Channel Capacity in the Presence of Feedback and Side Information

by

NEVROZ ŞEN

A thesis submitted to the
Department of Mathematics and Statistics
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen's University
Kingston, Ontario, Canada
July 2013

Abstract

This thesis deals with the Shannon-theoretic fundamental limits of channel coding for single-user channels with memory and feedback and for multi-user channels with side information. We first consider the feedback capacity of a class of symmetric channels with memory modelled as finite-state Markov channels. The symmetry yields the existence of a hidden Markov noise process that facilitates the channel description as a function of input and noise, where the function satisfies a desirable invertibility property. We show that feedback does not increase capacity for such class of finite-state channels and that both their non-feedback and feedback capacities are achieved by an independent and uniformly distributed input. As a result, the capacity is given as a difference of output and noise entropy rates, where the output is also a hidden Markov process; hence, capacity can be approximated via well known algorithms.

We then consider the memoryless state-dependent multiple-access channel (MAC) where the encoders and the decoder are provided with various degrees of asymmetric noisy channel state information (CSI). For the case where the encoders observe causal, asymmetric noisy CSI and the decoder observes complete CSI, inner and outer bounds to the capacity region, which are tight for the sum-rate capacity, are provided. Next, single-letter characterizations for the channel capacity regions under each of the following settings are established: (a) the CSI at the encoders are

non-causal and asymmetric deterministic functions of the CSI at the decoder (b) the encoders observe asymmetric noisy CSI with asymmetric delays and the decoder observes complete CSI; (c) a degraded message set scenario with asymmetric noisy CSI at the encoders and complete and/or noisy CSI at the decoder.

Finally, we consider the above state-dependent MAC model and identify what is required to be provided to the receiver in order to get a tight converse for the sum-rate capacity. Inspired by the coding schemes of the lossless CEO problem as well as of a recently proposed achievable region, we provide an inner bound which demonstrates the rate required to transmit this information to the receiver.

Bibliographic Note

Parts of Chapter 3 were presented at the Forty-Seventh Annual Allerton Conference on Communication, Control and Computing, 2009 [ŞAY09a, ŞAY09b] and appeared in the IEEE Transactions on Information Theory, July 2011, [ŞAY11].

Parts of Chapter 4 were presented at the Forty-Ninth Annual Allerton Conference on Communication, Control and Computing, 2011 [ŞCYA11], at the IEEE International Symposium on Information Theory, 2012 [ŞAYC12] and is now under revision [ŞAYC].

Acknowledgments

I have had wonderful support from many people along the completion of my dissertation. This acknowledgement is dedicated to those.

First of all, my parents, Mehmet Emin Şen and Leyla Şen, are one of the main reasons that this thesis exists. Every step that I took during this process, starting from the first one, is initiated with the trust and confidence that are provided by my parents. It is my father's wisdom that let me to think about Ph.D. and it is the wish of being as successful as him that has resulted in this degree. Having this opportunity, please accept my deepest appreciation for all the sacrifices that you have done for me.

I am deeply grateful to my advisors Professor Fady Alajaji and Professor Serdar Yüksel for their trust, their encouragement and for their continuous support on all professional and personal issues. I have learned a lot from you and thank you for sharing your knowledge, expertise and insights with me during the years that we have worked together.

I would like to thank to Professor Tamas Linder for accepting to be my thesis examiner, his constructive comments and for accepting to be in my supervisory committee. I am very thankful to Professor Abdol-Reza Mansouri for both accepting to be in my supervisory committee and for his valuable advices on many issues that I needed to talk with him. Furthermore, I would also like to thank my exam committee, Professor Stark C. Draper, Professor Gregory G. Smith and Professor Steven D. Blostein for their generous time and effort and for their detailed comments. Thank you Ms. Jennifer Read for your tremendous help during my years at Queen's.

My mother-in-law, Besi Doğan's, assistance during this period is invaluable. I am deeply indebted to you for being with us whenever we needed and most importantly, for making my kids never feel alone.

The support from my friends has never been reduced even though we are all in different continents. İlker, Mervan, Miçe and Çağatay; thank you for keeping your friendship with me. Your emails and calls had always reduced the stress of the graduate life. One of my biggest luck at Kingston is to find a friend that I believe we have many common values; my deepest gratitude to Pantelis Isaias for his friendship.

My brothers, Mahmut, Rojhat and Fırat, my sister, Seher, my brothers-in-law, Düzgün and Engin, and my sister-in-law, Diren; I am thankful to all of you for being so caring and as a result of this, among many others, visiting us in Kingston many times. I have no doubt that I would have much more difficult period if I had not had the support from all of you.

Finally, and most importantly, my wife Dilek; it is due to your patience, constant encouragement and belief in me that I have been able to complete this dissertation. My daughters Arjen and Delal, my constant source of hope, you have always been very generous and tolerant whenever I asked your time. I don't think that I can express my appreciation to you with words; this thesis is dedicated to you.

To my parents, Mehmet Emin and Leyla

To my wife, Dilek

and

To my daughters, Arjen Viyan and Delal Zilan

Table of Contents

List of Figures

Chapter 1

Introduction

Availability of past channel outputs, called feedback, at the encoder and availability of channel state information (CSI) in various degrees at the encoder and/or at the decoder have several benefits in a communication system. In particular, both feedback and state information can increase capacity, decrease the complexity of the encoder and decoder, and reduce latency.

The current state-of-the art of feedback and side information in information theory differs considerably in single vs. multi-user setups. As an example, there exists a general formula for the capacity of channels with feedback [CA95], [TM09] for the single-user case, whereas the same problem for a two-user multiple access channel (MAC) is still open. Similarly, the characterization of the capacity with various degrees of CSI at the transmitter (CSIT) and at the receiver (CSIR) is well understood for single-user channels. However, for multi-user channels, availability of CSI at the encoders and/or at the decoder reveals many difficult problems, especially if the information available at the encoders are asymmetric.

In both feedback and side information problems, mainly due to the nature of the information structures available to the decision makers (encoders and decoders), tools from stochastic and decentralized control theory have been popular in recent years. Furthermore, as there is an increased tendency towards cooperative communications and networked systems, it is believed that both of these fields will play an important role to understand information theoretic questions in these areas. We herein investigate coding schemes that use stochastic and decentralized control theory to some extent. We first present the most relevant contributions in the literature.

1.1 Literature Review

1.1.1 Feedback

When a single-user channel has no memory, i.e., the noise process corrupting the channel input has no statistical dependence over time, it is known that feedback does not help to increase the capacity [Sha56]. Although, in general, feedback increases the capacity of channels when there is memory, in [Ala95], it is shown that feedback does not increase the capacity of discrete channels with modulo additive noise where the noise process has arbitrary memory (not necessarily stationary or ergodic). It is also shown that for any channel with memory satisfying the symmetry conditions defined in [AF94], feedback does not increase its capacity. Recently, it has been shown that feedback does not increase the capacity of the compound Gilbert-Elliot channel [SP09], which is a family of finite-state Markov (FSM) channels.

In a more recent work, it has been shown that it is possible to formulate the computation of feedback capacity as a dynamic programming problem and therefore it can be solved by using the value iteration algorithm under information stability conditions [TM09],[Tat00]. In [PWG09], finite-state channels with feedback, where feedback is a time-invariant deterministic function of the output samples, is considered. It is shown that if the state of the channel is known both at the encoder and the decoder then feedback does not increase capacity. In [YKT05] and [CB05], directed information is used to calculate the feedback capacity of some classes of FSM channels. In particular, the channel state is assumed in [YKT05] to be a deterministic function of the previous state and input; whereas in [CB05] the channel state is assumed to be a deterministic function of the output. In addition to these results, it has also been shown that feedback does not increase the capacity for a binary erasure channel with Markovian state [DG06].

Although feedback does not help increase the capacity of discrete memoryless channels (DMCs), it does increase the capacity of memoryless MACs [GW75], [CL81] due to the user cooperation through channel outputs. However, there exists no single letter characterization, i.e., an expression that does not depend on the block length of the coding scheme (see also [CK81, pg.259-261] for a detailed discussion), of the capacity region of MAC with feedback even for the simplest setup; two-user, memoryless and perfect feedback. This problem has attracted much attention in recent years and many achievability results, i.e., inner bounds to the capacity region, as well as an interpretation of an existing inner bound from a stochastic control point of view [AS12] have been established.

1.1.2 State Side Information

Modeling communication channels with a state process, which governs the channel behaviour, fits well for many physical scenarios and in addition to channel output feedback, communication with state feedback has been widely motivated in both single and multi-user communications. For single-user channels, the characterization of the capacity with various degrees of channel state information at the transmitter (CSIT) and at the receiver (CSIR) is well understood. Among them, Shannon [Sha58] provides the capacity formula for a discrete memoryless channel with causal noiseless CSIT, where the state process is independent and identically distributed (i.i.d.), in terms of Shannon strategies (random functions from the state space to the channel input space). In [GP80], Gel'fand and Pinsker consider the same problem with non-causal side information and establish a single-letter capacity formula. In [Sal92], noisy state observation available at both the transmitter and the receiver is considered and the capacity under such a setting is derived. Later, in [CS99], this result is shown to be a special case of Shannon's model and the authors also prove that when CSIT is a deterministic function of CSIR, optimal codes can be constructed directly on the input alphabet. In [EZ00], the authors examine the discrete modulo-additive noise channel with causal CSIT which governs the noise distribution, and they determine the optimal strategies that achieve channel capacity. In [GV97], fading channels with perfect channel state information at the transmitter are considered and it is shown that with instantaneous and perfect CSI, the transmitter can adjust the data rates for each channel state to maximize the average transmission rate. In [YT07], a single letter characterization of the capacity region for single-user finite-state Markovian channels with quantized state information available at the transmitter and full state information at the decoder is provided.

The literature on state dependent multiple access channels with different assumptions of CSIR and CSIT (such as causal vs non-causal, perfect vs imperfect) is extensive and the main contributions of the current thesis have several interactions with the available results in the literature, which we present in Section 1.2. Hence, we believe that in order to suitably highlight the contributions of this thesis, it is worth to discuss the relevant literature for the multi-user setting in more detail. To start, [DN02] provides a multi-letter characterization of the capacity region of time-varying MACs with general channel statistics (with/without memory) under a general state process (not necessarily stationary or ergodic) and with various degrees of CSIT and CSIR. In [DN02], it is also shown that when the channel is memoryless, if the encoders use only the past k asymmetric partial (but not noisy) CSI and the decoder has complete CSI, then it is possible to simplify the multi-letter characterization to a single letter one [DN02, Theorem 4]. In [Jaf06], a general framework for the capacity region of MACs with causal and non-causal CSI is presented. More explicitly, an achievable rate region is presented for the memoryless state-dependent MAC with correlated CSI and the sum-rate capacity is established under the condition that the state information available to each encoder are independent. In [CS05], MACs with complete CSIR and noncausal, partial, rate limited CSITs are considered. In particular, for the degraded case, i.e., the case where the CSI available at one of the encoders is a subset of the CSI available at the other encoder, a single letter formula for the capacity region is provided and when the CSITs are not degraded, inner and outer bounds are derived, see [CS05, Theorems 1, 2]. In [CY11] state-dependent MAC in which transmitters observe asymmetric partial quantized CSI causally, and the receiver has full CSI is considered and a single letter characterization of the capacity region is obtained. In [LS13b], memoryless state-dependent MACs with two independent states (see also [LS13a] for the common state), each known causally and strictly causally to one encoder, is considered and an achievable rate region, which is shown to contain an achievable region where each user applies Shannon strategies, is proposed. In [LSY13], another achievable rate region for the same problem is proposed and in [LS] it is shown that this region can be strictly larger than the one proposed in [LS13b]. In [LS13b], it is also shown that strictly causal CSI does not increase the sum-rate capacity. In [BSP12], the finite-state Markovian MAC with asymmetric delayed CSITs is studied and its capacity region is determined. In [SK05], the capacity region of some multiple-user channels with causal CSI is established and inner and outer capacity bounds are provided for the MAC. Another active research direction on the state-dependent MAC regards the so-called cooperative state-dependent MAC where there exists a degraded condition on the message sets. In particular, [SBSV08] and [KL07] characterize the capacity region of the cooperative state-dependent MAC with states non-causally and causally available at the transmitters. More recent results on the cooperative state-dependent MAC problem include [ZPS11], [ZPSS] and [PSSB11].

1.2 Contributions and Connections with the Literature

1.2.1 Feedback

Considering the structure in typical communication channels and the results in the literature that we presented above, it is worth to look for the most general notion of symmetry for channels with memory under which feedback does not increase capacity. With this motivation, we first study the feedback capacity of a class of symmetric FSM channels, which we call "quasi-symmetric" FSM channels, and prove that feedback does not help increase their capacity. This result is shown by demonstrating that for an FSM channel satisfying the symmetry conditions defined in the thesis, its feedback capacity is achieved by an independent and uniformly distributed (i.u.d.) input which implies that its non feedback capacity is also achieved by uniform input distribution. The symmetry conditions for this result is then relaxed by allowing the receiver to observe full CSI. These results are demonstrated in Chapter 3.

A by-product contribution of this result is that the channel capacity is given as a difference of the output and noise entropy rates, where the output is driven by the i.u.d. input and is also hidden Markovian. Thus, the capacity can be easily evaluated using existing algorithms for the computation of entropy and information rates in hidden Markov channels (e.g., see [ALV$^+$06]).

1.2.2 State Side Information

The succeeding chapters, Chapters 4 and 5, focus on multi-user models with asymmetric CSI. In particular, Chapter 4 considers several scenarios where the encoders and the decoder observe various degrees of noisy CSI. The essential requirement we impose is that the noisy CSI available to the decision makers is realized via the corruption of CSI by different noise processes, which gives a realistic physical structure of the communication setup. We herein note that the asymmetric noisy CSI assumption is acceptable as typically the feedback links are imperfect and sufficiently far from each other so that the information carried through them is corrupted by different (independent) noise processes. It should also be noted that asymmetric side information has many applications in different multi-user models. Finally, what makes (asymmetric) noisy setups particularly interesting are the facts that

(*i*) No transmitter CSI contains the CSI available to the other one;

(*ii*) CSI available to the decoder does not contain any of the CSI available to the two encoders.

When existing results, which provide a single letter capacity formulation, are examined, it can be observed that most of them do not satisfy (*i*) or (*ii*) or both (e.g., [CY11], [DN02], [Jaf06], [CS05], [BSP12]). Nonetheless, among these, [DN02] discusses the situation with noisy CSI and makes the observation that the situation where the CSITs and CSIR are noisy versions of the state S_t can be accommodated by their models. However, they also note that if the noises corrupting transmitters and receiver CSI are different, then the encoder CSI will, in general, not be contained in the decoder CSI. Hence, motivated by similar observations in the literature (e.g., [Jaf06]), we partially treat the scenarios below and provide inner and outer bounds, which are tight for the sum-rate capacity, for scenario (1) below and provide a single-letter characterization for the capacity region of the latter scenarios:

(1) The state-dependent MAC in which each of the transmitters has an asymmetric causal noisy CSI and the receiver has complete CSI (Theorems 4.2.1, 4.2.2 and Corollary 4.2.1).

(2) The state-dependent MAC in which each of the transmitters has an asymmetric non-causal noisy CSIT which is a deterministic function of the CSIR at the receiver (Theorem 4.2.3).

(3) The state-dependent MAC in which each of the transmitters has an asymmetrically delayed and asymmetric noisy CSI and the receiver has complete CSI (Theorem 4.2.4).

(4) The state-dependent MAC with degraded message set where both transmitters transmit a common message and one transmitter (informed transmitter) transmits a private message. The informed transmitter has causal noisy CSI, the other encoder has a delayed noisy CSI and the receiver has various degrees of CSI (Theorems 4.2.5 and 4.2.6).

Let us now briefly position these contributions with respect to the available results in the literature. The sum-rate capacity determined in (1) can be thought as an extension of [Jaf06, Theorem 4] to the case where the encoders have correlated CSI. The causal setup of (2) is solved in [CY11]. The solution that we provide to the non-causal case partially solves [CS05] and extends [Jaf06, Theorem 5] to the case where the encoders have correlated CSI. Furthermore, since the causal and non-causal capacities are identical for scenario (2), the causal solution can be considered as an extension of [CS99, Proposition 1] to a noisy multi-user case. Finally, (4) is an extension of [SBSV08, Theorem 4] to a noisy setup.

As it is mentioned above, for the multi-user state-dependent channels, the capacity region is not known in general but, what is known is that in the causal CSIT case Shannon strategies are in general suboptimal (e.g., see [LS13a]). On the other hand, for only few scenarios, it is known that Shannon strategies are optimal for the sum-rate capacity. When these scenarios are examined, it can be seen that the optimality is realized under the situations of either CSITs are independent (e.g., see [Jaf06], [LS13b]) or whenever CSITs are correlated, full CSI is available at the receiver (e.g., see Chapter 4). Hence, a natural question to ask is what is the most general condition under which Shannon strategies are optimal in terms of sum-rate capacity. Chapter 5 explores this condition and shows that when the state processes are asymmetric, Shannon strategies are optimal if the decoder is provided with some information which makes the CSITs conditionally independent.

With this result at hand, the next step is to investigate what is the minimum rate required to transmit such information to the receiver when there is no CSIR. This chapter also characterizes the rate required to transmit this information by using the lossless CEO approach [GP79] and by adopting the recent proof technique of [LS13b, Theorem 1].

1.3 Organization of Thesis

We proceed by introducing a short background chapter. We discuss the feedback capacity of a class of symmetric channels with memory in Chapter 3. Chapter 4 presents the results on the state dependent multiple access channel where the encoders and the decoder have several degrees of asymmetric noisy state information. In Chapter 5 we present the loss of optimality of using memoryless strategies for the sum-rate capacity in a state dependent multiple access channel scenario when there is no CSI at the receiver. Chapter 6 concludes and outlines future work.

Chapter 2

Background and Fundamental Results

This chapter contains basic material on channel coding and capacity. In Section 2.1, we first give an overview for the notation and conventions which will be used throughout the thesis. In Section 2.2, we introduce standard typicality definitions that will be used in the proofs throughout the thesis and in Section 2.4, we discuss fundamental results on channel coding under feedback and side information.

2.1 Notations and Conventions

Throughout the thesis, we will use the following notations. A random variable will be denoted by an upper case letter X and its particular realization by a lower case letter x. For a vector v, and a positive integer i, v_i will denote the i-th entry of v, while $v_{[i]} = (v_1, \cdots, v_i)$ will denote the vector of the first i entries and $v_{[i,j]} = (v_i, \cdots, v_j)$, $i \leq j$ will denote the vector of entries between i, j of v. For a finite set $\mathcal{A}$ and $n \in \mathbb{N}$ the n-fold Cartesian product $\mathcal{A}$'s will be denoted by $\mathcal{A}^n$ and $\mathcal{P}(\mathcal{A})$ will denote the simplex of probability distributions over $\mathcal{A}$. For a positive integer n, we shall denote by $\mathcal{A}^{(n)} := \bigcup_{0<s<n} \mathcal{A}^s$ the set of $\mathcal{A}$-strings of length smaller than n. Probability distributions are denoted by $P(\cdot)$ and subscripted by the name of the random variables and conditioning, e.g., $P_{U,T|V,S}(u,t|v,s)$ is the conditional probability of $(U=u,T=t)$ given $(V=v,S=s)$. For any set $\mathcal{A}$, $\mathcal{A}^c$ will denote its complement and its size will be denoted by $|\mathcal{A}|$. We denote the indicator function of an event E by $1_{\{E\}}$. The probability of some events, say $\Pr(E)$, taken under the distribution $P_{X_{[n]}}(\cdot)$ on $\mathcal{X}^n$ shall be interpreted as $\Pr(E) = \sum_{x_{[n]}\in\mathcal{X}^n} P_{X_{[n]}}(x_{[n]})1_{\{E\}}$. All sets considered hereafter are finite.

2.2 Relevant Definitions

In a broad sense, a communication system consists of three parts: The source(s), the destination(s) and the channel(s) which consist of noisy (in general) transmission mediums to transfer the signal from the source(s) to the destination(s) and characterised by the triplet $(\mathcal{X}, p_{\mathcal{C}}(y|x), \mathcal{Y})$ where $\mathcal{X}$ and $\mathcal{Y}$ are the input and output alphabets, respectively. The goal is to reconstruct the source(s) at some or all destination(s) with arbitrary low error probability. Let us make these notions precise in a single source, single destination and single channel scenario. Let M be the source which is uniformly distributed in the finite set $\mathcal{M}$.

Definition 2.2.1. *An $(n, 2^{nR})$ code with blocklength n and rate pair R for a channel, $(\mathcal{X}, p_{\mathcal{C}}(y|x), \mathcal{Y})$, consists of*

(1) A sequence of mappings for the encoder

$$\phi_t : \mathcal{M} \to \mathcal{X}, \;\; t = 1, 2, ..., n;$$

2) An associated decoding function

$$\psi : \mathcal{Y}^n \to \mathcal{M}.$$

The system's probability of error, $P_e^{(n)}$, is given by

$$P_e^{(n)} = \frac{1}{2^{nR}} \sum_{\mathrm{m}=1}^{2^{nR}} \Pr\left(\psi(Y_{[n]}) \neq \mathrm{m} | \mathrm{M} = \mathrm{m}\right).$$

A rate R is achievable if for any $\epsilon > 0$, there exists, for all n sufficiently large an $(n, 2^{nR})$ code such that $\frac{1}{n} \log |\mathcal{M}| \geq R > 0$ and at the same time $P_e^{(n)} \leq \epsilon$. The capacity, C, is defined to be the supremum of all achievable rates (for extending this definition to more than one sources and destinations see, for example, Definition 4.2.1).

For discrete memoryless channels (DMCs), Shannon's noisy channel coding theorem [Sha48] shows that

$$C = \max_{P_X(\cdot)} I(X;Y) \tag{2.1}$$

where $I(X;Y)$ is the mutual information between the random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ where the distribution of the random variable Y, $P_Y(\cdot)$, is induced by $P_X(\cdot)$ and $p_{\mathcal{C}}(y|x)$.

It should be observed that this characterization is complete in the sense that it shows that for $R < \max_{P_X(\cdot)} I(X;Y)$, there exists a code achieving an arbitrarily low error probability for sufficiently large n and for $R > \max_{P_X(\cdot)} I(X;Y)$ there does not exist a code satisfying $\frac{1}{n} \log |\mathcal{M}| \geq R > 0$ and at the same time $P_e^{(n)} \leq \epsilon$ for $n \to \infty$ and $\epsilon \to 0$.

One of the standard tools in obtaining inner bounds, i.e., a set of achievable rates, is to look at the behaviour of the sequences generated under some distribution in the large blocklength regime and to classify these sequences under this behaviour. Particular sets of interest in this domain are referred to as"typical sets". Let us recall two notions of typical sets.

Definition 2.2.2. *[CT06, Section 15.2] Let $k \geq 1$ be finite and $(X^1, \cdots, X^k)$ denote a collection of random variables with some fixed joint distribution $P_{X^1,\cdots,X^k}(x^1, \cdots, x^k)$. Let U denote an ordered subset of $\{X^1, \cdots, X^k\}$ with respect to the indices. Consider n independent copies of U and denote these by $\boldsymbol{U}$. Thus, $\Pr(\boldsymbol{U} = \boldsymbol{u}) = \prod_{i=1}^n P(U_i = u_i)$, $\boldsymbol{u} \in \mathcal{U}^n$. As an example, if $U = (X^j, X^l)$ then*

$$\begin{aligned} \Pr(\boldsymbol{U} = \boldsymbol{u}) &= \Pr(X^j_{[n]}, X^l_{[n]} = x^j_{[n]}, x^l_{[n]}) \\ &= \prod_{i=1}^n P_{X^j_i, X^l_i}(x^j_i, x^l_i). \end{aligned}$$

Then, the set A_ϵ^n of ϵ-typical n-sequences $\{(x^1_{[n]}, \cdots, x^k_{[n]})\}$ with respect to the distribution $P_{X^1,\cdots,X^k}(x^1, \cdots, x^k)$ is defined by

$$\begin{aligned} &\mathrm{A}_\epsilon^n \left(X^1, \cdots, X^k\right) := \mathrm{A}_\epsilon^n \\ &\quad = \Bigg\{(x^1_{[n]}, \cdots, x^k_{[n]}) \in \mathcal{X}_1^n \times \cdots \mathcal{X}_k^n : \\ &\qquad\qquad |-\frac{1}{n}\log\left(P_{\boldsymbol{U}}(\boldsymbol{u})\right) - H(U)| < \epsilon, \forall U \subseteq \{X^1, \cdots, X^k\}\Bigg\} \end{aligned}$$

where $H(\cdot)$ denotes the entropy.

To give an example for this definition, let $\mathrm{A}_\epsilon^n(U)$ denotes the restriction of A_ϵ^n to the coordinates of U and when $U = (X^1, X^2)$, we have

$$\begin{aligned} \mathrm{A}_\epsilon^n(X^1, X^2) = \Bigg\{&(x^1_{[n]}, x^2_{[n]}) : |-\frac{1}{n}\log P_{X^1_{[n]}, X^2_{[n]}}(x^1_{[n]}, x^2_{[n]}) - H(X^1, X^2)| < \epsilon, \\ &|-\frac{1}{n}\log P_{X^1_{[n]}}(x^1_{[n]}) - H(X^1)| < \epsilon,\ \ |-\frac{1}{n}\log P_{X^2_{[n]}}(x^2_{[n]}) - H(X^2)| < \epsilon\Bigg\}. \end{aligned}$$

Definition 2.2.2 can be further weakened and replaced with a classification according to the empirical behaviour of the sequences.

