1})$$

for any $x_t \in \mathcal{X}$, $y_t \in \mathcal{Y}$, and $s_{t-1}, s_t \in \mathcal{S}$; moreover, $P_{Y_t|X_tS_{t-1}}(\cdot|\cdot,1)$ is a $\mathrm{BSC}(q_1)$ (i.e., a binary symmetric channel with crossover probability q_1) while $P_{Y_t|X_tS_{t-1}}(\cdot|\cdot,2)$ is a $\mathrm{BSC}(q_2)$.

For this channel, it is clear that the maximizer is given by P_{X_1} with $P_{X_1}(1) = P_{X_1}(2) = \frac{1}{2}$. Therefore, we have

$$C_1^U = 1 - \frac{1}{\gamma+1} H_b(q_1) - \frac{\gamma}{\gamma+1} H_b(q_2) + H_b\left(\frac{1}{\gamma+1}\right)$$
$$C_1^L = 1 - \frac{1}{\gamma+1} H_b(q_1) - \frac{\gamma}{\gamma+1} H_b(q_2) - H_b\left(\frac{1}{\gamma+1}\right)$$

where $H_b(\cdot)$ is the binary entropy function. In contrast, we have

$$\overline{C}_1 = \max(2 - H_b(q_1), 2 - H_b(q_2)) \geq 1$$
$$\underline{C}_1 = \min(-H_b(q_1), -H_b(q_2)) \leq 0$$

yielding trivial capacity bounds for this channel.

2) The channel transition probability is of the form

$$P_{Y_tS_t|X_tS_{t-1}}(y_t,s_t|x_t,s_{t-1}) = P_{Y_t|X_tS_t}(y_t|x_t,s_t) P_{S_t|X_tS_{t-1}}(s_t|x_t,s_{t-1})$$

for any $x_t \in \mathcal{X}$, $y_t \in \mathcal{Y}$, and $s_{t-1}, s_t \in \mathcal{S}$; moreover, $P_{Y_t|X_tS_t}(\cdot|\cdot,1)$ is a $\mathrm{BSC}(q_1)$ while $P_{Y_t|X_tS_t}(\cdot|\cdot,2)$ is a $\mathrm{BSC}(q_2)$. Note that $Y_t - (X_t, S_{t-1}) - S_t$ form a Markov chain in Case 1) while $Y_t - (X_t, S_t) - S_{t-1}$ form a Markov chain in Case 2).

Let $P_{X_1}(2) = p$. The conditional probability distribution $P_{X_1Y_1S_1|S_0}$ is given by

x,y,s	$P_{X_1Y_1S_1\|S_0}(x,y,s\|1)$	$P_{X_1Y_1S_1\|S_0}(x,y,s\|2)$
$1,1,1$	$(1-p)(1-q_1)(1-a)$	$(1-p)(1-q_1)b$
$1,1,2$	$(1-p)(1-q_2)a$	$(1-p)(1-q_2)(1-b)$
$1,2,1$	$(1-p)q_1(1-a)$	$(1-p)q_1b$
$1,2,2$	$(1-p)q_2a$	$(1-p)q_2(1-b)$
$2,1,1$	$pq_1(1-c)$	pq_1d
$2,1,2$	pq_2c	$pq_2(1-d)$
$2,2,1$	$p(1-q_1)(1-c)$	$p(1-q_1)d$
$2,2,2$	$p(1-q_2)c$	$p(1-q_2)(1-d)$

Therefore, we have the equation shown at the bottom of the page. Now it can be computed that

$$I(X_1;Y_1|S_0=1) = H(X_1) - H(X_1|Y_1,S_0=1) = \phi(p,q_1,q_2,1-a,1-c)$$
$$I(X_1;Y_1|S_0=2) = H(X_1) - H(X_1|Y_1,S_0=2) = \phi(p,q_1,q_2,b,d)$$

$$P_{Y_1|S_0}(1|1) = (1-p)(1-q_1)(1-a) + (1-p)(1-q_2)a + pq_1(1-c) + pq_2c$$
$$P_{Y_1|S_0}(1|2) = (1-p)(1-q_1)b + (1-p)(1-q_2)(1-b) + pq_1d + pq_2(1-d)$$
$$P_{X_1|Y_1S_0}(1|1,1) = \frac{(1-p)(1-q_1)(1-a) + (1-p)(1-q_2)a}{(1-p)(1-q_1)(1-a) + (1-p)(1-q_2)a + pq_1(1-c) + pq_2c}$$
$$P_{X_1|Y_1S_0}(1|1,2) = \frac{(1-p)(1-q_1)b + (1-p)(1-q_2)(1-b)}{(1-p)(1-q_1)b + (1-p)(1-q_2)(1-b) + pq_1d + pq_2(1-d)}$$
$$P_{X_1|Y_1S_0}(1|2,1) = \frac{(1-p)q_1(1-a) + (1-p)q_2a}{(1-p)q_1(1-a) + (1-p)q_2a + p(1-q_1)(1-c) + p(1-q_2)c}$$
$$P_{X_1|Y_1S_0}(1|2,2) = \frac{(1-p)q_1b + (1-p)q_2(1-b)}{(1-p)q_1b + (1-p)q_2(1-b) + p(1-q_1)d + p(1-q_2)(1-d)}$$

where [see (17), shown at the bottom of the page]. Therefore, we have

$$C_1^U = \max_{p\in[0,1]} \frac{1}{\gamma+1}\phi(p,q_1,q_2,1-a,1-c) + \frac{\gamma}{1+\gamma}\phi(p,q_1,q_2,b,d) + H_b\left(\frac{1}{\gamma+1}\right)$$
$$C_1^L = \max_{p\in[0,1]} \frac{1}{\gamma+1}\phi(p,q_1,q_2,1-a,1-c) + \frac{\gamma}{1+\gamma}\phi(p,q_1,q_2,b,d) - H_b\left(\frac{1}{\gamma+1}\right).$$

If $q_1 = q_2 = q$, then $Y_t - X_t - (S_t, S_{t-1})$ form a Markov chain, which is subsumed by Case 1). So the maximizer p is equal to $\frac{1}{2}$, and consequently

$$C_1^U = 1 - H_b(q) + H_b\left(\frac{1}{\gamma+1}\right)$$
$$C_1^L = 1 - H_b(q) - H_b\left(\frac{1}{\gamma+1}\right).$$

Example 5.3: Let $\mathcal{X} = \mathcal{Y} = \mathcal{S} = \{1,2\}$. Denote the transition probability matrices $P_{S_t|X_tS_{t-1}}(\cdot|1,\cdot)$ and $P_{S_t|X_tS_{t-1}}(\cdot|2,\cdot)$ by T_1 and T_2, respectively, where

$$T_1 = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}, \quad T_2 = \begin{pmatrix} 1-c & c \\ d & 1-d \end{pmatrix}.$$

We assume $0 \le \min(a,b,c,d) \le \max(a,b,c,d) \le 1$, $0 < a+b < 2$, and $0 < c+d < 2$. The condition in Theorem 2.10 holds under this assumption, which implies that $\text{conv}(\mathcal{A}_s)$ is a line segment joining $(1-e,e)$ and $(1-f,f)$ with $e \ge f$, where the expressions of e and f can be found in Example 5.1. Denote $P_{X_1S_0}(1,1), P_{X_1S_0}(1,2), P_{X_1S_0}(2,1)$, and $P_{X_1S_0}(2,2)$ by θ_{11}, θ_{12}, θ_{21}, and θ_{22}, respectively. It can be verified that $P_{X_1S_0} \in \text{conv}(\mathcal{P}(\mathcal{X})\times\mathcal{A}_s)$ if and only if $(\theta_{11},\theta_{12},\theta_{21},\theta_{22}) \in \Theta$, where $\Theta = \{(\theta'_{11},\theta'_{12},\theta'_{21},\theta'_{22}) : e\theta'_{11} \ge (1-e)\theta'_{12}, f\theta'_{11} \le (1-f)\theta'_{12}, (1-e)(1-\theta'_{11}-\theta'_{12}) \le \theta'_{21} \le (1-f)(1-\theta'_{11}-\theta'_{12}), \theta'_{11}+\theta'_{12}+\theta'_{21}+\theta'_{22} = 1\}$ (see Fig. 1).

Consider the following two cases:

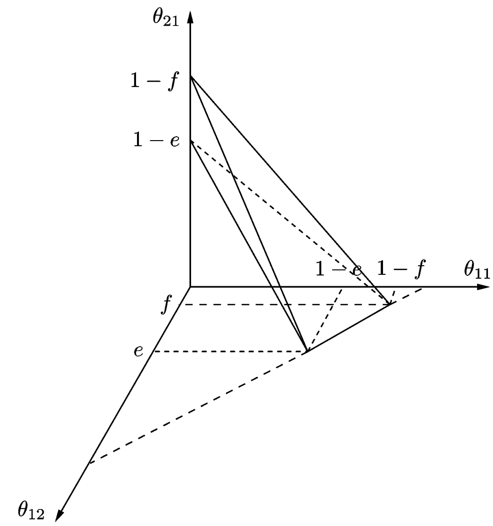

Fig. 1. Plot of $\text{conv}(\mathcal{P}(\mathcal{X})\times\mathcal{A}_{s_0})$. Note that $\theta_{22} = 1-\theta_{11}-\theta_{12}-\theta_{21}$.

1) The channel transition probability is of the form

$$P_{Y_tS_t|X_tS_{t-1}}(y_t,s_t|x_t,s_{t-1}) = P_{Y_t|X_tS_{t-1}}(y_t|x_t,s_{t-1})P_{S_t|X_tS_{t-1}}(s_t|x_t,s_{t-1})$$

for any $x_t \in \mathcal{X}$, $y_t \in \mathcal{Y}$, and $s_{t-1}, s_t \in \mathcal{S}$; moreover, $P_{Y_t|X_tS_{t-1}}(\cdot|\cdot,1)$ is a BSC(q_1) while $P_{Y_t|X_tS_{t-1}}(\cdot|\cdot,2)$ is a BSC(q_2).

Let $P_{S_0}(2) = \tau$. It can be computed that

$$C_1^U = \max_{P_{X_1S_0}\in\text{conv}(\mathcal{P}(\mathcal{X})\times\mathcal{A}_s)} I(X_1;Y_1|S_0) + H(S_0) \quad (18)$$
$$= \max_{\tau\in[f,e]} (1-\tau)(1-H_b(q_1)) + \tau(1-H_b(q_2)) + H_b(\tau)$$

where the second equality follows from the fact that the maximization in (18) is achieved when X_1 is independent of S_0 and $P_{X_1}(1) = P_{X_1}(2) = \frac{1}{2}$. Therefore, we have the second equation shown at the bottom of the page, where

$$\tau_{\max} = \frac{2^{H_b(q_1)-H_b(q_2)}}{2^{H_b(q_1)-H_b(q_2)}+1}.$$

$$\begin{aligned}\phi(p,q_1,q_2,\alpha,\beta) = H_b(p) &- [(1-p)(1-q_1)\alpha + (1-p)(1-q_2)(1-\alpha) + pq_1\beta + pq_2(1-\beta)] \\ &\times H_b\left(\frac{(1-p)(1-q_1)\alpha + (1-p)(1-q_2)(1-\alpha)}{(1-p)(1-q_1)\alpha + (1-p)(1-q_2)(1-\alpha) + pq_1\beta + pq_2(1-\beta)}\right) \\ &- [(1-p)q_1\alpha + (1-p)q_2(1-\alpha) + p(1-q_1)\beta + p(1-q_2)(1-\beta)] \\ &\times H_b\left(\frac{(1-p)q_1\alpha + (1-p)q_2(1-\alpha)}{(1-p)q_1\alpha + (1-p)q_2(1-\alpha) + p(1-q_1)\beta + p(1-q_2)(1-\beta)}\right)\end{aligned} \quad (17)$$

$$C_1^U = \begin{cases} 1-H_b(q_1) + e(H_b(q_1)-H_b(q_2)) + H_b(e), & e < \tau_{\max} \\ 1-H_b(q_1) + \tau_{\max}(H_b(q_1)-H_b(q_2)) + H_b(\tau_{\max}), & e \ge \tau_{\max} \ge f \\ 1-H_b(q_1) + f(H_b(q_1)-H_b(q_2)) + H_b(f), & f > \tau_{\max} \end{cases}$$

To compute C_1^L, the joint probability distribution $P_{X_1S_0}$ is restricted in $\mathcal{P}(\mathcal{X}) \times \operatorname{conv}(\mathcal{A}_s)$, and we have

$$\begin{aligned} C_1^L &= \max_{P_{X_1} \in \mathcal{P}(\mathcal{X})} \min_{P_{S_0} \in \operatorname{conv}(\mathcal{A}_s)} I(X_1;Y_1|S_0) - H(S_0) \\ &= \min_{\tau \in [f,e]} (1-\tau)(1-H_b(q_1)) + \tau(1-H_b(q_2)) - H_b(\tau) \end{aligned}$$

which yields the first equation shown at the bottom of the page, where

$$\tau_{\min} = \frac{1}{2^{H_b(q_1)-H_b(q_2)}+1}.$$

2) The channel transition probability is of the form

$$\begin{aligned} &P_{Y_tS_t|X_tS_{t-1}}(y_t,s_t|x_t,s_{t-1}) \\ &\qquad = P_{Y_t|X_tS_t}(y_t|x_t,s_t)P_{S_t|X_tS_{t-1}}(s_t|x_t,s_{t-1}) \end{aligned}$$

for any $x_t \in \mathcal{X}$, $y_t \in \mathcal{Y}$, and $s_{t-1}, s_t \in \mathcal{S}$; moreover, $P_{Y_t|X_tS_t}(\cdot|\cdot,1)$ is a BSC(q_1) while $P_{Y_t|X_tS_t}(\cdot|\cdot,2)$ is a BSC(q_2). We shall first compute C_1^U. Note that the joint probability distribution $P_{X_1Y_1S_0}$ is given by

x,y,s	$P_{X_1Y_1S_0}(x,y,s)$
$1,1,1$	$\theta_{11}(1-q_1)(1-a)+\theta_{11}(1-q_2)a$
$1,1,2$	$\theta_{12}(1-q_1)b+\theta_{12}(1-q_2)(1-b)$
$1,2,1$	$\theta_{11}q_1(1-a)+\theta_{11}q_2a$
$1,2,2$	$\theta_{12}q_1b+\theta_{12}q_2(1-b)$
$2,1,1$	$\theta_{21}q_1(1-c)+\theta_{21}q_2c$
$2,1,2$	$\theta_{22}q_1d+\theta_{22}q_2(1-d)$
$2,2,1$	$\theta_{21}(1-q_1)(1-c)+\theta_{21}(1-q_2)c$
$2,2,2$	$\theta_{22}(1-q_1)d+\theta_{22}(1-q_2)(1-d)$

Therefore, we have the second equation shown at the bottom of the page.

To compute C_1^L, the joint probability distribution $P_{X_1S_0}$ is restricted in $\mathcal{P}(\mathcal{X}) \times \operatorname{conv}(\mathcal{A}_s)$. Let $P_{X_1}(2) = p$ and $P_{S_0}(1) = \tau$. We have

$$\begin{aligned} C_1^L = \max_{p\in[0,1]} \min_{\tau\in[f,e]} &(1-\tau)\phi(p,q_1,q_2,1-a,1-c) \\ &+\tau\phi(p,q_1,q_2,b,d) - H_b(\tau) \end{aligned}$$

$$C_1^L = \begin{cases} 1-H_b(q_1)+e(H_b(q_1)-H_b(q_2))-H_b(e), & e<\tau_{\min} \\ 1-H_b(q_1)+\tau_{\min}(H_b(q_1)-H_b(q_2))-H_b(\tau_{\min}), & e\geq\tau_{\min}\geq f \\ 1-H_b(q_1)+f(H_b(q_1)-H_b(q_2))-H_b(f), & f>\tau_{\min} \end{cases}$$

$$\begin{aligned} C_1^U &= \max_{P_{X_1S_0}\in\operatorname{conv}(\mathcal{P}(\mathcal{X})\times\mathcal{A}_s)} I(X_1;Y_1|S_0)+H(S_0) \\ &= \max_{P_{X_1S_0}\in\operatorname{conv}(\mathcal{P}(\mathcal{X})\times\mathcal{A}_s)} H(X_1,S_0)-H(X_1|Y_1,S_0) \\ &= \max_{(\theta_{11},\theta_{12},\theta_{21},\theta_{22})\in\Theta} -\theta_{11}\log\theta_{11}-\theta_{12}\log\theta_{12}-\theta_{21}\log\theta_{21}-\theta_{22}\log\theta_{22} \\ &\quad -[\theta_{11}(1-q_1)(1-a)+\theta_{11}(1-q_2)a+\theta_{21}q_1(1-c)+\theta_{21}q_2c] \\ &\quad \times H_b\left(\frac{\theta_{11}(1-q_1)(1-a)+\theta_{11}(1-q_2)a}{\theta_{11}(1-q_1)(1-a)+\theta_{11}(1-q_2)a+\theta_{21}q_1(1-c)+\theta_{21}q_2c}\right) \\ &\quad -[\theta_{12}(1-q_1)b+\theta_{12}(1-q_2)(1-b)+\theta_{22}q_1d+\theta_{22}q_2(1-d)] \\ &\quad \times H_b\left(\frac{\theta_{12}(1-q_1)b+\theta_{12}(1-q_2)(1-b)}{\theta_{12}(1-q_1)b+\theta_{12}(1-q_2)(1-b)+\theta_{22}q_1d+\theta_{22}q_2(1-d)}\right) \\ &\quad -[\theta_{11}q_1(1-a)+\theta_{11}q_2a+\theta_{21}(1-q_1)(1-c)+\theta_{21}(1-q_2)c] \\ &\quad \times H_b\left(\frac{\theta_{11}q_1(1-a)+\theta_{11}q_2a}{\theta_{11}q_1(1-a)+\theta_{11}q_2a+\theta_{21}(1-q_1)(1-c)+\theta_{21}(1-q_2)c}\right) \\ &\quad -[\theta_{12}q_1b+\theta_{12}q_2(1-b)+\theta_{22}(1-q_1)d+\theta_{22}(1-q_2)(1-d)] \\ &\quad \times H_b\left(\frac{\theta_{12}q_1b+\theta_{12}q_2(1-b)}{\theta_{12}q_1b+\theta_{12}q_2(1-b)+\theta_{22}(1-q_1)d+\theta_{22}(1-q_2)(1-d)}\right) \end{aligned}$$

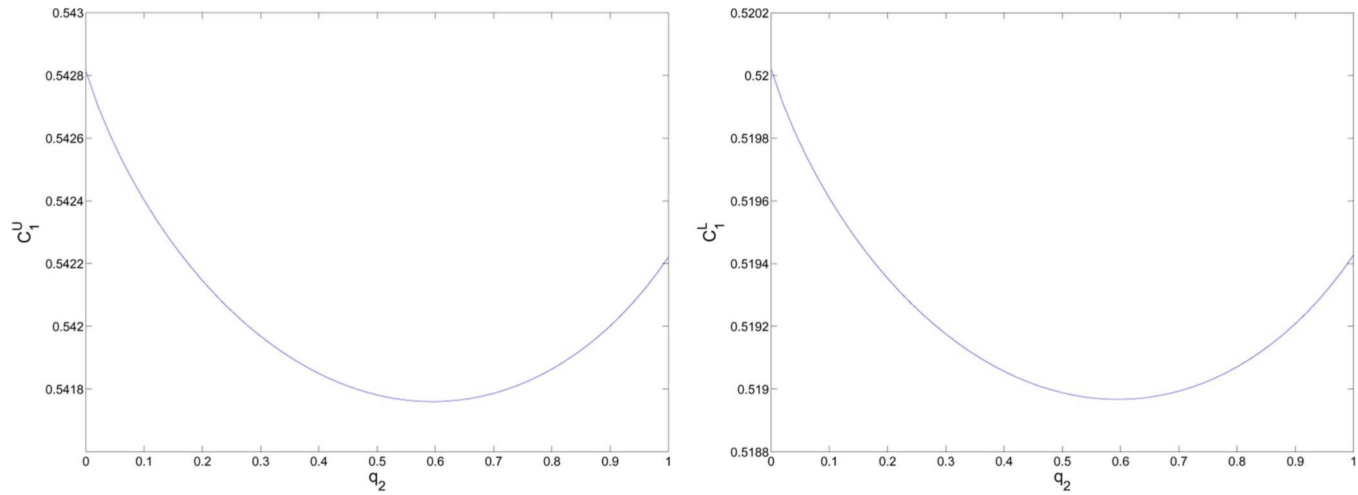

Fig. 2. Plots of C_1^U and C_1^L with $a = 0.0001$, $b = 0.1$, $c = 0.00005$, $d = 0.1$, $e = \frac{a}{a+b}$, $f = \frac{c}{c+d}$, $q_1 = 0.1$, and $q_2 \in [0, 1]$.

where $\phi(\cdot)$ is defined in (17). Note that (see the equation shown at the bottom of the page), where

$$\tau^* = \frac{1}{2^{\phi(p,q_1,q_2,b,d)-\phi(p,q_1,q_2,1-a,1-c)} + 1}.$$

Therefore, we have

$$C_1^L = \max_{p\in[0,1]} \eta(\tau^*, p, q_1, q_2, a, b, c, d).$$

See Fig. 2 for plots of C_1^U and C_1^L.