Definition 2.2.3. *[CK81] Let $N(a|k_{[n]})$ denote the number of occurrences of the letter a in the vector $k_{[n]}$. Then, for a given P_K defined over $\mathcal{K}$, the δ-typical set is defined as*

$$\begin{aligned} \mathcal{T}_K^{\delta} \;:=\; \Big\{ k_{[n]} \in \mathcal{K}^n : \;\; | \, n^{-1} N(a|k_{[n]}) - P_K(a) \, | \;\; &\leq \;\delta, \;\; \forall a \in \mathcal{K}, \\ N(a|k_{[n]}) = 0, \;\; \textit{if} \;\; P_K(a) = 0 \Big\} \, . \end{aligned}$$

Similarly, for a given joint distribution $P_{K,L}$, the conditional δ-typical set is

$$\begin{aligned} \mathcal{T}_{K|L}^{\delta} := \Big\{ k_{[n]} \in \mathcal{K}^n : \;\; | \, n^{-1} N(a,b|k_{[n]},l_{[n]}) - n^{-1} N(b|l_{[n]}) P_{K|L}(a|b) \, | \;\; &\leq \;\delta, \\ \forall (a,b) \in \mathcal{K} \times \mathcal{L}, \;\; N(a,b|k_{[n]},l_{[n]}) = 0, \;\; \textit{if} \;\; P_{K|L}(a|b) = 0 \Big\} \, . \end{aligned}$$

We have the following lemma.

Lemma 2.2.1. *[CK81, Lemma 2.10] If $l_{[n]} \in \mathcal{T}_L^{\delta}$ and $k_{[n]} \in \mathcal{T}_{K|L}^{\delta'}$ then $(l_{[n]}, k_{[n]}) \in \mathcal{T}_{K,L}^{\delta+\delta'}$ and consequently $k_{[n]} \in \mathcal{T}_K^{\delta''}$ where $\delta'' := (\delta + \delta')|\mathcal{L}|$.*

Both of the typical sets, i.e., Definition 2.2.2 and Definition 2.2.3, will be used in the thesis and following [CK81], δ in the Definition 2.2.3 will depend on n such that

$$\delta_n \to 0, \;\; \sqrt{n}\delta_n \to \infty, \;\; \text{and} \;\; n \to \infty. \tag{2.2}$$

Furthermore, the following convention will be used throughout the thesis.

Delta-Convention [CK81, Convention 2.11]: To every set $\mathcal{X}$ and respectively ordered pair of sets $(\mathcal{X}, \mathcal{Y})$, there is a given sequence $\{\delta_n\}_{n=1}^{\infty}$ satisfying (2.2). Typical sequences are understood with these δ_n's. The sequences $\{\delta_n\}$ are considered as fixed and dependence on them will be omitted. Accordingly, the δ will be omitted from the notation. In the situations where typical sequences should generate typical ones, we assume that the corresponding δ_n's are chosen according to Lemma 2.2.1.

When a rate pair, (R_a, R_b), is shown to be achievable under typicality decoding, one needs to investigate the behaviour of the probability of the error events of the encoding and decoding processes under the joint typicality coding and decoding. A considerably rich set of results is available in the literature and for the purpose of this thesis, we only present the most relevant ones.

2.3 Two Typicality Lemmas

Arguably, one of the main intuitions under the joint typicality results can be summarized by the fact that when the rate of codebook construction satisfy some criterion, the asymptotic behaviour of the sequences, which are independently and identically distributed generated, are identical. The following result has such an interpretation.

Lemma 2.3.1. *Covering Lemma, [EGK11, Lemma 3.3] Let* $(U, X, \hat{X})$ *distributed according to* $P_{U,X,\hat{X}}(u, x, \hat{x})$ *and* $\delta < \tilde{\delta}$. *Let* $(U_{[n]}, X_{[n]})$, *distributed according to* $P_{U_{[n]},X_{[n]}}(u_{[n]}, x_{[n]})$, *be a pair of arbitrarily distributed random variables such that* $\Pr\left((U_{[n]}, X_{[n]}) \in \mathcal{T}^{\delta}_{U,X}\right) \to 1$ *as* $n \to \infty$ *and let* $\hat{X}_{[n]}(m)$, $m \in \mathcal{A}$, $|\mathcal{A}| \geq 2^{nR}$, *be random sequences conditionally independent of each other and of* $X_{[n]}$ *given* $U_{[n]}$, *each distributed according to* $\prod_{i=1}^{n} P_{\hat{X}_i|U_i}(\hat{x}_i|u_i)$. *Then, there exists* $\eta(\tilde{\delta}) \to 0$ *as* $\tilde{\delta} \to 0$ *such that*

$$\Pr\left((U_{[n]}, X_{[n]}, \hat{X}_{[n]}(m)) \notin \mathcal{T}^{\tilde{\delta}}_{U,X,\hat{X}},\ \ \forall m \in \mathcal{A}\right) \to 0$$

as $n \to 0$, *if* $R > I(X; \hat{X}|U) + \eta(\tilde{\delta})$.

This lemma shows that if we consider the random sequences $\hat{X}_{[n]}(m)$, $m \in \mathcal{A}$, $|\mathcal{A}| \geq 2^{nR}$, as the reproduction of the source sequences $X_{[n]}$ and if $R > I(X; \hat{X}|U)$ then there is at least one reproduction sequence which is jointly typical with the source sequence. The proof follows from the properties of joint typical sets; see [EGK11], which we omit.

One can alternatively characterize the probability of the event that independent (or conditionally independent) sequences being jointly typical. By using the properties of mutual information and joint entropy, the following result can be obtained.

Lemma 2.3.2. *[Gas04, Lemma 8] If* $(\tilde{X}_{[n]}, \tilde{Y}_{[n]}, \tilde{Z}_{[n]}) \sim P_{X_{[n]}}(x_{[n]})P_{Y_{[n]}}(y_{[n]})P_{Z_{[n]}}(z_{[n]})$, *i.e., they have the same marginals as* $P_{X_{[n]},Y_{[n]},Z_{[n]}}(x_{[n]}, y_{[n]}, z_{[n]})$ *but they are independent, then*

$$\Pr\left(\left(\tilde{X}_{[n]}, \tilde{Y}_{[n]}, \tilde{Z}_{[n]}\right) \in \mathrm{A}_{\epsilon}^{n}\right) \leq 2^{-n[I(Z,X;Y)+I(Z,Y;X)-I(X;Y|Z)-\delta]}$$

for some $\delta \to 0$ *as* $\epsilon \to 0$ *and* $n \to \infty$.

It is worth to present a short proof of the above lemma, which can be shown, for example, following similar steps as in [CT06, Theorem 15.2.3], for a better interpretation of the result. Note that

$$\begin{aligned}\Pr\left(\left(\tilde{X}_{[n]}, \tilde{Y}_{[n]}, \tilde{Z}_{[n]}\right) \in \mathcal{A}_{\epsilon}^{n}\right) &= \sum_{(x_{[n]},y_{[n]},z_{[n]})\in\mathcal{A}_{\epsilon}^{n}} P_{X_{[n]}}(x_{[n]})P_{Y_{[n]}}(y_{[n]})P_{Z_{[n]}}(z_{[n]}) \\ &\overset{(i)}{\leq} 2^{n(H(X,Y,Z)+\epsilon)}2^{-n(H(X)-\epsilon)}2^{-n(H(Y)-\epsilon)}2^{-n(H(Z)-\epsilon)} \\ &= 2^{-n(H(X)+H(Y)+H(Z)-H(X,Y,Z)-4\epsilon)}\end{aligned}$$

where (i) follows from [CT06, Theorem 15.2.1]. The proof is complete by observing that the exponent in the last step can be written as the exponent in the lemma.

The above two lemmas will be invoked in Chapter 5.

While the above typicality definition and related lemmas are required for the achievability proof, the following theorem, which relates the error probability to the entropy, is used in the converse proofs, i.e., the proofs that show one cannot do better, of many coding theorems as well as in the results of this thesis.

Theorem 2.3.1 (Fano's Inequlaity). *[EGK11, Section 2.1] Let (X, Y) be distributed according to $P_{X,Y}(x, y)$ and $P_e := \Pr(X \neq Y)$. Then*

$$H(X|Y) \leq H(P_e) + P_e \log |\mathcal{X}| \leq 1 + P_e \log |\mathcal{X}|.$$

2.4 Channel Coding Results

2.4.1 Channel Coding with Feedback

The extension of the Definition 2.2.1 to the other scenarios is done by letting the decision makers, i.e., the encoder(s) and the decoder(s), use the information provided to them to generate their output. In this manner, feedback information theory considers the situation that the encoder(s) are provided with past channel outputs. Since the encoder can learn and estimate the channel by observing the channel outputs, it is expected that feedback may increase capacity. However, Shannon [Sha56] showed that for DMCs, the feedback and non-feedback capacities are identical and hence, feedback does not increase the capacity. This result can be shown as follows. Consider a DMC defined by $(\mathcal{X}, p_{\mathcal{C}}(y|x), \mathcal{Y})$ and let C_{FB} denote its feedback capacity. The feedback capacity is defined to be the supremum of all achievable rates under the feedback coding policy; $X_i = \psi_i\left(W, Y_{[i-1]}\right)$ where $\{\psi_i : \{1, 2, ..., 2^{nR}\} \times \mathcal{Y}^{i-1} \to \mathcal{X}\}_{i=1}^n$ are the encoding functions, W is the message (governed by a uniform distribution), $Y_{[i-1]}$ are the past channel outputs and X_i is the channel input at time i. It is obvious that $C_{FB} \geq C$. On the other hand, from Fano's inequality and following the standard steps of the converse to the channel coding theorem, we have for any achievable R

$$nR = H(W) = H(W|Y_{[n]}) + I(W; Y_{[n]}) \leq 1 + {P_e}^{(n)} nR + I(W; Y_{[n]}).$$

Observe now that

$$\begin{aligned} I(W;Y_{[n]}) &= \sum_{i=1}^{n} H(Y_i|Y_{[i-1]}) - H(Y_i|Y_{[i-1]},W) \\ &\overset{(i)}{=} \sum_{i=1}^{n} H(Y_i|Y_{[i-1]}) - H(Y_i|Y_{[i-1]},W,X_i) \\ &\overset{(ii)}{=} \sum_{i=1}^{n} H(Y_i|Y_{[i-1]}) - H(Y_i|X_i) \\ &\overset{(iii)}{\leq} \sum_{i=1}^{n} H(Y_i) - H(Y_i|X_i) \\ &\overset{(iv)}{\leq} nC \end{aligned}$$

where (i) is due to $X_i = \psi_i \left(W, Y_{[i-1]}\right)$, (ii) is due to channel being memoryless, (iii) is due to the fact that conditioning does not increase the entropy and (iv) follows from (2.1). Hence, $nR \leq 1 + P_e{}^{(n)}nR + nC$ and dividing both sides by n and taking the $n \to \infty$ yields $R \leq C$ and $C_{FB} = C$. Note that the critical step in the above derivation is that the term $Y_{[i-1]}$ could be ignored without loss of optimality. When the channel has memory, this action can not be performed. For a derivation of a converse for channels with memory see Chapter 3.

Notice that the situation for the memoryless multi-user channels with output feedback is completely different. Feedback can increase the capacity in the multi-user scenarios; e.g., see [GW75] and [CL81]. This is because in the multi-user setup the encoders can cooperate via the channel output feedback to remove the uncertainty of the decoder about the messages. Although this cooperation based coding schemes yields nice inner bounds, for many of the multi-user channels with feedback, complete characterizations for the capacity regions are not available.

In some situations, such as the case where the feedback channel is restrictive, the decoder might wish actively to send his local information on the channel to the encoder who may not have access to the same information. These types of problems have been widely studied and in the next section we provide a brief discussion for a set of these results.

2.4.2 Channel Coding with Side Information

Consider a channel whose conditional output probability distribution is controlled by a process, called state. These type of channels can help model many problems, depending on some assumptions regarding the channel state and on the availability and quality (complete or partial) of the side information (CSI) at the transmitter (CSIT) and/or the receiver (CSIR) [KSM07]. From the transmitter's perspective, a fundamental difference appears if the side information is available in a causal or non-causal manner. In the causal case, at each time instant, the encoder can only use the past and the current CSI (which can be complete or partial), whereas in the non-causal case, the transmitter knows in advance the realization of the entire state sequence from the beginning to the end of the block. The causal model is first introduced by Shannon [Sha58] and he derived an expression for the capacity. We now discuss this result in more detail. We consider the channel depicted in Figure 2.1. The channel input, output and state process belong to the finite sets, $\mathcal{X}, \mathcal{Y}$ and $\mathcal{S}$, respectively. The setup considered in [Sha58] assumes the state is memoryless, independent of W and hence, we have

$$P_{Y_{[n]}|X_{[n]},S_{[n]}}(y_{[n]}|x_{[n]}, s_{[n]}) = \prod_{t=1}^{n} P_{Y_t|X_t,S_t}(y_t|x_t, s_t) \tag{2.3}$$

$$P_{S_{[n]}}(s_{[n]}) = \prod_{t=1}^{n} P_{S_t}(s_t). \tag{2.4}$$

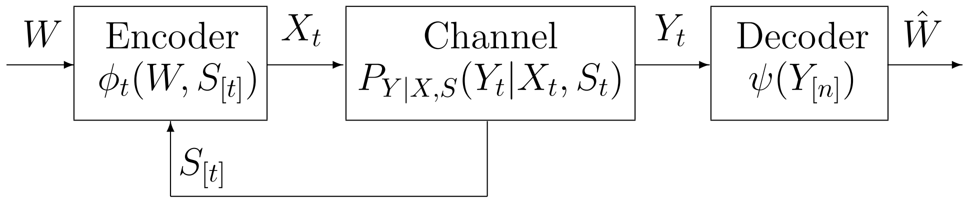

Figure 2.1: Single-user channel with causal side information

For this communication a code is defined similar to the Definition 2.2.1 except the encoding function is changed to

$$\phi_t : \mathcal{W} \times \mathcal{S}^t \rightarrow \mathcal{X}, \ \ t = 1, 2, ...n.$$

The capacity of this channel is given by

$$C = \max_{P_T(\cdot)} I(T; Y) \tag{2.5}$$

where the random variable $T \in \mathcal{T}$ represents a random mapping from the channel state process to the channel input process and $\mathcal{T}$ denotes the set of all possible mappings; $|\mathcal{T}| = |\mathcal{X}|^{|\mathcal{S}|}$, $P_T(\cdot)$ is a distribution on $\mathcal{T}$ and the joint distribution of (Y, S, X, T) satisfies

$$P_{Y,S,X,T}(y, s, x, t) = P_{Y|X,S}(y|t(s), s)1_{\{x=t(s)\}}P_S(s)P_T(t). \tag{2.6}$$

Once we have the capacity result for an ordinary DMC, this result follows from the following observation. Note that $X_t = T_t(S_t)$ and assuming that T is independent of S, we have

$$P_{Y_{[n]}|T_{[n]}}(y_{[n]}|t_{[n]}) \ \ = \ \ \sum_{s_{[n]}} P_{Y_{[n]}|S_{[n]},T_{[n]}}(y_{[n]}|s_{[n]}, t_{[n]})P_{S_{[n]}}(s_{[n]})$$

$$= \sum_{s_{[n]}} \prod_{t=1}^{n} P_{Y_t|Y_{[t-1]},S_{[n]},T_{[n]}}(y_t|y_{[t-1]}, s_{[n]}, t_{[n]}) P_{S_t}(s_t)$$

$$= \sum_{s_{[n]}} \prod_{t=1}^{n} P_{Y_t|S_t,X_t}(y_t|s_t, t_t(s_t)) P_{S_t}(s_{[t]}) \tag{2.7}$$

$$= \prod_{t=1}^{n} P_{Y_t|T_t}(y_t|t_t). \tag{2.8}$$

Therefore, we can define an equivalent (memoryless) channel between $T \in \mathcal{T}$ and $Y \in \mathcal{Y}$ and described explicitly by $P_{Y_t|T_t}(y_t|t_t) = \sum_{s_t \in \mathcal{S}} P_{Y_t|S_t,X_t}(y_t|s_t, t_t(s_t)) P_{S_t}(s_{[t]})$. In order to complete the proof, which follows from (2.1), we need to identify the random variable T. This follows from the fact that since for all $t \geq 1$, $X_t = \phi_t \left(\mathrm{W}, S_{[t]}\right) = \phi_t \left(\mathrm{W}, S_{[t-1]}, S_t\right)$, we can define $T_t \in \mathcal{T}$ satisfying (2.6) by putting, for every $s \in \mathcal{S}$,

$$T_t(s) := \phi_t \left(\mathrm{W}, S_{[t-1]}, s\right). \tag{2.9}$$

Based on this, the above result demonstrates that the capacity of this channel is equal to the capacity of an ordinary DMC, with the same output alphabet but with an extended input alphabet. A particular realization of T, say $t : \mathcal{S} \to \mathcal{X}$, is termed as *Shannon strategy.*

In [Sal92], Shannon's result is extended to the case where the transmitter and the receiver has access to noisy causal CSI and the capacity is derived. However, as shown in [CS99], one can define an equivalent channel that is governed by the process available at the transmitter, and therefore, the result in [Sal92] can be shown to reduce to the Shannon's result. More explicitly, assuming that (2.3) and (2.4) still hold, $X_t = \phi_t \left(\mathrm{W}, \hat{S}_{[t]}\right)$, $\hat{\mathrm{W}} = \psi \left(Y_{[n]}, \check{S}_{[n]}\right)$, i.e, the receiver has also access to the noisy CSI, and

$$P_{\hat{S}_{[n]},\check{S}_{[n]},S_{[n]}}(\hat{s}_{[n]}, \check{s}_{[n]}, s_{[n]}) = \prod_{i=1}^{n} P_{\hat{S}_t,\check{S}_t,S_t}(\hat{s}_t, \check{s}_t, s_t), \tag{2.10}$$

where $\hat{S}_t \in \hat{\mathcal{S}}, \check{S}_t \in \check{\mathcal{S}}$. In [Sal92] the capacity for this setup is shown to be

$$C = \max_{P_{\hat{T}}(\cdot)} I(T; Y|\check{S}) \tag{2.11}$$

where $\hat{T} \in \hat{\mathcal{T}}$, $|\hat{\mathcal{T}}| = |\mathcal{X}|^{|\hat{\mathcal{S}}|}$, represents a random mapping between $\hat{\mathcal{S}}$ and $\mathcal{X}$, and the joint distribution of $(Y, S, \hat{S}, \check{S}, X, T)$ satisfies

$$P_{Y,S,\hat{S},\check{S},X,T}(y, s, \hat{s}, \check{s}, x, t) = P_{Y|X,S}(y|x, s)1_{\{x=t(\hat{s})\}}P_{S,\hat{S},\check{S}}(s, \hat{s}, \check{s})P_T(t). \tag{2.12}$$

In order to demonstrate that the above setup is no more general than Shannon's setup in which perfect CSI is available at the transmitter, define $\check{Y} := (Y, \check{S})$ as the modified output and consider the following channel conditional distribution

$$\begin{aligned} P_{Y,\check{S}|\hat{S},X}(y, \check{s}|\hat{s}, x) &= \sum_{s\in\mathcal{S}} P_{Y,\check{S}|\hat{S},X,S}(y, \check{s}|\hat{s}, x, s)P_{S|\hat{S},X}(s|\hat{s}, x) \\ &= \sum_{s\in\mathcal{S}} P_{Y|X,S}(y|x, s)P_{S,\check{S}|\hat{S}}(s, \check{s}|\hat{s}). \end{aligned} \tag{2.13}$$

Observing that $I(T; \check{S}) = 0$, it is now clear to see that the channel described by (2.13) is the same type as that is studied by Shannon [Sha58].

In addition to assuming that the transmitter has causal side information, there might be some scenarios where the transmitter has non-causal side information. This problem is first motivated by the modelling of defective cells, where the positions of the defective cells are considered to be the CSI, in computer memory [KT74] and later the capacity for this model is determined in [GP80].

For the problem definition, we keep the setup of causal CSI, (2.3) and (2.4) also hold, but now we have $X_t = \phi_t\left(W, S_{[n]}\right)$. Then, the capacity for this channel is given by

$$C = \max_{P_{U,X|S}} [I(U; Y) - I(U; S)] \tag{2.14}$$

where U is an auxiliary random variable with cardinality $\mathcal{U} \leq |\mathcal{X}||\mathcal{S}|+1$ and the joint distribution of (Y, U, S, X) satisfies the Markov condition $U \to (X, S) \to Y$. We now briefly describe the achievability and the converse proof of this result.

The novel approach in the converse of (2.14) is to introduce an auxiliary random variable. More explicitly, from Fano's inequality and following standard steps one can obtain $R \leq \frac{1}{n}\sum_{i=1}^{n} I(\mathrm{W}; Y_i|Y_{[i-1]}) + \epsilon_n$ where $\epsilon_n \to 0$ as $n \to 0$. Then

$$\begin{aligned}
\sum_{i=1}^{n} I(\mathrm{W}; Y_i|Y_{[i-1]}) &\leq \sum_{i=1}^{n} I(\mathrm{W}, Y_{[i-1]}; Y_i) \\
&= \sum_{i=1}^{n} \left[I(\mathrm{W}, Y_{[i-1]}, S_{[i+1,n]}; Y_i) - I(S_{[i+1,n]}, Y_i|\mathrm{W}, Y_{[i-1]})\right] \\
&= \sum_{i=1}^{n} \left[I(\mathrm{W}, Y_{[i-1]}, S_{[i+1,n]}; Y_i) - I(S_i; Y_{[i-1]}|\mathrm{W}, S_{[i+1,n]})\right] \\
&= \sum_{i=1}^{n} \left[I(\mathrm{W}, Y_{[i-1]}, S_{[i+1,n]}; Y_i) - I(S_i; Y_{[i-1]}, \mathrm{W}, S_{[i+1,n]})\right]
\end{aligned}$$

where the third step follows from the Csiszár sum identity [EGK11] and the last step is valid since S_i is independent of $(\mathrm{W}, S_{[i+1,n]})$. Now, let $U_i := (\mathrm{W}, S_{[i+1,n]})$ and hence,

$$R \leq \frac{1}{n}\sum_{i=1}^{n} \left[I(U_i; Y_i) - I(U_i; S_i)\right] + \epsilon_n \tag{2.15}$$

where as desired $U_i \to (X_i, S_i) \to Y_i \ \forall i = 1, \cdots, n$. This implies

$$R \leq \max_{P_{U,X|S}} \left[I(U; Y) - I(U; S)\right] + \epsilon_n. \tag{2.16}$$

The proof is completed by invoking the support lemma [CK81] to make the cardinality of U finite.

For the achievability of (2.14), the idea of random binning is used. This scheme was applied to many problems in information theory and the main motivation of the binning idea is to distribute a set of codewords into subsets (called bins) and these bins themselves are also randomly constructed according to a specified rate which are indexed by a different message. To make this argument more precise, let us show that a rate $R_{nc} \leq \max_{P_{U,X|S}}[I(U;Y) - I(U;S)]$ is achievable via random binning for the above problem. We first construct the codewords; for each message $\mathrm{w} \in \{1, \cdots, 2^{nR_{nc}}\}$, generate 2^{nR_b} independent and identically distributed codewords $u_{[n]}$ according to $P_U(u)$. These codewords form the bins. Index these codewords by $\{u_{[n]}[\mathrm{w}, 1], \cdots, u_{[n]}[\mathrm{w}, 2^{nR_b}]\}$. These codewords are distributed to the encoder and the decoder. Given the state sequence $s_{[n]}$, in order to send the message w, the encoder looks for a codeword in bin w that is jointly typical with $s_{[n]}$, say $u_{[n]}[\mathrm{w}, \mathrm{l}]$. In the case when there is no codeword that is jointly typical, the encoder declares an error. In the case there are more than one codewords, the encoder picks the codeword with the smallest bin index. The encoder then generates the channel input according to the $x_t = \phi_t(u_i[\mathrm{w}, \mathrm{l}], s_i)$, $i = 1, \cdots, n$. The decoding is done by joint typical decoding. In particular, after observing $y_{[n]}$, the decoder looks for the pair of indices $(\hat{\mathrm{w}}, \hat{\mathrm{l}})$ such that $u_{[n]}[\hat{\mathrm{w}}, \hat{\mathrm{l}}]$ is jointly typical with $y_{[n]}$. If there are more than one pair $(\hat{\mathrm{w}}, \hat{\mathrm{l}})$ or there are no indices that are jointly typical then the decoder declares an error. Observe now that there are two error events; the encoder and the decoder error event: By the covering lemma, Lemma 2.3.1, the encoder's error event goes to zero if $R_b > I(U;S)$ and by the joint typicality lemmas the error event at the decoder goes to zero if $R_{nc} + R_b < I(U;Y)$. These two together imply that $R_{nc} \leq \max_{P_{U,X|S}}[I(U;Y) - I(U;S)]$ is achievable.

Finally, it is worth to comment on the extension of the results for the single-user channels that we presented above to the case of multi-user channels. Foremost, in multi-user channels, the associated problems are mostly open. However, it is known that the optimal structures of the solutions, such as Shannon strategis, do not hold in general when there are more than one user. Secondly, depending on the relation between the CSI at the encoders as well as at the CSI at decoder, there might be an initiative for the encoders to cooperate and hence increase the rates. Chapters 4 and 5 of the thesis discuss such scenarios and results in more detail.

Chapter 3

Single User Channels with Feedback

In this chapter, we discuss feedback capacity of a class of finite-state Markov (FSM) channels [Gal68] which encompass symmetry in their channel transition matrices. Along this way, we first define symmetry for channels with memory, referred to as quasi-symmetric FSM channels, and then show the existence of a hidden Markov noise process, due to the symmetry characteristics of the channel, which is conditionally independent of the input given the sate. As a result, the FSM channel can be succinctly described as a function of input and noise, where the function is an invertible map between the noise and output alphabets for a fixed input. With this fact, the feedback capacity problem reduces to the maximization of entropy of the output process. In the second step, we show that this entropy is maximized by a uniform input distribution. It should be noted that for quasi-symmetric FSM channels, uniform inputs do not necessarily yield uniform outputs; this is a key symmetry property used in previous works for showing that feedback does not increase capacity for symmetric channels with memory (e.g., [Ala95],[AF94]). Throughout this chapter, we will represent a finite-state Markov source by a pair $[\mathcal{S}, P]$, where $\mathcal{S}$ is the state set and P is the state transition probability matrix. We will also be assuming that the Markov processes in the paper are stationary, aperiodic and irreducible (hence ergodic).

3.1 Feedback Capacity of Symmetric FSM Channels

When the communication channel is modeled via a state process, such as FSM channels, it might be assumed that some degree of CSI is available at the encoder and/or decoder. In the rest of this section, it is assumed that CSI is not available at the encoder and decoder. We treat the scenario where CSI is fully known at the decoder in the next secrtion.

3.1.1 Quasi-Symmetric FSM Channel

A finite-state Markov (FSM) channel is defined by a pentad $[\mathcal{X}, \mathcal{Y}, \mathcal{S}, P_S, \mathcal{C}]$, where $\mathcal{X}$ is the input alphabet, $\mathcal{Y}$ is the output alphabet, and the Markov process $\{S_n\}_{n=1}^{\infty}$, $S_n \in \mathcal{S}$ is represented by the pair $[\mathcal{S}, P_S]$ where $\mathcal{S}$ is the state set and P_S is the state transition probability matrix. We assume that the sets $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{S}$ are all finite. The set $\mathcal{C}$ is a collection of transition probability distributions, $p_{\mathcal{C}}(y|x, s)$, on $\mathcal{Y}$ for each $x \in \mathcal{X}$, $s \in \mathcal{S}$. We consider the problem of communicating message $W \in \{1, 2, \cdots, 2^{nR}\}$ over the FSM channel (without or with the use of feedback) via a code of rate R and blocklength n,[1] where W is uniformly distributed over $\{1, 2, \cdots, 2^{nR}\}$ and independent of $S_{[n]}$. We assume that the FSM channel satisfies the following properties under both the absence and presence of feedback:

(I) Markov Property: For any integer $i \geq 1$.

$$P_{S_i|S_{[i-1]},Y_{[i-1]},X_{[i-1]},W}(s_i|s_{[i-1]}, y_{[i-1]}, x_{[i-1]}, w) = P_{S_i|S_{i-1}}(s_i|s_{i-1}). \tag{3.1}$$

(II) For any integer $i \geq 1$,

$$P_{Y_i|,S_i,X_i,S_{[i-1]},X_{[i-1]},Y_{[i-1]},W}(y_i|s_i, x_i, s_{[i-1]}, x_{[i-1]}, y_{[i-1]}, w) = p_{\mathcal{C}}(y_i|s_i, x_i) \tag{3.2}$$

where $p_{\mathcal{C}}(.|., .)$ is defined by $\mathcal{C}$. When the channel is without feedback, we also assume that the FSM channel satisfies:

(II.b) For any integer $i \geq 1$,

$$P_{Y_{[i-1]}|X_{[i]},S_{[i]}}(y_{[i-1]}|x_{[i]}, s_{[i]}) = P_{Y_{[i-1]}|X_{[i-1]},S_{[i-1]}}(y_{[i-1]}|x_{[i-1]}, s_{[i-1]}). \tag{3.3}$$

Note that due to Properties II and II.b, $P_{Y_{[n]}|X_{[n]},S_{[n]}}(y_{[n]}|x_{[n]}, s_{[n]}) = \prod_{i=1}^{n} p_{\mathcal{C}}(y_i|s_i, x_i)$ when the channel is without feedback. Furthermore, the non-feedback codewords $X_{[n]}$ at the channel input are only a function of W (which is independent of $S_{[n]}$); hence, in the non-feedback scenario, the channel input $\{X_i\}$ is also independent of $S_{[n]}$.

We are interested in a subclass of FSM channels where the channel transition matrices, $Q^s \stackrel{\triangle}{=} [p_{\mathcal{C}}(y|s, x)]_{xy}$, $s \in \mathcal{S}$, carry some notion of symmetry which is similar to the symmetry defined for DMCs as in the following.

Definition 3.1.1. *A DMC with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and channel transition matrix $Q = [p_{\mathcal{C}}(y|x)]_{xy}$ is symmetric if the rows of Q are permutations of each other and the columns are permutations of each other [CT06].*

[1] Both feedback and non-feedback codes of rate R and blocklength n, which yield up to 2^{nR} codewords $X_{[n]} \in \mathcal{X}^n$ for transmission over the channel, are explicitly defined in Section 3.1.2 in terms of a pair of encoding and decoding functions.

Definition 3.1.2. *A DMC with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and channel transition matrix $Q = [p_\mathcal{C}(y|x)]_{xy}$ is weakly-symmetric if the rows of Q are permutations of each other and all the column sums $\sum_x p_\mathcal{C}(y|x)$ are identically equal to a constant [CT06].*

Definition 3.1.3. *A DMC with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and channel transition matrix $Q = [p_\mathcal{C}(y|x)]_{xy}$ is quasi-symmetric if Q can be partitioned along its columns into weakly-symmetric* sub-arrays, $\tilde{Q}_1, \tilde{Q}_2, \ldots, \tilde{Q}_m$, with each $\tilde{Q}_i$ having size $|\mathcal{X}| \times |\mathcal{Y}_i|$, where $\mathcal{Y}_1 \cup \cdots \cup \mathcal{Y}_m = \mathcal{Y}$ and $\mathcal{Y}_i \cap \mathcal{Y}_j = \emptyset$, $\forall i \neq j$ [Ala99]. A *weakly-symmetric* sub-array is a matrix whose rows are permutations of each other and whose column sums are all identically equal to a constant.

Note that for a quasi-symmetric DMC, the rows of its entire transition matrix, Q, are also permutations of each other. It is also worth pointing out that the above quasi-symmetry[2] notion for DMCs encompasses Gallager's symmetry definition [Gal68, p.94]. A simple example of a quasi-symmetric DMC can be given by the following (stochastic, i.e., with row sums equal to 1) transition matrix, Q, for which $a_1 + a_2 = 2a_3$ and $a_4 + a_5 = 2a_6$, and it can be partitioned along its columns into two weakly-symmetric sub-arrays

$$Q = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ a_3 & a_2 & a_1 & a_6 & a_5 & a_4 \\ a_2 & a_1 & a_3 & a_5 & a_4 & a_6 \\ a_3 & a_1 & a_2 & a_6 & a_4 & a_5 \end{bmatrix}, \quad \tilde{Q}_1 = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_3 & a_2 & a_1 \\ a_2 & a_1 & a_3 \\ a_3 & a_1 & a_2 \end{bmatrix}, \text{ and } \tilde{Q}_2 = \begin{bmatrix} a_4 & a_5 & a_6 \\ a_6 & a_5 & a_4 \\ a_5 & a_4 & a_6 \\ a_6 & a_4 & a_5 \end{bmatrix}.$$

[2] The capacity of a quasi-symmetric DMC is achieved by a uniform input distribution and it can be expressed via a simple closed-form formula [Ala99]: $C = \sum_{i=1}^m \alpha_i C_i$ where $\alpha_i \triangleq \sum_{y \in \mathcal{Y}_i} P(y|x) =$ sum of any row in $\tilde{Q}_i$, $i = 1, \cdots, m$, and $C_i = \log_2 |\mathcal{Y}_i| - H\left(\text{any row in the matrix} \frac{1}{\alpha_i}\tilde{Q}_i\right)$, $i = 1, \cdots, m$.

We can now define similar notions of symmetry for FSM channels.

Definition 3.1.4. *(e.g., [Rez06, GV96]) An FSM channel is symmetric if for each state $s \in \mathcal{S}$, the rows of Q^s are permutations of each other such that the row permutation pattern is identical for all states, and similarly, if for each $s \in \mathcal{S}$ the columns of Q^s are permutations of each other with an identical column permutation pattern across all states.*

By considering the identical permutation pattern across the states, the above definition can be extended to the other types of symmetries as follows.

Definition 3.1.5. *An FSM channel is weakly-symmetric if for each state $s \in \mathcal{S}$, Q^s is weakly-symmetric and the row permutation pattern is identical for all states.*

Definition 3.1.6. *An FSM channel is quasi-symmetric if for each state $s \in \mathcal{S}$, Q^s is quasi-symmetric and the row permutation pattern is identical for all states.*

To illustrate these definitions, let us consider the following conditional probability matrices of a two-state quasi-symmetric FSM channel with $\mathcal{X} = \{1,2,3,4\}$, $\mathcal{Y} = \{1,2,3,4,5,6\}$ and $\mathcal{S} = \{1,2\}$:

$$Q^1 = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ a_3 & a_2 & a_1 & a_6 & a_5 & a_4 \\ a_2 & a_1 & a_3 & a_5 & a_4 & a_6 \\ a_3 & a_1 & a_2 & a_6 & a_4 & a_5 \end{bmatrix}, \quad Q^2 = \begin{bmatrix} a'_1 & a'_2 & a'_3 & a'_4 & a'_5 & a'_6 \\ a'_3 & a'_2 & a'_1 & a'_6 & a'_5 & a'_4 \\ a'_2 & a'_1 & a'_3 & a'_5 & a'_4 & a'_6 \\ a'_3 & a'_1 & a'_2 & a'_6 & a'_4 & a'_5 \end{bmatrix}, \tag{3.4}$$

where Q^1 and Q^2 are stochastic matrices. As it can be seen, Q^1 and Q^2 have the same row permutation pattern and are both quasi-symmetric.

It directly follows by definition that symmetric and weakly symmetric FSM channels are special cases of quasi-symmetric FSM channels. Therefore, we focus on quasi-symmetric FSM channels for the sake of generality.

Let us define $\mathcal{Z}$ (which will serve as a noise alphabet) such that $|\mathcal{Y}| = |\mathcal{Z}|$. Then for each state s, since the rows of Q^s are permutations of each other, we can find functions $f_s(.) : \mathcal{Z} \to [0,1]$ and $\Phi_s(.,.) : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ that are onto given x (i.e., for each $x \in \mathcal{X}$, $\Phi_s(x,.) : \mathcal{Y} \to \mathcal{Z}$ is onto), such that

$$f_s(\Phi_s(x,y)) = p_{\mathcal{C}}(y|x,s). \tag{3.5}$$

Note that since each function $\Phi_s(x,.) : \mathcal{Y} \to \mathcal{Z}$ is onto given x and since $|\mathcal{Y}| = |\mathcal{Z}|$, then it is also one-to-one given x; i.e., $\Phi_s(x,y) = \Phi_s(x,y') \Rightarrow y = y'$. Thus $\Phi_s(x,.) : \mathcal{Y} \to \mathcal{Z}$ is invertible for each $x \in \mathcal{X}$.

For the sake of completeness, we herein provide an explicit construction for the functions $f_s(.)$ and $\Phi_s(.,.)$. The construction is basically as follows: for each (x,y) pair having identical channel conditional probability $p_{\mathcal{C}}(y|x,s)$ under state s, $\Phi_s(x,y)$ returns the same value z with $f_s(z)$ set to equal $p_{\mathcal{C}}(y|x,s)$. More explicitly, let $\mathcal{X} = \{x_{(1)}, x_{(2)}, \cdots, x_{(k)}\}$, $\mathcal{Y} = \{y_{(1)}, y_{(2)}, \cdots, y_{(|\mathcal{Y}|)}\}$, $\mathcal{Z} = \{z_{(1)}, z_{(2)}, \cdots, z_{(|\mathcal{Y}|)}\}$, $I = \{1, 2, \cdots, k\}$ and $J = \{1, 2, \cdots, |\mathcal{Y}|\}$. For $s \in \mathcal{S}$, let $q^s_{i,j} \stackrel{\triangle}{=} p_{\mathcal{C}}(y_{(j)}|x_{(i)}, s)$, $i \in I$ and $j \in J$, be the entries of Q^s. Since Q^s is quasi-symmetric, then for each $i = 1, 2, \cdots, k$, there exists a permutation $\pi^s_i : J \to J$ on the column indices of the entries of the ith row of Q^s such that the first row of Q^s is a permutation of every other row. The row permutations are as follows. The first permutation π_1 is set as the identity function: $\pi^s_1(j) = j$ for all $j \in J$. The remaining permutations for $i = 2, \cdots, k$, are given by $\pi^s_i(1) = k$ where k is the smallest integer in J for which $q^s_{1,k} = q^s_{i,1}$, and for $j = 2, \cdots, |\mathcal{Y}|$, $\pi^s_i(j) = k'$ where k' is the smallest available (not yet assigned for values $1, 2, \ldots, j-1$) integer in J for which $q^s_{1,k'} = q^s_{i,j}$. This assignment rule is valid whether or not the rows of Q^s contain identical entries. Specifically, if the ith row of Q^s $(i \geq 2)$ has d identical entries $q^s_{i,j_1} = q^s_{i,j_2} = \cdots = q^s_{i,j_d}$ with $j_1 < j_2 < \cdots < j_d$

in J, then (by the channel's row symmetry) there exist integers $l_1 < l_2 < \cdots < l_d$ in J with $q^s_{i,j_1} = q^s_{i,j_2} = \cdots = q^s_{i,j_d} = q^s_{1,l_1} = q^s_{1,l_2} = \cdots = q^s_{1,l_d}$. In this case we set: $\pi^s_i(j_t) = l_t$ for $t = 1, 2, \ldots, d$, and $\pi^s_i(j) = \tilde{k}$ where $\tilde{k}$ is the *unique* integer in J for which $q^s_{1,\tilde{k}} = q^s_{i,j}$ for $j \in J \setminus \{j_1, j_2, \cdots, j_d\}$. Then, $f_s(.)$ and $\Phi_s(.,.)$ are given as follows: $\Phi_s(x_{(i)}, y_{(j)}) = z_{(\pi^s_i(j))}$ and $f_s(z_{(j)}) = q^s_{1,j}$, $i \in I$, $j \in J$.

Lemma 3.1.1. *The function $\Phi_s(.,.)$, as defined above together with $f_s(.)$ to satisfy (3.5), is invariant with s.*[3]

Proof. It directly follows from the construction that $\Phi_s(x_{(i)}, y_{(j)}) = z_{(\pi^s_i(j))} = z_{(\pi^{\tilde{s}}_i(j))} = \Phi_{\tilde{s}}(x_{(i)}, y_{(j)})$, $\forall s, \tilde{s} \in \mathcal{S}$ and $\forall x_{(i)} \in \mathcal{X}, y_{(j)} \in \mathcal{Y}$ since by Definition 3.1.6, $\pi^s_i(j)$ is identical for all states. □

Therefore, for a quasi-symmetric FSM channel, there exists a function $\Phi(.,.) : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ that is invertible given x (i.e., for each $x \in \mathcal{X}$, $\Phi(x,.) : \mathcal{Y} \to \mathcal{Z}$ is invertible) such that the random variable $Z = \Phi(X, Y)$ has the conditional distribution

$$\begin{aligned} P_{Z|X,S}(z|x,s) &= \frac{P_{Z,X,S}(z,x,s)}{P_{X,S}(x,s)} = \frac{P_{Y,Z,X,S}(y,z,x,s)}{P_{X,S}(x,s)} \\ &= \frac{P_{Z|X,Y,S}(z|x,y,s)p_{\mathcal{C}}(y|x,s)P_{X,S}(x,s)}{P_{X,S}(x,s)} \\ &\overset{(a)}{=} p_{\mathcal{C}}(y|x,s) = f_s(z). \end{aligned} \tag{3.6}$$

where $y = \nu(x, z)$ and $\nu(.,.) : \mathcal{X} \times \mathcal{Z} \to \mathcal{Y}$ is the inverse of Φ in the sense that $\nu(x,\cdot) = \Phi(x,\cdot)^{-1}$ for each $x \in \mathcal{X}$, and (a) is due to the fact that $P_{Z|X,Y,S}(z|x,y,s) = 1$. This important observation first given in [Rez06], reduces the set of conditional probability distributions which identifies the quasi-symmetric FSM channel to an $|\mathcal{S}| \times |\mathcal{Z}|$ matrix

[3] In the rest of thesis we use $\Phi(\cdot,\cdot)$ instead of $\Phi_s(\cdot,\cdot)$.

T defined by

$$T[s, z] = f_s(z). \tag{3.7}$$

Therefore, for quasi-symmetric FSM channels, we have that for any n,

$$P_{Z_n|X_n,S_n}(z_n|x_n, s_n) = P_{Z_n|S_n}(z_n|s_n) = T\left[s_n, z_n\right]. \tag{3.8}$$

To make this statement explicit, let us consider the FSM channel given in (3.4). For this channel, we can derive the functions $z = \Phi(x, y)$ and $f_s(z)$, as explicitly shown above; for e.g., we have $\Phi(1,1) = \Phi(2,3) = \Phi(3,2) = \Phi(4,2) = 1$ and $f_1(1) = a_1$ and $f_2(1) = a'_1$. Therefore, the channel conditional probabilities for each state can now be defined by Φ and the matrix T, where

$$T = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ a'_1 & a'_2 & a'_3 & a'_4 & a'_5 & a'_6 \end{bmatrix}.$$

Hence, the fundamental property for quasi-symmetric FSM channels is the existence of a noise process $\{Z_n\}$ given by $Z_n = \Phi(X_n, Y_n)$ such that Z_n is independent of X_n given S_n. The class of FSM channels having this property, when there is no feedback, are termed variable noise channels [GV96].

The features that we have developed so far are valid for any quasi-symmetric FSM channel. However, while discussing the feedback capacity of these channels we assume that the channels also satisfy the following assumption.

Assumption 3.1.1. *We assume that for a fixed $y \in \mathcal{Y}$, the column sum $\sum_x f_s(\Phi(x, y))$ is invariant with $s \in \mathcal{S}$:*

$$\sum_x f_s(\Phi(x, y)) = \sum_x f_{s'}(\Phi(x, y)) \;\; \forall s, s' \in \mathcal{S}, \tag{3.9}$$

where $f_s(\Phi(x, y)) = p_{\mathcal{C}}(y|x, s)$.

In other words, the assumption requires that for each output value y, the $|\mathcal{S}|$ column sums corresponding to output y in the channel transition matrices are all identical; i.e.,

$$\sum_{x\in\mathcal{X}} p_{\mathcal{C}}(y|x,s_1) = \sum_{x\in\mathcal{X}} p_{\mathcal{C}}(y|x,s_2) = \cdots = \sum_{x\in\mathcal{X}} p_{\mathcal{C}}(y|x,s_{|\mathcal{S}|}), \quad \forall y \in \mathcal{Y}.$$

However, for a fixed $s \in \mathcal{S}$, $\sum_x p_{\mathcal{C}}(y|x,s)$ is not necessarily invariant with $y \in \mathcal{Y}$, and as such, a uniform input does not yield a uniform output in general. This requirement will be needed in our dynamic programming approach which we use to determine the optimal feedback control action (as will be seen in the next section).[4]

3.1.2 Feedback Capacity of Quasi-Symmetric FSM Channels

In this section, we will show that feedback does not increase the capacity of quasi-symmetric FSM channels defined in the previous section. By feedback, we mean that there exists a channel from the receiver to the transmitter which is noiseless and delayless. Thus at any given time, all previously received outputs are unambiguously known by the transmitter and can be used for encoding the message into the next code symbol.

A feedback code with blocklength n and rate R consists of a sequence of mappings

$$\psi_i : \{1, 2, ..., 2^{nR}\} \times \mathcal{Y}^{i-1} \to \mathcal{X}$$

for $i = 1, 2, ...n$ and an associated decoding function

$$\Upsilon : \mathcal{Y}^n \to \{1, 2, ..., 2^{nR}\}.$$

Thus, when the transmitter wants to send message $W \in \mathcal{W} = \{1, 2, ..., 2^{nR}\}$, where W is uniformly distributed over $\mathcal{W}$ and is independent of $S_{[n]}$, it sends the

[4]Note for our main results to hold, we require the FSM channel as defined via properties (I) and (II) to be quasi-symmetric, in addition to satisfying Assumption 3.1.1.

codeword $X_{[n]}$, where $X_1 = \psi_1(W)$ and $X_i = \psi_i(W, Y_{[i-1]})$, for $i = 2, \cdots, n$. In the case when there is no feedback, the codeword $X_{[n]}$, where $X_1 = \psi_1(W)$ and $X_i = \psi_i(W)$, for $i = 2, \cdots, n$ is transmitted; and thus a non-feedback code is a special case of a feedback code. For a received $Y_{[n]}$ at the channel output, the receiver uses the decoding function to estimate the transmitted message as $\hat{W} = \Upsilon(Y_{[n]})$. A decoding error is made when $\hat{W} \neq W$. The probability of error is given by

$$P_e^{(n)} = \frac{1}{2^{nR}} \sum_{k=1}^{2^{nR}} P\left\{\Upsilon(Y_{[n]}) \neq W | W = k\right\}.$$

It should also be observed that when communicating with feedback, Property (3.1.1.b) does not hold, since X_i is a function of $Y_{[i-1]}$ (in addition to W); also $X_{[n]}$ and $S_{[n]}$ are no longer independent as X_i causally depends on $Z_{[i-1]}$ and hence $S_{[i-1]}$, for $i = 1, 2, \cdots, n$.

The capacity with feedback, C_{FB}, is the supremum of all admissible rates; i.e., rates for which there exists sequences of feedback codes with asymptotically vanishing probability of error. The (classical) non-feedback capacity, C_{NFB}, is defined similarly (by replacing feedback codes with non-feedback codes). Since a non-feedback code is a special case of a feedback code, we always have $C_{FB} \geq C_{NFB}$.

The main result of this section is as follows.

Theorem 3.1.1. *For a quasi-symmetric FSM channel* $[\mathcal{X}, \mathcal{Y}, \mathcal{S}, P_S, Z, T, \Phi]$ *satisfying Assumption 3.1.1, its feedback capacity is given by*

$$C_{FB} = \mathcal{H}(\tilde{Y}) - \mathcal{H}(Z)$$

where $\mathcal{H}(\tilde{Y})$ *is the entropy rate of the output process* $\{\tilde{Y}_i\}$ *driven by an i.u.d. input and* $\mathcal{H}(Z)$ *is the entropy rate of the channel's noise (hidden Markovian) process* $\{Z_i\}_{i=1}^{\infty}$.

We devote the remainder of the section to prove this theorem and deduce that feedback does not help increasing the capacity of quasi-symmetric FSM channels satisfying Assumption 3.1.1.

From Fano's inequality, we have

$$H(W|Y_n) \leq h_b({P_e}^{(n)}) + {P_e}^{(n)} \log_2(2^{nR} - 1) \leq 1 + {P_e}^{(n)} nR$$

where the first inequality holds since $h_b(P_e^{(n)}) \leq 1$, where $h_b(\cdot)$ is the binary entropy function. Since W is uniformly distributed,

$$nR = H(W) = H(W|Y_{[n]}) + I(W; Y_{[n]}) \leq 1 + {P_e}^{(n)} nR + I(W; Y_{[n]})$$

where R is any admissible rate. Dividing both sides by n and taking the lim inf yields

$$C_{FB} \leq \liminf_{n\to\infty} \sup_{\{\psi_i\}_{i=1}^n} \frac{1}{n} I(W; Y_{[n]}). \tag{3.10}$$

For every coding policy with feedback $\{\psi_i, 1 \leq i \leq n\}$, there are induced maps $\{\eta_i, 1 \leq i \leq n\}$ such that

$$\eta_i : \mathcal{X}^{i-1} \times \mathcal{Y}^{i-1} \to \mathbb{P}(\mathcal{X}), \tag{3.11}$$

with

$$\eta_i(x_{[i-1]}, y_{[i-1]}) = \big(\beta_i(x_{(1)}), \beta_i(x_{(2)}), \cdots, \beta_i(x_{(k)})\big) \tag{3.12}$$

and

$$\beta_i(x_{(j)}) = \sum_{w \in \mathcal{W}} P_{W|X_{[i-1]},Y_{[i-1]}}(w|x_{[i-1]}, y_{[i-1]}) 1_{\{x_{(j)} = \psi_i(w, y_{[i-1]})\}} \tag{3.13}$$

for $j = 1, 2, \cdots, k$, where $\mathcal{X} = \{x_{(1)}, x_{(2)}, \cdots, x_{(k)}\}$ with $k = |\mathcal{X}|$, $1_{\{\cdot\}}$ denotes the indicator function and $\mathbb{P}(\mathcal{X})$ denotes the space of probability distributions on $\mathcal{X}$. Every η_i can also be identified by the collection of control actions at time i:

$$\mathcal{D}_i \triangleq \{P_{X_i|X_{[i-1]},Y_{[i-1]}}(x_i|x_{[i-1]}, y_{[i-1]}) : x_{[i-1]} \in \mathcal{X}^{i-1}, y_{[i-1]} \in \mathcal{Y}^{i-1}\}. \tag{3.14}$$

In view of this discussion, following [TM09], we have

$$
\begin{aligned}
&\liminf_{n\to\infty} \sup_{\{\psi_i\}_{i=1}^n} \frac{1}{n} I(W;Y_{[n]}) \\
&\quad = \liminf_{n\to\infty} \sup_{\{\psi_i\}_{i=1}^n} \frac{1}{n} \sum_{i=1}^{n} \left[H(Y_i|Y_{[i-1]}) - H(Y_i|W,Y_{[i-1]})\right] \\
&\quad = \liminf_{n\to\infty} \sup_{\{\psi_i\}_{i=1}^n} \frac{1}{n} \sum_{i=1}^{n} \left[H(Y_i|Y_{[i-1]}) - H(Y_i|W,Y_{[i-1]},X_{[i]})\right] \\
&\quad = \liminf_{n\to\infty} \sup_{\{\psi_i\}_{i=1}^n} \frac{1}{n} \sum_{i=1}^{n} \left[H(Y_i|Y_{[i-1]}) - H(Y_i|Y_{[i-1]},X_{[i]})\right] \qquad (3.15) \\
&\quad \leq \liminf_{n\to\infty} \sup_{\{\mathcal{D}_i\}_{i=1}^n} \frac{1}{n} \sum_{i=1}^{n} \left[H(Y_i|Y_{[i-1]}) - H(Y_i|Y_{[i-1]},X_{[i]})\right] \qquad (3.16)
\end{aligned}
$$

where (3.15) is shown below and (3.16) holds since for every collection of feedback encoding functions, $\{\psi_i\}_{i=1}^n$, there exists a collection of control actions $\{\mathcal{D}_i\}_{i=1}^n$ which is demonstrated in (3.11)-(3.14). Note that the right-hand side of (3.16) is the directed information whose supremum has been shown to be the feedback capacity under information stability conditions [TM09].