Example 5.4: The setting is the same as that of Example 5.2 with the only difference that the state process is assumed to be known at the receiver. Again, we shall consider the following two cases:

1) The channel transition probability is of the form

$$P_{Y_tS_t|X_tS_{t-1}}(y_t, s_t|x_t, s_{t-1}) = P_{Y_t|X_tS_{t-1}}(y_t|x_t, s_{t-1})P_{S_t|X_tS_{t-1}}(s_t|x_t, s_{t-1})$$

for any $x_t \in \mathcal{X}$, $y_t \in \mathcal{Y}$, and $s_{t-1}, s_t \in \mathcal{S}$; moreover, $P_{Y_t|X_tS_{t-1}}(\cdot|\cdot, 1)$ is a BSC(q_1) while $P_{Y_t|X_tS_{t-1}}(\cdot|\cdot, 2)$ is a BSC(q_2).

Let $P_{X_1}(2) = p$. The conditional probability distribution $P_{X_1Y_1S_1|S_0}$ is given by

x, y, s	$P_{X_1Y_1S_1\|S_0}(x, y, s\|1)$	$P_{X_1Y_1S_1\|S_0}(x, y, s\|2)$
$1, 1, 1$	$(1-p)(1-q_1)(1-a)$	$(1-p)(1-q_2)b$
$1, 1, 2$	$(1-p)(1-q_1)a$	$(1-p)(1-q_2)(1-b)$
$1, 2, 1$	$(1-p)q_1(1-a)$	$(1-p)q_2b$
$1, 2, 2$	$(1-p)q_1a$	$(1-p)q_2(1-b)$
$2, 1, 1$	$pq_1(1-c)$	pq_2d
$2, 1, 2$	pq_1c	$pq_2(1-d)$
$2, 2, 1$	$p(1-q_1)(1-c)$	$p(1-q_2)d$
$2, 2, 2$	$p(1-q_1)c$	$p(1-q_2)(1-d)$

and the induced condition probability distributions $P_{Y_1S_1|S_0}$ and $P_{X_1|Y_1S_1S_0}$ are given by the tables shown at the bottom of the next page. Now it can be computed that

$$I(X_1; Y_1, S_1|S_0 = 1) = H(X_1) - H(X_1|Y_1, S_1, S_0 = 1) = \varphi(p, q_1, 1-a, 1-c)$$

$$I(X_1; Y_1, S_1|S_0 = 2) = H(X_1) - H(X_1|Y_1, S_1, S_0 = 2) = \varphi(p, q_2, b, d)$$

$$\begin{aligned}
&\min_{\tau\in[f,e]} (1-\tau)\phi(p, q_1, q_2, 1-a, 1-c) + \tau\phi(p, q_1, q_2, b, d) - H_b(\tau) \\
&= \begin{cases} (1-e)\phi(p, q_1, q_2, 1-a, 1-c) + e\phi(p, q_1, q_2, b, d) - H_b(e), & e < \tau^* \\ (1-\tau^*)\phi(p, q_1, q_2, 1-a, 1-c) + \tau^*\phi(p, q_1, q_2, b, d) - H_b(\tau^*), & e \geq \tau^* \geq f \\ (1-f)\phi(p, q_1, q_2, 1-a, 1-c) + f\phi(p, q_1, q_2, b, d) - H_b(f), & f > \tau^* \end{cases} \\
&\triangleq \eta(\tau^*, p, q_1, q_2, a, b, c, d),
\end{aligned}$$

where

$$\begin{aligned}\varphi(p,q,\alpha,\beta) = H_b(p) &- [(1-p)(1-q)\alpha + pq\beta] \\ &\times H_b\left(\frac{(1-p)(1-q)\alpha}{(1-p)(1-q)\alpha + pq\beta}\right) \\ &- [(1-p)(1-q)(1-\alpha) + pq(1-\beta)] \\ &\times H_b\left(\frac{(1-p)(1-q)(1-\alpha)}{(1-p)(1-q)(1-\alpha) + pq(1-\beta)}\right) \\ &- [(1-p)q\alpha + p(1-q)\beta] \\ &\times H_b\left(\frac{(1-p)q\alpha}{(1-p)q\alpha + p(1-q)\beta}\right) \\ &- [(1-p)q(1-\alpha) + p(1-q)(1-\beta)] \\ &\times H_b\left(\frac{(1-p)q(1-\alpha)}{(1-p)q(1-\alpha) + p(1-q)(1-\beta)}\right). \end{aligned} \tag{19}$$

Therefore, by Corollary 3.14, we have

$$C^S = \max_{p\in[0,1]} \frac{1}{\gamma+1}\varphi(p,q_1,1-a,1-c) + \frac{\gamma}{1+\gamma}\varphi(p,q_2,b,d).$$

If $\gamma = 1$ and $a + c = 1$, then

$$C^S = \max_{p\in[0,1]} \frac{1}{2}\varphi(p,q_1,1-a,a) + \frac{1}{2}\varphi(p,q_2,a,1-a). \tag{20}$$

Note that $\varphi(p,q,\alpha,\beta)$ is a concave function of p. Moreover, it is easy to verify that $\varphi(p,q,\alpha,\beta) = \varphi(1-p,q,\alpha,\beta)$ if $\alpha+\beta = 1$. Therefore, the maximum in (20) is achieved at $p = \frac{1}{2}$, which yields

$$C^S = \frac{1}{2}\varphi\left(\frac{1}{2},q_1,1-a,a\right) + \frac{1}{2}\varphi\left(\frac{1}{2},q_2,a,1-a\right).$$

It is interesting to further specialize to the case $q_1 = q_2 = q$. Now C^S is given by

$$\begin{aligned}C^S(q,a) = 1 &- [(1-q)(1-a) + qa] \\ &\times H_b\left(\frac{(1-q)(1-a)}{(1-q)(1-a)+qa}\right) \\ &- [(1-q)a + q(1-a)] \\ &\times H_b\left(\frac{(1-q)a}{(1-q)a + q(1-a)}\right). \end{aligned} \tag{21}$$

The following is easy to verify by direct evaluation of the expression in (21):

a) $C^S\left(q,\frac{1}{2}\right) = 1 - H_b(q)$. When $a = \frac{1}{2}$, no information can be conveyed via the state transitions, and the channel is simply a $\mathrm{BSC}(q)$.

b) $C^S\left(\frac{1}{2},a\right) = 1 - H_b(a)$. When $q = \frac{1}{2}$, no information can be conveyed via the relationship between the channel input and the channel output. The only information that can be conveyed is via the state transitions, for which the effective channel is a $\mathrm{BSC}(a)$.

c) More generally, we have the symmetry relation

$$C^S(\mu,\nu) = C^S(\nu,\mu). \tag{22}$$

Indeed, by simple operations at the receiver, one can convert a channel with parameters $(q=\mu, a=\nu)$ to one with parameters $(q=\nu, a=\mu)$. To see this, we define $Z_t = 1$ if $S_t = S_{t-1}$ and $Z_t = 2$ if $S_t \neq S_{t-1}$, $t = 1,2\ldots$. It is easy to verify that in this special case, the finite-state channel $P_{Y_tS_t|X_tS_{t-1}}$ is equivalent to the memoryless channel $P_{Y_tZ_t|X_t}$ with the form

$$P_{Y_tZ_t|X_t}(y,z|x) = P_{Y_t|X_t}(y|x)P_{Z_t|X_t}(z|x)$$

for any $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and $z \in \mathcal{Z} \triangleq \{1,2\}$, where $P_{Y_t|X_t}$ is a $\mathrm{BSC}(q)$ and $P_{Z_t|X_t}$ is a $\mathrm{BSC}(a)$. Moreover, the symmetry relation (22) follows from the symmetric roles of Y_t and Z_t in the memoryless channel $P_{Y_tZ_t|X_t}$. Fig. 3 presents a plot of $C^S(q,a)$.

2) The channel transition probability is of the form

$$\begin{aligned}P_{Y_tS_t|X_tS_{t-1}}&(y_t,s_t|x_t,s_{t-1}) \\ &= P_{Y_t|X_tS_t}(y_t|x_t,s_t)P_{S_t|X_tS_{t-1}}(s_t|x_t,s_{t-1})\end{aligned}$$

for any $x_t \in \mathcal{X}$, $y_t \in \mathcal{Y}$, and $s_{t-1}, s_t \in \mathcal{S}$; moreover, $P_{Y_t|X_tS_t}(\cdot|\cdot,1)$ is a $\mathrm{BSC}(q_1)$ while $P_{Y_t|X_tS_t}(\cdot|\cdot,2)$ is a

y,s	$P_{Y_1S_1\|S_0}(y,s\|1)$	$P_{Y_1S_1\|S_0}(y,s\|2)$
$1,1$	$(1-p)(1-q_1)(1-a)+pq_1(1-c)$	$(1-p)(1-q_2)b+pq_2d$
$1,2$	$(1-p)(1-q_1)a+pq_1c$	$(1-p)(1-q_2)(1-b)+pq_2(1-d)$
$2,1$	$(1-p)q_1(1-a)+p(1-q_1)(1-c)$	$(1-p)q_2b+p(1-q_2)d$
$2,2$	$(1-p)q_1a+p(1-q_1)c$	$(1-p)q_2(1-b)+p(1-q_2)(1-d)$

x,y,s	$P_{X_1\|Y_1S_1S_0}(x\|y,s,1)$	$P_{X_1\|Y_1S_1S_0}(x\|y,s,2)$
$1,1,1$	$\frac{(1-p)(1-q_1)(1-a)}{(1-p)(1-q_1)(1-a)+pq_1(1-c)}$	$\frac{(1-p)(1-q_2)b}{(1-p)(1-q_2)b+pq_2d}$
$1,1,2$	$\frac{(1-p)(1-q_1)a}{(1-p)(1-q_1)a+pq_1c}$	$\frac{(1-p)(1-q_2)(1-b)}{(1-p)(1-q_2)(1-b)+pq_2(1-d)}$
$1,2,1$	$\frac{(1-p)q_1(1-a)}{(1-p)q_1(1-a)+p(1-q_1)(1-c)}$	$\frac{(1-p)q_2b}{(1-p)q_2b+p(1-q_2)d}$
$1,2,2$	$\frac{(1-p)q_1a}{(1-p)q_1a+p(1-q_1)c}$	$\frac{(1-p)q_2(1-b)}{(1-p)q_2(1-b)+p(1-q_2)(1-d)}$

BSC(q_2). Let $P_{X_1}(2) = p$. The conditional probability distribution $P_{X_1Y_1S_1|S_0}$ is computed in Example 5.2. The induced conditional probability distributions $P_{Y_1S_1|S_0}$ and $P_{X_1|Y_1S_1S_0}$ are given by the tables at the bottom of the page.

Now it can be computed that

$$\begin{aligned} I(X_1;Y_1,S_1|S_0=1) &= H(X_1) - H(X_1|Y_1,S_1,S_0=1) \\ &= \psi(p,q_1,q_2,1-a,1-c) \\ I(X_1;Y_1,S_1|S_0=2) &= H(X_1) - H(X_1|Y_1,S_1,S_0=2) \\ &= \psi(p,q_1,q_2,b,d) \end{aligned}$$

where [see (23), shown at the bottom of the page]. Therefore, by Corollary 3.14, we have

$$C^S = \max_{p\in[0,1]} \frac{1}{\gamma+1}\psi(p,q_1,q_2,1-a,1-c) + \frac{\gamma}{1+\gamma}\psi(p,q_1,q_2,b,d). \quad (24)$$

The maximization in (24) is easy to perform numerically. In general, it is hard to find the maximizing p explicitly. Even for $\gamma = 1$, the maximizing value of p is, in general, not $\frac{1}{2}$. For example, when $(q_1,q_2,a,c,\gamma) = (0.1,0.5,0.1,0.5,1)$, the maximizing value of p is approximately 0.49218365 (and the associated capacity is approximately 0.36663024).

a) If $q_1 = q_2 = q$, then $Y_t - X_t - (S_t, S_{t-1})$ form a Markov chain, which is subsumed by Case 1).

b) Consider the special case where $\gamma = 1$ and $a = 1-c$. We have

$$C^S = \max_{p\in[0,1]} \frac{1}{2}\psi(p,q_1,q_2,1-a,a) + \frac{1}{2}\psi(p,q_1,q_2,a,1-a). \quad (25)$$

Note that $\psi(p,q_1,q_2,\alpha,\beta)$ is a concave function of p. Moreover, it is easy to verify that $\psi(1-p,q_1,q_2,1-a,a) + \psi(1-p,q_1,q_2,a,1-a) = \psi(p,q_1,q_2,1-a,a)+\psi(p,q_1,q_2,a,1-a)$. Therefore, the maximum in (25) is achieved at $p = \frac{1}{2}$, which yields

$$C^S = \frac{1}{2}\psi\left(\frac{1}{2},q_1,q_2,1-a,a\right) + \frac{1}{2}\psi\left(\frac{1}{2},q_1,q_2,a,1-a\right).$$

Example 5.5: Let $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y} = \mathcal{S} = \{1,2\}$. Define $W_t = 1$ if $X_{1,t} = X_{2,t}$, and $W_t = 2$ if $X_{1,t} \neq X_{2,t}$. Suppose the finite-state multiple access channel has the property that $(Y_t,S_t) - (W_t,S_{t-1}) - (X_{1,t},X_{2,t})$ form a Markov chain.

y,s	$P_{Y_1S_1\|S_0}(y,s\|1)$	$P_{Y_1S_1\|S_0}(y,s\|2)$
$1,1$	$(1-p)(1-q_1)(1-a)+pq_1(1-c)$	$(1-p)(1-q_1)b+pq_1d$
$1,2$	$(1-p)(1-q_2)a+pq_2c$	$(1-p)(1-q_2)(1-b)+pq_2(1-d)$
$2,1$	$(1-p)q_1(1-a)+p(1-q_1)(1-c)$	$(1-p)q_1b+p(1-q_1)d$
$2,2$	$(1-p)q_2a+p(1-q_2)c$	$(1-p)q_2(1-b)+p(1-q_2)(1-d)$

x,y,s	$P_{X_1\|Y_1S_1S_0}(x\|y,s,1)$	$P_{X_1\|Y_1S_1S_0}(x\|y,s,2)$
$1,1,1$	$\frac{(1-p)(1-q_1)(1-a)}{(1-p)(1-q_1)(1-a)+pq_1(1-c)}$	$\frac{(1-p)(1-q_1)b}{(1-p)(1-q_1)b+pq_1d}$
$1,1,2$	$\frac{(1-p)(1-q_2)a}{(1-p)(1-q_2)a+pq_2c}$	$\frac{(1-p)(1-q_2)(1-b)}{(1-p)(1-q_2)(1-b)+pq_2(1-d)}$
$1,2,1$	$\frac{(1-p)q_1(1-a)}{(1-p)q_1(1-a)+p(1-q_1)(1-c)}$	$\frac{(1-p)q_1b}{(1-p)q_1b+p(1-q_1)d}$
$1,2,2$	$\frac{(1-p)q_2a}{(1-p)q_2a+p(1-q_2)c}$	$\frac{(1-p)q_2(1-b)}{(1-p)q_2(1-b)+p(1-q_2)(1-d)}$

$$\begin{aligned} \psi(p,q_1,q_2,\alpha,\beta) &= H_b(p) - [(1-p)(1-q_1)\alpha+pq_1\beta]H_b\left(\frac{(1-p)(1-q_1)\alpha}{(1-p)(1-q_1)\alpha+pq_1\beta}\right) \\ &\quad - [(1-p)(1-q_2)(1-\alpha)+pq_2(1-\beta)]H_b\left(\frac{(1-p)(1-q_2)(1-\alpha)}{(1-p)(1-q_2)(1-\alpha)+pq_2(1-\beta)}\right) \\ &\quad - [(1-p)q_1\alpha+p(1-q_1)\beta]H_b\left(\frac{(1-p)q_1\alpha}{(1-p)q_1\alpha+p(1-q_1)\beta}\right) \\ &\quad - [(1-p)q_2(1-\alpha)+p(1-q_2)(1-\beta)]H_b\left(\frac{(1-p)q_2(1-\alpha)}{(1-p)q_2(1-\alpha)+p(1-q_2)(1-\beta)}\right) \end{aligned} \quad (23)$$

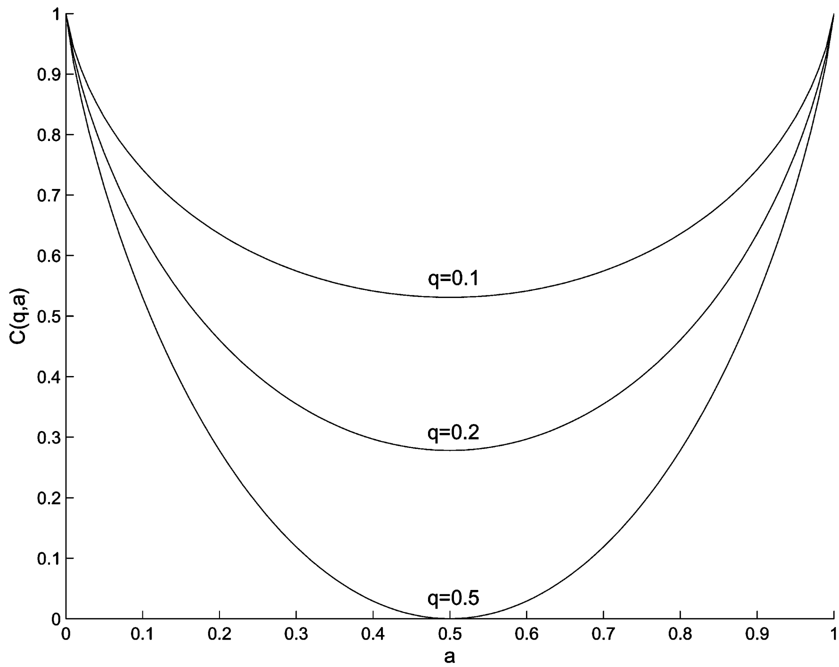

Fig. 3. Plots of $C^S(q,a)$, for the 3 values $q = 0.1, 0.2, 0.5$ (corresponding to the curves from top to bottom) and $0 \le a \le 1$. Note, for the upper two curves, that the capacity is positive even when $q = 0.5$, since information is communicated through the state transitions. By (22), the curves shown also plot $C^S(q,a)$, for the 3 values $a = 0.1, 0.2, 0.5$, and $0 \le q \le 1$. Note that the bottom curve coincides with that of the capacity of the BSC (as a function of the crossover probability). Remark: strictly speaking, Corollary 2.8 is not applicable if $a = 0$ or $a = 1$; however, since in this extreme case the channel inputs can be reconstructed from the state process, the capacity is clearly 1.