Proof of Equation (3.15). We need to show that

$$P_{Y_i|W,X_{[i]},Y_{[i-1]}}(y_i|w,x_{[i]},y_{[i-1]}) = P_{Y_i|X_{[i]},Y_{[i-1]}}(y_i|x_{[i]},y_{[i-1]}), \text{ for } i = 1,2,\cdots,n.$$

Note that

$$
\begin{aligned}
&P_{Y_i|W,X_{[i]},Y_{[i-1]}}(y_i|w,x_{[i]},y_{[i-1]}) \\
&\overset{(a)}{=} \sum_{s_i} p_{\mathcal{C}}(y_i|x_i,s_i) \frac{P_{X_i|X_{[i-1]},Y_{[i-1]}}(x_i|x_{[i-1]},y_{[i-1]}) P_{S_i,W,X_{[i-1]},Y_{[i-1]}}(s_i,w,x_{[i-1]},y_{[i-1]})}{P_{X_i|X_{[i-1]},Y_{[i-1]}}(x_i|x_{[i-1]},y_{[i-1]}) P_{W,X_{[i-1]},Y_{[i-1]}}(w,x_{[i-1]},y_{[i-1]})} \\
&= \sum_{s_i} p_{\mathcal{C}}(y_i|x_i,s_i) P_{S_i|W,X_{[i-1]},Y_{[i-1]}}(s_i|w,x_{[i-1]},y_{[i-1]}) \\
&\overset{(b)}{=} \sum_{s_i} p_{\mathcal{C}}(y_i|x_i,s_i) P_{S_i|X_{[i-1]},Y_{[i-1]}}(s_i|x_{[i-1]},y_{[i-1]}) \\
&= P_{Y_i|X_{[i]},Y_{[i-1]}}(y_i|x_{[i]},y_{[i-1]})
\end{aligned}
$$

where (a) follows from Property (II), and (b) is valid since

$$\begin{aligned}
&P_{S_i|W,X_{[i-1]},Y_{[i-1]}}(s_i|w,x_{[i-1]},y_{[i-1]})\\
&\overset{(i)}{=} \sum_{s_{i-1}} P_{S_i|S_{i-1}}(s_i|s_{i-1}) P_{S_{i-1}|W,X_{[i-1]},Y_{[i-1]}}(s_{i-1}|w,x_{[i-1]},y_{[i-1]})\\
&\overset{(ii)}{=} \sum_{s_{i-1}} P_{S_i|S_{i-1}}(s_i|s_{i-1}) P_{S_{i-1}|X_{[i-1]},Y_{[i-1]}}(s_{i-1}|x_{[i-1]},y_{[i-1]})\\
&= P_{S_i|X_{[i-1]},Y_{[i-1]}}(s_i|x_{[i-1]},y_{[i-1]})
\end{aligned}$$

where (i) is due to (I) and (ii) can be shown recursively as follows:

$$\begin{aligned}
P_{S|W,X,Y}(s_1|w,x_1,y_1) &= \frac{P_{S,W,X,Y}(s_1,w,x_1,y_1)}{\sum_{s_1} P_{S,W,X,Y}(s_1,w,x_1,y_1)}\\
&= \frac{P_{Y|X,S}(y_1|x_1,s_1)P_{X,S,W}(x_1,s_1,w)}{\sum_{s_1} P_{Y|X,S}(y_1|x_1,s_1)P_{X,S,W}(x_1,s_1,w)}\\
&\overset{(iii)}{=} \frac{P_{Y|X,S}(y_1|x_1,s_1)P_S(s_1)P_{X,W}(x_1,w)}{\sum_{s_1} P_{Y|X,S}(y_1|x_1,s_1)P_S(s_1)P_{X,W}(x_1,w)}\\
&= \frac{P_{Y|X,S}(y_1|x_1,s_1)P_{S|X}(s_1|x_1)}{\sum_{s_1} P_{Y|X,S}(y_1|x_1,s_1)P_{S|X}(s_1|x_1)}\\
&= P_{S|X,Y}(s_1|x_1,y_1) \qquad (3.17)
\end{aligned}$$

where (iii) is valid since s_1 is independent of w and x_1 (as x_1 is only a function of w). Similarly,

$$\begin{aligned}
&P_{S_{[2]}|W,X_{[2]},Y_{[2]}}(s_2|w,x_{[2]},y_{[2]})\\
&= \frac{P_{S_{[2]},W,X_{[2]},Y_{[2]}}(s_2,w,x_{[2]},y_{[2]})}{\sum_{s_2} P_{S_{[2]},W,X_{[2]},Y_{[2]}}(s_2,w,x_{[2]},y_{[2]})}\\
&= \frac{P_{Y|X,S}(y_2|x_2,s_2)P_{X_{[2]},Y_1,S_2,W}(x_{[2]},y_1,s_2,w)}{\sum_{s_2} P_{Y|X,S}(y_2|x_2,s_2)P_{X_{[2]},Y_1,S_2,W}(x_{[2]},y_1,s_2,w)}\\
&\overset{(iv)}{=} \frac{P_{Y|X,S}(y_2|x_2,s_2)P_{X_2|X_1,Y_1,W}(x_2|x_1,y_1,w)P_{S_2,X_1,Y_1,W}(s_2,x_1,y_1,w)}{\sum_{s_2} P_{Y|X,S}(y_2|x_2,s_2)P_{X_2|X_1,Y_1,W}(x_2|x_1,y_1,w)P_{S_2,X_1,Y_1,W}(s_2,x_1,y_1,w)}\\
&= \frac{P_{Y|X,S}(y_2|x_2,s_2)P_{S_2|X_1,Y_1,W}(s_2|x_1,y_1,w)P_{X,Y,W}(x_1,y_1,w)}{\sum_{s_2} P_{Y|X,S}(y_2|x_2,s_2)P_{S_2|X_1,Y_1,W}(s_2|x_1,y_1,w)P_{X,Y,W}(x_1,y_1,w)}\\
&\overset{(v)}{=} \frac{P_{Y|X,S}(y_2|x_2,s_2)\sum_{s_1} P_{S_2|S_1}(s_2|s_1)P_{S_1|X_1,Y_1}(s_1|x_1,y_1)}{\sum_{s_2} P_{Y|X,S}(y_2|x_2,s_2)\sum_{s_1} P_{S_2|S_1}(s_2|s_1)P_{S_1|X_1,Y_1}(s_1|x_1,y_1)}
\end{aligned}$$

$$= \quad P_{S_2|X_{[2]},Y_{[2]}}(s_2|x_{[2]},y_{[2]})$$

where (iv) is valid since x_2 is a function of x_1, y_1 and w and (v) is due to (3.17). Using these steps recursively for $i = 1, 2, \cdots, n$ yields (ii) and completes the proof. □

Now, let us consider the following equation

$$\sup_{\{\mathcal{D}_i\}_{i=1}^n} \frac{1}{n} \sum_{i=1}^{n} \left[H(Y_i|Y_{[i-1]}) - H(Y_i|Y_{[i-1]}, X_{[i]}) \right]. \tag{3.18}$$

In the following, we first show that $H(Y_i|Y_{[i-1]}, X_{[i]}) = H(Z_i|Z_{[i-1]})$ and next we show that $\sum_{i=1}^n H(Y_i|Y_{[i-1]})$ is maximized when $\left\{P_{X_i|X_{[i-1]},Y_{[i-1]}}(x_i|x_{[i-1]},y_{[i-1]})\right\}_{i=1}^n$ is uniformly distributed.

Lemma 3.1.2. *The quasi-symmetric FSM channel satisfies*

$$H(Y_i|Y_{[i-1]}, X_{[i]}) = H(Z_i|Z_{[i-1]}), \;\; \forall i = 1, \cdots, n.$$

Proof. Observe first that

$$\begin{aligned} P_{Y_i|X_{[i]},Y_{[i-1]}}(y_i|x_{[i]},y_{[i-1]}) &\overset{(i)}{=} P_{Z_i|X_{[i]},Y_{[i-1]}}(z_i|x_{[i]},y_{[i-1]}) \\ &\overset{(ii)}{=} P_{Z_i|X_{[i]},Y_{[i-1]},Z_{[i-1]}}(z_i|x_{[i]},y_{[i-1]},z_{[i-1]}) \\ &\overset{(iii)}{=} P_{Z_i|X_{[i]},Z_{[i-1]}}(z_i|x_{[i]},z_{[i-1]}) \end{aligned}$$

where (i) and (ii) is valid since $z_{[i]} = \Phi\left(x_{[i]}, y_{[i]}\right)$ and (iii) is valid since $y_i = \nu(x_i, z_i)$ where $\nu(x,\cdot) = \Phi^{-1}(x,\cdot)$. We next show that $Z_i \to Z_{[i-1]} \to X_{[i]}$ form a Markov chain. Note that

$$\begin{aligned} &P_{Z_i|X_{[i]},Z_{[i-1]}}(z_i|x_{[i]},z_{[i-1]}) \\ &\quad = \frac{P_{Z_i,X_{[i]},Z_{[i-1]}}(z_i,x_{[i]},z_{[i-1]})}{P_{X_{[i]},Z_{[i-1]}}(x_{[i]},z_{[i-1]})} \\ &\quad \overset{(iv)}{=} \frac{P_{X_i|X_{[i-1]},Z_{[i-1]}}(x_i|x_{[i-1]},z_{[i-1]})P_{X_{[i-1]},Z_i,Z_{[i-1]}}(x_{[i-1]},z_i,z_{[i-1]})}{P_{X_i|X_{[i-1]},Z_{[i-1]}}(x_i|x_{[i-1]},z_{[i-1]})P_{X_{[i-1]},Z_{[i-1]}}(x_{[i-1]},z_{[i-1]})} \\ &\quad = P_{Z_i|X_{[i-1]},Z_{[i-1]}}(z_i|x_{[i-1]},z_{[i-1]}) \end{aligned}$$

where (iv) is valid since the feedback input depends (causally) only on $\left(x_{[i-1]}, y_{[i-1]}\right)$, or equivalently on $\left(x_{[i-1]}, z_{[i-1]}\right)$. Similarly, we get

$$\begin{aligned}
&P_{Z_i|X_{[i-1]},Z_{[i-1]}}(z_i|x_{[i-1]}, z_{[i-1]})\\
&\quad= \frac{P_{X_{i-1}|X_{[i-2]},Z_{[i-2]}}(x_{i-1}|x_{[i-2]}, z_{[i-2]})P_{X_{[i-2]},Z_i,Z_{[i-1]}}(x_{[i-2]}, z_i, z_{[i-1]})}{P_{X_{i-1}|X_{[i-2]},Z_{[i-2]}}(x_{i-1}|x_{[i-2]}, z_{[i-2]})P_{X_{[i-2]},Z_{[i-1]}}(x_{[i-2]}, z_{[i-1]})}\\
&\quad= P_{Z_i|X_{[i-2]},Z_{[i-1]}}(z_i|x_{[i-2]}, z_{[i-1]}).
\end{aligned}$$

Using these steps recursively, we get

$$\begin{aligned}
&P_{Z_i|X_{[i-2]},Z_{[i-1]}}(z_i|x_{[i-2]}, z_{[i-1]})\\
&\quad= \frac{P_{Z_i,X_{[i-2]},Z_{[i-1]}}(z_i, x_{[i-2]}, z_{[i-1]})}{P_{X_{[i-2]},Z_{[i-1]}}(x_{[i-2]}, z_{[i-1]})}\\
&\quad\overset{(v)}{=} \frac{P_{X_{i-2}|X_{[i-3]},Z_{[i-3]}}(x_{i-2}|x_{[i-3]}, z_{[i-3]})P_{Z_i,X_{[i-3]},Z_{[i-i]}}(z_i, x_{[i-3]}, z_{[i-1]})}{P_{X_{i-2}|X_{[i-3]},Z_{[i-3]}}(x_{i-2}|x_{[i-3]}, z_{[i-3]})P_{X_{[i-3]},Z_{[i-i]}}(x_{[i-3]}, z_{[i-1]})}\\
&\quad\vdots\\
&\quad\overset{(vi)}{=} \frac{P_{X_2|X_1,Z_1}(x_2|x_1, z_1)P_{Z_i,X_1,Z_{[i-1]}}(z_i, x_1, z_{[i-1]})}{P_{X_2|X_1,Z_1}(x_2|x_1, z_1)P_{X_1,Z_{[i-1]}}(x_1, z_{[i-1]})}\\
&\quad\overset{(vii)}{=} \frac{P_X(x_1)P_{Z_{[i]}}(z_i, z_{[i-1]})}{P(x_1)P_{Z_{[i-1]}}(z_{[i-1]})}\\
&\quad= P_{Z_i|Z_{[i-1]}}(z_i|z_{[i-1]})
\end{aligned}$$

where $(v), (vi)$ and (vii) are valid due to the same reasoning above. □

We next show that all of the terms $H(Y_i|Y_{[i-1]})$ in (3.18) are maximized by uniform feedback control actions. We solve this problem using dynamic programming [Ber01]. The optimization problem can be written as:

$$\max_{\{\mathcal{D}_1,\cdots,\mathcal{D}_n\}} \{H(Y_n|Y_{[n-1]}) + H(Y_{n-1}|Y_{[n-2]}) + \cdots + H(Y_1)\}. \tag{3.19}$$

Let

$$V_i\left(P_{Y_{[i-1]}}(y_{[i-1]}), \mathcal{D}_1, \cdots, \mathcal{D}_{i-1}\right) = \max_{\mathcal{D}_i}\left[H(Y_i|Y_{[i-1]}) + V_{i+1}\left(P_{Y_{[i]}}(y_{[i]}), \mathcal{D}_1, \cdots, \mathcal{D}_i\right)\right]$$

where $V_{n+1}\left(P_{Y_{[n]}}(y_{[n]}),\mathcal{D}_1,\cdots,\mathcal{D}_n\right)=0$ and the $V_i\left(P_{Y_{[i-1]}}(y_{[i-1]}),\mathcal{D}_1,\cdots,\mathcal{D}_{i-1}\right)$ terms are explicitly given for $i=1,\cdots,n$ as follows:

$$\begin{aligned}
V_n\left(P_{Y_{[n-1]}}(y_{[n-1]}),\mathcal{D}_1,\cdots,\mathcal{D}_{n-1}\right) &= \max_{\mathcal{D}_n} H(Y_n|Y_{[n-1]})\\
V_{n-1}\left(P_{Y_{[n-2]}}(y_{[n-2]}),\mathcal{D}_1,\cdots,\mathcal{D}_{n-2}\right) &= \max_{\mathcal{D}_{n-1}}\left\{H(Y_{n-1}|Y_{[n-2]})+\max_{\mathcal{D}_n}\left\{H(Y_n|Y_{[n-1]})\right\}\right\}\\
V_{n-2}\left(P_{Y_{[n-3]}}(y_{[n-3]}),\mathcal{D}_1,\cdots,\mathcal{D}_{n-3}\right) &= \max_{\mathcal{D}_{n-2}}\left\{H(Y_{n-2}|Y_{[n-3]})+\max_{\mathcal{D}_{n-1}}\left\{H(Y_{n-1}|Y_{[n-2]})\right.\right.\\
&\qquad\left.\left.+\max_{\mathcal{D}_n}\left\{H(Y_n|Y_{[n-1]})\right\}\right\}\right\}\\
&\ \ \vdots\\
V_1 &= \max_{\mathcal{D}_1}\left\{H(Y_1)+\cdots+\max_{\mathcal{D}_{n-1}}\left\{H(Y_{n-1}|Y_{[n-2]})\right.\right.\\
&\qquad\left.\left.+\max_{\mathcal{D}_n}\left\{H(Y_n|Y_{[n-1]})\right\}\cdots\right\}\right\}.
\end{aligned} \tag{3.20}$$

Here, $V_{i+1}\left(P_{Y_{[i]}}(y_{[i]}),\mathcal{D}_1,\cdots,\mathcal{D}_i\right)$ denotes the reward-to-go at time i, which is the future reward generated by the control action at time i.

Thus (3.19) is given by V_1 in (3.20), which indicates that the optimization problem is nested and dynamic. It is nested since the actions and the action outcomes, that are the realizations of the channel inputs and outputs, are available in future time stages. It is dynamic, since the control actions applied at time k affect the future reward value realizations at time stages $i>k$. Thus an optimal selection of the actions, should maximize both the current reward $H(Y_i|Y_{[i-1]})$ and the reward-to-go $V_{i+1}\left(P_{Y_{[i]}}(y_{[i]}),\mathcal{D}_1,\cdots,\mathcal{D}_i\right)$ (see (3.20)).

Therefore, the optimization problem turns out to be finding the best induced policies $\{\eta_i, 1\le i\le n\}$; that is the best collection of functions used to generate the set of control actions $\{\mathcal{D}_i, 1\le i\le n\}$ which achieve V_1. We next show that the optimal set of control actions achieving V_1 is composed of uniform input distributions for $i = 1, \cdots, n$. Toward this goal, we find a condition such that the control actions taken at times $(i-1), \cdots, 1$ do not affect the reward value attained at time i, when the control action at time i is uniform. Specifically, we find that a sufficient condition to manage this problem is requiring $\sum_x f_s(\Phi(x,y))$ to be invariant with $s \in \mathcal{S}$, i.e., Assumption 3.1.1 . This will be explicitly shown in Lemma 3.1.4. We first have the following.

Lemma 3.1.3. *For the quasi-symmetric FSM channel, each conditional output entropy $H(Y_i|Y_{[i-1]})$, $i = 1, \cdots, n$ in (3.18), given the past sets of control actions $\mathcal{D}_1, \mathcal{D}_2, \cdots, \mathcal{D}_{i-1}$, is maximized by uniform feedback control actions:*

$$\begin{aligned} \mathcal{D}_i^\star &\triangleq \operatorname*{argmax}_{\mathcal{D}_i} H(Y_i|Y_{[i-1]}) \\ &= \left\{ P_{X_i|X_{[i-1]},Y_{[i-1]}}(x_i|x_{[i-1]}, y_{[i-1]}) = \frac{1}{|\mathcal{X}|}, \forall (x_{[i-1]}, y_{[i-1]}) \in \mathcal{X}^{i-1} \times \mathcal{Y}^{i-1} \right\} \end{aligned}$$

for all $x_i \in \mathcal{X}$ and for all $i = 1, \cdots, n$.

Proof. Note first that

$$H(Y_i|Y_{[i-1]}) = \sum_{y_{[i-1]}} P_{Y_{[i-1]}}(y_{[i-1]}) H(Y_i|Y_{[i-1]} = y_{[i-1]}) \tag{3.21}$$

where

$$H(Y_i|Y_{[i-1]} = y_{[i-1]}) = -\sum_{y_i} P_{Y_i|Y_{[i-1]}}(y_i|y_{[i-1]}) \log P_{Y_i|Y_{[i-1]}}(y_i|y_{[i-1]}). \tag{3.22}$$

To show that $H(Y_i|Y_{[i-1]})$ in (3.21) is maximized by a uniform input distribution, it is enough to show that such a uniform distribution maximizes each of the $H(Y_i|Y_{[i-1]} = y_{[i-1]})$ terms. We now expand $P_{Y_i|Y_{[i-1]}}(y_i|y_{[i-1]})$ as follows

$$P_{Y_i|Y_{[i-1]}}(y_i|y_{[i-1]})$$

$$
\begin{aligned}
&= \sum_{x_i} \sum_{x_{[i-1]}} \sum_{s_i} \sum_{s_{[i-1]}} P_{Y_i,X_{[i]},S_{[i]}|Y_{[i-1]}}(y_i, x_i, x_{[i-1]}, s_i, s_{[i-1]}|y_{[i-1]}) \\
&= \sum_{x_i,x_{[i-1]}} \sum_{s_i,s_{[i-1]}} P_{Y_i|X_{[i]},S_{[i]},Y_{[i-1]}}(y_i|x_i, x_{[i-1]}, s_i, s_{[i-1]}, y_{[i-1]}) \\
&\qquad\qquad P_{X_{[i]},S_{[i]}|Y_{[i-1]}}(x_i, x_{[i-1]}, s_i, s_{[i-1]}|y_{[i-1]}) \\
&\overset{(i)}{=} \sum_{x_i,x_{[i-1]}} \sum_{s_i,s_{[i-1]}} p_{\mathcal{C}}(y_i|x_i, s_i) P_{X_{[i]},S_{[i]}|Y_{[i-1]}}(x_i, x_{[i-1]}, s_i, s_{[i-1]}|y_{[i-1]}) \\
&= \sum_{x_i,x_{[i-1]}} \sum_{s_i,s_{[i-1]}} p_{\mathcal{C}}(y_i|x_i, s_i) P_{X_{[i]},S_{[i-1]}|Y_{[i-1]}}(x_i, x_{[i-1]}, s_{[i-1]}|y_{[i-1]}) \\
&\qquad\qquad P_{S_i|X_{[i]},S_{[i-1]},Y_{[i-1]}}(s_i|x_i, x_{[i-1]}, s_{[i-1]}, y_{[i-1]}) \\
&\overset{(ii)}{=} \sum_{x_i,x_{[i-1]}} \sum_{s_i,s_{[i-1]}} p_{\mathcal{C}}(y_i|x_i, s_i) P_{X_{[i-1]},S_{[i-1]}|Y_{[i-1]}}(x_{[i-1]}, s_{[i-1]}|y_{[i-1]}) P_{S_i|S_{[i-1]}}(s_i|s_{[i-1]}) \\
&\qquad\qquad P_{X_i|X_{[i-1]},S_{[i-1]},Y_{[i-1]}}(x_i|x_{[i-1]}, s_{[i-1]}, y_{[i-1]}) \\
&\overset{(iii)}{=} \sum_{x_i,x_{[i-1]}} \sum_{s_i,s_{[i-1]}} p_{\mathcal{C}}(y_i|x_i, s_i) P_{X_i|X_{[i-1]},Y_{[i-1]}}(x_i|x_{[i-1]}, y_{[i-1]}) P_{S_i|S_{[i-1]}}(s_i|s_{[i-1]}) \\
&\qquad\qquad P_{X_{[i-1]},S_{[i-1]}|Y_{[i-1]}}(x_{[i-1]}, s_{[i-1]}|y_{[i-1]}) \qquad (3.23)
\end{aligned}
$$

where (i) follows by (3.2), (ii) is valid due to the Property (I) and finally (iii) is due to the fact that the feedback input depends only on $(x_{[i-1]}, y_{[i-1]})$.

The key observation in equation (3.23) is the existence of an equivalent channel. More specifically, $\sum_{s_i} p_{\mathcal{C}}(y_i|x_i, s_i) P_{S_i|S_{[i-1]}}(s_i|s_{[i-1]})$ represents a quasi-symmetric channel transition matrix such that its entries are determined by the entries of the channel transition matrices of each state and the transition distribution of state probabilities. To continue, by (3.5),

$$
\begin{aligned}
P_{Y_i|Y_{[i-1]}}(y_i|y_{[i-1]}) = &\sum_{x_i,x_{[i-1]}} \sum_{s_i,s_{[i-1]}} f_{s_i}(\Phi(x_i, y_i)) P_{S_i|S_{[i-1]}}(s_i|s_{[i-1]}) \\
&P_{X_i|X_{[i-1]},Y_{[i-1]}}(x_i|x_{[i-1]}, y_{[i-1]}) P_{X_{[i-1]},S_{[i-1]}|Y_{[i-1]}}(x_{[i-1]}, s_{[i-1]}|y_{[i-1]}).
\end{aligned}
$$

By definition of quasi-symmetry, there exists m weakly symmetric sub-arrays in the channel transition matrix at each state s_i. Among these sub-arrays, let us pick $\tilde{Q}_j^{s_i}$ of size $|\mathcal{X}| \times |\mathcal{Y}_j|$. (We assume that the partition of $\mathcal{Y}$ is identical across all states.) Let y_{j_t}, for $t = 1, \ldots, |\mathcal{Y}_j|$, denote the output values in sub-array j. Therefore, we obtain

$$\begin{aligned} P_{Y_i|Y_{[i-1]}}(y_{j_t}|y_{[i-1]}) &= \sum_{x_i, x_{[i-1]}, s_i, s_{[i-1]}} f_{s_i}(\Phi(x_i, y_{j_t})) P_{S_i|S_{[i-1]}}(s_i|s_{[i-1]}) \\ & P_{X_i|X_{[i-1]}, Y_{[i-1]}}(x_i|x_{[i-1]}, y_{[i-1]}) P_{X_{[i-1]}, S_{[i-1]}|Y_{[i-1]}}(x_{[i-1]}, s_{[i-1]}|y_{[i-1]}). \end{aligned} \tag{3.24}$$

For $\mathcal{X} = \{x_{(1)}, \cdots, x_{(k)}\}$, $k = |\mathcal{X}|$, let $\kappa(i-1) = P_{S_i|S_{[i-1]}}(s_i|s_{[i-1]})$, $\chi(i-1) = P_{X_{[i-1]}, S_{[i-1]}|Y_{[i-1]}}(x_{[i-1]}, s_{[i-1]}|y_{[i-1]})$ and $P_{X_i|X_{[i-1]}, Y_{[i-1]}}(x_{(l)}|x_{[i-1]}, y_{[i-1]}) = \varphi_i(x_{(l)})$ for $l = 1, \ldots, k$. Then, for $t = 1, \ldots, |\mathcal{Y}_j|$, we can write

$$\begin{aligned} P_{Y_i|Y_{[i-1]}}(y_{j_1}|y_{[i-1]}) &= \sum_{x_{[i-1]}, s_{[i-1]}} \chi(i-1) \sum_{s_i} \kappa(i-1) \left(\varphi_i(x_{(1)}) f_{s_i}(\Phi(x_{(1)}, y_{j_1})) + \right. \\ & \left. \cdots + \varphi_i(x_{(k)}) f_{s_i}(\Phi(x_{(k)}, y_{j_1}))\right), \end{aligned}$$

$$\begin{aligned} P_{Y_i|Y_{[i-1]}}(y_{j_2}|y_{[i-1]}) &= \sum_{s_{[i-1]}, x_{[i-1]}} \chi(i-1) \sum_{s_i} \kappa(i-1) \left(\varphi_i(x_{(1)}) f_{s_i}(\Phi(x_{(1)}, y_{j_2})) + \right. \\ & \left. \cdots + \varphi_i(x_{(k)}) f_{s_i}(\Phi(x_{(k)}, y_{j_2}))\right), \\ P_{Y_i|Y_{[i-1]}}(y_{j_{|\mathcal{Y}_j|}}|y_{[i-1]}) &= \sum_{s_{[i-1]}, x_{[i-1]}} \chi(i-1) \sum_{s_i} \kappa(i-1) \left\{\varphi_i(x_{(1)}) f_{s_i}(\Phi(x_{(1)}, y_{j_{|\mathcal{Y}_j|}})) + \right. \\ & \left. \cdots + \varphi_i(x_{(k)}) f_{s_i}(\Phi(x_{(k)}, y_{j_{|\mathcal{Y}_j|}}))\right\}. \end{aligned}$$