We denote the transition probability matrices $P_{S_t|W_tS_{t-1}}(\cdot|1,\cdot)$ and $P_{S_t|W_tS_{t-1}}(\cdot|2,\cdot)$ by T_1 and T_2, respectively, where

$$T_1 = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}, \quad T_2 = \begin{pmatrix} 1-c & c \\ d & 1-d \end{pmatrix}.$$

Assume $0 \le \min(a,b,c,d) \le \max(a,b,c,d) \le 1$, $0 < a+b < 2$, $0 < c+d < 2$, and

$$\frac{a}{b} = \frac{c}{d} = \gamma.$$

Due to the special structure of this multiple access channel, the results derived in Example 5.4 are directly applicable. We shall consider the following two cases:

1) The channel transition probability is of the form

$$P_{Y_tS_t|W_tS_{t-1}}(y_t, s_t|w_t, s_{t-1}) = P_{Y_t|W_tS_{t-1}}(y_t|w_t, s_{t-1})P_{S_t|W_tS_{t-1}}(s_t|w_t, s_{t-1})$$

for any $w_t \in \mathcal{W} \triangleq \{0,1\}$, $y_t \in \mathcal{Y}$, and $s_{t-1}, s_t \in \mathcal{S}$; moreover, $P_{Y_t|W_tS_{t-1}}(\cdot|\cdot,1)$ is a $\mathrm{BSC}(q_1)$ while $P_{Y_t|W_tS_{t-1}}(\cdot|\cdot,2)$ is a $\mathrm{BSC}(q_2)$. We have

$$\mathcal{R}^S = \Bigg\{(R_1,R_2) \in \mathbb{R}_+^2 : R_1 + R_2 \le \max_{p\in[0,1]} \frac{1}{\gamma+1}\varphi(p, q_1, 1-a, 1-c) + \frac{\gamma}{1+\gamma}\varphi(p, q_2, b, d)\Bigg\}$$

where $\varphi(\cdot)$ is defined in (19).

2) The channel transition probability is of the form

$$P_{Y_tS_t|W_tS_{t-1}}(y_t, s_t|w_t, s_{t-1}) = P_{Y_t|W_tS_t}(y_t|w_t, s_t)P_{S_t|W_tS_{t-1}}(s_t|w_t, s_{t-1})$$

for any $w_t \in \mathcal{W}$, $y_t \in \mathcal{Y}$, and $s_{t-1}, s_t \in \mathcal{S}$; moreover, $P_{Y_t|W_tS_t}(\cdot|\cdot,1)$ is a $\mathrm{BSC}(q_1)$ while $P_{Y_t|W_tS_t}(\cdot|\cdot,2)$ is a $\mathrm{BSC}(q_2)$. We have

$$\mathcal{R}^S = \Bigg\{(R_1,R_2) \in \mathbb{R}_+^2 : R_1 + R_2 \le \max_{p\in[0,1]} \frac{1}{\gamma+1}\psi(p, q_1, q_2, 1-a, 1-c) + \frac{\gamma}{1+\gamma}\psi(p, q_1, q_2, b, d)\Bigg\}$$

where $\psi(\cdot)$ is defined in (23).

VI. Conclusion

We have used the theory of Markov set-chains to derive new finite-letter upper and lower bounds on the capacity of finite-state channels. Compared with the existing capacity bounds, the new bounds can more effectively capture the long term behavior of the state process. In particular, these bounds coincide and yield single-letter capacity characterizations for a class of channels with the state process known at the receiver, including channels whose long-term marginal state distribution is independent of the input process. Analogous results are derived for finite-state multiple access channels. A natural future direction is to see whether the approach of the present work can be applied also to obtain bounds on the capacity of finite-state channels with feedback that would improve on those of [22].

APPENDIX A PROOF OF COROLLARY 2.7

1) It is clear that $\mathcal{A}_s\mathcal{T} \subseteq \mathcal{B}\mathcal{T}$ if $\mathcal{A}_s \subseteq \mathcal{B}$. By Theorem 2.6, if all finite products of matrices from $\mathcal{T}$ are regular, then $\mathcal{A}_s\mathcal{T} = \mathcal{A}_s$, which yields the desired result.
2) If $\mathcal{B}\mathcal{T} \subseteq \mathcal{B}$, then $\mathcal{B}\mathcal{T}^n \subseteq \mathcal{B}$ for all n. By Theorem 2.6, $\mathcal{B}\mathcal{T}^n$ converges to $\mathcal{A}_s$ in the Hausdorff metric as n goes to infinity. Therefore, we must have $\mathcal{A}_s \subseteq \mathcal{B}$.
3) If $\mathcal{B} \subseteq \mathcal{B}\mathcal{T}$, then $\mathcal{B} \subseteq \mathcal{B}\mathcal{T}^n$ for all n. It follows from Theorem 2.6 that $\mathcal{B}\mathcal{T}^n$ converges to $\mathcal{A}_s$ in the Hausdorff metric as n goes to infinity. Therefore, we must have $\mathcal{B} \subseteq \mathcal{A}_s$.

APPENDIX B PROOF OF THEOREM 2.10

Firstly, we shall show that $\mathrm{conv}(\mathcal{A}_s) = \mathrm{conv}(\mathrm{conv}(\mathcal{A}_s)\mathcal{T})$. Since $\mathcal{A}_s = \mathcal{A}_s\mathcal{T}$, it follows that $\mathrm{conv}(\mathcal{A}_s) = \mathrm{conv}(\mathcal{A}_s\mathcal{T}) \subseteq \mathrm{conv}(\mathrm{conv}(\mathcal{A}_s)\mathcal{T})$. Now we proceed to show the other direction. Let $\mathcal{E}$ be the set of extreme points of $\mathrm{conv}(\mathcal{A}_s)$, i.e., the set of points in $\mathrm{conv}(\mathcal{A}_s)$ which do not lie in any open line segment joining two distinct points of $\mathrm{conv}(\mathcal{A}_s)$. Clearly, we have $\mathcal{E} \subseteq \mathcal{A}_s$ and $\mathcal{E}\mathcal{T} \subseteq \mathcal{A}_s\mathcal{T} = \mathcal{A}_s$. Furthermore, let $\mathcal{E}'$ be the set of extreme points of $\mathrm{conv}(\mathrm{conv}(\mathcal{A}_s)\mathcal{T})$. Note that $\mathcal{E}' \subseteq \mathrm{conv}(\mathcal{A}_s)\mathcal{T} = \mathrm{conv}(\mathcal{E})\mathcal{T}$. By the definition of $\mathcal{E}'$, it is easy to

$$
\begin{aligned}
&I\left(X_1^{nk};Y_1^{nk}|S_0=s_0\right)\\
&\quad\le I\left(X_1^{nk};Y_1^{nk},\{S_{mk}\}_{m=1}^{n-1}|S_0=s_0\right)\\
&\quad= H\left(Y_1^{nk},\{S_{mk}\}_{m=1}^{n-1}|S_0=s_0\right)-H\left(Y_1^{nk},\{S_{mk}\}_{m=1}^{n-1}|X_1^{nk},S_0=s_0\right)\\
&\quad=\sum_{i=1}^{n}\left[H\left(Y_{(i-1)k+1}^{ik},S_{(i-1)k}|Y_1^{(i-1)k},\{S_{mk}\}_{m=1}^{i-2},S_0=s_0\right)\right.\\
&\qquad\qquad\left.-H\left(Y_{(i-1)k+1}^{ik},S_{(i-1)k}|X_1^{nk},Y_1^{(i-1)k},\{S_{mk}\}_{m=1}^{i-2},S_0=s_0\right)\right]\\
&\quad=\sum_{i=1}^{n}\left[H\left(Y_{(i-1)k+1}^{ik}|Y_1^{(i-1)k},\{S_{mk}\}_{m=1}^{i-1},S_0=s_0\right)-H\left(Y_{(i-1)k+1}^{ik}|X_1^{nk},Y_1^{(i-1)k},\{S_{mk}\}_{m=1}^{i-1},S_0=s_0\right)\right.\\
&\qquad\qquad\left.+H\left(S_{(i-1)k}|Y_1^{(i-1)k},\{S_{mk}\}_{m=1}^{i-2},S_0=s_0\right)-H\left(S_{(i-1)k}|X_1^{nk},Y_1^{(i-1)k},\{S_{mk}\}_{m=1}^{i-2},S_0=s_0\right)\right]\\
&\quad\le\sum_{i=1}^{n}\left[H\left(Y_{(i-1)k+1}^{ik}|Y_1^{(i-1)k},\{S_{mk}\}_{m=1}^{i-1},S_0=s_0\right)-H\left(Y_{(i-1)k+1}^{ik}|X_1^{nk},Y_1^{(i-1)k},\{S_{mk}\}_{m=1}^{i-1},S_0=s_0\right)\right.\\
&\qquad\qquad\left.+H\left(S_{(i-1)k}|Y_1^{(i-1)k},\{S_{mk}\}_{m=1}^{i-2},S_0=s_0\right)\right]\\
&\quad\le\sum_{i=1}^{n}\left[H\left(Y_{(i-1)k+1}^{ik}|Y_1^{(i-1)k},\{S_{mk}\}_{m=1}^{i-1},S_0=s_0\right)-H\left(Y_{(i-1)k+1}^{ik}|X_1^{nk},Y_1^{(i-1)k},\{S_{mk}\}_{m=1}^{i-1},S_0=s_0\right)\right.\\
&\qquad\qquad\left.+H\left(S_{(i-1)k}|S_0=s_0\right)\right]\\
&\quad=\sum_{i=1}^{n}\left[H\left(Y_{(i-1)k+1}^{ik}|Y_1^{(i-1)k},\{S_{mk}\}_{m=1}^{i-1},S_0=s_0\right)-H\left(Y_{(i-1)k+1}^{ik}|X_{(i-1)k+1}^{ik},S_{(i-1)k},S_0=s_0\right)\right.\\
&\qquad\qquad\left.+H\left(S_{(i-1)k}|S_0=s_0\right)\right] &(26)\\
&\quad\le\sum_{i=1}^{n}\left[H\left(Y_{(i-1)k+1}^{ik}|S_{(i-1)k},S_0=s_0\right)-H\left(Y_{(i-1)k+1}^{ik}|X_{(i-1)k+1}^{ik},S_{(i-1)k},S_0=s_0)+H(S_{(i-1)k}|S_0=s_0\right)\right]\\
&\quad=\sum_{i=1}^{n}\left[I\left(X_{(i-1)k+1}^{ik};Y_{(i-1)k+1}^{ik}|S_{(i-1)k},S_0=s_0\right)+H\left(S_{(i-1)k}|S_0=s_0\right)\right] &(27)
\end{aligned}
$$

see that $\mathcal{E}' \subseteq \mathcal{E}\mathcal{T}$. Therefore, we have $\text{conv}(\text{conv}(\mathcal{A}_s)\mathcal{T}) = \text{conv}(\mathcal{E}') \subseteq \text{conv}(\mathcal{E}\mathcal{T}) \subseteq \text{conv}(\mathcal{A}_s)$.

Let $\mathcal{B}$ be any nonempty compact convex set satisfying $\mathcal{B} = \text{conv}(\mathcal{B}\mathcal{T})$. Let $\mathcal{E}''$ be the set of extreme points of $\mathcal{B}$. Since $\mathcal{E}''$ is also the set of extreme points of $\text{conv}(\mathcal{B}\mathcal{T})$, it follows that $\mathcal{E}'' \subseteq \mathcal{E}''\mathcal{T}$, which further implies $\mathcal{E}'' \subseteq \mathcal{E}''\mathcal{T}^n$ for all n. Therefore, we have $\mathcal{B} = \text{conv}(\mathcal{E}'') \subseteq \text{conv}(\mathcal{E}''\mathcal{T}^n) \subseteq \text{conv}(\mathcal{B}\mathcal{T}^n)$ for all n. On the other hand, it is easy to see that $\text{conv}(\mathcal{B}\mathcal{T}^n) \subseteq \text{conv}(\ldots\text{conv}(\text{conv}(\mathcal{B}\mathcal{T})\mathcal{T})\ldots\mathcal{T}) = \mathcal{B}$. Therefore, we have $\mathcal{B} = \text{conv}(\mathcal{B}\mathcal{T}^n)$ for all n. Since $\mathcal{B}\mathcal{T}^n$ converges to $\mathcal{A}_s$ in the Hausdorff metric, it implies that $\mathcal{B} = \text{conv}(\mathcal{A}_s)$. The proof is complete.

Appendix C
Proof of Theorem 3.1

We shall compute an upper bound on the channel capacity by considering a genie-aided system in which the state information $\{S_{mk}\}_{m=1}^{\infty}$ is provided to the receiver. Specifically, we have (26)–(27), shown at the bottom of the previous page, where (26) follows from the property of finite-state channels.

Note that for any $x_{(i-1)k+1}^{ik} \in \mathcal{X}^k$, and $s_0, s_{(i-1)k} \in \mathcal{S}$, see the equation shown at the bottom of the page, where the last equality follows from the fact that $X_{(i-1)k+1}^{ik} - \left(X_1^{(i-1)k}, S_0\right) - S_{(i-1)k}$ form a Markov chain. It is clear that $P_{X_{(i-1)k+1}^{ik}|X_1^{(i-1)k}S_0}\left(\cdot|x_1^{(i-1)k}, s_0\right) \in \mathcal{P}(\mathcal{X}^k)$ and $P_{S_{(i-1)k}|X_1^{(i-1)k}S_0}\left(\cdot|x_1^{(i-1)k}, s_0\right) \in \mathcal{G}_{s_0,(i-1)k}$. Therefore, we have

$$P_{X_{(i-1)k+1}^{ik}S_{(i-1)k}|S_0}\left(\cdot,\cdot|s_0\right) \in \text{conv}(\mathcal{P}(\mathcal{X}^k) \times \mathcal{G}_{s_0,(i-1)k}). \tag{28}$$

In view of (27), (28), and the fact that $\lim_{i\to\infty} \delta(\mathcal{G}_{s_0,(i-1)k}, \mathcal{A}_{s_0}) = 0$, the proof is complete.

Appendix D
Proof of Lemma 3.3

Since $P_{X_1^k S_0} \in \text{conv}\left(\mathcal{P}(\mathcal{X}^k) \times \mathcal{A}_s\right)$, there exist $m \in \mathbb{N}$, $\mu_i \in [0,1]$, $P_{X_1^k}^{(i)} \in \mathcal{P}(\mathcal{X}^k)$, $P_{S_0}^{(i)} \in \mathcal{A}_s$, $i = 1,\ldots,m$, such that $\sum_{i=1}^m \mu_i = 1$ and

$$P_{X_1^k S_0}\left(x_1^k, s_0\right) = \sum_{i=1}^m \mu_i P_{X_1^k}^{(i)}\left(x_1^k\right) P_{S_0}^{(i)}(s_0)$$
$$x_1^k \in \mathcal{X}^k, \quad s_0 \in \mathcal{S}.$$

Note that for any $x_{k_1+1}^k \in \mathcal{X}^{k_2}$ and $s_{k_1} \in \mathcal{S}$, see (29) shown at the bottom of the next page, where (29) follows from the fact that $S_{k_1} - \left(X_1^{k_1}, S_0\right) - X_{k_1+1}^k$ form a Markov chain. See (30), shown at the bottom of the next page. In light of the fact that that $P_{S_0}^{(i)} \in \mathcal{A}_s$ and $\mathcal{A}_s\mathcal{T} \subseteq \mathcal{A}_s$, we have $P_{S_{k_1}|X_1^{k_1}}^{(i)}\left(\cdot|x_1^{k_1}\right) \in \mathcal{A}_s\mathcal{T}^{k_1} \subseteq \mathcal{A}_s$ for any fixed $x_1^{k_1}$. Furthermore, it is easy to see that $P_{X_{k_1+1}^k|X_1^{k_1}}^{(i)}\left(\cdot|x_1^{k_1}\right) \in \mathcal{P}(\mathcal{X}^{k_2})$ for any fixed $x_1^{k_1}$. Therefore, we have $P_{X_{k_1+1}^k S_{k_1}} \in \text{conv}\left(\mathcal{P}(\mathcal{X}^{k_2}) \times \mathcal{A}_s\right)$.

For any $x_1^{k_1} \in \mathcal{X}^{k_1}$ and $s_0 \in \mathcal{S}$

$$\begin{aligned} P_{X_1^{k_1}S_0}\left(x_1^{k_1}, s_0\right) &= \sum_{x_{k_1+1}^k} P_{X_1^k S_0}\left(x_1^k, s_0\right) \\ &= \sum_{x_{k_1+1}^k}\sum_{i=1}^m \mu_i P_{X_1^k}^{(i)}\left(x_1^k\right) P_{S_0}^{(i)}(s_0) \\ &= \sum_{i=1}^m \mu_i P_{X_1^{k_1}}^{(i)}\left(x_1^{k_1}\right) P_{S_0}^{(i)}(s_0). \end{aligned}$$

Therefore, we have $P_{X_1^{k_1}S_0} \in \text{conv}\left(\mathcal{P}(\mathcal{X}^{k_1}) \times \mathcal{A}_s\right)$. The proof is complete.

Appendix E
Proof of Theorem 3.4

Note that [see (31)–(32), shown at the bottom of the next page], where (31) follows from the fact that $Y_1^{k_1} - \left(X_1^{k_1}, S_0\right) -$

$$\begin{aligned} &P_{X_{(i-1)k+1}^{ik}S_{(i-1)k}|S_0}\left(x_{(i-1)k+1}^{ik}, s_{(i-1)k}|s_0\right) \\ &\quad= \sum_{x_1^{(i-1)k}} P_{X_1^{ik}S_{(i-1)k}|S_0}\left(x_1^{ik}, s_{(i-1)k}|s_0\right) \\ &\quad= \sum_{x_1^{(i-1)k}} P_{X_{(i-1)k+1}^{ik}|X_1^{(i-1)k}S_{(i-1)k}S_0}\left(x_{(i-1)k}^{ik}|x_1^{(i-1)k}, s_{(i-1)k}, s_0\right) \\ &\quad\quad\times P_{S_{(i-1)k}|X_1^{(i-1)k}S_0}\left(s_{(i-1)k}|x_1^{(i-1)k}, s_0\right) P_{X_1^{(i-1)k}|S_0}\left(x_1^{(i-1)k}|s_0\right) \\ &\quad= \sum_{x_1^{(i-1)k}} P_{X_{(i-1)k+1}^{ik}|X_1^{(i-1)k}S_0}\left(x_{(i-1)k}^{ik}|x_1^{(i-1)k}, s_0\right) P_{S_{(i-1)k}|X_1^{(i-1)k}S_0}\left(s_{(i-1)k}|x_1^{(i-1)k}, s_0\right) P_{X_1^{(i-1)k}|S_0}\left(x_1^{(i-1)k}|s_0\right) \end{aligned}$$

$X_{k_1+1}^k$ form a Markov chain, and (32) follows from the property of finite-state channels. Therefore, see (33), shown at the bottom of the next page, where (33) follows from Lemma 3.3.

Appendix F
Proof of Theorem 3.5

We shall compute a lower bound on the channel capacity by coupling the problem with a genie-aided system in which the state information $\{S_{mk}\}_{m=1}^{\infty}$ is provided to the receiver. Let $\left\{X_{(i-1)k+1}^{ik}\right\}_{i=1}^{\infty}$ be a stationary and memoryless vector

$$\begin{aligned}
&P_{X_{k_1+1}^k S_{k_1}}\left(x_{k_1+1}^k, s_{k_1}\right)\\
&= \sum_{x_1^{k_1}, s_0} P_{X_1^k S_{k_1} S_0}\left(x_1^k, s_{k_1}, s_0\right)\\
&= \sum_{x_1^{k_1}, s_0} P_{S_{k_1}|X_1^k S_0}\left(s_{k_1}|x_1^k, s_0\right) P_{X_1^k S_0}\left(x_1^k, s_0\right)\\
&= \sum_{x_1^{k_1}, s_0} P_{S_{k_1}|X_1^{k_1} S_0}\left(s_{k_1}|x_1^{k_1}, s_0\right) P_{X_1^k S_0}\left(x_1^k, s_0\right)\\
&= \sum_{x_1^{k_1}, s_0} P_{S_{k_1}|X_1^{k_1} S_0}\left(s_{k_1}|x_1^{k_1}, s_0\right) \sum_{i=1}^{m} \mu_i P_{X_1^k}^{(i)}\left(x_1^k\right) P_{S_0}^{(i)}(s_0)\\
&= \sum_{x_1^{k_1}, s_0} P_{S_{k_1}|X_1^{k_1} S_0}\left(s_{k_1}|x_1^{k_1}, s_0\right) \sum_{i=1}^{m} \mu_i P_{X_1^{k_1}}^{(i)}\left(x_1^{k_1}\right) P_{X_{k_1+1}^k|X_1^{k_1}}^{(i)}\left(x_{k_1+1}^k|x_1^{k_1}\right) P_{S_0}^{(i)}(s_0)\\
&= \sum_{i=1}^{m} \sum_{x_1^{k_1}} \left[\mu_i P_{X_1^{k_1}}^{(i)}\left(x_1^{k_1}\right)\right] \left[P_{X_{k_1+1}^k|X_1^{k_1}}^{(i)}\left(x_{k_1+1}^k|x_1^{k_1}\right)\right] \left[\sum_{s_0} P_{S_{k_1}|X_1^{k_1} S_0}\left(s_{k_1}|x_1^{k_1}, s_0\right) P_{S_0}^{(i)}(s_0)\right]
\end{aligned} \tag{29}$$

$$P_{S_{k_1}|X_1^{k_1}}^{(i)}\left(s_{k_1}|x_1^{k_1}\right) = \sum_{s_0} P_{S_{k_1}|X_1^{k_1} S_0}\left(s_{k_1}|x_1^{k_1}, s_0\right) P_{S_0}^{(i)}(s_0), \quad x_1^{k_1} \in \mathcal{X}^{k_1}, \;\; s_{k_1} \in \mathcal{S} \tag{30}$$

$$\begin{aligned}
&I\left(X_1^k; Y_1^k|S_0\right)\\
&\le I\left(X_1^k; Y_1^k, S_{k_1}|S_0\right)\\
&= H\left(Y_1^k, S_{k_1}|S_0\right) - H\left(Y_1^k, S_{k_1}|X_1^k, S_0\right)\\
&= H\left(Y_1^{k_1}|S_0\right) + H\left(Y_{k_1+1}^k, S_{k_1}|Y_1^{k_1}, S_0\right) - H\left(Y_1^{k_1}|X_1^k, S_0\right) - H\left(Y_{k_1+1}^k, S_{k_1}|Y_1^{k_1}, X_1^k, S_0\right)\\
&= H\left(Y_1^{k_1}|S_0\right) + H\left(S_{k_1}|Y_1^{k_1}, S_0\right) + H\left(Y_{k_1+1}^k|Y_1^{k_1}, S_{k_1}, S_0\right) - H\left(Y_1^{k_1}|X_1^k, S_0\right) - H\left(S_{k_1}|Y_1^{k_1}, X_1^k, S_0\right)\\
&\quad - H\left(Y_{k_1+1}^k|Y_1^{k_1}, X_1^k, S_{k_1}, S_0\right)\\
&\le H\left(Y_1^{k_1}|S_0\right) + H(S_{k_1}) + H\left(Y_{k_1+1}^k|S_{k_1}\right) - H\left(Y_1^{k_1}|X_1^k, S_0\right) - H\left(Y_{k_1+1}^k|Y_1^{k_1}, X_1^k, S_{k_1}, S_0\right)\\
&= H\left(Y_1^{k_1}|S_0\right) + H(S_{k_1}) + H\left(Y_{k_1+1}^k|S_{k_1}\right) - H\left(Y_1^{k_1}|X_1^{k_1}, S_0\right) - H\left(Y_{k_1+1}^k|Y_1^{k_1}, X_1^k, S_{k_1}, S_0\right) &(31)\\
&= H\left(Y_1^{k_1}|S_0\right) + H\left(S_{k_1}\right) + H\left(Y_{k_1+1}^k|S_{k_1}\right) - H\left(Y_1^{k_1}|X_1^{k_1}, S_0\right) - H\left(Y_{k_1+1}^k|X_{k_1}^k, S_{k_1}\right) &(32)\\
&= I\left(X_1^{k_1}; Y_1^{k_1}|S_0\right) + I\left(X_{k_1+1}^k; Y_{k_1+1}^k|S_{k_1}\right) + H(S_{k_1})
\end{aligned}$$

process independent of the initial state S_0. We have (34)–(36), shown at the bottom of the page, where (34) is because $Y_{(i-1)k+1}^{ik} - (S_{(i-1)k}, S_0) - \left(Y_1^{(i-1)k}, \{S_{mk}\}_{m=1}^{i-2}\right)$ form a Markov chain (which is further due to the fact that $X_{(i-1)k+1}^{ik}$ is independent of $\left(Y_1^{(i-1)k}, \{S_{mk}\}_{m=0}^{i-1}\right)$), and (35) follows from the property of finite-state channels.

Note that for any $x_{(i-1)k+1}^{ik} \in \mathcal{X}^k$, and $s_0, s_{(i-1)k} \in \mathcal{S}$

$$\begin{aligned} P_{X_{(i-1)k+1}^{ik} S_{(i-1)k}|S_0}&\left(x_{(i-1)k+1}^{ik}, s_{(i-1)k}|s_0\right) \\ &= P_{X_{(i-1)k+1}^{ik}}\left(x_{(i-1)k+1}^{ik}\right) P_{S_{(i-1)k}|S_0}\left(s_{(i-1)k}|s_0\right) \\ &= P_{X_{(i-1)k+1}^{ik}}\left(x_{(i-1)k+1}^{ik}\right) \end{aligned}$$

$$\begin{aligned} kC_k^U &= \max_s \max_{P_{X_1^k S_0} \in \mathrm{conv}(\mathcal{P}(\mathcal{X}^k)\times\mathcal{A}_s)} I\left(X_1^k; Y_1^k | S_0\right) + H(S_0) \\ &\leq \max_s \max_{P_{X_1^k S_0} \in \mathrm{conv}(\mathcal{P}(\mathcal{X}^k)\times\mathcal{A}_s)} I\left(X_1^{k_1}; Y_1^{k_1} | S_0\right) + H(S_0) + I\left(X_{k_1+1}^k; Y_{k_1+1}^k | S_{k_1}\right) + H(S_{k_1}) \\ &\leq \max_s \max_{P_{X_1^{k_1} S_0} \in \mathrm{conv}\left(\mathcal{P}(\mathcal{X}^{k_1})\times\mathcal{A}_s\right)} I\left(X_1^{k_1}; Y_1^{k_1} | S_0\right) + H(S_0) \\ &\quad + \max_s \max_{P_{X_{k_1+1}^k S_{k_1}} \in \mathrm{conv}\left(\mathcal{P}\left(\mathcal{X}^{k_2}\right)\times\mathcal{A}_s\right)} I\left(X_{k_1+1}^k; Y_{k_1+1}^k | S_{k_1}\right) + H(S_{k_1}) \\ &= k_1 C_{k_1}^U + k_2 C_{k_2}^U \end{aligned} \tag{33}$$

$$\begin{aligned} &I\left(X_1^{nk}; Y_1^{nk} | S_0 = s_0\right) \\ &\geq I\left(X_1^{nk}; Y_1^{nk}, \{S_{mk}\}_{m=1}^{n-1} | S_0 = s_0\right) - H\left(\{S_{mk}\}_{m=1}^{n-1} | S_0\right) \\ &= \sum_{i=1}^n \Bigg[H\left(Y_{(i-1)k+1}^{ik}, S_{(i-1)k} | Y_1^{(i-1)k}, \{S_{mk}\}_{m=1}^{i-2}, S_0 = s_0\right) \\ &\qquad - H\left(Y_{(i-1)k+1}^{ik}, S_{(i-1)k} | X_1^{nk}, Y_1^{(i-1)k}, \{S_{mk}\}_{m=1}^{i-2}, S_0 = s_0\right) - H\left(S_{(i-1)k} | \{S_{mk}\}_{m=1}^{i-2}, S_0 = s_0\right) \Bigg] \\ &= \sum_{i=1}^n \Bigg[H\left(Y_{(i-1)k+1}^{ik} | Y_1^{(i-1)k}, \{S_{mk}\}_{m=1}^{i-1}, S_0 = s_0\right) + H\left(S_{(i-1)k} | Y_1^{(i-1)k}, \{S_{mk}\}_{m=1}^{i-2}, S_0 = s_0\right) \\ &\qquad - H\left(Y_{(i-1)k+1}^{ik} | X_1^{nk}, Y_1^{(i-1)k}, \{S_{mk}\}_{m=1}^{i-1}, S_0 = s_0\right) - H\left(S_{(i-1)k} | X_1^{nk}, Y_1^{(i-1)k}, \{S_{mk}\}_{m=1}^{i-2}, S_0 = s_0\right) \\ &\qquad - H\left(S_{(i-1)k} | \{S_{mk}\}_{m=1}^{i-2}, S_0 = s_0\right) \Bigg] \\ &\geq \sum_{i=1}^n \Bigg[H\left(Y_{(i-1)k+1}^{ik} | Y_1^{(i-1)k}, \{S_{mk}\}_{m=1}^{i-1}, S_0 = s_0\right) - H\left(Y_{(i-1)k+1}^{ik} | X_1^{nk}, Y_1^{(i-1)k}, \{S_{mk}\}_{m=1}^{i-1}, S_0 = s_0\right) \\ &\qquad - H\left(S_{(i-1)k} | S_0 = s_0\right) \Bigg] \\ &= \sum_{i=1}^n \Bigg[H\left(Y_{(i-1)k+1}^{ik} | S_{(i-1)k}, S_0 = s_0\right) - H\left(Y_{(i-1)k+1}^{ik} | X_1^{nk}, Y_1^{(i-1)k}, \{S_{mk}\}_{m=1}^{i-1}, S_0 = s_0\right) \\ &\qquad - H(S_{(i-1)k} | S_0 = s_0) \Bigg] \qquad (34) \\ &= \sum_{i=1}^n \left[H\left(Y_{(i-1)k+1}^{ik} | S_{(i-1)k}, S_0 = s_0\right) - H\left(Y_{(i-1)k+1}^{ik} | X_{(i-1)k+1}^{ik}, S_{(i-1)k}, S_0 = s_0\right) - H\left(S_{(i-1)k} | S_0 = s_0\right) \right] \qquad (35) \\ &= \sum_{i=1}^n \left[I\left(X_{(i-1)k+1}^{ik}; Y_{(i-1)k+1}^{ik} | S_{(i-1)k}, S_0 = s_0\right) - H(S_{(i-1)k} | S_0 = s_0) \right] \qquad (36) \end{aligned}$$

$$\times \sum_{x_1^{(i-1)k}} P_{S_{(i-1)k} X_1^{(i-1)k}|S_0}\left(s_{(i-1)k}, x_1^{(i-1)k}|s_0\right)$$
$$= P_{X_{(i-1)k+1}^{ik}}\left(x_{(i-1)k+1}^{ik}\right) \sum_{x_1^{(i-1)k}} P_{X_1^{(i-1)k}}\left(x_1^{(i-1)k}\right)$$
$$\times P_{S_{(i-1)k}|X_1^{(i-1)k} S_0}\left(s_{(i-1)k}|x_1^{(i-1)k}, s_0\right)$$

which, together with the fact that $P_{S_{(i-1)k}|X_1^{(i-1)k} S_0}\left(\cdot|x_1^{(i-1)k}, s_0\right) \in \mathcal{G}_{s_0,(i-1)k}$ for any fixed $x_1^{(i-1)k} \in \mathcal{X}^{(i-1)k}$, implies

$$P_{X_{(i-1)k+1}^{ik} S_{(i-1)k}|S_0}(\cdot,\cdot|s_0) \in \mathcal{P}(\mathcal{X}^k) \times \mathrm{conv}(\mathcal{G}_{s_0,(i-1)k}). \tag{37}$$

In view of (36), (37), and the fact that $\lim_{i\to\infty} \delta(\mathcal{G}_{s_0,(i-1)k}, \mathcal{A}_{s_0}) = 0$, the proof is complete.

Appendix G
Proof of Lemma 3.7

Since $P_{S_0} \in \mathrm{conv}(\mathcal{A}_s)$, there exist $m \in \mathbb{N}$, $\mu_i \in [0,1]$, and $P_{S_0}^{(i)} \in \mathcal{A}_s$, $i = 1,\dots,m$, such that $\sum_{i=1}^m \mu_i = 1$ and $P_{S_0}(s_0) = \sum_{i=1}^m \mu_i P_{S_0}^{(i)}(s_0)$, $s_0 \in \mathcal{S}$. For any $s_k \in \mathcal{S}$

$$\begin{aligned}
P_{S_k}(s_k) &= \sum_{x_1^k, s_0} P_{X_1^k S_k S_0}\left(x_1^k, s_k, s_0\right)\\
&= \sum_{x_1^k, s_0} P_{S_k|X_1^k S_0}\left(s_k|x_1^k, s_0\right) P_{X_1^k S_0}\left(x_1^k, s_0\right)\\
&= \sum_{x_1^k, s_0} P_{S_k|X_1^k S_0}\left(s_k|x_1^k, s_0\right) P_{X_1^k}\left(x_1^k\right) P_{S_0}(s_0)\\
&= \sum_{x_1^k, s_0} P_{S_k|X_1^k S_0}\left(s_k|x_1^k, s_0\right) P_{X_1^{k_1}}\left(x_1^k\right)\\
&\quad \times \sum_{i=1}^m \mu_i P_{S_0}^{(i)}(s_0)\\
&= \sum_{i=1}^m \sum_{x_1^k} \mu_i P_{X_1^k}\left(x_1^k\right)\\
&\quad \times \left[\sum_{s_0} P_{S_k|X_1^k S_0}\left(s_k|x_1^k, s_0\right) P_{S_0}^{(i)}(s_0)\right]\\
&= \sum_{i=1}^m \sum_{x_1^k} \left[\mu_i P_{X_1^k}\left(x_1^k\right)\right] \left[P_{S_k|X_1^k}^{(i)}\left(s_k|x_1^k\right)\right]
\end{aligned}$$

where $P_{S_k|X_1^k}^{(i)}$ is defined in (30). Since $P_{S_k|X_1^k}^{(i)}\left(\cdot|x_1^k\right) \in \mathcal{A}_s$ for any fixed x_1^k, we have $P_{S_k} \in \mathrm{conv}(\mathcal{A}_s)$.