It should be noted that, each $f_{s_i}(\Phi(x_{(l)}, y_{j_t}))$ above corresponds to an entry in the channel transition matrix $\tilde{Q}^{s_i}$ at state s_i. Note also that the rows of the sub-array $\tilde{Q}_j^{s_i}$ are permutations of each other. In other words, each $f_{s_i}(\Phi(x_{(l)}, y_{j_t}))$ value appears exactly k times (once in each row) in the sub-array $Q_j^{s_i}$. Thus, the feedback control action $\varphi_i(x_{(l)})$ is multiplied by a different $f_{s_i}(\Phi(x_{(l)}, y_{j_t}))$ value for each $t = 1, \ldots, |\mathcal{Y}_j|$ in the $P_{Y_i|Y_{[i-1]}}(y_{j_t}|y_{[i-1]})$ given above. Hence, $\sum_{t=1}^{|\mathcal{Y}_j|} P_{Y_i|Y_{[i-1]}}(y_{j_t}|y_{[i-1]})$ is equal to

$$\sum_{t=1}^{|\mathcal{Y}_j|} P_{Y_i|Y_{[i-1]}}(y_{j_t}|y_{[i-1]})$$

$$= \sum_{s_{[i-1]},x_{[i-1]}} \chi(i-1) \sum_{s_i} \kappa(i-1) \sum_{l=1}^{k} \varphi_i(x_{(l)}) \sum_{t=1}^{|\mathcal{Y}_j|} f_{s_i}(\Phi(x_{(l)}, y_{j_t}))$$

$$= \sum_{s_{[i-1]},x_{[i-1]}} \chi(i-1) \sum_{s_i} \kappa(i-1) \sum_{l=1}^{k} \varphi_i(x_{(l)}) \sum_{t=1}^{|\mathcal{Y}_j|} p_c(y_{j_t}|x_{(l)}, s_i) \quad (3.25)$$

$$= \sum_{s_{[i-1]},x_{[i-1]}} \chi(i-1) \sum_{s_i} \kappa(i-1) \sum_{t=1}^{|\mathcal{Y}_j|} p_c(y_{j_t}|x_{(l)}, s_i) \quad (3.26)$$

where (3.25) is due to (3.5) and (3.26) is valid since $\tilde{Q}_j^{s_i}$ is weakly symmetric and as such $\sum_{t=1}^{|\mathcal{Y}_j|} p_c(y_{j_t}|x_{(l)}, s_i)$ is identical for each $x_{(l)}$, and noting that $\sum_{l=1}^{k} \varphi_i(x_{(l)}) = 1$ verifies (3.26). The critical observation is that the value attained by (3.26) is independent of the feedback control actions. Similarly, for all the other $m-1$ sub-arrays, their conditional output sums will be independent of the feedback control actions. Let us denote these sums by $\Omega_1, \ldots, \Omega_m$. More specifically for sub-array j, let $\Omega_j = \sum_{t=1}^{|\mathcal{Y}_j|} P_{Y_i|Y_{[i-1]}}(y_{i_t}|y_{[i-1]})$. Then the maximization of (3.22) now becomes,

$$\operatorname*{argmax}_{\Omega_{j,t}} - \sum_{j=1}^{m} \sum_{t=1}^{|\mathcal{Y}_j|} \Omega_{j,t} \log \Omega_{j,t} \quad (3.27)$$

where $\sum_{j=1}^{m} \sum_{t=1}^{|\mathcal{Y}_j|} \Omega_{j,t} = 1$ and $\Omega_{j,t} = P_{Y_i|Y_{[i-1]}}(y_{j_t}|y_{[i-1]})$, $t = 1, \ldots, |\mathcal{Y}_j|$, $j = 1, \ldots, m$. For each sub-array j, we need to find the $\Omega_{j,t}$ values that maximize $\sum_{t=1}^{|\mathcal{Y}_j|} \Omega_{j,t} \log \Omega_{j,t}$. By the log-sum inequality, we have that

$$- \sum_{t=1}^{|\mathcal{Y}_j|} \Omega_{j,t} \log \Omega_{j,t} \leq - \sum_{t=1}^{|\mathcal{Y}_j|} \Omega_{j,t} \log \frac{\sum_{t=1}^{|\mathcal{Y}_j|} \Omega_{j,t}}{|\mathcal{Y}_j|} \quad (3.28)$$

with equality if and only if

$$\Omega_{j,t} \;=\; \Omega_{s,w} \;\; \forall s, w \in \{1, \ldots, |\mathcal{Y}_j|\}. \quad (3.29)$$

In other words, for the sub-array j, the conditional entropy is maximized if and only if the conditional output probabilities in this sub-array are identical. Since this fact is valid for the other sub-arrays, to maximize the conditional entropy we need to (3.29) to be valid for all sub-arrays.

At this point, we have shown that the conditional output entropy is maximized if the conditional output probabilities are identical for each sub-array. In order to complete this step, we have to show that this is achieved by uniform input distributions.

Now, let us consider two conditional output probabilities, $P_{Y_i|Y_{[i-1]}}(y_{j_s}|y_{[i-1]})$ and $P_{Y_i|Y_{[i-1]}}(y_{j_t}|y_{[i-1]})$, in sub-array j. Then $P_{Y_i|Y_{[i-1]}}(y_{j_s}|y_{[i-1]}) = P_{Y_i|Y_{[i-1]}}(y_{j_t}|y_{[i-1]})$ implies that

$$\sum_{l=1}^{k} \varphi_i(x_{(l)}) f_{s_i}(\Phi(x_{(l)}, y_{j_s})) = \sum_{l=1}^{k} \varphi_i(x_{(l)}) f_{s_i}(\Phi(x_{(l)}, y_{j_t})). \tag{3.30}$$

However, for a fixed output $\sum_{l=1}^{k} f_{s_i}(\Phi(x_{(l)}, y_{j_s}))$ is equal to the sum of the column corresponding to output y_{j_s} (similarly for y_{j_t}) and since sub-array j is weakly symmetric, the column sums are identical. Therefore, (3.30) can be achieved if $\varphi_i(x_{(l)}) = \varphi_i(x_{(m)}) = \frac{1}{k}\ \forall\ l, m = 1, \ldots, k$, by which we get $P_{Y_i|Y_{[i-1]}}(y_{j_s}|y_{[i-1]}) = P_{Y_i|Y_{[i-1]}}(y_{j_t}|y_{[i-1]}) = \frac{1}{|\mathcal{X}|} \sum_{l=1}^{k} f_{s_i}(\Phi(x_{(l)}, y_{j_s}))$. Thus for other sub-arrays since they are also weakly-symmetric, the uniform feedback control action will also satisfy the equivalence of conditional output probabilities. □

With this lemma, we have shown that for each i, $H(Y_i|Y_{[i-1]})$ is maximized by the uniform input distribution. However, this is not sufficient to conclude that the optimal set of control actions attaining V_1, i.e., the optimal set of control actions maximizing $\sum_{i=1}^{n} H(Y_i|Y_{[i-1]})$, consists of a sequence of uniform input distributions for $i = 1, \cdots, n$. This is because Lemma 3.1.3 only maximizes the current conditional entropy via a uniform input (that is it is optimal in a myopic sense); however, it is still possible that a non-uniform input might result in a higher value function through the rewards-to-go. Let us now look at $P_{Y_i|Y_{[i-1]}}(y_i|y_{[i-1]})$ when we apply a uniform distribution at time i (current time). We obtain using (3.23) that

$$\begin{aligned}
&P_{Y_i|Y_{[i-1]}}(y_i|y_{[i-1]}) \\
&= \sum_{x_i,x_{[i-1]}} \sum_{s_i,s_{[i-1]}} p_{\mathcal{C}}(y_i|x_i,s_i) P_{X_i|X_{[i-1]},Y_{[i-1]}}(x_i|x_{[i-1]},y_{[i-1]}) \\
&\qquad P_{S_i|S_{[i-1]}}(s_i|s_{[i-1]}) P_{X_{[i-1]},S_{[i-1]}|Y_{[i-1]}}(x_{[i-1]},s_{[i-1]}|y_{[i-1]}) \\
&\stackrel{(i)}{=} \frac{1}{|\mathcal{X}|} \sum_{x_i,x_{[i-1]}} \sum_{s_i,s_{[i-1]}} p_{\mathcal{C}}(y_i|x_i,s_i) P_{S_i|S_{[i-1]}}(s_i|s_{[i-1]}) \\
&\qquad P_{X_{[i-1]},S_{[i-1]}|Y_{[i-1]}}(x_{[i-1]},s_{[i-1]}|y_{[i-1]}) \\
&= \frac{1}{|\mathcal{X}|} \sum_{x_i} \sum_{s_i} \sum_{s_{[i-1]}} p_{\mathcal{C}}(y_i|x_i,s_i) P_{S_i|S_{[i-1]}}(s_i|s_{[i-1]}) \\
&\qquad \sum_{x_{[i-1]}} P_{X_{[i-1]},S_{[i-1]}|Y_{[i-1]}}(x_{[i-1]},s_{[i-1]}|y_{[i-1]}) \\
&= \frac{1}{|\mathcal{X}|} \sum_{x_i} \sum_{s_i} \sum_{s_{[i-1]}} p_{\mathcal{C}}(y_i|x_i,s_i) P_{S_i|S_{[i-1]}}(s_i|s_{[i-1]}) P_{S_{[i-1]}|Y_{[i-1]}}(s_{[i-1]}|y_{[i-1]}) \\
&= \frac{1}{|\mathcal{X}|} \sum_{x_i} \sum_{s_i} p_{\mathcal{C}}(y_i|x_i,s_i) P_{S_i|Y_{[i-1]}}(s_i|y_{[i-1]})
\end{aligned}$$

where (i) is valid since $P_{X_i|X_{[i-1]},Y_{[i-1]}}(x_i|x_{[i-1]},y_{[i-1]})$ is uniform. Note that the dependency on past input control actions comes through $P_{S_i|Y_{[i-1]}}(s_i|y_{[i-1]})$ which includes transition probabilities between states, on which we have no control.

Lemma 3.1.4. *Assume that the feedback control action $P_{X_i|X_{[i-1]},Y_{[i-1]}}(x_i|x_{[i-1]},y_{[i-1]})$, at (current) time i, is uniform. Then the value of $H(Y_i|Y_{[i-1]})$ is independent of past feedback control actions at times $(i-1),\cdots,1$ if $\sum_x f_s(\Phi(x,y))$ is invariant with $s \in \mathcal{S}$ (i.e., if Assumption 3.1.1 holds).*

Proof. We have the following:

$$P_{Y_i|Y_{[i-1]}}(y_i|y_{[i-1]}) = \frac{1}{|\mathcal{X}|} \sum_{x_i} \sum_{s_i} p_{\mathcal{C}}(y_i|x_i,s_i) P_{S_i|Y_{[i-1]}}(s_i|y_{[i-1]})$$

$$= \frac{1}{|\mathcal{X}|}\sum_{s_i} P_{S_i|Y_{[i-1]}}(s_i|y_{[i-1]}) \sum_{x_i} p_{\mathcal{C}}(y_i|x_i, s_i)$$
$$= \frac{1}{|\mathcal{X}|}\sum_{s_i} P_{S_i|Y_{[i-1]}}(s_i|y_{[i-1]}) \underbrace{\sum_{x_i} f_s(\Phi(x_i, y_i))}.$$

Since the underbraced term is invariant with s, the proof is complete as the final sum will be $\frac{1}{|\mathcal{X}|}\sum_{x_i} f_s(\Phi(x_i, y_i))$. □

We have so far shown that $H(Y_i|X_{[i]}, Y_{[i-1]}) = H(Z_i|Z_{[i-1]})$ and that uniform input distributions maximize $\sum_{i=1}^{n} H(Y_i|Y_{[i-1]})$. With these results in hand, we have thus shown the following upperbound for the feedback capacity

$$C_{FB} \leq \liminf_{n\to\infty} \frac{1}{n}[H(\tilde{Y}_{[n]}) - H(Z_{[n]})] \tag{3.31}$$

where $H(\tilde{Y}_{[n]})$ is the output entropy when the input is uniform.

Let us now define a Hidden Markov Process (HMP) [EM02] which we will use while discussing the ergodicity of the noise and output processes. An HMP is denoted by a quadruple $[\mathcal{S}, P, \mathcal{Z}, T]$ in which $[\mathcal{S}, P]$ is a Markov process and T is the observation matrix defined by (3.7). The non-Markov process $\{Z_i\}_{i=1}^{\infty}$ with alphabet $\mathcal{Z}$ is called HMP and it is the noisy version of the state process observed through a DMC determined by T.

Lemma 3.1.5. *For the quasi-symmetric FSM channel with feedback, the noise process is an HMP with parameters* $[\mathcal{S}, P, \mathcal{Z}, T]$.

Proof. To show this result, it suffices to show that $P_{Z_i|S_i,Z_{[i-1]}}(z_i|s_i, z_{[i-1]}) = P_{Z|S}(z_i|s_i)$. Since $\{S_i\}_{i=1}^{\infty}$ is Markovian, it directly implies that $P_{S_i|S_{i-1},Z_{[i-1]}}(s_i|s_{i-1}, z_{[i-1]}) = P_{S_i|S_{i-1}}(s_i|s_{i-1})$. Note that

$$P_{Z_i|S_i,Z_{[i-1]}}(z_i|s_i, z_{[i-1]})$$

$$
\begin{aligned}
&= \sum_{x_{[i-1]}} \sum_{\{(x_i,y_i):z_i=\Phi(x_i,y_i)\}} P_{Y_i,X_{[i]}|S_i,Z_{[i-1]}}(y_i, x_i, x_{[i-1]}|s_i, z_{[i-1]}) \\
&\overset{(i)}{=} \sum_{\{(x_i,y_i):z_i=\Phi(x_i,y_i)\}} p_{\mathcal{C}}(y_i|x_i, s_i) P_{X_i|X_{[i-1]},S_i,Z_{[i-1]}}(x_i|x_{[i-1]}, s_i, z_{[i-1]}) \\
&\qquad \sum_{x_{[i-1]}} P_{X_{[i-1]}|S_i,Z_{[i-1]}}(x_{[i-1]}|s_i, z_{[i-1]}) \\
&\overset{(ii)}{=} \sum_{\{(x_i,y_i):z_i=\Phi(x_i,y_i)\}} f_{s_i}(\Phi(x_i, y_i)) P_{X_i|X_{[i-1]},Z_{[i-1]}}(x_i|x_{[i-1]}, z_{[i-1]}) \\
&\qquad \sum_{x_{[i-1]}} P_{X_{[i-1]}|S_i,Z_{[i-1]}}(x_{[i-1]}|s_i, z_{[i-1]}) \\
&= \sum_{x_{[i-1]}} P_{X_{[i-1]}|S_i,Z_{[i-1]}}(x_{[i-1]}|s_i, z_{[i-1]}) f_{s_i}(z_i) \\
&\qquad \left(\sum_{\{(x_i,y_i):z_i=\Phi(x_i,y_i)\}} P_{X_i|X_{[i-1]},Z_{[i-1]}}(x_i|x_{[i-1]}, z_{[i-1]}) \right) \\
&\overset{(iii)}{=} f_{s_i}(z_i) \overset{(iv)}{=} P_{Z|S}(z_i|s_i) \qquad (3.32)
\end{aligned}
$$

where (i) follows from (3.2) of Property (II) and the fact that $y_{[i-1]} = \nu(x_{[i-1]}, z_{[i-1]})$ is one-to-one with $z_{[i-1]}$ given $x_{[i-1]}$, (ii) is valid by (3.5) and by the fact that feedback input depends on $(x_{[i-1]}, z_{[i-1]})$, (iii) is valid since each z_i is satisfied by $|\mathcal{X}|$ number of (x_i, y_i) pairs where each x_i is different and (iv) follows from (3.5), (3.6) and (3.8). □

It should also be noted that, the output process, $\{\tilde{Y}_i\}_{i=1}^{\infty}$, for an i.u.d. input $\{X_i\}_{i=1}^{\infty}$ is also an HMP since

$$
\begin{aligned}
P_{Y_i|S_i,Y_{[i-1]}}(\tilde{y}_i|s_i, \tilde{y}_{[i-1]}) &= \sum_{x_i} P_{Y_i,X_i|S_i,Y_{[i-1]}}(\tilde{y}_i, x_i|s_i, \tilde{y}_{[i-1]}) \\
&\overset{(a)}{=} \sum_{x_i} p_{\mathcal{C}}(\tilde{y}_i|x_i, s_i) P_{X_i|S_i,Y_{[i-1]}}(x_i|s_i, \tilde{y}_{[i-1]}) \\
&\overset{(b)}{=} \sum_{x_i} p_{\mathcal{C}}(\tilde{y}_i|x_i, s_i) P_{X|S}(x_i|s_i) = P_{Y|S}(\tilde{y}_i|s_i) \qquad (3.33)
\end{aligned}
$$

where (a) is due to (3.2) and (b) is due to the fact that X_i is uniformly distributed. The channel associated with the HMP is memoryless and as such it is stationary.

Therefore, since the state process is stationary and ergodic both the output and noise processes are stationary and ergodic; this is stated in the following lemma:

Lemma 3.1.6. *For the quasi-symmetric FSM channel* $[\mathcal{X}, \mathcal{Y}, \mathcal{S}, P_S, Z, T, \Phi]$*, the noise process is stationary ergodic. Also the output process is stationary ergodic under an i.u.d. input.*

We can now complete the proof of Theorem 3.1.1 and conclude that feedback does not increase capacity for the class of quasi-symmetric FSM channels satisfying Assumption 3.1.1.

Proof of Theorem 3.1.1: With (3.31) we already have a converse for the feedback capacity. We need to show that this bound is achievable. We first note that by Lemma 3.1.6 the noise and output processes are stationary which imply that

$$\begin{aligned} C_{FB} &\leq \liminf_{n\to\infty} \sup_{\left\{P_{X|X_{[i-1]},Y_{[i-1]}}(x_i|x_{[i-1]},y_{[i-1]})\right\}_{i=1}^{n}} \frac{1}{n}\sum_{i=1}^{n} H(Y_i|Y_{[i-1]}) - H(Y_i|Y_{[i-1]}, X_{[i]}) \\ &= \liminf_{n\to\infty} \frac{1}{n}[H(\tilde{Y}_{[n]}) - H(Z_{[n]})] \\ &= \lim_{n\to\infty} \frac{1}{n}[H(\tilde{Y}_{[n]}) - H(Z_{[n]})] = \mathcal{H}(\tilde{Y}) - \mathcal{H}(Z). \end{aligned} \tag{3.34}$$

It is sufficient to show that the bound in (3.34) is achievable. We now remark that there exists a coding policy which achieves this bound. Note that since the noise process is stationary and ergodic, it can be shown that $\mathcal{H}(\tilde{Y}) - \mathcal{H}(Z)$ is an admissible rate (e.g. see [TM09, Theorem 5.3] and [VH94, Theorem 2]). Thus,

$$C_{FB} \geq \lim_{n\to\infty} \frac{1}{n}[H(\tilde{Y}_{[n]}) - H(Z_{[n]})] = \mathcal{H}(\tilde{Y}) - \mathcal{H}(Z)$$

and this completes the proof. □

Corollary 3.1.1. *Feedback does not increase capacity of quasi-symmetric FSM channels satisfying Assumption 3.1.1 (i.e., for which* $\sum_x f_s(\Phi(x, y))$ *is invariant with*

$s \in \mathcal{S}$).

Proof. The result follows by noting that a non-feedback code is a special case of a feedback code and that the non-feedback capacity is also achieved by uniform input distributions. This can be shown more explicitly as follows

$$\begin{aligned} C_{FB} &= \lim_{n\to\infty} \frac{1}{n} H(\tilde{Y}_{[n]}) - \lim_{n\to\infty} \frac{1}{n} H(Z_{[n]}) \\ &\stackrel{(i)}{=} \lim_{n\to\infty} \frac{1}{n} H(Y_{[n]}) \Big|_{P_{X_{[n]}}(x_{[n]}) = \frac{1}{|\mathcal{X}|^n}} - \lim_{n\to\infty} \frac{1}{n} H(Z_{[n]}) \\ &\leq \lim_{n\to\infty} \frac{1}{n} \sup_{P_{X_{[n]}}(x_{[n]})} I(X_{[n]}; Y_{[n]}) = C_{NFB} \end{aligned}$$

where C_{NFB} is the non-feedback capacity and (i) is valid since the input process is i.u.d. Finally, since $C_{FB} \geq C_{NFB}$, we obtain that $C_{FB} = C_{NFB}$. □

3.1.3 Examples of Quasi-Symmetric FSM Channels

In this section, we present examples of quasi-symmetric FSM channels which satisfy Assumption 3.1.1 and hence have identical feedback and non-feedback capacities. We also provide their feedback capacity expression which, when not given in single-letter form, can be simulated using existing algorithms (e.g., see [ALV+06]) for the computation of entropy rates of HMPs.

***A. Gilbert-Elliot Channel* (e.g., [MD89]):** One of the widely used FSM channels is the Gilbert-Elliot channel denoted by $[\mathcal{X}, \mathcal{Y}, \mathcal{S}, \mathcal{P}, \mathcal{C}]$, where $\mathcal{X} = \mathcal{Y} = \mathcal{S} = \{0, 1\}$. The two states are called "bad" state and "good" state, respectively, and the state transition matrix is given by:

$$P = \begin{bmatrix} 1-g & g \\ b & 1-b \end{bmatrix},$$

where $0 < g < 1, 0 < b < 1$ and in either of these two states, the channel is a binary symmetric channel (BSC) with the following transition matrices for states $s = 0$ and $s = 1$, respectively:

$$Q^0 = \begin{bmatrix} 1-p_G & p_G \\ p_G & 1-p_G \end{bmatrix}, Q^1 = \begin{bmatrix} 1-p_B & p_B \\ p_B & 1-p_B \end{bmatrix}.$$

From the above channel transition matrices, it can be observed that the Gilbert-Elliot channel is a symmetric FSM channel by Definition 3.1.4. Then, there exists a random variable $Z = \Phi(X,Y)$ with alphabet $\mathcal{Z} = \{0,1\}$ and a function $f_s(z)$ such that, $f_0(0) = 1-p_G$ and $f_0(1) = p_G$, $f_1(0) = 1-p_B$ and $f_0(1) = p_B$. Therefore, one can define the $T[s,z]$ matrix for this channel as

$$T = \begin{bmatrix} 1-p_G & p_G \\ 1-p_B & p_B \end{bmatrix},$$

and we obtain that $\Phi(X,Y) = X \oplus Y$, where $\oplus$ represents modulo-2 addition, and $T[s,z]$ defined above. By Corollary 3.1.1, feedback does not increase the capacity of the Gilbert-Elliot channel and it should be noted that this result is a special case of [Ala95] and [SP09]. Since $|\mathcal{X}| = 2$, the feedback capacity of the Gilbert-Elliot channel can be found as

$$C_{FB} = C_{NFB} = 1 - \mathcal{H}(Z),$$

where $\mathcal{H}(Z)$ is the entropy rate of the HMP $\{Z_i\}_{i=1}^{\infty}$ and can be computed as shown in [MD89] or [ALV$^+$06].

***B. Discrete Modulo Additive Channel with Markovian Noise*:** Consider the discrete channel with a common alphabet $\mathcal{A} = \{0,1,\ldots,q-1\}$ for the input, output and noise processes. The channel is described by the equation $Y_n = X_n \oplus Z_n$, for $n = 1,2,3,\ldots$, and Y_n, X_n and Z_n denotes the output, input and noise processes respectively. The noise process, $\{Z_n\}_{n=1}^{n=\infty}$, is Markovian and it is independent of the input process. It is straightforward to see that the channel transition matrix for this channel is symmetric for each state, where the state is given by the previous noise variable: $S_i = Z_{i-1}$. For simplicity, let us assume that $q = 3$. Then, the channel transition matrix at state s_i, Q^{s_i}, will be as follows:

$$Q^{s_i} = \begin{bmatrix} P(Z_i = 0|Z_{i-1} = s_i) & P(Z_i = 1|Z_{i-1} = s_i) & P(Z_i = 2|Z_{i-1} = s_i) \\ P(Z_i = 2|Z_{i-1} = s_i) & P(Z_i = 0|Z_{i-1} = s_i) & P(Z_i = 1|Z_{i-1} = s_i) \\ P(Z_i = 1|Z_{i-1} = s_i) & P(Z_i = 2|Z_{i-1} = s_i) & P(Z_i = 0|Z_{i-1} = s_i) \end{bmatrix}.$$

For each state, the channel transition matrix will still be symmetric with the same row permutation order. Furthermore, it also satisfies Assumption 3.1.1 since column sums are always one. Therefore, the discrete modulo additive channel is a symmetric FSM channel with $\mathcal{A} = \{0, 1, 2\}$ and $\Phi(X, Y) = X \oplus Y$. Hence, by Corollary 3.1.1, feedback does not increase the capacity of the discrete modulo additive channel with Markovian noise. Note that for this channel uniform input gives uniform output and therefore, feedback capacity of this channel is $C_{FB} = C_{NFB} = \log 3 - \mathcal{H}(Z) = H(Z_2|Z_1)$ where $\mathcal{H}(Z) = H(Z_2|Z_1)$ is the entropy rate of Markov noise $\{Z_i\}_{i=1}^{\infty}$. This example can be readily extended for the case of Mth order Markovian noise; in that case the state S_i is given by $S_i = (Z_{i-1}, \cdots, Z_{i-M})$ and the noise entropy rate is $\mathcal{H}(Z) = H(Z_{M+1}|Z^M)$.

This result is a special case of [Ala95]. It has been recently extended to finite-state multiple access channels in [PWC09].