Appendix H
Proof of Theorem 3.8

Let $P_1 \in \mathcal{P}(\mathcal{X}^{k_1})$ and $P_2 \in \mathcal{P}(\mathcal{X}^{k_2})$ be the two input distributions that achieve $C_{k_1}^L$ and $C_{k_2}^L$, respectively. Let $X_1^{k_1}$ and $X_{k_1+1}^k$ be two independent random vectors with $P_{X_1^{k_1}} = P_1$ and $P_{X_{k_1+1}^k} = P_2$. Assume both $X_1^{k_1}$ and $X_{k_1+1}^k$ are independent of S_0. By this construction, if $P_{S_0} \in \mathrm{conv}(\mathcal{A}_s)$, then $P_{X_1^{k_1} S_0} \in \mathcal{P}(\mathcal{X}^{k_1}) \times \mathrm{conv}(\mathcal{A}_s)$; moreover, it follows from Lemma 3.7 that $P_{X_{k_1+1}^k S_{k_1}} \in \mathcal{P}(\mathcal{X}^{k_2}) \times \mathrm{conv}(\mathcal{A}_s)$. Note that [see (38), shown at the bottom of the page], where (38) is because $Y_{k_1+1}^k - S_{k_1} - \left(Y_1^{k_1}, S_0\right)$ form a Markov chain (which is further due to the fact that $X_{k_1+1}^k$ is independent of $\left(Y_1^{k_1}, S_{k_1}, S_0\right)$). Therefore, we have

$$\begin{aligned}
kC_k^L &\geq \min_s \min_{P_{S_0}\in \mathrm{conv}(\mathcal{A}_s)} I\left(X_1^k; Y_1^k|S_0\right) - H(S_0)\\
&\geq \min_s \min_{P_{S_0}\in \mathrm{conv}(\mathcal{A}_s)} I\left(X_1^{k_1}; Y_1^{k_1}|S_0\right) - H(S_0)\\
&\quad + \min_s \min_{P_{S_{k_1}}\in \mathrm{conv}(\mathcal{A}_s)} I\left(X_{k_1+1}^k; Y_{k_1+1}^k|S_{k_1}\right) - H(S_{k_1})\\
&= k_1 C_{k_1}^L + k_2 C_{k_2}^L.
\end{aligned}$$

The proof is complete.

$$\begin{aligned}
I\left(X_1^k; Y_1^k|S_0\right) &= H\left(Y_1^k|S_0\right) - H\left(Y_1^k|X_1^k, S_0\right)\\
&= H\left(Y_1^{k_1}|S_0\right) + H\left(Y_{k_1+1}^k|Y_1^{k_1}, S_0\right) - H\left(Y_1^k|X_1^k, S_0\right)\\
&\geq H\left(Y_1^{k_1}|S_0\right) + H\left(Y_{k_1+1}^k|Y_1^{k_1}, S_{k_1}, S_0\right) - H\left(Y_1^k|X_1^k, S_0\right)\\
&= H\left(Y_1^{k_1}|S_0\right) + H\left(Y_{k_1+1}^k|S_{k_1}\right) - H\left(Y_1^k|X_1^k, S_0\right)\\
&= H\left(Y_1^{k_1}|S_0\right) + H\left(Y_{k_1+1}^k|S_{k_1}\right) - H\left(Y_1^{k_1}|X_1^k, S_0\right) - H\left(Y_{k_1+1}^k|X_1^k, Y_1^{k_1}, S_0\right)\\
&\geq H\left(Y_1^{k_1}|S_0\right) + H\left(Y_{k_1+1}^k|S_{k_1}\right) - H\left(Y_1^{k_1}|X_1^{k_1}, S_0\right) - H\left(Y_{k_1+1}^k, S_{k_1}|X_1^k, Y_1^{k_1}, S_0\right)\\
&\geq I\left(X_1^{k_1}; Y_1^{k_1}|S_0\right) + H\left(Y_{k_1+1}^k|S_{k_1}\right) - H\left(Y_{k_1+1}^k|X_1^k, Y_1^{k_1}, S_{k_1}, S_0\right) - H(S_{k_1})\\
&= I\left(X_1^{k_1}; Y_1^{k_1}|S_0\right) + H\left(Y_{k_1+1}^k|S_{k_1}\right) - H\left(Y_{k_1+1}^k|X_{k_1+1}^k, S_{k_1}\right) - H(S_{k_1})\\
&= I\left(X_1^{k_1}; Y_1^{k_1}|S_0\right) + I\left(X_{k_1+1}^k; Y_{k_1+1}^k|S_{k_1}\right) - H(S_{k_1})
\end{aligned} \tag{38}$$

APPENDIX I
PROOF OF THEOREM 3.11

1) It is easy to verify that $\overline{C}_k^S \geq C_k^{S,U}$. Therefore, we only need to prove $C_k^{S,U} \geq \overline{C}^S$.
Note that [see (39), shown at the bottom of the page], where (39) follows from the property of finite-state channels. The rest of the proof is similar to that of Theorem 3.1, and thus is omitted.
2) It is easy to see that $\underline{C}_k^S \leq C_k^{S,L}$. Therefore, we only need to prove $C_k^{S,L} \leq \underline{C}^S$.
Let $\left\{X_{(i-1)k+1}^{ik}\right\}_{i=1}^{\infty}$ be a stationary and memoryless vector process independent of the initial state S_0. Note that [see (40)–(41), shown at the bottom of the page], where (40) is because $Y_{(i-1)k+1}^{ik} - (S_{(i-1)k}, S_0) - \left(Y_1^{(i-1)k}, S_1^{(i-1)k-1}\right)$ form a Markov chain (which is further due to the fact that $X_{(i-1)k+1}^{ik}$ is independent of $\left(Y_1^{(i-1)k}, S_0^{(i-1)k}\right)$), and (41) follows from the property of finite-state channels. The rest of the proof is similar to that of Theorem 3.5, and thus is omitted.

APPENDIX J
PROOF OF LEMMA 4.1

1) For any $P_{X_{1,1}^k X_{2,1}^k S_0} \in \Xi_k(\mathcal{B})$, there exist $m_1, m_2 \in \mathbb{N}$, $P_{X_{1,1}^k}^{(i)} \in \mathcal{P}\left(\mathcal{X}_1^k\right)$, $P_{X_{2,1}^k}^{(i)} \in \mathcal{P}\left(\mathcal{X}_2^k\right)$, $P_{S_0}^{(i,j)} \in \mathcal{B}$, $\mu_{1,i}, \mu_{2,j} \in [0,1]$, $i = 1, \ldots, m_1$, $j = 1, \ldots, m_2$, such that $\sum_{i=1}^{m_1} \mu_{1,i} = \sum_{j=1}^{m_2} \mu_{2,j} = 1$, and
$$P_{X_{1,1}^k X_{2,1}^k S_0}\left(x_{1,1}^k, x_{2,1}^k, s_0\right) = \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} \mu_{1,i}\mu_{2,j} P_{X_{1,1}^k}^{(i)}\left(x_{1,1}^k\right) P_{X_{2,1}^k}^{(j)}\left(x_{2,1}^k\right) P_{S_0}^{(i,j)}(s_0)$$
for all $x_{1,1}^k \in \mathcal{X}_1^k$, $x_{2,1}^k \in \mathcal{X}_2^k$, $s_0 \in \mathcal{S}$. Therefore, we have the equation shown at the bottom of the next page, which implies $P_{X_{1,1}^k X_{2,1}^k} \in \mathcal{P}\left(\mathcal{X}_1^k\right) \times \mathcal{P}\left(\mathcal{X}_2^k\right)$.
2) By Part 1) of Lemma 4.1, it suffices to show $\Big\{P_{X_{1,1}^k X_{2,1}^k S_0} \in \mathcal{P}\left(\mathcal{X}_1^k \times \mathcal{X}_2^k \times \mathcal{S}\right):$

$$\begin{aligned}
I&\left(X_1^{nk}; Y_1^{nk}, S_1^{nk} | S_0 = s_0\right)\\
&= H\left(Y_1^{nk}, S_1^{nk} | S_0 = s_0\right) - H\left(Y_1^{nk}, S_1^{nk} | X_1^{nk}, S_0 = s_0\right)\\
&= \sum_{i=1}^{n}\left[H\left(Y_{(i-1)k+1}^{ik}, S_{(i-1)k+1}^{ik} | Y_1^{(i-1)k}, S_1^{(i-1)k}, S_0 = s_0\right) - H\left(Y_{(i-1)k+1}^{ik}, S_{(i-1)k+1}^{ik} | X_1^{nk}, Y_1^{(i-1)k}, S_1^{(i-1)k}, S_0 = s_0\right)\right]\\
&\leq \sum_{i=1}^{n}\left[H\left(Y_{(i-1)k+1}^{ik}, S_{(i-1)k+1}^{ik} | S_{(i-1)k}, S_0 = s_0\right) - H\left(Y_{(i-1)k+1}^{ik}, S_{(i-1)k+1}^{ik} | X_1^{nk}, Y_1^{(i-1)k}, S_1^{(i-1)k}, S_0 = s_0\right)\right]\\
&= \sum_{i=1}^{n}\left[H\left(Y_{(i-1)k+1}^{ik}, S_{(i-1)k+1}^{ik} | S_{(i-1)k}, S_0 = s_0\right) - H\left(Y_{(i-1)k+1}^{ik}, S_{(i-1)k+1}^{ik} | X_{(i-1)k+1}^{ik}, S_{(i-1)k}, S_0 = s_0\right)\right] \qquad (39)\\
&= \sum_{i=1}^{n} I\left(X_{(i-1)k+1}^{ik}; Y_{(i-1)k+1}^{ik}, S_{(i-1)k+1}^{ik} | S_{(i-1)k}, S_0 = s_0\right)
\end{aligned}$$

$$\begin{aligned}
I&\left(X_1^{nk}; Y_1^{nk}, S_1^{nk} | S_0 = s_0\right)\\
&= H\left(Y_1^{nk}, S_1^{nk} | S_0 = s_0\right) - H\left(Y_1^{nk}, S_1^{nk} | X_1^{nk}, S_0 = s_0\right)\\
&= \sum_{i=1}^{n}\left[H\left(Y_{(i-1)k+1}^{ik}, S_{(i-1)k+1}^{ik} | Y_1^{(i-1)k}, S_1^{(i-1)k}, S_0 = s_0\right) - H\left(Y_{(i-1)k+1}^{ik}, S_{(i-1)k+1}^{ik} | X_1^{nk}, Y_1^{(i-1)k}, S_1^{(i-1)k}, S_0 = s_0\right)\right]\\
&= \sum_{i=1}^{n}\left[H\left(Y_{(i-1)k+1}^{ik}, S_{(i-1)k+1}^{ik} | S_{(i-1)k}, S_0 = s_0\right) - H\left(Y_{(i-1)k+1}^{ik}, S_{(i-1)k+1}^{ik} | X_1^{nk}, Y_1^{(i-1)k}, S_1^{(i-1)k}, S_0 = s_0\right)\right] \qquad (40)\\
&= \sum_{i=1}^{n}\left[H\left(Y_{(i-1)k+1}^{ik}, S_{(i-1)k+1}^{ik} | S_{(i-1)k}, S_0 = s_0\right) - H\left(Y_{(i-1)k+1}^{ik}, S_{(i-1)k+1}^{ik} | X_{(i-1)k+1}^{ik}, S_{(i-1)k}, S_0 = s_0\right)\right] \qquad (41)\\
&= \sum_{i=1}^{n} I\left(X_{(i-1)k+1}^{ik}; Y_{(i-1)k+1}^{ik}, S_{(i-1)k+1}^{ik} | S_{(i-1)k}, S_0 = s_0\right)
\end{aligned}$$

$P_{X_{1,1}^k X_{2,1}^k} \in \mathcal{P}\left(\mathcal{X}_1^k\right) \times \mathcal{P}\left(\mathcal{X}_2^k\right)\Big\} \subseteq \Xi_k(\mathcal{P}(\mathcal{S}))$. For any $P_{X_{1,1}^k X_{2,1}^k S_0}$ with $P_{X_{1,1}^k X_{2,1}^k} \in \mathcal{P}\left(\mathcal{X}_1^k\right) \times \mathcal{P}\left(\mathcal{X}_2^k\right)$

$$P_{X_{1,1}^k X_{2,1}^k S_0}\left(x_{1,1}^k, x_{2,1}^k, s_0\right) = \sum_{\widetilde{x}_{1,1}^k, \widetilde{x}_{2,1}^k} P_{X_{1,1}^k}\left(\widetilde{x}_{1,1}^k\right) P_{X_{2,1}^k}\left(\widetilde{x}_{2,1}^k\right) \times \mathbb{I}_{\widetilde{x}_{1,1}^k}\left(x_{1,1}^k\right) \mathbb{I}_{\widetilde{x}_{2,1}^k}\left(x_{2,1}^k\right) P_{S_0|X_{1,1}^k X_{2,1}^k}\left(s_0|\widetilde{x}_{1,1}^k, \widetilde{x}_{2,1}^k\right)$$

where $\mathbb{I}_{\widetilde{x}_{1,1}^k}(\cdot)$ and $\mathbb{I}_{\widetilde{x}_{2,1}^k}(\cdot)$ are the indicator functions. Note that $\mathbb{I}_{\widetilde{x}_{1,1}^k}(\cdot)\mathbb{I}_{\widetilde{x}_{2,1}^k}(\cdot)P_{S_0|X_{1,1}^k X_{2,1}^k}(\cdot|\widetilde{x}_{1,1}^k, \widetilde{x}_{2,1}^k) \in \mathcal{P}\left(\mathcal{X}_1^k\right) \times \mathcal{P}\left(\mathcal{X}_2^k\right) \times \mathcal{P}(\mathcal{S})$ for any fixed $\widetilde{x}_{1,1}^k \in \mathcal{X}_1^k$, $\widetilde{x}_{2,1}^k \in \mathcal{X}_2^k$. Therefore, we have $P_{X_{1,1}^k X_{2,1}^k S_0} \in \Xi_k(\mathcal{P}(\mathcal{S}))$.

3) For any $P_{X_{1,1}^k X_{2,1}^k S_0} \in \Xi_k(\mathcal{B})$, there exist $m_1, m_2 \in \mathbb{N}$, $P_{X_{1,1}^k}^{(i)} \in \mathcal{P}\left(\mathcal{X}_1^k\right)$, $P_{X_{2,1}^k}^{(i)} \in \mathcal{P}\left(\mathcal{X}_2^k\right)$, $P_{S_0}^{(i,j)} \in \mathcal{B}$, $\mu_{1,i}, \mu_{2,j} \in [0,1]$, $i = 1, \ldots, m_1$, $j = 1, \ldots, m_2$, such that $\sum_{i=1}^{m_1} \mu_{1,i} = \sum_{j=1}^{m_2} \mu_{2,j} = 1$, and

$$P_{X_{1,1}^k X_{2,1}^k S_0}\left(x_{1,1}^k, x_{2,1}^k, s_0\right) = \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} \mu_{1,i}\mu_{2,j} P_{X_{1,1}^k}^{(i)}\left(x_{1,1}^k\right) \times P_{X_{2,1}^k}^{(j)}\left(x_{2,1}^k\right) P_{S_0}^{(i,j)}(s_0)$$

for all $x_{1,1}^k \in \mathcal{X}_1^k$, $x_{2,1}^k \in \mathcal{X}_2^k$, $s_0 \in \mathcal{S}$.

Define $\Xi_k\left(\mathcal{B}, \mu_{2,j}, P_{X_{2,1}^k}^{(j)}, j = 1, \ldots, m_2\right) = \Big\{P'_{X_{1,1}^k X_{2,1}^k S_0} \in \mathcal{P}\left(\mathcal{X}_1^k \times \mathcal{X}_2^k \times \mathcal{S}\right)$: there exist $\widetilde{P}_{X_{1,1}^k} \in \mathcal{P}\left(\mathcal{X}_1^k\right), \widetilde{P}_{S_0}^{(j)} \in \mathcal{B}, j = 1, \ldots, m_2$, such that $P'_{X_{1,1}^k X_{2,1}^k S_0}\left(x_{1,1}^k, x_{2,1}^k, s_0\right) = \sum_{j=1}^{m_2} \mu_{2,j}\widetilde{P}_{X_{1,1}^k}\left(x_{1,1}^k\right) P_{X_{2,1}^k}^{(j)}\left(x_{2,1}^k\right) \widetilde{P}_{S_0}^{(j)}(s_0)$ for all $x_{1,1}^k \in \mathcal{X}_1^k, x_{2,1}^k \in \mathcal{X}_2^k, s_0 \in \mathcal{S}\Big\}$. It is easy to see that $P_{X_{1,1}^k X_{2,1}^k S_0} \in \mathrm{conv}\left(\Xi_k\left(\mathcal{B}, \mu_{2,j}, P_{X_{2,1}^k}^{(j)}, j = 1, \ldots, m_2\right)\right)$. Therefore, by Carathéodory's theorem, there exist $\widetilde{P}_{X_{1,1}^k}^{(i)} \in \mathcal{P}\left(\mathcal{X}_1^k\right)$, $\widetilde{P}_{S_0}^{(i,j)} \in \mathcal{B}$, $\widetilde{m}_1 \leq |\mathcal{X}_1|^k|\mathcal{X}_2|^k|\mathcal{S}|$, $\widetilde{\mu}_{1,i} \in [0,1]$, $i = 1, \ldots, \widetilde{m}_1$, such that

$$P_{X_{1,1}^k X_{2,1}^k S_0}\left(x_{1,1}^k, x_{2,1}^k, s_0\right) = \sum_{i=1}^{\widetilde{m}_1}\sum_{j=1}^{m_2} \widetilde{\mu}_{1,i}\mu_{2,j} \widetilde{P}_{X_{1,1}^k}^{(i)}\left(x_{1,1}^k\right) \times P_{X_{2,1}^k}^{(j)}\left(x_{2,1}^k\right) \widetilde{P}_{S_0}^{(i,j)}(s_0)$$

for all $x_{1,1}^k \in \mathcal{X}_1^k$, $x_{2,1}^k \in \mathcal{X}_2^k$, $s_0 \in \mathcal{S}$.