***C. A Symmetric Discrete Channel with Markovian Noise*:** Consider a discrete, not necessarily additive, channel with Markovian noise [AF94]. More precisely, consider the channel given by $Y_i = f(X_i, Z_i)$ for $i = 1, 2, \cdots$ where X_i, Z_i and Y_i are the input, noise and output of the channel, respectively, and $f : \mathcal{X} \times \mathcal{Z} \to \mathcal{Y}$

is a given function. Assume also that $\{X_i\}$ and $\{Z_i\}$ are independent from each other and the channel satisfies the following properties.[5]

1. $|\mathcal{X}| = |\mathcal{Y}| = |\mathcal{Z}| = q$.

2. Given the input x, $f(x,.)$ is one-to-one; i.e., $\forall x \in \mathcal{X}\ \ f(x,z) = f(x,\bar{z}) \Rightarrow z = \bar{z}$.

3. f^{-1} exists such that $z = f^{-1}(x,y)$ and given y, $f^{-1}(.,y)$ is one-to-one; i.e., $\forall y \in \mathcal{Y}\ \ f^{-1}(x,y) = f^{-1}(\bar{x},y) \Rightarrow x = \bar{x}$.

We note that a channel satisfying these conditions has a symmetric channel transition matrix for each state, where the state is given by the previous noise variable: $S_i = Z_{i-1}$. Therefore, this channel is a symmetric FSM channel with the same permutation order determined by the function f. It also satisfies Assumption 3.1.1 as the column sums are one for each state. Therefore, by Corollary 3.1.1, feedback does not increase the capacity of these channels. This result is first shown in [AF94], where the noise process may be non-Markovian and non-ergodic in general. Similar to the previous example, uniform input yields uniform output for this channel and therefore, feedback capacity of this channel is $C_{FB} = C_{NFB} = \log q - \mathcal{H}(Z) = \log q - H(Z_2|Z_1)$. As in the previous example, this example can be extended for the case of Mth order Markov noise.

We next present two different channels which illustrate the result of this chapter when the column sums for each state are different than one.

[5] In [AF94], it is stated that $|\mathcal{X}| = |\mathcal{Z}| = q$. However, following the proof, it can be evidently seen that $|\mathcal{Y}| = q$ is also assumed.

***D. Binary Channel with Erasures, Errors and Markovian State*:** Consider the two-state channel given by $\mathcal{X} = \{0, 1\}$, $\mathcal{S} = \{s_1, s_2\}$, where $\{S_i\}$ is Markovian, $\mathcal{Y} = \{0, E, 1\}$ with the following channel transition matrices

$$Q^{s_1} = \begin{bmatrix} 1-\varepsilon-\xi & \xi & \varepsilon \\ \varepsilon & \xi & 1-\varepsilon-\xi \end{bmatrix}, \quad Q^{s_2} = \begin{bmatrix} 1-\varepsilon'-\xi' & \xi' & \varepsilon' \\ \varepsilon' & \xi' & 1-\varepsilon'-\xi' \end{bmatrix}$$

where $0 < \varepsilon, \xi, \varepsilon', \xi' < 1$ are fixed. We first note that this channel is a two-state quasi-symmetric FSM channel, since we can partition Q^{s_1} and Q^{s_2} in two symmetric sub-arrays given by

$$\tilde{Q}^{s_1}_{\mathcal{Y}_1} = \begin{bmatrix} 1-\varepsilon-\xi & \varepsilon \\ \varepsilon & 1-\varepsilon-\xi \end{bmatrix}, \quad \tilde{Q}^{s_1}_{\mathcal{Y}_2} = \begin{bmatrix} \xi \\ \xi \end{bmatrix}$$

and

$$\tilde{Q}^{s_2}_{\mathcal{Y}_1} = \begin{bmatrix} 1-\varepsilon'-\xi' & \varepsilon' \\ \varepsilon' & 1-\varepsilon'-\xi' \end{bmatrix}, \quad \tilde{Q}^{s_2}_{\mathcal{Y}_2} = \begin{bmatrix} \xi' \\ \xi' \end{bmatrix}$$

respectively, where $\mathcal{Y}_1 = \{0, 1\}$ and $\mathcal{Y}_1 = \{E\}$ with identical permutation order between states. For this channel, if we set $\xi = \xi'$, then we automatically satisfy Assumption 3.1.1 since the column sums in both Q^{s_1} and Q^{s_2} will be $1-\xi$, 2ξ and $1-\xi$ respectively. In other words, although the error probabilities are different across the states ($\epsilon \neq \epsilon'$ in general), we still have identical column sums. Therefore, by Corollary 3.1.1, feedback does not increase the capacity of this channel. Furthermore, since both the output and noise process are HMPs the value of feedback capacity can be computed using [ALV$^+$06].

***E. Non-Binary Noise Discrete Channel with Markovian Noise*:** We now present a binary-input 2^q-ary output communication channel with memory which was recently introduced in [PA09], [PAM12] (in the absence of feedback). This channel, which we refer to as the non-binary noise channel (NBNDC), is explicitly described by the following equation

$$Y_k = (2^q - 1)X_k + (-1)^{X_k} Z_k \tag{3.35}$$

for $k = 1, 2, \cdots$, where $X_k \in \mathcal{X} = \{0, 1\}$ is the input, $Y_k, Z_k \in \mathcal{Z} = \mathcal{Y} = \{0, 1, \cdots, 2^q - 1\}$ is the output and the noise processes, respectively. The noise and input processes are independent from each other and we assume that the noise process is Markovian (an Mth order Markov process can also be considered as examined in [PA09] for modeling the underlying fading channel). For the sake of simplicity, we consider the NBNDC channel with $q = 2$. Let $\Lambda = [\lambda_{s_i,j}]_{i=1,\cdots,4;j=1,\cdots,4}$, where $\lambda_{s_i,j} \triangleq P(Z_i = j|Z_{i-1} = s_i)$, denotes the transition probability matrix of the noise process. Then, with the state $S_i = Z_{i-1}$, the channel transition matrix at state s_i, Q^{s_i}, is given by

$$Q^{s_i} = \begin{bmatrix} \lambda_{s_i,0} & \lambda_{s_i,1} & \lambda_{s_i,2} & \lambda_{s_i,3} \\ \lambda_{s_i,3} & \lambda_{s_i,2} & \lambda_{s_i,1} & \lambda_{s_i,0} \end{bmatrix}.$$

Note that NBNDC is a quasi-symmetric FSM channel but it does not necessarily satisfy Assumption 3.1.1 . However, it can be easily shown that for any Λ satisfying that both $\sum_{j=0,3} \lambda_{i,j}$ and $\sum_{j=1,2} \lambda_{i,j}$ do not change with different i values, Assumption 3.1.1 is satisfied; therefore, by Corollary 3.1.1, feedback does not increase capacity of such NBNDC channels. Furthermore, the non-feedback capacity of NBNDC is given in [PA09] as $C_{NFB} = 1 + \mathcal{H}(W) - H(Z_2|Z_1)$, where $\mathcal{H}(W)$ is the entropy rate of the process $\{W_k\}$ which is defined on the alphabet $\mathcal{W} = \{0, 1, \cdots, 2^{q-1} - 1\}$ with $W_k = \min\{Z_k, 2^q - 1 - Z_k\}$. Therefore, if Λ satisfies the condition that both $\sum_{j=0,3} \lambda_{i,j}$ and $\sum_{j=1,2} \lambda_{i,j}$ do not change with different i values, we then have $C_{FB} = C_{NFB} = 1 + \mathcal{H}(W) - H(Z_2|Z_1)$. Note that $\{W_k\}$ is an HMP and as such $\mathcal{H}(W)$ can be computed as shown in [ALV$^+$06].

There is one more quasi-symmetric FSM channel that needs further attention. We now investigate how its channel properties directly satisfy the condition that the previous feedback control actions do not affect the current value of the conditional output entropy. In other words, the example below satisfies Lemma 3.1.4 without having the condition that the column sums are identical among different states, (i.e., it does not satisfy Assumption 3.1.1).

***F. Simplified Binary Erasure Channel with Markovian State*:** Consider the following binary erasure channel [DG06], which is a simplified (special) case of the erasure channel of Example D and has been used to model packet losses in a packet communication network, such as the Internet. The channel has binary input and ternary output; $\mathcal{X} = \{0, 1\}, \mathcal{Y} = \{0, E, 1\}$. Let S_i denote the state of the erasure channel when the packet i arrives such that when $S_i = 1$, the packet is erased, and when $S_i = 0$, the packet gets through. For a given input, the channel output is identical to the input if there is no erasure, and it is equal to the erasure symbol (E) if an erasure occurs. Therefore, the channel transition matrices at states $0, 1$ will be as follows

$$Q^0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } Q^1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

This channel can be considered as a special case of deletion channel in which the erased packet is assumed to be known by the decoder. Therefore, in an erasure channel, the receiver has also the side information about the state. In [DG06], it is shown that feedback does not increase the capacity of this channel. We herein note that the approach presented in this thesis gives the same result.

Proposition 3.1.1. *Feedback does not increase capacity of simplified binary erasure channel with Markovian state and the feedback capacity is achieved by an i.u.d. input.*

Proof. We first note that since the channel is quasi-symmetric for each state, the conditional output entropy is maximized by uniform input distributions. What we further need to show is the independence of the value attained by $H(Y_i|Y_{[i-1]} = y_{[i-1]})$ from previous input control actions. In particular, we need to show that $P_{S_i|Y_{[i-1]}}(s_i|y_{[i-1]})$ is independent of past input control actions (see Lemma 3.1.4). It should be noted that

$$P_{S_i|Y_{[i-1]}}(s_i|y_{[i-1]}) = \sum_{s_{i-1}} P_{S_i|S_{i-1}}(s_i|s_{i-1}) P_{S_{i-1}|Y_{[i-1]}}(s_{i-1}|y_{[i-1]}).$$

Thus, given $y_{[i-1]}$, s_{i-1} is deterministic and independent of $x_{[i-1]}$. Integrating this fact in our approach proves the desired result. □

It has been shown that [DG06, Proposition 3.1] the capacity of this channel, with and without feedback, is given by $C_{FB} = C_{NFB} = (1 - p_e)$ where p_e is the erasure probability.

This particular example has the benefit of learning the state deterministically by only observing the output. We should remark that availability of both the state information and output feedback, where we discuss in more detail in the next section, has also been considered within different setups in some other works and the situations for which feedback does not help increasing capacity are determined (see [PWG09, Theorem 19] and [Vis99]).

3.2 Feedback Capacity of Symmetric FSM Channels with CSIR

In the previous section, we show that if a quasi-symmetric finite-state Markov channel satisfies Assumption 3.1.1 then feedback cannot increase the capacity. We now show that if complete CSI is available at the decoder of a quasi-symmetric FSM channel then feedback can not increase the capacity even in the absence of this condition.

Note that by solely modifying the decoding function, a feedback code now consists of a sequence of encoding functions

$$\psi_i : \{1, 2, ..., 2^{nR}\} \times \mathcal{Y}^{i-1} \to \mathcal{X}$$

for $i = 1, 2, ...n$ and an associated decoding function

$$\Upsilon : \mathcal{Y}^n \times \mathcal{S}^n \to \{1, 2, ..., 2^{nR}\}.$$

The main result of this section is as follows.

Theorem 3.2.1. *For a quasi-symmetric FSM channel $[\mathcal{X}, \mathcal{Y}, \mathcal{S}, P_S, Z, T, \Phi]$ with complete CSI available at the decoder its feedback capacity is given by*

$$C^s_{FB} = \mathcal{H}(\tilde{Y}) - \mathcal{H}(Z|S)$$

where $\mathcal{H}(\tilde{Y})$ is the entropy rate of the output process $\{\tilde{Y}_i\}$ driven by an i.u.d. input and $\mathcal{H}(Z|S)$ is the conditional entropy rate of the process $\{Z_i\}_{i=1}^{\infty}$.

We devote the remainder of the section to prove this theorem. To start, with Fano's inequality, we can show that $C^s_{FB} \leq \liminf_{n\to\infty} \sup_{\{\psi_i\}_{i=1}^n} \frac{1}{n} I(W; Y_{[n]}, S_{[n]})$. Furthermore,

$$\liminf_{n\to\infty} \sup_{\{\psi_i\}_{i=1}^n} \frac{1}{n} I(W; Y_{[n]}, S_{[n]})$$

$$= \liminf_{n\to\infty} \sup_{\{\psi_i\}_{i=1}^n} \frac{1}{n}\sum_{i=1}^{n} \left[H(Y_i, S_i|Y_{[i-1]}, S_{[i-1]}) - H(Y_i, S_i|W, Y_{[i-1]}, S_{[i-1]})\right]$$

$$= \liminf_{n\to\infty} \sup_{\{\psi_i\}_{i=1}^n} \frac{1}{n}\sum_{i=1}^{n} \left[H(Y_i|S_i, Y_{[i-1]}, S_{[i-1]}) - H(Y_i|W, S_i, Y_{[i-1]}, S_{[i-1]})\right] \quad (3.36)$$

$$= \liminf_{n\to\infty} \sup_{\{\psi_i\}_{i=1}^n} \frac{1}{n}\sum_{i=1}^{n} \left[H(Y_i|S_i, Y_{[i-1]}, S_{[i-1]}) - H(Y_i|W, S_i, Y_{[i-1]}, S_{[i-1]}, X_{[i]})\right] \quad (3.37)$$

$$\leq \liminf_{n\to\infty} \sup_{\{\mathcal{D}_i\}_{i=1}^n} \frac{1}{n}\sum_{i=1}^{n} \left[H(Y_i|S_i, Y_{[i-1]}, S_{[i-1]}) - H(Y_i|S_i, Y_{[i-1]}, X_{[i]})\right] \quad (3.38)$$

$$= \liminf_{n\to\infty} \sup_{\{\mathcal{D}_i\}_{i=1}^n} \frac{1}{n}\sum_{i=1}^{n} \left[H(Y_i|S_i, Y_{[i-1]}, S_{[i-1]}) - H(Z_i|S_i)\right] \quad (3.39)$$

$$= \liminf_{n\to\infty} \sup_{\{\mathcal{D}_i\}_{i=1}^n} \frac{1}{n}\sum_{i=1}^{n} \left[H(Y_i|S_i, Y_{[i-1]}) - H(Z_i|S_i)\right] \quad (3.40)$$

where (3.36) is due to (3.1), (3.38) is due to (3.2) and (3.39) is due to Lemmas 3.1.2 and 3.1.5. Recall also that

$$\mathcal{D}_i \triangleq \{P_{X_i|X_{[i-1]},Y_{[i-1]}}(x_i|x_{[i-1]}, y_{[i-1]}) : x_{[i-1]} \in \mathcal{X}^{i-1}, y_{[i-1]} \in \mathcal{Y}^{i-1}\}.$$

Finally, (3.40) follows since

$$\begin{aligned}
&P_{Y_i|S_i,Y_{[i-1]},S_{[i-1]}}(y_i|s_i, y_{[i-1]}, s_{[i-1]}) \\
&= \sum_{x_i}\sum_{x_{[i-1]}} p_{\mathcal{C}}(y_i|x_i, s_i) P_{X_i|X_{[i-1]},Y_{[i-1]}}(x_i|x_{[i-1]}, y_{[i-1]}) P_{X_{[i-1]}|S_{[i]},Y_{[i-1]}}(x_{[i-1]}|s_{[i]}, y_{[i-1]}) \\
&= \sum_{x_i}\sum_{x_{[i-1]}} p_{\mathcal{C}}(y_i|x_i, s_i) P_{X_i|X_{[i-1]},Y_{[i-1]}}(x_i|x_{[i-1]}, y_{[i-1]}) \\
&\qquad \prod_{j=1}^{i-1} P_{X_j|X_{[j-1]},Y_{[j-1]}}(x_j|x_{[j-1]}, y_{[j-1]}) \qquad (3.41) \\
&= P_{Y_i|S_i,Y_{[i-1]}}(y_i|s_i, y_{[i-1]})
\end{aligned}$$

where (3.41) follows due to the fact that the feedback input at time j depends only on $(x_{[j-1]}, y_{[j-1]})$.

Now, let us consider the following equation

$$\sup_{\{\mathcal{D}_i\}_{i=1}^n} \frac{1}{n} \sum_{i=1}^{n} \left[H(Y_i|S_i, Y_{[i-1]}) - H(Z_i|S_i) \right]. \tag{3.42}$$

Since $H(Z_i|S_i)$ is independent of $\{\mathcal{D}_i\}_{i=1}^n$, it is sufficient to show that the term $\sum_{i=1}^n H(Y_i|S_i, Y_{[i-1]})$ is maximized by uniform $\{P_{X_i|X_{[i-1]},Y_{[i-1]}}(x_i|x_{[i-1]}, y_{[i-1]})\}_{i=1}^n$.

Observe first that

$$H(Y_i|S_i, Y_{[i-1]}) = \sum_{s_i, y_{[i-1]}} P_{S_i,Y_{[i-1]}}(s_i, y_{[i-1]}) H(Y_i|Y_{[i-1]} = y_{[i-1]}, S_i = s_i) \tag{3.43}$$

where

$$H(Y_i|Y_{[i-1]} = y_{[i-1]}, S_i = s_i) = -\sum_{y_i} P_{Y_i|Y_{[i-1]},S_i}(y_i|y_{[i-1]}, s_i) \log P_{Y_i|Y_{[i-1]},S_i}(y_i|y_{[i-1]}, s_i).$$

Therefore, the analysis in Lemma 3.1.3 still holds and hence,

$$\begin{aligned} \mathcal{D}_i^\star &\stackrel{\triangle}{=} \operatorname*{argmax}_{\mathcal{D}_i} H(Y_i|S_i, S_{[i-1]}, Y_{[i-1]}) \\ &= \left\{ P_{X_i|X_{[i-1]},Y_{[i-1]}}(x_i|x_{[i-1]}, y_{[i-1]}) = \frac{1}{|\mathcal{X}|}, \forall x_{[i-1]} \in \mathcal{X}^{i-1}, \forall y_{[i-1]} \in \mathcal{Y}^{i-1} \right\} \end{aligned}$$

for all $x_i \in \mathcal{X}$ and for all $i = 1, \cdots, n$. However, in order to claim the optimality of uniform input distributions we still need to eliminate the effects of past control actions as satisfied via Assumption 3.1.1 in the previous section.

Hence, consider $P_{Y_i|S_i,Y_{[i-1]}}(y_i|s_i,_{[i-1]})$ when $P_{X_i|X_{[i-1]},Y_{[i-1]}}(x_i|x_{[i-1]}, y_{[i-1]})$ is uniform. We have

$$\begin{aligned} P_{Y_i|S_i,Y_{[i-1]}}(y_i|s_i,_{[i-1]}) &= \sum_{x_i, x_{[i-1]}} p_{\mathcal{C}}(y_i|x_i, s_i) P_{X_i|X_{[i-1]},Y_{[i-1]}}(x_i|x_{[i-1]}, y_{[i-1]}) \\ &\qquad P_{X_{[i-1]}|S_i,Y_{[i-1]}}(x_{[i-1]}|s_i, y_{[i-1]}) \\ &= \frac{1}{|\mathcal{X}|} \sum_{x_i, x_{[i-1]}} p_{\mathcal{C}}(y_i|x_i, s_i) P_{X_{[i-1]}|S_i,Y_{[i-1]}}(x_{[i-1]}|s_i, y_{[i-1]}) \\ &= \frac{1}{|\mathcal{X}|} \sum_{x_{[i-1]}} P_{X_{[i-1]}|S_i,Y_{[i-1]}}(x_{[i-1]}|s_i, y_{[i-1]}) \sum_{x_i} p_{\mathcal{C}}(y_i|x_i, s_i) \end{aligned}$$

$$= \frac{1}{|\mathcal{X}|} \sum_{x_i} f_{s_i}(\Phi(x_i, y_i))$$

Note that the term $\frac{1}{|\mathcal{X}|} \sum_{x_i} f_{s_i}(\Phi(x_i, y_i))$ is independent of past control actions and therefore, we can conclude that $\sum_{i=1}^{n} H(Y_i|S_i, Y_{[i-1]})$ is maximized by uniform feedback control actions.

So far we have $H(Y_i|S_i, X_{[i]}, Y_{[i-1]}) = H(Z_i|S_i)$ and that $\sum_{i=1}^{n} H(Y_i|S_i, Y_{[i-1]})$ is maximized by uniform input distributions. Note that as shown in (3.33) under uniform distribution $H(Y_i|S_i, Y_{[i-1]}) = H(Y_i|S_i)$. The proof of Theorem 3.2.1 can now be completed directly by following the proof of Theorem 3.1.1 and hence, we can conclude that feedback does not increase capacity of quasi-symmetric FSM channels when CSI is available at the receiver.

3.3 Conclusion and Remarks

In this chapter, we presented a class of symmetric channels which encapsulates a variety of discrete channels with memory. Motivated by several results in the literature, we established a class of symmetric FSM channels for which feedback does not increase their capacity. We showed this result by first reformulating the optimization problem in terms of dynamic programming and then proving that, under feedback, the capacity achieving distribution is uniform. An important observation should be highlighted again; when feedback exists, one can learn the channel via the past control actions and as such may apply a nonuniform distribution which will result in a higher output entropy and capacity. We present a sufficient condition, Assumption 3.1.1, under which it is still possible to learn the channel via these past control actions; however, this learning does not affect the optimal distribution. We then observe that

Assumption 3.1.1 is not required when channel state information is available at the receiver. It is also worth observing that even though we have emphasized finite-state channels with Markovian state (i.e., FSM channels) due to their wide use in the literature, our result also holds when the state process is not Markovian but still stationary ergodic.[6] Finally, although this result covers a large class of discrete channels with memory, we believe that by adopting the approach of this thesis, it is possible to show a similar result for a further general class of both symmetric and asymmetric channels whose feedback capacity is achieved by an i.i.d. input, both in the single-user and multiple-user settings.

[6] In this case, Property (I) is modified by replacing $P_{S_i|S_{i-1}}(s_i|s_{i-1})$ with $P_{S_i|S_{[i-1]}}(s_i|s_{[i-1]})$ and the noise process is no longer an HMP but remains stationary ergodic.

Chapter 4

Multiple Access Channel with Receiver Side Information

In this chapter, we study the problem of reliable communication over multiple-access channels where the channel is driven by an independent and identically distributed state process and the encoders and the decoder are provided with various degrees of asymmetric noisy channel state information (CSI). The essential requirement we impose is that the noisy CSI available to the encoders is realized via the corruption of CSI by different noise processes, which gives a realistic physical structure of the communication setup. We also allow the receiver to observe full CSI in the majority of the results (the treatment for the case when there is no CSI at the receiver is discussed in the next chapter). We consider several scenarios. For the case where the encoders observe causal, asymmetric noisy CSI and the decoder observes complete CSI, inner and outer bounds to the capacity region, which are tight for the sum-rate capacity, are provided. Next, single-letter characterizations for the channel capacity regions under each of the following system settings are established: (a) the CSI at the encoders are asymmetric deterministic functions of the CSI at the decoder and the encoders have non-causal noisy CSI (its causal version is recently solved in [CY11]); (b) the encoders observe asymmetric noisy CSI with asymmetric delays and the decoder observes complete CSI; (c) a degraded message set scenario with asymmetric noisy CSI at the encoders and complete and/or noisy CSI at the decoder. The main component in these results is a generalization of a converse coding approach, recently introduced in [CY11] for the MAC with asymmetric quantized CSI at the encoders and herein considerably extended and adapted for the noisy CSI setup.

4.1 The Converse Coding Approach

The most relevant paper to this work is [CY11] which provides a converse coding approach for the state-dependent MAC where asymmetric partial state information is available at the encoders. In this work, we adopt and expand on the converse technique of this paper and use it in a noisy setup. The converse coding approach of [CY11] is based on team decision theoretic methods [Wit98] (see also [Yük13], [MT09] and [NT11] for recent team decision and control theoretic approaches as well as [YB13] for team decision theory as well as review of information structures) where the authors use *memoryless stationary team policies* which play a key role in showing that the past information is irrelevant. As the authors mention in [CY11, Remark 2], for the validity of their arguments, it would suffice that the state information available at the decoder contains the ones available at the two transmitters. In this way, the decoder does not need to estimate the coding policies used in decentralized time-sharing.

For the noisy setup, we need to modify this approach to account for the fact that the decoder does not have access to the state information at the encoders, and that the past state information does not lead to a tractable recursion. This difficulty is overcome by showing that a product form on the team policies exists in the noisy setup as well.

The rest of the chapter is organized as follows. In Section 4.2, we formally state scenarios (1)-(4), and present the main results and several observations. In Section 4.3, we provide two examples in one of which (the modulo-additive state-dependent MAC) we apply the result of [EZ00] and get the full capacity region by only considering the tightness of the sum-rate capacity. Finally, in Section 4.4, we present concluding remarks.

4.2 Main Results

We consider a two-user memoryless state-dependent MAC, with two encoders, a, b, and two independent message sources W_a and W_b which are uniformly distributed in the finite sets $\mathcal{W}_a$ and $\mathcal{W}_b$, respectively. The channel inputs from the encoders are $X^a \in \mathcal{X}_a$ and $X^b \in \mathcal{X}_b$, respectively, and the channel output is $Y \in \mathcal{Y}$. The channel state process is modeled as a sequence $\{S_t\}_{t=1}^{\infty}$ of i.i.d. random variables in some finite space $\mathcal{S}$. Let (S_t^a, S_t^b) denote a pair of random variables available at two encoders, a, b, respectively, at time t. Throughout the paper, by asymmetric side information we will refer to the case where $\{S_t^a \neq S_t^b\}$, $\forall t$. Furthermore, by noisy side information will refer to the case where (S_t^a, S_t^b, S_t) are correlated according to a given joint distribution $P_{S^a,S^b,S}(s^a, s^b, s)$.

4.2.1 Asymmetric Causal Noisy CSIT

Let the two encoders have access to a causal noisy version of the state information S_t at each time $t \geq 1$, modeled by $S_t^a \in \mathcal{S}_a$, $S_t^b \in \mathcal{S}_b$, respectively, where the joint distribution of (S_t, S_t^a, S_t^b) factorizes as

$$P_{S_t^a, S_t^b, S_t}(s_t^a, s_t^b, s_t) = P_{S_t^a|S_t}(s_t^a|s_t) P_{S_t^b|S_t}(s_t^b|s_t) P_{S_t}(s_t). \tag{4.1}$$

The system is depicted in Fig. 4.1. Let $\mathbf{W} := (W_a, W_b)$ and S_t be fully available at the receiver and assume that $\{(S_t, S_t^a, S_t^b)\}_{t=1}^{\infty}$ is a sequence of i.i.d. triples, independent from (W_a, W_b). Therefore, we have that for any $n \geq 1$,

$$P_{S_{[n]}, S_{[n]}^a, S_{[n]}^b, \mathbf{W}}(s_{[n]}, s_{[n]}^a, s_{[n]}^b, \mathbf{w}) = \prod_{t=1}^{n} \frac{1}{|\mathcal{W}_a|} \frac{1}{|\mathcal{W}_b|} P_{S_t^a|S_t}(s_t^a|s_t) P_{S_t^b|S_t}(s_t^b|s_t) P_{S_t}(s_t). \tag{4.2}$$

The channel inputs at time t, i.e., X_t^a and X_t^b, are functions of the locally available information $(W_a, S_{[t]}^a)$ and $(W_b, S_{[t]}^b)$, respectively. Let $\mathbf{X_t} := (X_t^a, X_t^b)$. Then, the laws governing n-sequences of state, input and output letters are given by

$$P_{Y_{[n]}|\mathbf{W}, \mathbf{X}_{[n]}, S_{[n]}, S_{[n]}^a, S_{[n]}^b}(y_{[n]}|\mathbf{w}, \mathbf{x}_{[n]}, s_{[n]}, s_{[n]}^a, s_{[n]}^b) = \prod_{t=1}^{n} P_{Y_t|X_t^a, X_t^b, S_t}(y_t|x_t^a, x_t^b, s_t), \tag{4.3}$$

where the channel's transition probability distribution, $P_{Y_t|X_t^a, X_t^b, S_t}(y_t|x_t^a, x_t^b, s_t)$, is given a priori.

Definition 4.2.1. *An $(n, 2^{nR_a}, 2^{nR_b})$ code with block length n and rate pair (R_a, R_b) for a state-dependent MAC with causal noisy state information consists of*

(1) A sequence of mappings for each encoder

$$\phi_t^{(a)} : \mathcal{S}_a^t \times \mathcal{W}_a \to \mathcal{X}_a, \;\; t = 1, 2, ...n;$$

$$\phi_t^{(b)} : \mathcal{S}_b^t \times \mathcal{W}_b \to \mathcal{X}_b, \;\; t = 1, 2, ...n.$$

2) An associated decoding function

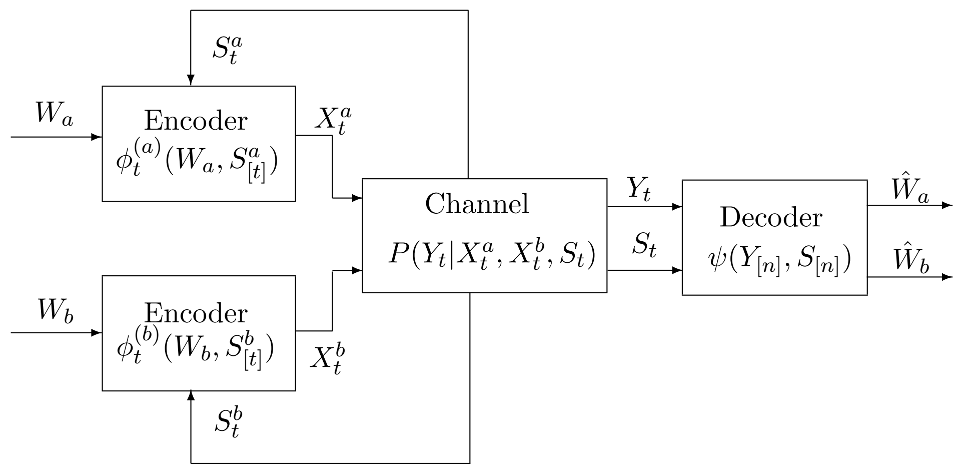

Figure 4.1: The multiple-access channel with asymmetric causal noisy state information.

$$\psi : \mathcal{S}^n \times \mathcal{Y}^n \to \mathcal{W}_a \times \mathcal{W}_b.$$

The system's probability of error, $P_e^{(n)}$, is given by

$$P_e^{(n)} = \frac{1}{2^{n(R_a+R_b)}} \sum_{w_a=1}^{2^{nR_a}} \sum_{w_b=1}^{2^{nR_b}} \Pr\left(\psi(Y_{[n]}, S_{[n]}) \neq (w_a, w_b) | \mathbf{W} = \mathbf{w}\right).$$

A rate pair (R_a, R_b) is achievable if for any $\epsilon > 0$, there exists, for all n sufficiently large an $(n, 2^{nR_a}, 2^{nR_b})$ code such that $\frac{1}{n}\log|\mathcal{W}_a| \geq R_a > 0$, $\frac{1}{n}\log|\mathcal{W}_b| \geq R_b > 0$ and $P_e^{(n)} \leq \epsilon$. The capacity region of the state-dependent MAC, $\mathcal{C}_{FS}$, is the closure of the set of all achievable rate pairs (R_a, R_b) and the sum-rate capacity is defined as $\mathcal{C}_{FS}^{\sum} := \max_{(R_a,R_b)\in\mathcal{C}_{FS}}(R_a + R_b)$.

Before proceeding with the main result, we introduce *memoryless stationary team policies* [CY11] and their associated rate regions. Let the set of all possible functions from $\mathcal{S}_a$ to $\mathcal{X}_a$ and $\mathcal{S}_b$ to $\mathcal{X}_b$ be denoted by $\mathcal{T}_a := \mathcal{X}_a^{|\mathcal{S}_a|}$ and $\mathcal{T}_b := \mathcal{X}_b^{|\mathcal{S}_b|}$, respectively. We shall refer to $\mathcal{T}_a$-valued and $\mathcal{T}_b$-valued random vectors as Shannon strategies.

Definition 4.2.2. *[CY11] A memoryless stationary (in time) team policy is a family*

$$\Pi = \{\pi = (\pi_{T^a}(\cdot), \pi_{T^b}(\cdot)) \in \mathcal{P}(\mathcal{T}_a) \times \mathcal{P}(\mathcal{T}_b)\} \tag{4.4}$$

of probability distribution pairs on $(\mathcal{T}_a, \mathcal{T}_b)$.

For every memoryless stationary team policy π*, let* $\mathcal{R}_{FS}(\pi)$ *denote the region of all rate pairs* $R = (R_a, R_b)$ *satisfying*

$$R_a < I(T^a; Y|T^b, S) \tag{4.5}$$

$$R_b < I(T^b; Y|T^a, S) \tag{4.6}$$

$$R_a + R_b < I(T^a, T^b; Y|S) \tag{4.7}$$

where S*,* T^a*,* T^b *and* Y *are random variables taking values in* $\mathcal{S}$*,* $\mathcal{T}_a$*,* $\mathcal{T}_b$ *and* $\mathcal{Y}$*, respectively, and whose joint probability distribution factorizes as*

$$P_{S,T^a,T^b,Y}(s, t^a, t^b, y) = P_S(s) P_{Y|T^a,T^b,S}(y|t^a, t^b, s) \pi_{T^a}(t^a) \pi_{T^b}(t^b). \tag{4.8}$$

Let $\mathcal{C}_{IN} := \overline{co}\left(\bigcup_\pi \mathcal{R}_{FS}(\pi) \right)$ denote the closure of the convex hull of the rate regions $\mathcal{R}_{FS}(\pi)$ given by (4.5)-(4.7) associated to all possible memoryless stationary team polices as defined in (4.4).

Theorem 4.2.1 (Inner Bound to $\mathcal{C}_{FS}$)**.** $\mathcal{C}_{IN} \subseteq \mathcal{C}_{FS}$.

Proof of Theorem 4.2.1. Fix $(R_a, R_b) \in \mathcal{R}_{FS}(\pi)$.

Codebook Generation Fix $\pi_{T^a}(t^a)$ and $\pi_{T^b}(t^b)$. For each $w_a \in \{1, \cdots, 2^{nR_a}\}$, randomly generate its corresponding n-tuple $t^a_{[n],w_a}$, each according to $\prod_{i=1}^n \pi_{T^a_i}(t^a_{i,w_a})$. Similarly, for each $w_b \in \{1, \cdots, 2^{nR_b}\}$, randomly generate its corresponding n-tuple $t^b_{[n],w_b}$, each according to $\prod_{i=1}^n \pi_{T^b_i}(t^b_{i,w_b})$. The set of these codeword pairs form the codebook, which is revealed to the decoder while codewords t^l_{i,w_l} are revealed to encoder l, $l = \{a, b\}$.

Encoding Define the encoding functions as follows: $x_i^a(w_a) = \phi_i^a(w_a, s_{[i]}^a) = t_{i,w_a}^a(s_i^a)$ and $x_i^b(w_b) = \phi_i^b(w_b, s_{[i]}^b) = t_{i,w_b}^b(s_i^b)$ where t_{i,w_a}^a and t_{i,w_b}^b denote the ith component of $t_{[n],w_a}^a$ and $t_{[n],w_b}^b$, respectively, and s_i^a and s_i^b denote the last components of $s_{[i]}^a$ and $s_{[i]}^b$, respectively, $i = 1, \cdots, n$. Therefore, to send the messages w_a and w_b, we simply transmit the corresponding $t_{[n],w_a}^a$ and $t_{[n],w_b}^b$, respectively.

Decoding After receiving $(y_{[n]}, s_{[n]})$, the decoder looks for the only (w_a, w_b) pair such that $(t_{[n],w_a}^a, t_{[n],w_b}^b, y_{[n]}, s_{[n]})$ are jointly $\epsilon-$typical and declares this pair as its estimate $(\hat{w}_a, \hat{w}_b)$.

Error Analysis Without loss of generality, we can assume that $(w_a, w_b) = (1, 1)$ was sent. An error occurs, if the correct codewords are not typical with the received sequence or there is a pair of incorrect codewords that are typical with the received sequence. Define the events $E_{\alpha,\beta} \triangleq \big\{(T_{[n],\alpha}^a, T_{[n],\beta}^b, Y_{[n]}, S_{[n]}) \in \mathrm{A}_\epsilon^n\big\}$, $\alpha \in \{1, \cdots, 2^{nR_a}\}$ and $\beta \in \{1, \cdots, 2^{nR_b}\}$. Then, by the union bound we get

$$\begin{aligned} P_e^n &= P\big(E_{1,1}^c \bigcup_{(\alpha,\beta)\neq(1,1)} E_{\alpha,\beta}\big) \\ &\leq P(E_{1,1}^c) + \sum_{\alpha=1,\beta\neq1} P(E_{\alpha,\beta}) + \sum_{\alpha\neq1,\beta=1} P(E_{\alpha,\beta}) + \sum_{\alpha\neq1,\beta\neq1} P(E_{\alpha,\beta}) \qquad (4.9) \end{aligned}$$

where $E_{1,1}^c$ denotes the complement set of $E_{1,1}$. Since $\{Y_i, S_i, T_i^a, T_i^b\}_{i=1}^\infty$ is an i.i.d. sequence and by [CT06, Theorem 15.2.1], $P(E_{1,1}^c) \to 0$ as $n \to \infty$. Next, let us consider the second term

$$\begin{aligned} \sum_{\alpha=1,\beta\neq1} P(E_{\alpha=1,\beta\neq1}) &= \sum_{\alpha=1,\beta\neq1} P((T_{[n],1}^a, T_{[n],\beta}^b, Y_{[n]}, S_{[n]}) \in \mathrm{A}_\epsilon^n) \\ &\overset{(i)}{=} \sum_{\alpha=1,\beta\neq1} \sum_{(t_{[n]}^a, t_{[n]}^b, y_{[n]}, s_{[n]}) \in \mathrm{A}_\epsilon^n} P_{T_{[n]}^b}(t_{[n]}^b) P_{T_{[n]}^a, Y_{[n]}, S_{[n]}}(t_{[n]}^a, y_{[n]}, s_{[n]}) \\ &\leq \sum_{\alpha=1,\beta\neq1} |A_\epsilon^n| 2^{-n[H(T^b)-\epsilon]} 2^{-n[H(T^a,Y,S)-\epsilon]} \qquad (4.10) \\ &\leq 2^{nR_b} 2^{-n[H(T^b)+H(T^a,Y,S)-H(T^a,T^b,Y,S)-3\epsilon]} \end{aligned}$$

$$\stackrel{(ii)}{=} \quad 2^{n[R_b - I(T^b;Y|S,T^a) - 3\epsilon]} \tag{4.11}$$

where (i) holds since for $\beta \neq 1$, $T^b_{[n],\beta}$ is independent of $(T^a_{[n],1}, Y_{[n]}, S_{[n]})$ and (ii) follows since T^b and (T^a, S) are independent and $I(T^b; Y, T^a, S) = I(T^b; T^a, S) + I(T^b; Y|T^a, S) = I(T^b; Y|T^a, S)$, where $I(T^b; T^a, S) = 0$. Following the same steps for $(\alpha \neq 1, \beta = 1)$ and $(\alpha \neq 1, \beta \neq 1)$ we get

$$\begin{aligned} \sum_{\alpha \neq 1, \beta = 1} P(E_{\alpha,\beta}) &\leq 2^{n[R_a - I(T^a;Y|T^b,S) - 3\epsilon]} \\ \sum_{\alpha \neq 1, \beta \neq 1} P(E_{\alpha,\beta}) &\leq 2^{n[R_a + R_b - I(T^a,T^b;Y|S) - 3\epsilon]}, \end{aligned} \tag{4.12}$$

and the rate conditions of the $\mathcal{R}_{FS}(\pi)$ imply that each term tends in (4.9) tends to zero as $n \to \infty$. This shows the achievability of a rate pair $(R_a, R_b) \in \mathcal{R}_{FS}(\pi)$. Achievability of any rate pair in $\mathcal{C}_{IN}$ follows from a standard time-sharing argument.

□

Let $\mathcal{C}_{OUT} := \left\{ (R_a, R_b) \in \mathbb{R}^+ \times \mathbb{R}^+ : R_a + R_b \leq \sup_{\pi_{T^a}(t^a)\pi_{T^b}(t^b)} I(T^a, T^b; Y|S) \right\}$ where $\mathbb{R}^+$ is the set of positive reals.

Theorem 4.2.2 (Outer Bound to $\mathcal{C}_{FS}$)**.** $\mathcal{C}_{FS} \subseteq \mathcal{C}_{OUT}$.

Proof of Theorem 4.2.2. We need to show that all achievable rates satisfy

$$R_a + R_b \leq \sup_{\pi_{T^a}(t^a)\pi_{T^b}(t^b)} I(T^a, T^b; Y|S),$$

i.e., a converse for the sum-rate capacity. Following [CY11], for $1 \leq t \leq n$, let

$$\alpha_\mu := \frac{1}{n} P_{S_{[t-1]}}(\mu) \text{ and } \eta(\epsilon) := \frac{\epsilon}{1-\epsilon} \log |\mathcal{Y}| + \frac{H(\epsilon)}{1-\epsilon}. \tag{4.13}$$

Observe that $\lim_{\epsilon \to 0} \eta(\epsilon) = 0$ and

$$\sum_{\mu \in \mathcal{S}^{(n)}} \alpha_\mu = \frac{1}{n} \sum_{1 \leq t \leq n} \sum_{\mu \in \mathcal{S}^{t-1}} P_{S_{[t-1]}}(\mu) = 1, \tag{4.14}$$

where $\mathcal{S}^{(n)}$ is the set of all $\mathcal{S}$-strings of length less than n. Let

$$\sum_{\mu\in\mathcal{S}^{(n)}} \alpha_\mu I(T_t^a, T_t^b; Y_t|S_t, S_{[t-1]} = \mu) := \sum_{1\le t\le n}\ \sum_{\mu\in\mathcal{S}^{t-1}} \alpha_\mu I(T_t^a, T_t^b; Y_t|S_t, S_{[t-1]} = \mu).$$

First recall that, for all $t \geq 1$, $X_t^a = \phi_t^{(a)}\left(W_a, S_{[t]}^a\right) = \phi_t^{(a)}\left(W_a, S_{[t-1]}^a, S_t^a\right)$ and $X_t^b = \phi_t^{(b)}\left(W_b, S_{[t]}^b\right) = \phi_t^{(b)}\left(W_b, S_{[t-1]}^b, S_t^b\right)$. Then, we can define the Shannon strategies $T_t^a \in \mathcal{T}_a$ and $T_t^b \in \mathcal{T}_b$ by putting, for every $s_a \in \mathcal{S}_a$ and $s_b \in \mathcal{S}_b$,

$$T_t^a(s_a) := \phi_t^{(a)}\left(W_a, S_{[t-1]}^a, s_a\right), \qquad T_t^b(s_b) := \phi_t^{(b)}\left(W_b, S_{[t-1]}^b, s_b\right). \tag{4.15}$$

We now show that the sum of any achievable rate pair can be written as the convex combinations of mutual information terms which are indexed by the realization of past complete CSI.

Lemma 4.2.1. *Let $T_t^a \in \mathcal{T}_a$ and $T_t^b \in \mathcal{T}_b$ be the Shannon strategies induced by $\phi_t^{(a)}$ and $\phi_t^{(b)}$, respectively, as shown in (4.15). Assume that a rate pair $R = (R_a, R_b)$, with block length $n \geq 1$ and a constant $\epsilon \in (0, 1/2)$, is achievable. Then,*

$$R_a + R_b \leq \sum_{\mu\in\mathcal{S}^{(n)}} \alpha_\mu I(T_t^a, T_t^b; Y_t|S_t, S_{[t-1]} = \mu) + \eta(\epsilon). \tag{4.16}$$

Proof. Let $\mathbf{T}_t := (T_t^a, T_t^b)$. By Fano's inequality, we get

$$H(\mathbf{W}|Y_{[n]}, S_{[n]}) \leq H(\epsilon) + \epsilon \log(|\mathcal{W}_a||\mathcal{W}_b|). \tag{4.17}$$

Observing that

$$\begin{aligned} I(\mathbf{W}; Y_{[n]}, S_{[n]}) &= H(\mathbf{W}) - H(\mathbf{W}|Y_{[n]}, S_{[n]}) \\ &= \log(|\mathcal{W}_a||\mathcal{W}_b|) - H(\mathbf{W}|Y_{[n]}, S_{[n]}). \end{aligned} \tag{4.18}$$

Combining (4.17) and (4.18) gives

$$(1-\epsilon)\log(|\mathcal{W}_a||\mathcal{W}_b|) \leq I(\mathbf{W}; Y_{[n]}, S_{[n]}) + H(\epsilon)$$

and

$$R_a + R_b \leq \frac{1}{n}\log(|\mathcal{W}_a||\mathcal{W}_b|) \leq \frac{1}{1-\epsilon}\frac{1}{n}\big(I(\mathbf{W};Y_{[n]},S_{[n]}) + H(\epsilon)\big)\,. \quad (4.19)$$

Furthermore,

$$\begin{aligned} I(\mathbf{W};Y_{[n]},S_{[n]}) &= \sum_{t=1}^{n}\big[H(Y_t,S_t|S_{[t-1]},Y_{[t-1]}) - H(Y_t,S_t|\mathbf{W},S_{[t-1]},Y_{[t-1]})\big] \\ &\overset{(i)}{=} \sum_{t=1}^{n}\big[H(Y_t|S_{[t]},Y_{[t-1]}) - H(Y_t|\mathbf{W},S_{[t]},Y_{[t-1]})\big] \\ &\overset{(ii)}{\leq} \sum_{t=1}^{n}\big[H(Y_t|S_{[t]}) - H(Y_t|\mathbf{W},S_{[t]},Y_{[t-1]},\mathbf{T}_t)\big] \\ &\overset{(iii)}{=} \sum_{t=1}^{n}\big[H(Y_t|S_{[t]}) - H(Y_t|S_{[t]},\mathbf{T}_t)\big] \\ &= \sum_{t=1}^{n} I(\mathbf{T}_t;Y_t|S_{[t]}) \end{aligned} \quad (4.20)$$

where (i) is implied by (4.2), in (ii) $\mathbf{T}_t := (T_t^a, T_t^b)$ are Shannon strategies whose realizations are mappings $t_t^i : S_t^i \to X_t^i$ for $i = \{a,b\}$ and thus (ii) holds since conditioning does not increase entropy. Finally, (iii) follows since

$$\begin{aligned} &P_{Y_t|\mathbf{W},S_t,S_{[t-1]},Y_{[t-1]},T_t^a,T_t^b}(y_t|\mathbf{w},s_t,s_{[t-1]},y_{[t-1]},t_t^a,t_t^b) \\ &\quad= \sum_{s_t^a,s_t^b} P_{Y_t|S_t,S_t^a,S_t^b,T_t^a,T_t^b}(y_t|s_t,s_t^a,s_t^b,t_t^a,t_t^b)P_{S_t^a,S_t^b|S_t}(s_t^a,s_t^b|s_t) \\ &\quad= P_{Y_t|S_t,T_t^a,T_t^b}(y_t|s_t,t_t^a,t_t^b) \end{aligned} \quad (4.21)$$

where the first equality is verified by (4.3) and (4.2), where $x_t^i = t_t^i(s_t^i)$ for $i = \{a,b\}$. At this point, it is worth to note that by (4.21), one can remove $S_{[t-1]}$ from (4.20) in the conditioning. However, we will soon observe why it is crucial to keep it when we prove the product form. Now, let $\chi(\epsilon) := \frac{H(\epsilon)}{n(1-\epsilon)}$ and combining (4.19)-(4.20) gives

$$R_a + R_b \leq \frac{1}{n}\log(|\mathcal{W}_a||\mathcal{W}_b|)$$

$$
\begin{aligned}
&\le \left(\frac{1}{1-\epsilon}\frac{1}{n}\sum_{t=1}^{n} I(T_t^a, T_t^b; Y_t|S_{[t]})\right) + \chi(\epsilon) + (n-1)\chi(\epsilon) \\
&\overset{(a)}{\le} \frac{1}{1-\epsilon}\frac{1}{n}\sum_{t=1}^{n} I(T_t^a, T_t^b; Y_t|S_{[t]}) + \eta(\epsilon) - \frac{\epsilon}{1-\epsilon}\frac{1}{n}\sum_{t=1}^{n} I(T_t^a, T_t^b; Y_t|S_{[t]}) \\
&= \frac{1}{n}\sum_{t=1}^{n} I(T_t^a, T_t^b; Y_t|S_{[t]}) + \eta(\epsilon) \qquad (4.22)
\end{aligned}
$$

where (a) is valid since $I(T_t^a, T_t^b; Y_t|S_{[t]}) \le \log|\mathcal{Y}|$. Furthermore,

$$
I(T_t^a, T_t^b; Y_t|S_{[t]}) = n \sum_{\mu\in\mathcal{S}^{t-1}} \alpha_\mu I(T_t^a, T_t^b; Y_t|S_t, S_{[t-1]} = \mu), \qquad (4.23)
$$

and substituting the above into (4.22) yields (4.16). □

Note that, for any $t \ge 1$, $I(T_t^a, T_t^b; Y_t|S_t, S_{[t-1]} = \mu)$ is a function of the joint conditional distribution of channel state S_t, inputs T_t^a, T_t^b and output Y_t given the past realization $(S_{[t-1]} = \mu)$. Hence, to complete the proof of the outer bound, we need to show that $P_{T_t^a,T_t^b,Y_t,S_t|S_{[t-1]}}(t^a, t^b, y, s|\mu)$ factorizes as in (4.8). This is done in the lemma below. In particular, it is crucial to observe that the knowledge of the past state at the decoder, $S_{[t-1]}$, is enough to provide a product form on T^a and T^b. Let

$$
\begin{aligned}
\Upsilon_{\mu_{\mathbf{a}}}^a(t^a) &:= \{w_a : \phi_t^{(a)}(w_a, s_{[t-1]}^a = \mu_{\mathbf{a}}) = t^a\} \\
\Upsilon_{\mu_{\mathbf{b}}}^b(t^b) &:= \{w_b : \phi_t^{(b)}(w_b, s_{[t-1]}^b = \mu_{\mathbf{b}}) = t^b\} \qquad (4.24) \\
\pi_{T^a}^{\mu_{\mathbf{a}}}(t^a) &:= \sum_{w_a\in\Upsilon_{\mu_{\mathbf{a}}}^a(t^a)} \frac{1}{|\mathcal{W}_a|} \\
\pi_{T^b}^{\mu_{\mathbf{b}}}(t^b) &:= \sum_{w_b\in\Upsilon_{\mu_{\mathbf{b}}}^b(t^b)} \frac{1}{|\mathcal{W}_b|} \\
\pi_{T^a}^{\mu}(t^a) &:= \sum_{\mu_{\mathbf{a}}} \pi_{T^a}^{\mu_{\mathbf{a}}}(t^a) P_{S_{[t-1]}^a|S_{[t-1]}}(\mu_{\mathbf{a}}|\mu), \\
\pi_{T^b}^{\mu}(t^b) &:= \sum_{\mu_{\mathbf{b}}} \pi_{T^b}^{\mu_{\mathbf{b}}}(t^b) P_{S_{[t-1]}^b|S_{[t-1]}}(\mu_{\mathbf{b}}|\mu), \qquad (4.25)
\end{aligned}
$$

where $\mu_{\mathbf{a}}$ and $\mu_{\mathbf{b}}$ denote particular realizations of $S^a_{[t-1]}$ and $S^b_{[t-1]}$, respectively.