Define $\Xi_k\left(\mathcal{B}, \widetilde{\mu}_{1,i}, \widetilde{P}_{X_{1,1}^k}^{(i)}, i = 1, \ldots, \widetilde{m}_1\right) = \Big\{P''_{X_{1,1}^k X_{2,1}^k S_0} \in \mathcal{P}\left(\mathcal{X}_1^k \times \mathcal{X}_2^k \times \mathcal{S}\right)$: there exist $\widetilde{P}_{X_{2,1}^k} \in \mathcal{P}\left(\mathcal{X}_2^k\right), \overline{P}_{S_0}^{(i)} \in \mathcal{B}, i = 1, \ldots, \widetilde{m}_1$, such that $P''_{X_{1,1}^k X_{2,1}^k S_0}\left(x_{1,1}^k, x_{2,1}^k, s_0\right) = \sum_{i=1}^{\widetilde{m}_1} \widetilde{\mu}_{1,i}\widetilde{P}_{X_{1,1}^k}^{(i)}\left(x_{1,1}^k\right) \widetilde{P}_{X_{2,1}^k}\left(x_{2,1}^k\right) \overline{P}_{S_0}^{(i)}(s_0)$ for all $x_{1,1}^k \in \mathcal{X}_1^k, x_{2,1}^k \in \mathcal{X}_2^k, s_0 \in \mathcal{S}\Big\}$. Since $P_{X_{1,1}^k X_{2,1}^k S_0} \in \mathrm{conv}\left(\Xi_k\left(\mathcal{B}, \widetilde{\mu}_{1,j}, \widetilde{P}_{X_{1,1}^k}^{(i)}, i = 1, \ldots, \widetilde{m}_1\right)\right)$, by Carathéodory's theorem, there exist $\widetilde{P}_{X_{2,1}^k}^{(j)} \in \mathcal{P}\left(\mathcal{X}_2^k\right)$, $\overline{P}_{S_0}^{(i,j)} \in \mathcal{B}$, $\widetilde{m}_2 \leq |\mathcal{X}_1|^k|\mathcal{X}_2|^k|\mathcal{S}|$, $\widetilde{\mu}_{2,j} \in [0,1]$, $j = 1, \ldots, \widetilde{m}_2$, such that

$$P_{X_{1,1}^k X_{2,1}^k S_0}\left(x_{1,1}^k, x_{2,1}^k, s_0\right) = \sum_{i=1}^{\widetilde{m}_1}\sum_{j=1}^{\widetilde{m}_2} \widetilde{\mu}_{1,i}\widetilde{\mu}_{2,j} \widetilde{P}_{X_{1,1}^k}^{(i)}\left(x_{1,1}^k\right) \times \widetilde{P}_{X_{2,1}^k}^{(j)}\left(x_{2,1}^k\right) \overline{P}_{S_0}^{(i,j)}(s_0)$$

for all $x_{1,1}^k \in \mathcal{X}_1^k$, $x_{2,1}^k \in \mathcal{X}_2^k$, $s_0 \in \mathcal{S}$. Assuming $\widetilde{m}_1 < \widetilde{m}_2$, we can write

$$P_{X_{1,1}^k X_{2,1}^k S_0}\left(x_{1,1}^k, x_{2,1}^k, s_0\right) = \sum_{i=1}^{\widetilde{m}_2}\sum_{j=1}^{\widetilde{m}_2} \widetilde{\mu}_{1,i}\widetilde{\mu}_{2,j} \widetilde{P}_{X_{1,1}^k}^{(i)}\left(x_{1,1}^k\right) \times \widetilde{P}_{X_{2,1}^k}^{(j)}\left(x_{2,1}^k\right) \overline{P}_{S_0}^{(i,j)}(s_0)$$

where $\widetilde{\mu}_{1,i} = 0$, and $\widetilde{P}_{X_{1,1}^k}^{(i)}$, $\overline{P}_{S_0}^{(i,j)}$ can be arbitrary probability distributions from $\mathcal{P}\left(\mathcal{X}_1^k\right)$ and $\mathcal{B}$, $i = \widetilde{m}_1 + 1, \ldots, \widetilde{m}_2$. Therefore, there is no loss of generality to assume $m_1 = m_2$ and $\max(m_1, m_2) \leq |\mathcal{X}_1|^k|\mathcal{X}_2|^k|\mathcal{S}|$ in the definition of $\Xi_k(\mathcal{B})$.

4) For any $P_{X_{1,1}^k X_{2,1}^k S_0} \in \Xi_k(\mathrm{conv}(\mathcal{B}))$, there exist $m_1, m_2, m_3 \in \mathbb{N}$, $P_{X_{1,1}^k}^{(i)} \in \mathcal{P}\left(\mathcal{X}_1^k\right)$, $P_{X_{2,1}^k}^{(i)} \in \mathcal{P}\left(\mathcal{X}_2^k\right)$, $P_{S_0}^{(i,j,l)} \in \mathcal{B}$, $\mu_{1,i}, \mu_{2,j}, \mu_{3,l} \in [0,1]$, $i = 1, \ldots, m_1$, $j = 1, \ldots, m_2$, $l = 1, \ldots, m_3$, such that $\sum_{i=1}^{m_1} \mu_{1,i} = \sum_{j=1}^{m_2} \mu_{2,i} = \sum_{l=1}^{m_3} \mu_{3,i} = 1$ and

$$P_{X_{1,1}^k X_{2,1}^k S_0}\left(x_{1,1}^k, x_{2,1}^k, s_0\right) = \sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\sum_{l=1}^{m_3} \mu_{1,i}\mu_{2,j}\mu_{3,l} \times P_{X_{1,1}^k}^{(i)}\left(x_{1,1}^k\right) P_{X_{2,1}^k}^{(j)}\left(x_{2,1}^k\right) P_{S_0}^{(i,j,l)}(s_0)$$

$$P_{X_{1,1}^k}\left(x_{1,1}^k\right) = \sum_{x_{2,1}^k, s_0} P_{X_{1,1}^k X_{2,1}^k S_0}\left(x_{1,1}^k, x_{2,1}^k, s_0\right) = \sum_{i=1}^{m_1} \mu_{1,i} P_{X_{1,1}^k}^{(i)}\left(x_{1,1}^k\right)$$

$$P_{X_{2,1}^k}\left(x_{2,1}^k\right) = \sum_{x_{1,1}^k, s_0} P_{X_{1,1}^k X_{2,1}^k S_0}\left(x_{1,1}^k, x_{2,1}^k, s_0\right) = \sum_{j=1}^{m_2} \mu_{2,j} P_{X_{2,1}^k}^{(j)}\left(x_{2,1}^k\right)$$

$$P_{X_{1,1}^k X_{2,1}^k}\left(x_{1,1}^k, x_{2,1}^k\right) = \sum_{s_0} P_{X_{1,1}^k X_{2,1}^k S_0}\left(x_{1,1}^k, x_{2,1}^k, s_0\right) = \left(\sum_{i=1}^{m_1} \mu_{1,i} P_{X_{1,1}^k}^{(i)}\left(x_{1,1}^k\right)\right)\left(\sum_{j=1}^{m_2} \mu_{2,j} P_{X_{2,1}^k}^{(j)}\left(x_{2,1}^k\right)\right)$$

for all $x_{1,1}^k \in \mathcal{X}_1^k$, $x_{2,1}^k \in \mathcal{X}_2^k$, $s_0 \in \mathcal{S}$. For $j = 1, \ldots, m_2$, $l = 1, \ldots, m_3$, let $j' = m_3(j-1) + l$, $\widetilde{\mu}_{2,j'} = \mu_{2,j}\mu_{3,l}$, $\widetilde{P}_{X_{2,1}^k}^{(j')} = P_{X_{2,1}^k}^{(j)}$, and $\widetilde{P}_{S_0}^{(i,j')} = P_{S_0}^{(i,j,l)}$. We have

$$P_{X_{1,1}^k X_{2,1}^k S_0}\left(x_{1,1}^k, x_{2,1}^k, s_0\right) = \sum_{i=1}^{m_1} \sum_{j'=1}^{m_2 m_3} \mu_{1,i} \widetilde{\mu}_{2,j'} P_{X_{1,1}^k}^{(i)}\left(x_{1,1}^k\right) \times \widetilde{P}_{X_{2,1}^k}^{(j')}\left(x_{2,1}^k\right) \widetilde{P}_{S_0}^{(i,j')}(s_0)$$

for all $x_{1,1}^k \in \mathcal{X}_1^k$, $x_{2,1}^k \in \mathcal{X}_2^k$, $s_0 \in \mathcal{S}$, which implies $P_{X_{1,1}^k X_{2,1}^k S_0} \in \Xi_k(\mathcal{B})$.

5) Note that for any $x_{1,k_1+1}^k \in \mathcal{X}_1^{k_2}$, $x_{2,k_1+1}^k \in \mathcal{X}_2^{k_2}$, and $s_{k_1} \in \mathcal{S}$, see (42), shown at the bottom of the page, where (42) follows from the fact that $S_{k_1} - \left(X_{1,1}^{k_1}, X_{2,1}^{k_1}, S_0\right) - \left(X_{1,k_1+1}^k, X_{2,k_1+1}^k\right)$ form a Markov chain. For any $x_{1,1}^{k_1} \in \mathcal{X}_1^{k_1}$, $x_{2,1}^{k_1} \in \mathcal{X}_2^{k_1}$, and $x_{2,1}^{k_1} \in \mathcal{X}_2^{k_1}$, $s_{k_1} \in \mathcal{S}$, define

$$P_{S_{k_1}|X_{1,1}^{k_1} X_{2,1}^{k_1}}^{(i,j)}\left(s_{k_1}|x_{1,1}^{k_1}, x_{2,1}^{k_1}\right) = \sum_{s_0} P_{S_{k_1}|X_{1,1}^{k_1} X_{2,1}^{k_1} S_0}\left(s_{k_1}|x_{1,1}^{k_1}, x_{2,1}^{k_1}, s_0\right) P_{S_0}^{(i,j)}(s_0).$$

In view of the fact that that $P_{S_0}^{(i,j)} \in \mathcal{A}_s$ and $\mathcal{A}_s\mathcal{T} \subseteq \mathcal{A}_s$, we have $P_{S_{k_1}|X_{1,1}^{k_1}, X_{2,1}^{k_1}}^{(i,j)}\left(\cdot|x_{1,1}^{k_1}, x_{2,1}^{k_1}\right) \in \mathcal{A}_s\mathcal{T}^{k_1} \subseteq \mathcal{A}_s$ for any fixed $x_{1,1}^{k_1}$, $x_{2,1}^{k_1}$. It is also easy to see that $P_{X_{1,k_1+1}^k|X_{1,1}^{k_1}}^{(i)}\left(\cdot|x_{1,1}^{k_1}\right) \in \mathcal{P}\left(\mathcal{X}_1^{k_2}\right)$, $P_{X_{2,k_1+1}^k|X_{2,1}^{k_1}}^{(j)}\left(\cdot|x_{2,1}^{k_1}\right) \in \mathcal{P}\left(\mathcal{X}_2^{k_2}\right)$ for any fixed $x_{1,1}^{k_1}$, $x_{2,1}^{k_1}$. Therefore, we have $P_{X_{1,k_1+1}^k X_{2,k_1+1}^k S_{k_1}} \in \Xi_{k_2}(\mathcal{A}_s)$. For any $x_{1,1}^{k_1} \in \mathcal{X}_1^{k_1}$ and $s_0 \in \mathcal{S}$

$$\begin{aligned}
P_{X_{1,1}^{k_1} X_{2,1}^{k_1} S_0}&\left(x_{1,1}^{k_1}, x_{2,1}^{k_1}, s_0\right) \\
&= \sum_{x_{1,k_1+1}^k, x_{2,k_1+1}^k} P_{X_{1,1}^k X_{2,1}^k S_0}\left(x_{1,1}^k, x_{2,1}^k, s_0\right) \\
&= \sum_{x_{1,k_1+1}^k, x_{1,k_1+1}^k} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \mu_{1,i}\mu_{2,j} P_{X_{1,1}^k}^{(i)}\left(x_{1,1}^k\right) \times P_{X_{2,1}^k}^{(j)}\left(x_{2,1}^k\right) P_{S_0}^{(i,j)}(s_0) \\
&= \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \mu_{1,i}\mu_{2,j} P_{X_{1,1}^{k_1}}^{(i)}\left(x_{1,1}^{k_1}\right) P_{X_{2,1}^{k_1}}^{(j)}\left(x_{2,1}^{k_1}\right) P_{S_0}^{(i,j)}(s_0).
\end{aligned}$$

Therefore, we have $P_{X_{1,1}^{k_1} X_{2,1}^{k_1} S_0} \in \Xi_{k_1}(\mathcal{A}_s)$.

6) Since $P_{S_0} \in \text{conv}(\mathcal{A}_s)$, there exist $m \in \mathbb{N}$, $\mu_i \in [0,1]$, and $P_{S_0}^{(i)} \in \mathcal{A}_s$, $i = 1, \ldots, m$, such that $\sum_{i=1}^m \mu_i = 1$ and $P_{S_0}(s_0) = \sum_{i=1}^m \mu_i P_{S_0}^{(i)}(s_0)$, $s_0 \in \mathcal{S}$. For any $s_k \in \mathcal{S}$

$$\begin{aligned}
P_{S_k}(s_k) &= \sum_{x_{1,1}^k, x_{2,1}^k, s_0} P_{X_{1,1}^k X_{2,1}^k S_k S_0}\left(x_{1,1}^k, x_{2,1}^k, s_k, s_0\right) \\
&= \sum_{x_{1,1}^k, x_{2,1}^k, s_0} P_{S_k|X_{1,1}^k X_{2,1}^k S_0}\left(s_k|x_{1,1}^k, x_{2,1}^k, s_0\right)
\end{aligned}$$

$$\begin{aligned}
P_{X_{1,k_1+1}^k X_{2,k_1+1}^k S_{k_1}}&\left(x_{1,k_1+1}^k, x_{2,k_1+1}^k, s_{k_1}\right) \\
&= \sum_{x_{1,1}^{k_1}, x_{2,1}^{k_1}, s_0} P_{X_{1,1}^k X_{2,1}^k S_{k_1} S_0}\left(x_{1,1}^k, x_{2,1}^k, s_{k_1}, s_0\right) \\
&= \sum_{x_{1,1}^{k_1}, x_{2,1}^{k_1}, s_0} P_{S_{k_1}|X_{1,1}^k X_{2,1}^k S_0}\left(s_{k_1}|x_{1,1}^k, x_{2,1}^k, s_0\right) P_{X_{1,1}^k X_{2,1}^k S_0}\left(x_{1,1}^k, x_{2,1}^k, s_0\right) \\
&= \sum_{x_{1,1}^{k_1}, x_{2,1}^{k_1}, s_0} P_{S_{k_1}|X_{1,1}^{k_1} X_{2,1}^{k_1} S_0}\left(s_{k_1}|x_{1,1}^{k_1}, x_{2,1}^{k_1}, s_0\right) P_{X_{1,1}^k X_{2,1}^k S_0}\left(x_{1,1}^k, x_{2,1}^k, s_0\right) \\
&= \sum_{x_{1,1}^{k_1}, x_{2,1}^{k_1}, s_0} P_{S_{k_1}|X_{1,1}^{k_1} X_{2,1}^{k_1} S_0}\left(s_{k_1}|x_{1,1}^{k_1}, x_{2,1}^{k_1}, s_0\right) \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \mu_{1,i}\mu_{2,j} P_{X_{1,1}^k}^{(i)}\left(x_{1,1}^k\right) P_{X_{2,1}^k}^{(j)}\left(x_{2,1}^k\right) P_{S_0}^{(i,j)}(s_0) \\
&= \sum_{x_{1,1}^{k_1}, x_{2,1}^{k_1}, s_0} P_{S_{k_1}|X_{1,1}^{k_1} X_{2,1}^{k_1} S_0}\left(s_{k_1}|x_{1,1}^{k_1}, x_{2,1}^{k_1}, s_0\right) \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \mu_{1,i}\mu_{2,j} P_{X_{1,1}^{k_1}}^{(i)}\left(x_{1,1}^{k_1}\right) P_{X_{1,k_1+1}^k|X_{1,1}^{k_1}}^{(i)}\left(x_{1,k_1+1}^k|x_{1,1}^{k_1}\right) \\
&\quad \times P_{X_{2,1}^{k_1}}^{(j)}\left(x_{2,1}^{k_1}\right) P_{X_{2,k_1+1}^k|X_{2,1}^{k_1}}^{(j)}\left(x_{2,k_1+1}^k|x_{2,1}^{k_1}\right) P_{S_0}^{(i,j)}(s_0) \\
&= \sum_{i=1}^{m_1} \sum_{x_{1,1}^{k_1}} \sum_{j=1}^{m_2} \sum_{x_{2,1}^{k_1}} \left[\mu_{1,i} P_{X_{1,1}^{k_1}}^{(i)}\left(x_{1,1}^{k_1}\right)\right] \left[\mu_{2,j} P_{X_{2,1}^{k_1}}^{(j)}\left(x_{2,1}^{k_1}\right)\right] \left[P_{X_{1,k_1+1}^k|X_{1,1}^{k_1}}^{(i)}\left(x_{1,k_1+1}^k|x_{1,1}^{k_1}\right)\right] \\
&\quad \times \left[P_{X_{2,k_1+1}^k|X_{2,1}^{k_1}}^{(j)}\left(x_{2,k_1+1}^k|x_{2,1}^{k_1}\right)\right] \left[\sum_{s_0} P_{S_{k_1}|X_{1,1}^{k_1} X_{2,1}^{k_1} S_0}\left(s_{k_1}|x_{1,1}^{k_1}, x_{2,1}^{k_1}, s_0\right) P_{S_0}^{(i,j)}(s_0)\right]
\end{aligned} \tag{42}$$

$$\times P_{X_{1,1}^k X_{2,1}^k S_0}\left(x_{1,1}^k, x_{2,1}^k, s_0\right)$$
$$= \sum_{x_{1,1}^k, x_{2,1}^k, s_0} P_{S_k|X_{1,1}^k X_{2,1}^k S_0}\left(s_k|x_{1,1}^k, x_{2,1}^k, s_0\right)$$
$$\times P_{X_{1,1}^k}\left(x_{1,1}^k\right) P_{X_{2,1}^k}\left(x_{2,1}^k\right) P_{S_0}(s_0)$$
$$= \sum_{x_{1,1}^k, x_{2,1}^k, s_0} P_{S_k|X_{1,1}^k X_{2,1}^k S_0}\left(s_k|x_{1,1}^k, x_{2,1}^k, s_0\right)$$
$$\times P_{X_{1,1}^k}\left(x_{1,1}^k\right) P_{X_{2,1}^k}\left(x_{2,1}^k\right) \sum_{i=1}^m \mu_i P_{S_0}^{(i)}(s_0)$$
$$= \sum_{i=1}^m \sum_{x_{1,1}^k, x_{2,1}^k} \mu_i P_{X_{1,1}^k}\left(x_{1,1}^k\right) P_{X_{2,1}^k}\left(x_{2,1}^k\right)$$
$$\times \left[\sum_{s_0} P_{S_k|X_{1,1}^k X_{2,1}^k S_0}\left(s_k|x_{1,1}^k, x_{2,1}^k, s_0\right) P_{S_0}^{(i)}(s_0)\right]$$
$$= \sum_{i=1}^m \sum_{x_{1,1}^k, x_{2,1}^k} \left[\mu_i P_{X_{1,1}^k}\left(x_{1,1}^k\right) P_{X_{2,1}^k}\left(x_{2,1}^k\right)\right]$$
$$\times \left[P_{S_k|X_{1,1}^k X_{2,1}^k}^{(i)}\left(s_k|x_{1,1}^k, x_{2,1}^k\right)\right]$$

where we have defined the first equation shown at the bottom of the page. Since $P_{S_k|X_{1,1}^k X_{2,1}^k}^{(i)}\left(\cdot|x_1^k, x_{2,1}^k\right) \in \mathcal{A}_s$ for any fixed $x_{1,1}^k$ and $x_{2,1}^k$, we have $P_{S_k} \in \text{conv}(\mathcal{A}_s)$.

Appendix K
Proof of Theorem 4.2

1) Since $\overline{\mathcal{R}}_k \supseteq \mathcal{R}_k^O$ is obvious, it suffices to show $\mathcal{R}_k^O \supseteq \widetilde{\mathcal{R}}$. For $\alpha \in [0,1]$, define the supporting lines

$$\widetilde{\zeta}(\alpha) = \max_{(R_1,R_2)\in\widetilde{\mathcal{R}}} (1-\alpha)R_1 + \alpha R_2$$
$$\widetilde{\zeta}_k(\alpha) = \max_{(R_1,R_2)\in\widetilde{\mathcal{R}}_k} (1-\alpha)R_1 + \alpha R_2$$
$$\zeta_k^O(\alpha) = \max_{(R_1,R_2)\in\mathcal{R}_k^O} (1-\alpha)R_1 + \alpha R_2.$$

It is clear that

$$\widetilde{\zeta}(\alpha) = \limsup_{k\to\infty} \widetilde{\zeta}_k(\alpha), \quad \alpha \in [0,1].$$

It can be shown that for any $P_{X_{1,1}^k X_{2,1}^k} \in \mathcal{P}\left(\mathcal{X}_1^k\right)\times\mathcal{P}\left(\mathcal{X}_2^k\right)$ and $s_0 \in \mathcal{S}$, $\mathcal{R}_k\left(P_{X_{1,1}^k X_{2,1}^k}, s_0\right)$ is a polymatroid (see

---

$$P_{S_k|X_{1,1}^k X_{2,1}^k}^{(i)}\left(s_k|x_{1,1}^k, x_{2,1}^k\right) = \sum_{s_0} P_{S_k|X_{1,1}^k X_{2,1}^k S_0}\left(s_k|x_{1,1}^k, x_{2,1}^k, s_0\right) P_{S_0}^{(i)}(s_0), \quad x_{1,1}^k \in \mathcal{X}_1^k,\ x_{2,1}^k \in \mathcal{X}_2^k,\ s_k \in \mathcal{S}$$

$$\widetilde{\zeta}_k(\alpha) = \max_{s_0} \max_{P_{X_{1,1}^k X_{2,1}^k}\in\mathcal{P}(\mathcal{X}_1^k)\times\mathcal{P}(\mathcal{X}_2^k)} \frac{1-\alpha}{k} I\left(X_{1,1}^k; Y_1^k | X_{2,1}^k, S_0 = s_0\right) + \frac{\alpha}{k} I\left(X_{2,1}^k; Y_1^k | S_0 = s_0\right)$$
$$\zeta_k^O(\alpha) = \max_{s_0} \max_{P_{X_{1,1}^k X_{2,1}^k S_0}\in\Xi_k(\mathcal{A}_{s_0})} \frac{1-\alpha}{k} I\left(X_{1,1}^k; Y_1^k | X_{2,1}^k, S_0\right) + \frac{\alpha}{k} I\left(X_{2,1}^k; Y_1^k | S_0\right) + \frac{1}{k} H(S_0)$$

$$\widetilde{\zeta}_k(\alpha) = \max_{s_0} \max_{P_{X_{1,1}^k X_{2,1}^k}\in\mathcal{P}(\mathcal{X}_1^k)\times\mathcal{P}(\mathcal{X}_2^k)} \frac{1-\alpha}{k} I(X_{1,1}^k; Y_1^k | S_0 = s_0) + \frac{\alpha}{k} I\left(X_{2,1}^k; Y_1^k | X_{1,1}^k, S_0 = s_0\right)$$
$$\zeta_k^O(\alpha) = \max_{s_0} \max_{P_{X_{1,1}^k X_{2,1}^k S_0}\in\Xi_k(\mathcal{A}_{s_0})} \frac{1-\alpha}{k} I\left(X_{1,1}^k; Y_1^k | S_0\right) + \frac{\alpha}{k} I\left(X_{2,1}^k; Y_1^k | X_{1,1}^k, S_0\right) + \frac{1}{k} H(S_0)$$

$$I\left(X_{1,1}^{nk}; Y_1^{nk} | X_{2,1}^{nk}, S_0 = s_0\right) \le \sum_{i=1}^n \left[I\left(X_{1,(i-1)k+1}^{ik}; Y_{(i-1)k+1}^{ik} | X_{2,(i-1)k+1}^{ik}, S_{(i-1)k}, S_0 = s_0\right) + H(S_{(i-1)k} | S_0 = s_0)\right]$$
$$I\left(X_{1,1}^{nk}, X_{2,1}^{nk}; Y_1^{nk} | S_0 = s_0\right) \le \sum_{i=1}^n \left[I\left(X_{1,(i-1)k+1}^{ik}, X_{2,(i-1)k+1}^{ik}; Y_{(i-1)k+1}^{ik} | S_{(i-1)k}, S_0 = s_0\right) + H\left(S_{(i-1)k} | S_0 = s_0\right)\right]$$

[21] for the definition of polymatroid); moreover, for any $P_{X_{1,1}^k X_{2,1}^k S_0} \in \Xi_k(\mathcal{A}_{s_0})$, $\mathcal{R}_k^O\left(P_{X_{1,1}^k X_{2,1}^k S_0}\right)$ is also a polymatroid. In view of the polymatroid structure of $\mathcal{R}_k\left(P_{X_{1,1}^k X_{2,1}^k}, s_0\right)$ and $\mathcal{R}_k^O\left(P_{X_{1,1}^k X_{2,1}^k S_0}\right)$, we have, for $\alpha \in \left[0, \frac{1}{2}\right]$, see the second equation, shown at the bottom of the previous page, and for $\alpha \in \left(\frac{1}{2}, 1\right]$, see the third equation, shown at the bottom of the previous page.

By similar steps as in the proof of Theorem 3.1, one can easily verify that [see the fourth equation, shown at the bottom of the previous page].

Note that if $P_{X_{1,1}^{nk} X_{2,1}^{nk}|S_0}(\cdot,\cdot|s_0) \in \mathcal{P}\left(\mathcal{X}_1^{nk}\right) \times \mathcal{P}\left(\mathcal{X}_2^{nk}\right)$, then for any $x_{1,(i-1)k+1}^{ik} \in \mathcal{X}_1^k$, $x_{2,(i-1)k+1}^{ik} \in \mathcal{X}_2^k$, and $s_0, s_{(i-1)k} \in \mathcal{S}$, see the first equation, shown at the bottom of the page, where the last equality follows from the fact that

$$
\begin{aligned}
&P_{X_{1,(i-1)k+1}^{ik} X_{2,(i-1)k+1}^{ik} S_{(i-1)k}|S_0}\left(x_{1,(i-1)k+1}^{ik}, x_{2,(i-1)k+1}^{ik}, s_{(i-1)k}|s_0\right)\\
&\quad= \sum_{x_{1,1}^{(i-1)k}, x_{2,1}^{(i-1)k}} P_{X_{1,1}^{ik} X_{2,1}^{ik} S_{(i-1)k}|S_0}\left(x_{1,1}^{ik}, x_{2,1}^{ik}, s_{(i-1)k}|s_0\right)\\
&\quad= \sum_{x_{1,1}^{(i-1)k}, x_{2,1}^{(i-1)k}} P_{X_{1,1}^{ik} X_{2,1}^{ik}|S_0}\left(x_{1,1}^{ik}, x_{2,1}^{ik}|s_0\right) P_{S_{(i-1)k}|X_{1,1}^{ik} X_{2,1}^{ik} S_0}\left(s_{(i-1)k}|x_{1,1}^{ik}, x_{2,1}^{ik}, s_0\right)\\
&\quad= \sum_{x_{1,1}^{(i-1)k}, x_{2,1}^{(i-1)k}} P_{X_{1,1}^{ik}|S_0}\left(x_{1,1}^{ik}|s_0\right) P_{X_{2,1}^{ik}|S_0}\left(x_{2,1}^{ik}|s_0\right) P_{S_{(i-1)k}|X_{1,1}^{ik} X_{2,1}^{ik} S_0}\left(s_{(i-1)k}|x_{1,1}^{ik}, x_{2,1}^{ik}, s_0\right)\\
&\quad= \sum_{x_{1,1}^{(i-1)k}, x_{2,1}^{(i-1)k}} P_{X_{1,1}^{(i-1)k}|S_0}\left(x_{1,1}^{(i-1)k}|s_0\right) P_{X_{2,1}^{(i-1)k}|S_0}\left(x_{2,1}^{(i-1)k}|s_0\right) P_{X_{1,(i-1)k+1}^{ik}|X_{1,1}^{(i-1)k} S_0}\left(x_{1,(i-1)k+1}^{ik}|x_{1,1}^{(i-1)k}, s_0\right)\\
&\qquad \times P_{X_{2,(i-1)k+1}^{ik}|X_{2,1}^{(i-1)k} S_0}\left(x_{2,(i-1)k+1}^{ik}|x_{2,1}^{(i-1)k}, s_0\right) P_{S_{(i-1)k}|X_{1,1}^{ik} X_{2,1}^{ik} S_0}\left(s_{(i-1)k}|x_{1,1}^{ik}, x_{2,1}^{ik}, s_0\right)\\
&\quad= \sum_{x_{1,1}^{(i-1)k}, x_{2,1}^{(i-1)k}} P_{X_{1,1}^{(i-1)k}|S_0}\left(x_{1,1}^{(i-1)k}|s_0\right) P_{X_{2,1}^{(i-1)k}|S_0}\left(x_{2,1}^{(i-1)k}|s_0\right) P_{X_{1,(i-1)k+1}^{ik}|X_{1,1}^{(i-1)k} S_0}\left(x_{1,(i-1)k+1}^{ik}|x_{1,1}^{(i-1)k}, s_0\right)\\
&\qquad \times P_{X_{2,(i-1)k+1}^{ik}|X_{2,1}^{(i-1)k} S_0}\left(x_{2,(i-1)k+1}^{ik}|x_{2,1}^{(i-1)k}, s_0\right) P_{S_{(i-1)k}|X_{1,1}^{(i-1)k} X_{2,1}^{(i-1)k} S_0}\left(s_{(i-1)k}|x_{1,1}^{(i-1)k}, x_{2,1}^{(i-1)k}, s_0\right)
\end{aligned}
$$

$$
\begin{aligned}
\widetilde{\zeta}(\alpha) &= \limsup_{n\to\infty} \widetilde{\zeta}_{nk}(\alpha)\\
&= \limsup_{n\to\infty} \max_{s_0} \max_{P_{X_{1,1}^{nk} X_{2,1}^{nk}} \in \mathcal{P}\left(\mathcal{X}_1^{nk}\right)\times\mathcal{P}\left(\mathcal{X}_2^{nk}\right)} \frac{1-\alpha}{nk} I\left(X_{1,1}^{nk}; Y_1^{nk}|X_{2,1}^{nk}, S_0 = s_0\right) + \frac{\alpha}{nk} I\left(X_{2,1}^{nk}; Y_1^{nk}|S_0 = s_0\right)\\
&= \limsup_{n\to\infty} \max_{s_0} \max_{P_{X_{1,1}^{nk} X_{2,1}^{nk}} \in \mathcal{P}\left(\mathcal{X}_1^{nk}\right)\times\mathcal{P}\left(\mathcal{X}_2^{nk}\right)} \frac{1-2\alpha}{nk} I\left(X_{1,1}^{nk}; Y_1^{nk}|X_{2,1}^{nk}, S_0 = s_0\right) + \frac{\alpha}{nk} I\left(X_{1,1}^{nk}, X_{2,1}^{nk}; Y_1^{nk}|S_0 = s_0\right)\\
&\le \limsup_{i\to\infty} \max_{s_0} \max_{P_{X_{1,(i-1)k+1}^{ik} X_{2,(i-1)k+1}^{ik} S_{(i-1)k}|S_0}(\cdot,\cdot,\cdot|s_0) \in \Xi_k(\mathcal{G}_{s_0,(i-1)k})}\\
&\quad \frac{1-2\alpha}{k}\left[I\left(X_{1,(i-1)k+1}^{ik}; Y_{(i-1)k+1}^{ik}|X_{2,(i-1)k+1}^{ik}, S_{(i-1)k}, S_0 = s_0\right) + H(S_{(i-1)k}|S_0 = s_0)\right]\\
&\qquad + \frac{\alpha}{k}\left[I\left(X_{1,(i-1)k+1}^{ik}, X_{2,(i-1)k+1}^{ik}; Y_{(i-1)k+1}^{ik}|S_{(i-1)k}, S_0 = s_0\right) + H(S_{(i-1)k}|S_0 = s_0)\right]\\
&\le \limsup_{i\to\infty} \max_{s_0} \max_{P_{X_{1,(i-1)k+1}^{ik} X_{2,(i-1)k+1}^{ik} S_{(i-1)k}|S_0}(\cdot,\cdot,\cdot|s_0) \in \Xi_k(\mathcal{G}_{s_0,(i-1)k})}\\
&\quad \frac{1-\alpha}{k} I\left(X_{1,(i-1)k+1}^{ik}; Y_{(i-1)k+1}^{ik}|X_{2,(i-1)k+1}^{ik}, S_{(i-1)k}, S_0 = s_0\right)\\
&\qquad + \frac{\alpha}{k} I\left(X_{2,(i-1)k+1}^{ik}; Y_{(i-1)k+1}^{ik}|S_{(i-1)k}, S_0 = s_0\right) + \frac{1}{k} H(S_{(i-1)k}|S_0 = s_0)\\
&\le \zeta_k^O(\alpha)
\end{aligned}
$$

$\left(X_{1,(i-1)k+1}^{ik}, X_{2,(i-1)k+1}^{ik}\right) - \left(X_{1,1}^{(i-1)k}, X_{2,1}^{(i-1)k}, S_0\right) - S_{(i-1)k}$ form a Markov chain. It is clear that for any $x_{1,(i-1)k+1}^{ik} \in \mathcal{X}_1^k$, $x_{2,(i-1)k+1}^{ik} \in \mathcal{X}_2^k$, and $s_0 \in \mathcal{S}$, we have

$$P_{X_{1,(i-1)k+1}^{ik}|X_{1,1}^{(i-1)k}S_0}\left(\cdot|x_{1,1}^{(i-1)k}, s_0\right) \in \mathcal{P}\left(\mathcal{X}_1^k\right)$$
$$P_{X_{2,(i-1)k+1}^{ik}|X_{2,1}^{(i-1)k}S_0}\left(\cdot|x_{2,1}^{(i-1)k}, s_0\right) \in \mathcal{P}\left(\mathcal{X}_2^k\right)$$
$$P_{S_{(i-1)k}|X_1^{(i-1)k}S_0}\left(\cdot|x_{1,1}^{(i-1)k}, x_{2,1}^{(i-1)k}, s_0\right) \in \mathcal{G}_{s_0,(i-1)k}.$$

Therefore, $P_{X_{1,(i-1)k+1}^{ik}X_{2,(i-1)k+1}^{ik}S_{(i-1)k}|S_0}(\cdot,\cdot,\cdot|s_0) \in \Xi_k(\mathcal{G}_{s_0,(i-1)k})$. In view of the fact that $\lim_{i\to\infty} \delta(\mathcal{G}_{s_0,(i-1)k}, \mathcal{A}_{s_0}) = 0$, we have, for $\alpha \in \left[0, \frac{1}{2}\right]$, see the second equation, shown at the bottom of the previous page. By symmetry, $\widetilde{\zeta}(\alpha) \leq \zeta_k^O(\alpha)$ also holds for $\alpha \in \left(\frac{1}{2}, 1\right]$. Therefore, we have $\widetilde{\mathcal{R}} \subseteq \mathcal{R}_k^O$.

2) It is easy to verify that $\underline{\mathcal{R}}_k \subseteq \mathcal{R}_k^I$. So we only need to show $\mathcal{R}_k^I \subseteq \underline{\mathcal{R}}$. Let $\{\left(X_{1,(i-1)k+1}^{ik}, X_{2,(i-1)k+1}^{ik}\right)\}_{i=1}^{\infty}$ be a stationary and memoryless vector process independent of the initial state S_0. It can be verified that [see the first equation shown at the bottom of the page]. The rest of the proof is similar to that of Theorem 3.5, and thus is omitted.

Appendix L
Proof of Theorem 4.3

1) Since both $k_1\mathcal{R}_{k_1}^O$ and $k_2\mathcal{R}_{k_2}^O$ are convex, it follows that $k_1\mathcal{R}_{k_1}^O + k_2\mathcal{R}_{k_2}^O$ is also convex. Moreover, for $\alpha \in [0,1]$, we have

$$\max_{(R_1,R_2)\in k_1\mathcal{R}_{k_1}^O + k_2\mathcal{R}_{k_2}^O} (1-\alpha)R_1 + \alpha R_2 = k_1\zeta_{k_1}^O(\alpha) + k_2\zeta_{k_2}^O(\alpha).$$

$$I\left(X_{1,1}^{nk}; Y_1^{nk}|X_{2,1}^{nk}, S_0 = s_0\right) \geq \sum_{i=1}^{n}\left[I\left(X_{1,(i-1)k+1}^{ik}; Y_{(i-1)k+1}^{ik}|X_{2,(i-1)k+1}^{ik}, S_{(i-1)k}, S_0 = s_0\right) - H(S_{(i-1)k}|S_0 = s_0)\right]$$
$$I\left(X_{2,1}^{nk}; Y_1^{nk}|X_{1,1}^{nk}, S_0 = s_0\right) \geq \sum_{i=1}^{n}\left[I\left(X_{2,(i-1)k+1}^{ik}; Y_{(i-1)k+1}^{ik}|X_{1,(i-1)k+1}^{ik}, S_{(i-1)k}, S_0 = s_0\right) - H\left(S_{(i-1)k}|S_0 = s_0\right)\right]$$
$$I\left(X_{1,1}^{nk}, X_{2,1}^{nk}; Y_1^{nk}|S_0 = s_0\right) \geq \sum_{i=1}^{n}\left[I\left(X_{1,(i-1)k+1}^{ik}, X_{2,(i-1)k+1}^{ik}; Y_{(i-1)k+1}^{ik}|S_{(i-1)k}, S_0 = s_0\right) - H\left(S_{(i-1)k}|S_0 = s_0\right)\right]$$

$$I\left(X_{1,1}^{k}; Y_1^{k}|X_{2,1}^{k}, S_0\right) \leq I\left(X_{1,1}^{k_1}; Y_1^{k_1}|X_{2,1}^{k_1}, S_0\right) + I\left(X_{1,k_1+1}^{k}; Y_{k_1+1}^{k}|X_{2,k_1+1}^{k}, S_{k_1}\right) + H(S_{k_1})$$
$$I\left(X_{1,1}^{k}, X_{2,1}^{k}; Y_1^{k}|S_0\right) \leq I\left(X_{1,1}^{k_1}, X_{2,1}^{k_1}; Y_1^{k_1}|S_0\right) + I\left(X_{1,k_1+1}^{k}, X_{2,k_1+1}^{k}; Y_{k_1+1}^{k}|S_{k_1}\right) + H(S_{k_1})$$

$$\begin{aligned}
k\zeta_k^O(\alpha) &= \max_s \max_{P_{X_{1,1}^k X_{2,1}^k S_0}\in\Xi_k(\mathcal{A}_s)} (1-\alpha)I\left(X_{1,1}^k; Y_1^k|X_{2,1}^k, S_0\right) + \alpha I\left(X_{2,1}^k; Y_1^k|S_0\right) + H(S_0)\\
&= \max_s \max_{P_{X_{1,1}^k X_{2,1}^k S_0}\in\Xi_k(\mathcal{A}_s)} (1-2\alpha)I\left(X_{1,1}^k; Y_1^k|X_{2,1}^k, S_0\right) + \alpha I\left(X_{1,1}^k, X_{2,1}^k; Y_1^k|S_0\right) + H(S_0)\\
&\leq \max_s \max_{P_{X_{1,1}^k X_{2,1}^k S_0}\in\Xi_k(\mathcal{A}_s)} (1-2\alpha)\left[I\left(X_{1,1}^{k_1}; Y_1^{k_1}|X_{2,1}^{k_1}, S_0\right) + I\left(X_{1,k_1+1}^k; Y_{k_1+1}^k|X_{2,k_1+1}^k, S_{k_1}\right)\right]\\
&\quad + \alpha\left[I\left(X_{1,1}^{k_1}, X_{2,1}^{k_1}; Y_1^{k_1}|S_0\right) + I\left(X_{1,k_1+1}^k, X_{2,k_1+1}^k; Y_{k_1+1}^k|S_{k_1}\right)\right] + H(S_0) + H(S_{k_1})\\
&\leq \max_s \max_{P_{X_{1,1}^{k_1} X_{2,1}^{k_1} S_0}\in\Xi_{k_1}(\mathcal{A}_s)} (1-2\alpha)I\left(X_{1,1}^{k_1}; Y_1^{k_1}|X_{2,1}^{k_1}, S_0\right) + \alpha I\left(X_{1,1}^{k_1}, X_{2,1}^{k_1}; Y_1^{k_1}|S_0\right) + H(S_0)\\
&\quad + \max_s \max_{P_{X_{1,k_1+1}^k X_{2,k_1+1}^k S_{k_1}}\in\Xi_{k_2}(\mathcal{A}_s)} (1-2\alpha)I\left(X_{1,k_1+1}^k; Y_{k_1+1}^k|X_{2,k_1+1}^k, S_{k_1}\right)\\
&\quad + \alpha I\left(X_{1,k_1+1}^k, X_{2,k_1+1}^k; Y_{k_1+1}^k|S_{k_1}\right) + H(S_{k_1}) \qquad (43)\\
&= k_1\zeta_{k_1}^O(\alpha) + k_2\zeta_{k_2}^O(\alpha)
\end{aligned}$$

It can be verified that (cf. the proof of Theorem 3.4) (see the second equation shown at the bottom of the previous page). Therefore, for $\alpha \in \left[0, \frac{1}{2}\right]$, see (43), shown at the bottom of the previous page, where (43) follows from Part 5) of Lemma 4.1. By symmetry, $k\zeta_k^O(\alpha) \le k_1\zeta_{k_1}^O(\alpha) + k_2\zeta_{k_2}^O(\alpha)$ also holds for $\alpha \in \left(\frac{1}{2}, 1\right]$. Therefore, we have $k\mathcal{R}_k^O \subseteq k_1\mathcal{R}_{k_1}^O + k_2\mathcal{R}_{k_2}^O$.

2) Let $P_1 \in \mathcal{P}\left(\mathcal{X}_1^{k_1}\right) \times \mathcal{P}\left(\mathcal{X}_2^{k_1}\right)$ and $P_2 \in \mathcal{P}\left(\mathcal{X}_1^{k_2}\right) \times \mathcal{P}\left(\mathcal{X}_2^{k_2}\right)$ be the two arbitrary input distributions. Construct independent random vectors $X_{1,1}^{k_1}$, $X_{2,1}^{k_1}$, X_{1,k_1+1}^{k}, and X_{2,k_1+1}^{k} with $P_{X_{1,1}^{k_1}X_{2,1}^{k_1}} = P_1$ and $P_{X_{1,k_1+1}^{k}X_{2,k_1+1}^{k}} = P_2$. Assume $X_{1,1}^{k_1}$, $X_{2,1}^{k_1}$, X_{1,k_1+1}^{k}, and X_{2,k_1+1}^{k} are independent of S_0. By this construction, if $P_{S_0} \in \mathrm{conv}(\mathcal{A}_s)$, then $P_{X_{1,1}^{k_1}X_{2,1}^{k_1}S_0} \in \Xi'_{k_1}(\mathcal{A}_s)$; moreover, it follows from Part 6) of Lemma 4.1 that $P_{X_{1,k_1+1}^{k}X_{2,k_1+1}^{k}S_{k_1}} \in \Xi'_{k_2}(\mathcal{A}_s)$. It suffices to show

$$k\mathcal{R}_k^I\left(P_{X_{1,1}^k X_{2,1}^k}\right) \supseteq k_1\mathcal{R}_{k_1}^I(P_1) + k_2\mathcal{R}_{k_2}^I(P_2)$$

which boils down to the following three inequalities (see the first equation shown at the bottom of the page), where $P_{\widetilde{X}_{1,1}^{k_1}\widetilde{X}_{2,1}^{k_1}} = P_1$, $P_{\overline{X}_{1,1}^{k_2}\overline{X}_{2,1}^{k_2}} = P_2$, and $\widetilde{Y}_1^{k_1}$, $\overline{Y}_1^{k_2}$ are induced by $(\widetilde{X}_{1,1}^{k_1}\widetilde{X}_{2,1}^{k_1}, \widetilde{S}_0)$ and $(\overline{X}_{1,1}^{k_2}\overline{X}_{2,1}^{k_2}, \overline{S}_0)$, respectively. It can be verified that (cf. the proof of Theorem 3.8)

$$\begin{aligned} I\left(X_{1,1}^k; Y_1^k | X_{2,1}^k, S_0\right) \ge\, & I\left(X_{1,1}^{k_1}; Y_1^{k_1} | X_{2,1}^{k_1}, S_0\right) \\ & + I\left(X_{1,k_1+1}^k; Y_{k_1+1}^k | X_{2,k_1+1}^k, S_{k_1}\right) - H(S_{k_1}). \end{aligned}$$

Therefore, we have the second equation shown at the bottom of the page, which, together with the fact that (see the third equation shown at the bottom of the page),

$$\begin{aligned}
&\min_s \min_{P_{S_0}\in\mathrm{conv}(\mathcal{A}_s)} \left[I\left(X_{1,1}^k; Y_1^k | X_{2,1}^k, S_0\right) - H(S_0)\right]^+ \\
&\ge \min_s \min_{P_{\widetilde{S}_0}\in\mathrm{conv}(\mathcal{A}_s)} \left[I\left(\widetilde{X}_{1,1}^{k_1}; \widetilde{Y}_1^{k_1} | \widetilde{X}_{2,1}^{k_1}, \widetilde{S}_0\right) - H(\widetilde{S}_0)\right]^+ + \min_s \min_{P_{\overline{S}_0}\in\mathrm{conv}(\mathcal{A}_s)} \left[I\left(\overline{X}_{1,1}^{k_2}; \overline{Y}_1^{k_2} | \overline{X}_{2,1}^{k_2}, \overline{S}_0\right) - H(\overline{S}_0)\right]^+ \\
&\min_s \min_{P_{S_0}\in\mathrm{conv}(\mathcal{A}_s)} \left[I\left(X_{2,1}^k; Y_1^k | X_{1,1}^k, S_0\right) - H(S_0)\right]^+ \\
&\ge \min_s \min_{P_{\widetilde{S}_0}\in\mathrm{conv}(\mathcal{A}_s)} \left[I\left(\widetilde{X}_{2,1}^{k_1}; \widetilde{Y}_1^{k_1} | \widetilde{X}_{1,1}^{k_1}, \widetilde{S}_0\right) - H(\widetilde{S}_0)\right]^+ + \min_s \min_{P_{\overline{S}_0}\in\mathrm{conv}(\mathcal{A}_s)} \left[I(\overline{X}_{2,1}^{k_2}; \overline{Y}_1^{k_2} | \overline{X}_{1,1}^{k_2}, \overline{S}_0) - H(\overline{S}_0)\right]^+ \\
&\min_s \min_{P_{S_0}\in\mathrm{conv}(\mathcal{A}_s)} \left[I\left(X_{1,1}^k, X_{2,1}^k; Y_1^k | S_0\right) - H(S_0)\right]^+ \\
&\ge \min_s \min_{P_{\widetilde{S}_0}\in\mathrm{conv}(\mathcal{A}_s)} \left[I\left(\widetilde{X}_{1,1}^{k_1}, \widetilde{X}_{2,1}^{k_1}; \widetilde{Y}_1^{k_1} | \widetilde{S}_0\right) - H(\widetilde{S}_0)\right]^+ + \min_s \min_{P_{\overline{S}_0}\in\mathrm{conv}(\mathcal{A}_s)} \left[I\left(\overline{X}_{1,1}^{k_2}, \overline{X}_{2,1}^{k_2}; \overline{Y}_1^{k_2} | \overline{S}_0\right) - H(\overline{S}_0)\right]^+
\end{aligned}$$

$$\begin{aligned}
&\min_s \min_{P_{S_0}\in\mathrm{conv}(\mathcal{A}_s)} I\left(X_{1,1}^k; Y_1^k | X_{2,1}^k, S_0\right) - H(S_0) \\
&\ge \min_s \min_{P_{S_0}\in\mathrm{conv}(\mathcal{A}_s)} I\left(X_{1,1}^{k_1}; Y_1^{k_1} | X_{2,1}^{k_1}, S_0\right) + I\left(X_{1,k_1+1}^k; Y_{k_1+1}^k | X_{2,k_1+1}^k, S_{k_1}\right) - H(S_{k_1}) - H(S_0) \\
&\ge \min_s \min_{P_{\widetilde{S}_0}\in\mathrm{conv}(\mathcal{A}_s)} I\left(\widetilde{X}_{1,1}^{k_1}; \widetilde{Y}_1^{k_1} | \widetilde{X}_{2,1}^{k_1}, \widetilde{S}_0\right) - H(\widetilde{S}_0) + \min_s \min_{P_{\overline{S}_0}\in\mathrm{conv}(\mathcal{A}_s)} I\left(\overline{X}_{1,1}^{k_2}; \overline{Y}_1^{k_2} | \overline{X}_{2,1}^{k_2}, \overline{S}_0\right) - H(\overline{S}_0)
\end{aligned}$$

$$\begin{aligned}
&\min_s \min_{P_{S_0}\in\mathrm{conv}(\mathcal{A}_s)} I\left(X_{1,1}^k; Y_1^k | X_{2,1}^k, S_0\right) - H(S_0) \\
&\ge \max\left(\min_s \min_{P_{\widetilde{S}_0}\in\mathrm{conv}(\mathcal{A}_s)} I\left(\widetilde{X}_{1,1}^{k_1}; \widetilde{Y}_1^{k_1} | \widetilde{X}_{2,1}^{k_1}, \widetilde{S}_0\right) - H(\widetilde{S}_0), \min_s \min_{P_{\overline{S}_0}\in\mathrm{conv}(\mathcal{A}_s)} I\left(\overline{X}_{1,1}^{k_2}; \overline{Y}_1^{k_2} | \overline{X}_{2,1}^{k_2}, \overline{S}_0\right) - H(\overline{S}_0)\right)
\end{aligned}$$

$$\min_s \min_{P_{S_0} \in \mathrm{conv}(\mathcal{A}_s)} \left[I\left(X_{1,1}^k; Y_1^k | X_{2,1}^k, S_0\right) - H(S_0) \right]^+$$
$$\geq \min_s \min_{P_{\widetilde{S}_0} \in \mathrm{conv}(\mathcal{A}_s)} \left[I\left(\widetilde{X}_{1,1}^{k_1}; \widetilde{Y}_1^{k_1} | \widetilde{X}_{2,1}^{k_1}, \widetilde{S}_0\right) - H(\widetilde{S}_0) \right]^+ + \min_s \min_{P_{\overline{S}_0} \in \mathrm{conv}(\mathcal{A}_s)} \left[I\left(\overline{X}_{1,1}^{k_2}; \overline{Y}_1^{k_2} | \overline{X}_{2,1}^{k_2}, \overline{S}_0\right) - H(\overline{S}_0) \right]^+$$

implies (see the equation at the top of the page). The other two inequalities can be verified in a similar way. The details are omitted.

Acknowledgment

The authors would like to thank the Associate Editor and the two anonymous reviewers for their valuable comments and suggestions.

References

[1] D. Blackwell, L. Breiman, and A. J. Thomasian, "Proof of Shannon's transmission theorem for finite-state indecomposable channels," *Ann. Math. Statist.*, vol. 29, pp. 1209–1220, Dec. 1958.
[2] H. L. Yudkin, "On the exponential error bound and capacity for finite state channels," presented at the Int. Symp. Inform. Theory, San Remo, Italy, 1967.
[3] R. G. Gallager, *Information Theory and Reliable Communication*. Hoboken, NJ: Wiley, 1968.
[4] R. M. Gray, M. O. Dunham, and R. L. Gobbi, "Ergodicity of Markov channels," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 5, pp. 656–664, Sep. 1987.
[5] M. Mushkin and I. Bar-David, "Capacity and coding for the Gilbert-Elliott channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 6, pp. 1277–1290, Nov. 1989.
[6] A. J. Goldsmith and P. P. Varaiya, "Capacity, mutual information, and coding for finite-state Markov channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 868–886, May 1996.
[7] V. Sharma and S. K. Singh, "Entropy and channel capacity in the regenerative setup with applications to Markov channels," in *Proc. IEEE Int. Symp. Inform. Theory*, Washington, DC, Jun. 2001, pp. 283–283.
[8] P. O. Vontobel and D. M. Arnold, "An upper bound on the capacity of channels with memory and constraint input," in *Proc. IEEE Inf. Theory Workshop*, Cairns, Qld, Australia, Sep. 2001, pp. 147–149.
[9] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the achievable information rates of finite-state ISI channels," in *Proc. IEEE Globecom*, San Antonio, TX, Nov. 2001, pp. 2992–2996.
[10] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavcić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.
[11] T. Holliday, A. Goldsmith, and P. Glynn, "Capacity of finite state channels based on Lyapunov exponents of random matrices," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3509–3532, Aug. 2006.
[12] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 7, pp. 1147–1157, Jul. 1994.
[13] D. J. Hartfiel, "Sequential limits in Markov set-chains," *J. Appl. Probab.*, vol. 28, pp. 910–913, Dec. 1991.
[14] J. Hajnal, "Weak ergodicity in non-homogeneous Markov chains," *Proc. Combridge Philos. Soc.*, vol. 54, pp. 233–246, 1958.
[15] E. Seneta, "Coefficients of ergodicity: Structure and applications," *Adv. Appl. Probab.*, vol. 11, pp. 576–590, Sep. 1979.
[16] I. Daubechies and J. C. Lagarias, "Sets of matrices all infinite products of which converge," *Linear Algebra Appl.*, vol. 161, pp. 227–263, 1992.
[17] J. Wolfowitz, "Products of indecomposable, aperiodic, stochastic matrices," *Proc. Amer. Math. Soc.*, vol. 14, pp. 733–737, Oct. 1963.
[18] A. J. Thomasian, "A finite criterion for indecomposable channels," *Ann. Math. Statist.*, vol. 34, pp. 337–338, Mar. 1963.
[19] E. Seneta, "On the limiting set of nonnegative products," *Statist. Prob. Lett.*, vol. 2, pp. 159–163, May 1984.
[20] H. Permuter, T. Weissman, and J. Chen, "Capacity region of the finite-state multiple access channel with and without feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2455–2477, Jun. 2009.
[21] J. Edmonds, "Submodular functions, matroids and certain polyhedra," in *Combinatorial Structures and Their Applications*, R. Guy, H. Hanani, N. Sauer, and J. Schonheim, Eds. New York: Gordon and Breach, 1970, pp. 69–87.
[22] H. Permuter, T. Weissman, and A. Goldsmith, "Finite state channels with time-invariant deterministic feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 644–662, Feb. 2009.

Jun Chen (S'03–M'06) received the B.E. degree with honors in communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2001, and the M.S. and Ph.D. degrees in electrical and computer engineering from Cornell University, Ithaca, NY, in 2004 and 2006, respectively.

He was a Postdoctoral Research Associate in the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign, Urbana, from 2005 to 2006, and a Josef Raviv Memorial Postdoctoral Fellow at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, from 2006 to 2007. He is currently an Assistant Professor of Electrical and Computer Engineering at McMaster University, Hamilton, ON, Canada. He holds the Barber-Gennum Chair in Information Technology. His research interests include information theory, wireless communications, and signal processing.

Haim Permuter (S'08–M'09) received the B.Sc. (summa cum laude) and M.Sc. (summa cum laude) degrees in electrical and computer engineering from the Ben-Gurion University, Israel, in 1997 and 2003, respectively, and Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 2008.

Between 1997 and 2004, he was an officer with the research and development unit of the Israeli Defense Forces. He is currently a lecturer at Ben-Gurion university.

Dr. Permuter is a recipient of the Fullbright Fellowship, the Stanford Graduate Fellowship (SGF), Allon Fellowship, and Bergman award.

Tsachy Weissman (S'99–M'02–SM'07) received the Ph.D. degree in electrical engineering from the Technion—Israel Institute of Technology, Haifa, in 2001.

During 2002–2003, he was with the information theory research group at Hewlett-Packard Laboratories. He has been on the faculty of the Electrical Engineering Department, Stanford University, Stanford, CA, since 2003, spending the two academic years (2007–2009) on leave at the Technion Department of Electrical Engineering. His research is focused on information theory, statistical signal processing, the interplay between them, and their applications. He is the inventor of several patents and involved in a number of high-tech companies as a researcher or member of the technical board.

Dr. Weissman has received an NSF CAREER award, a joint IT/COM societies best paper award, a Horev fellowship for Leaders in Science and Technology, and a Henry Taub prize for excellence in research.