Lemma 4.2.2. *For every $1 \le t \le n$ and $\mu \in \mathcal{S}^{t-1}$, the following holds*

$$P_{T^a_t,T^b_t,Y_t,S_t|S_{[t-1]}}(t^a,t^b,y,s|\mu) \;=\; P_S(s)P_{Y|S,T^a,T^b}(y|s,t^a,t^b)\pi^{\mu}_{T^a}(t^a)\pi^{\mu}_{T^b}(t^b). \quad (4.26)$$

Proof. Let $\mathbf{S} := (S_t, S^a_t, S^b_t)$ and $\mathbf{s} := (s, s^a_t, s^b_t)$. Observe that

$$\begin{aligned}
&P_{T^a_t,T^b_t,Y_t,S_t|S_{[t-1]}}(t^a,t^b,y,s|\mu)\\
&\quad= \sum_{s^a_t\in\mathcal{S}^a}\sum_{s^b_t\in\mathcal{S}^b} P_{\mathbf{S},T^a_t,T^b_t,Y_t|S_{[t-1]}}(\mathbf{s},t^a,t^b,y|\mu)\\
&\quad= \sum_{s^a_t\in\mathcal{S}^a}\sum_{s^b_t\in\mathcal{S}^b} P_{Y|\mathbf{S},T^a_t,T^b_t}(y|\mathbf{s},t^a,t^b)P_{\mathbf{S},T^a_t,T^b_t|S_{[t-1]}}(\mathbf{s},t^a,t^b|\mu) \qquad (4.27)
\end{aligned}$$

where (4.27) is shown in (4.21). Let us now consider the term $P_{\mathbf{S},T^a_t,T^b_t|S_{[t-1]}}(\mathbf{s},t^a,t^b|\mu)$ above. We have the following

$$\begin{aligned}
&P_{\mathbf{S},T^a_t,T^b_t|S_{[t-1]}}(\mathbf{s},t^a,t^b|\mu)\\
&= \sum_{w_a\in\mathcal{W}_a}\sum_{w_b\in\mathcal{W}_b}\sum_{\mu_{\mathbf{a}}}\sum_{\mu_{\mathbf{b}}} P_{\mathbf{W},S^a_{[t-1]},S^b_{[t-1]},\mathbf{S},T^a_t,T^b_t|S_{[t-1]}}(\mathbf{w},\mu_{\mathbf{a}},\mu_{\mathbf{b}},\mathbf{s},t^a,t^b|\mu)\\
&\overset{(i)}{=} P_{\mathbf{S}}(\mathbf{s})\sum_{w_a\in\mathcal{W}_a}\sum_{w_b\in\mathcal{W}_b}\sum_{\mu_{\mathbf{a}}}\sum_{\mu_{\mathbf{b}}} P_{\mathbf{W},S^a_{[t-1]},S^b_{[t-1]},T^a_t,T^b_t|S_{[t-1]}}(\mathbf{w},\mu_{\mathbf{a}},\mu_{\mathbf{b}},t^a,t^b|\mu)\\
&\overset{(ii)}{=} P_{\mathbf{S}}(\mathbf{s})\sum_{w_a\in\mathcal{W}_a}\sum_{w_b\in\mathcal{W}_b}\sum_{\mu_{\mathbf{a}}}\sum_{\mu_{\mathbf{b}}} 1_{\{t^l=\phi^{(l)}_t(w_l,\mu_{\mathbf{l}}),\;\; l=a,b\}} P_{\mathbf{W},S^a_{[t-1]},S^b_{[t-1]}|S_{[t-1]}}(\mathbf{w},\mu_{\mathbf{a}},\mu_{\mathbf{b}}|\mu)\\
&\overset{(iii)}{=} P_{\mathbf{S}}(\mathbf{s})\sum_{w_a\in\mathcal{W}_a}\sum_{w_b\in\mathcal{W}_b}\sum_{\mu_{\mathbf{a}}}\sum_{\mu_{\mathbf{b}}} 1_{\{t^l=\phi^{(l)}_t(w_l,\mu_{\mathbf{l}}),\;\; l=a,b\}}\\
&\qquad\qquad \frac{1}{|\mathcal{W}_a|}\frac{1}{|\mathcal{W}_b|}P_{S^a_{[t-1]},S^b_{[t-1]}|S_{[t-1]}}(\mu_{\mathbf{a}},\mu_{\mathbf{b}}|\mu)\\
&\overset{(iv)}{=} P_{\mathbf{S}}(\mathbf{s})\sum_{\mu_{\mathbf{a}}} P_{S^a_{[t-1]}|S_{[t-1]}}(\mu_{\mathbf{a}}|\mu)\sum_{\mu_{\mathbf{b}}} P_{S^b_{[t-1]}|S_{[t-1]}}(\mu_{\mathbf{b}}|\mu)\\
&\qquad\qquad \sum_{w_a\in\mathcal{W}_a}\frac{1}{|\mathcal{W}_a|}1_{\{t^a=\phi^{(a)}_t(w_a,\mu_{\mathbf{a}})\}}\sum_{w_b\in\mathcal{W}_b}\frac{1}{|\mathcal{W}_b|}1_{\{t^b=\phi^{(b)}_t(w_b,\mu_{\mathbf{b}})\}}\\
&\overset{(v)}{=} P_{\mathbf{S}}(\mathbf{s})\sum_{\mu_{\mathbf{a}}} P_{S^a_{[t-1]}|S_{[t-1]}}(\mu_{\mathbf{a}}|\mu)\sum_{w_a\in\Upsilon^a_{\mu_{\mathbf{a}}}(t^a)}\frac{1}{|\mathcal{W}_a|}\sum_{\mu_{\mathbf{b}}} P_{S^b_{[t-1]}|S_{[t-1]}}(\mu_{\mathbf{b}}|\mu)\sum_{w_b\in\Upsilon^b_{\mu_{\mathbf{b}}}(t^b)}\frac{1}{|\mathcal{W}_b|}
\end{aligned}$$

$$\overset{(vi)}{=} \quad P_{\mathbf{S}}(\mathbf{s}) \sum_{\mu_{\mathbf{a}}} P_{S^a_{[t-1]}|S_{[t-1]}}(\mu_{\mathbf{a}}|\mu) \pi^{\mu_{\mathbf{a}}}_{T^a}(t^a) \sum_{\mu_{\mathbf{b}}} P_{S^b_{[t-1]}|S_{[t-1]}}(\mu_{\mathbf{b}}|\mu) \pi^{\mu_{\mathbf{b}}}_{T^b}(t^b)$$

$$\overset{(vii)}{=} \quad P_{\mathbf{S}}(\mathbf{s}) \pi^{\mu}_{T^a}(t^a) \pi^{\mu}_{T^b}(t^b) \tag{4.28}$$

where (i) is due to (4.2) and (4.15), (ii) is valid by (4.15), (iii) is due to (4.2), (iv) is valid by (4.1) and (4.15), (v) is valid due to (4.24) and $(vi) - (vii)$ is valid due to (4.25). Substituting (4.28) into (4.27) proves the lemma. □

We can now complete the proof of Theorem 4.2.2. We have

$$\begin{aligned} R_a + R_b &\leq \sum_{\mu \in \mathcal{S}^{(n)}} \alpha_\mu I(T^a_t, T^b_t; Y_t | S_t, S_{[t-1]} = \mu) + \eta(\epsilon) \\ &= \sum_{\mu \in \mathcal{S}^{(n)}} \alpha_\mu I(T^a_t, T^b_t; Y_t | S_t)_{\pi^{\mu}_{T^a}(t^a) \pi^{\mu}_{T^b}(t^b)} + \eta(\epsilon) \\ &\leq \sup_{\left(\pi_{T^a}(t^a) \pi_{T^b}(t^b) \in \Pi\right)} I(T^a_t, T^b_t; Y_t | S_t) + \eta(\epsilon), \end{aligned}$$

where $I(T^a_t, T^b_t; Y_t | S_t)_{\pi^{\mu}_{T^a}(t^a) \pi^{\mu}_{T^b}(t^b)}$ denotes the mutual information induced by the product distribution $\pi^{\mu}_{T^a}(t^a) \pi^{\mu}_{T^b}(t^b)$ and this step is valid since $I(T^a_t, T^b_t; Y_t | S_t, S_{[t-1]} = \mu)$ is a function of the joint conditional distribution of channel state S_t, inputs T^a_t, T^b_t and output Y_t given the past realization $(S_{[t-1]} = \mu)$. Hence, since $\lim_{\epsilon \to 0} \eta(\epsilon) = 0$, any achievable pair satisfies $R_a + R_b \leq \sup_{\pi_{T^a}(t^a) \pi_{T^b}(t^b)} I(T^a, T^b; Y | S)$. □

As a consequence of Theorems 4.2.1 and 4.2.2, we have the following corollary which can be thought of as an extension of [Jaf06, Theorem 4] to the case where the encoders have correlated CSI.

Corollary 4.2.1.

$$\mathcal{C}^{FS}_{\sum} = \sup_{\pi_{T^a}(t^a) \pi_{T^b}(t^b)} I(T^a, T^b; Y | S). \tag{4.29}$$

Proof of Corollary 4.2.1. We need to show that $\exists \ (R_a, R_b) \in \mathcal{C}_{IN}$ achieving (4.29).

We follows steps akin to [CT06, p.535] where discrete memoryless MACs are considered. Let us fix $\pi_{T^a}(t^a)\pi_{T^b}(t^b)$ and consider the rate constraints given in $\mathcal{C}_{IN}$

$$I(T^a;Y|T^b,S) = H(T^a|T^b,S) - H(T^a|T^b,Y,S) = H(T^a) - H(T^a|T^b,Y,S) \quad (4.30)$$

$$I(T^b;Y|T^a,S) = H(T^b|T^a,S) - H(T^b|T^a,Y,S) = H(T^b) - H(T^b|T^a,Y,S) \quad (4.31)$$

$$\begin{aligned} I(T^a,T^b;Y|S) &= H(T^a,T^b) - H(T^a,T^b|Y,S) \\ &= H(T^a) + H(T^b) - H(T^a|T^b,Y,S) - H(T^b|Y,S), \end{aligned} \quad (4.32)$$

where (4.30), (4.31) and (4.32) are valid since T^a and T^b are independent of each other and independent of S. Observe now that for any $\pi_{T^a}(t^a)\pi_{T^b}(t^b)$, $I(T^a;Y|T^b,S) + I(T^b;Y|T^a,S) \geq I(T^a,T^b;Y|S)$ since $H(T^b|Y,S) \geq H(T^b|T^a,Y,S)$. Therefore, the sum-rate constraint in $\mathcal{C}_{IN}$ is always active and hence, there exists $(R_a,R_b) \in \mathcal{C}_{IN}$ achieving (4.29). □

We conclude this section with a number of remarks.

Remark 4.2.1. *One essential step in the proof of Theorem 4.2.2 is that, once we have the complete CSI, conditioning on which allows a product form on T^a and T^b, there is no loss of optimality (for the sum-rate capacity) in using associated memoryless team policies instead of using all the past information at the receiver.*

Remark 4.2.2. *For the validity of Corollary 4.2.1, it is crucial to have the product form on (T^a,T^b). If this is not the case, we would get that*

$$\begin{aligned} & I(T^a;Y|T^b,S) + I(T^b;Y|T^a,S) \\ &\qquad = H(T^a|T^b) + H(T^b|T^a) - H(T^a|T^b,Y,S) - H(T^b|T^a,Y,S) \\ I(T^a,T^b;Y|S) &= H(T^a|T^b) + H(T^b) - H(T^a|T^b,Y,S) - H(T^b|Y,S). \end{aligned}$$

Therefore, it is possible to get an obsolete sum-rate constraint in $\mathcal{C}_{IN}$ and hence, *achievability of* $\mathcal{C}_{FS}^{\sum}$ *is not guaranteed. It should be noted that the channel inputs are not independent since* $X^a = T^a(S^a)$ *and* $X^b = T^b(S^b)$.

4.2.2 Partial Asymmetric CSITs: Non-Causal Case

In this section, we consider the situation where the transmitters have access to partial state information available at the decoder. In particular, let $S_t^i = f^i(S_t^r)$, where $f^i : \mathcal{S}_r \to \mathcal{S}_i$, $i = \{a, b\}$ and $S^r \in \mathcal{S}_r$ such that

$$P_{S_{[n]},S_{[n]}^a,S_{[n]}^b,S_{[n]}^r,\mathbf{W}}(s_{[n]}, s_{[n]}^a, s_{[n]}^b, s_{[n]}^r, \mathbf{w}) = \prod_{t=1}^{n} \frac{1}{|\mathcal{W}_a|}\frac{1}{|\mathcal{W}_b|} P_{S_t,S_t^a,S_t^b,S_t^r}(s_t, s_t^a, s_t^b, s_t^r). \quad (4.33)$$

Let $\mathbf{S} := (S_t, S_t^a, S_t^b, S_t^r)$ and $\mathbf{s} := (s, s_t^a, s_t^b, s_t^r)$. The channel is driven by the state process $\{S_t\}_{t=1}^{\infty}$ and hence,

$$P_{Y_{[n]}|\mathbf{W},\mathbf{X}_{[n]},\mathbf{S}_{[n]}}(y_{[n]}|\mathbf{w}, \mathbf{x}_{[n]}, \mathbf{s}_{[n]}) = \prod_{t=1}^{n} P_{Y_t|X_t^a,X_t^b,S_t}(y_t|x_t^a, x_t^b, s_t). \quad (4.34)$$

Note that one can define an equivalent channel with conditional output probability

$$P_{Y|X^a,X^b,S^r}^{eq}(y|x^a, x^b, s^r) = \sum_{s\in\mathcal{S}} P_{Y|X^a,X^b,S}(y|x^a, x^b, s) P_{S|S^r}(s|s^r). \quad (4.35)$$

Hence, the causal setup of this problem is no more general than the setup in [CY11] and the main contribution of this subsection is to show that the result of [CY11] also holds for non-causal coding.

We keep the channel codes definition identical for the causal and non-causal cases, except for the non-causal case we have; $\phi_t^{(i)} : \mathcal{S}_i^n \times \mathcal{W}_i \to \mathcal{X}_i^n$, $i = \{a, b\}$, $t = 1, \cdots, n$.

Let $\mathcal{C}^{nc}$ denote the capacity region. We need to modify Definition 4.2.2 in order to take the current CSI into account.

Definition 4.2.3. *A memoryless stationary (in time) team policy is a family*

$$\bar{\Pi} = \left\{ \bar{\pi} = \left(\pi_{X^a|S^a}(\cdot|f^a(s^r)), \pi_{X^b|S^b}(\cdot|f^b(s^r)) \right) \in \mathcal{P}(\mathcal{X}_a) \times \mathcal{P}(\mathcal{X}_b) \right\}. \quad (4.36)$$

For every $\bar{\pi}$ defined in (4.36), $\mathcal{R}^{nc}(\bar{\pi})$ denotes the region of all rate pairs $R = (R_a, R_b)$ satisfying

$$R_a < I(X^a; Y|X^b, S^r) \quad (4.37)$$

$$R_b < I(X^b; Y|X^a, S^r) \quad (4.38)$$

$$R_a + R_b < I(X^a, X^b; Y|S^r) \quad (4.39)$$

where S^r, X^a, X^b and Y are random variables taking values in $\mathcal{S}_r$, $\mathcal{X}_a$, $\mathcal{X}_b$ and $\mathcal{Y}$, respectively, and whose joint probability distribution factorizes as

$$P_{S^r,X^a,X^b,Y}(s^r, x^a, x^b, y) = P_{S^r}(s^r)P_{Y|X^a,X^b,S^r}(y|x^a, x^b, s^r)\pi_{X^a|S^a}(x^a|f^a(s^r))\pi_{X^b|S^b}(x^b|f^b(s^r)). \quad (4.40)$$

Let $\overline{co}\left(\bigcup_{\bar{\pi}} \mathcal{R}^{nc}(\bar{\pi})\right)$ denote the closure of the convex hull of the rate regions $\mathcal{R}^{nc}(\bar{\pi})$ given by (4.37)-(4.39) associated to all possible memoryless stationary team polices as defined in (4.36).

Theorem 4.2.3. $\mathcal{C}^{nc} = \overline{co}\left(\bigcup_{\bar{\pi}} \mathcal{R}^{nc}(\bar{\pi})\right)$.

For the achievability proof, see [CY11, Section III] and observe that any rate which is achievable with causal CSI is also achievable with non-causal CSI. The proof for the non-causal case is realized by observing that there is no loss of optimality if not only the past, as shown in [CY11], but also the future CSI is ignored given that the receiver is provided with complete CSI. A similar observation for independent CSIT is also made see [Jaf06, Theorem 5].

Converse Proof of Theorem 4.2.3. Let

$$\alpha_{\mu_{\mathbf{p}},\mathbf{f}} := \frac{1}{n} P_{S^r_{[1,t-1]},S^r_{[t+1,n]}}(\mu_{\mathbf{p}}, \mu_{\mathbf{f}}). \quad (4.41)$$

Observe that $(\mu_{\mathbf{p}} : \mu_{\mathbf{f}}) \in \mathcal{S}_r^{n-1}$, where $(v : w)$ denotes the concatenation of two vectors v and w, and

$$\sum_{(\mu_{\mathbf{p}}:\mu_{\mathbf{f}})} \alpha_{\mu_{\mathbf{p}},\mathbf{f}} := \frac{1}{n} \sum_{1 \le t \le n} \sum_{\mu_{\mathbf{p}},\mu_{\mathbf{f}}} P_{S^r_{[1,t-1]},S^r_{[t+1,n]}}(\mu_{\mathbf{p}}, \mu_{\mathbf{f}}) = 1.$$

Lemma 4.2.3. *Assume that a rate pair $R = (R_a, R_b)$, with block length $n \ge 1$ and a constant $\epsilon \in (0, 1/2)$, is achievable. Then,*

$$R_a \le \sum_{(\mu_{\mathbf{p}}:\mu_{\mathbf{f}})} \alpha_{\mu_{\mathbf{p}},\mathbf{f}} I(X_t^a; Y_t | X_t^b, S_t^r, S^r_{[t-1]} = \mu_{\mathbf{p}}, S^r_{[t+1,n]} = \mu_{\mathbf{f}}) + \eta(\epsilon) \quad (4.42)$$

$$R_b \le \sum_{(\mu_{\mathbf{p}}:\mu_{\mathbf{f}})} \alpha_{\mu_{\mathbf{p}},\mathbf{f}} I(X_t^b; Y_t | X_t^a, S_t^r, S^r_{[t-1]} = \mu_{\mathbf{p}}, S^r_{[t+1,n]} = \mu_{\mathbf{f}}) + \eta(\epsilon) \quad (4.43)$$

$$R_a + R_b \le \sum_{(\mu_{\mathbf{p}}:\mu_{\mathbf{f}})} \alpha_{\mu_{\mathbf{p}},\mathbf{f}} I(X_t^a, X_t^b; Y_t | S_t^r, S^r_{[t-1]} = \mu_{\mathbf{p}}, S^r_{[t+1,n]} = \mu_{\mathbf{f}}) + \eta(\epsilon) \quad (4.44)$$

Proof. Let us first consider the sum-rate. With standard steps, we get

$$R_a + R_b \le \frac{1}{1-\epsilon} \frac{1}{n} \left(I(\mathbf{W}; Y_{[n]}, S^r_{[n]}) + H(\epsilon) \right). \quad (4.45)$$

Note that since $S^r_{[n]}$ is independent of $\mathbf{W}$, we have $I(\mathbf{W}; Y_{[n]}, S^r_{[n]}) = I(\mathbf{W}; Y_{[n]} | S^r_{[n]})$ and

$$\begin{aligned}
I(\mathbf{W}; Y_{[n]} | S^r_{[n]}) &= \sum_{t=1}^{n} \left[H(Y_t | S^r_{[n]}, Y_{[t-1]}) - H(Y_t | \mathbf{W}, S^r_{[n]}, Y_{[t-1]}) \right] \\
&\overset{(i)}{\le} \sum_{t=1}^{n} \left[H(Y_t | S^r_{[n]}) - H(Y_t | \mathbf{W}, S^r_{[n]}, Y_{[t-1]}) \right] \\
&\overset{(ii)}{=} \sum_{t=1}^{n} \left[H(Y_t | S^r_{[n]}) - H(Y_t | \mathbf{W}, S^r_{[n]}, Y_{[t-1]}, \mathbf{X}_{[n]}) \right] \\
&\overset{(iii)}{=} \sum_{t=1}^{n} \left[H(Y_t | S^r_{[n]}) - H(Y_t | S^r_{[n]}, \mathbf{X}_t) \right] \\
&= \sum_{t=1}^{n} I(\mathbf{X}_t; Y_t | S^r_{[n]}) \qquad (4.46)
\end{aligned}$$

where (i) follows since conditioning does not increase entropy, (ii) holds since $X_t^i = \phi_t^{(i)}(W_i, f^i(S^r_{[n]}))$, $i = \{a, b\}$, and (iii) is due to (4.3). Combining (4.45) and (4.46) similar to (4.22), gives

$$R_a + R_b \leq \frac{1}{n}\sum_{t=1}^{n} I(X_t^a, X_t^b; Y_t | S_{[n]}^r) + \eta(\epsilon) \text{ and} \tag{4.47}$$

$$I(X_t^a, X_t^b; Y_t | S_{[n]}^r) = n \sum_{\mu_{\mathbf{p}},\mu_{\mathbf{f}}} \alpha_{\mu_{\mathbf{p},\mathbf{f}}} I(X_t^a, X_t^b; Y_t | S_t^r, S_{[t-1]}^r = \mu_{\mathbf{p}}, S_{[t+1,n]}^r = \mu_{\mathbf{f}}) \tag{4.48}$$

and substituting the above into (4.47) yields (4.44).

Let us now consider encoder a. Using Fano's inequality and standard steps we first get,

$$R_a \leq \frac{1}{1-\epsilon}\frac{1}{n}\left(I(W_a; Y_{[n]}, S_{[n]}^r) + H(\epsilon)\right). \tag{4.49}$$

Furthermore,

$$\begin{aligned}
I(W_a; Y_{[n]}, S_{[n]}^r) &\overset{(i)}{\leq} I(W_a; Y_{[n]} | S_{[n]}^r, W_b) \\
&= \sum_{t=1}^{n} \left[H(Y_t | S_{[n]}^r, Y_{[t-1]}, W_b) - H(Y_t | S_{[n]}^r, Y_{[t-1]}, \mathbf{W})\right] \\
&\overset{(ii)}{\leq} \sum_{t=1}^{n} \left[H(Y_t | S_{[n]}^r, W_b) - H(Y_t | S_{[n]}^r, Y_{[t-1]}, \mathbf{W})\right] \\
&\overset{(iii)}{=} \sum_{t=1}^{n} \left[H(Y_t | S_{[n]}^r, W_b, X_{[n]}^b) - H(Y_t | S_{[n]}^r, Y_{[t-1]}, \mathbf{W}, \mathbf{X}_{[n]})\right] \\
&\overset{(iv)}{\leq} \sum_{t=1}^{n} \left[H(Y_t | S_{[n]}^r, X_t^b) - H(Y_t | S_{[n]}^r, Y_{[t-1]}, \mathbf{W}, \mathbf{X}_{[n]})\right] \\
&\overset{(v)}{=} \sum_{t=1}^{n} \left[H(Y_t | S_{[n]}^r, X_t^b) - H(Y_t | S_{[n]}^r, X_t^b, X_t^a)\right] \\
&= \sum_{t=1}^{n} I(X_t^a; Y_t | X_t^b, S_{[n]}^r)
\end{aligned} \tag{4.50}$$

where (i) is due to (4.2) and conditioning does not increase entropy, (ii) holds since conditioning does not increase entropy, (iii) holds since $X_t^i = \phi^{(i)}(W_i, f^i(S_{[n]}^r))$, $i = \{a, b\}$, (iv) is valid since conditioning does not increase entropy and finally, (v) is valid due to (4.3) and S_t^i, $i = \{a, b\}$, being a function of S_t^r.

Now combining (4.49)-(4.50) and following steps akin to (4.47) and (4.48), we can verify (4.42). To verify (4.43) for encoder b it is enough to switch the roles of encoder a and (b). □

Observe now that for any $t \geq 1$, $I(X_t^a, X_t^b; Y_t | S_t^r, S_{[t-1]}^r = \mu_{\mathbf{p}}, S_{[t+1,n]}^r = \mu_{\mathbf{f}})$ is a function of $P_{X_t^a, X_t^b, Y_t, S_t^r | S_{[t-1]}^r, S_{[t+1,n]}^r}(x_t^a, x_t^b, y_t, s_t^r | \mu_{\mathbf{p}}, \mu_{\mathbf{f}})$. Hence, we need to show that this distribution factorizes as in (4.40). Let

$$\begin{aligned} \Upsilon^a_{\mu_{\mathbf{p}},\mu_{\mathbf{f}}}(x^a, f^a(s^r)) &:= \{w_a : \phi_t^{(a)}(w_a, f^a(\mu_{\mathbf{p}}, \mu_{\mathbf{f}}), f^a(s^r)) = x^a\} \\ \Upsilon^b_{\mu_{\mathbf{p}},\mu_{\mathbf{f}}}(x^b, f^b(s^r)) &:= \{w_b : \phi_t^{(b)}\left(w_b, f^b(\mu_{\mathbf{p}}, \mu_{\mathbf{f}}), f^b(s^r)\right) = x^b\} \qquad (4.51) \\ \pi^{\mu_{\mathbf{p}},\mu_{\mathbf{f}}}_{X^a|S^a}(x^a | f^a(s^r)) &:= \sum_{w_a \in \Upsilon^a_{\mu_{\mathbf{p}},\mu_{\mathbf{f}}}(x^a, f^a(s^r))} \frac{1}{|\mathcal{W}_a|}, \\ \pi^{\mu_{\mathbf{p}},\mu_{\mathbf{f}}}_{X^b|S^b}\left(x^b | f^b(s^r)\right) &:= \sum_{w_b \in \Upsilon^b_{\mu_{\mathbf{p}},\mu_{\mathbf{f}}}(x^b, f^b(s^r))} \frac{1}{|\mathcal{W}_b|}. \qquad (4.52) \end{aligned}$$

Lemma 4.2.4. *For every $1 \leq t \leq n$ and $(\mu_{\mathbf{p}} : \mu_{\mathbf{f}}) \in \mathcal{S}_r^{n-1}$, the following holds*

$$\begin{aligned} &P_{X_t^a, X_t^b, Y_t, S_t^r | S_{[t-1]}^r, S_{[t+1,n]}^r}(x^a, x^b, y, s^r | \mu_{\mathbf{p}}, \mu_{\mathbf{f}}) \\ &\quad = P_{S^r}(s^r) P_{Y|S^r, X^a, X^b}(y | s^r, x^a, x^b) \pi^{\mu_{\mathbf{p}},\mu_{\mathbf{f}}}_{X^a|S^a}(x^a | f^a(s^r)) \pi^{\mu_{\mathbf{p}},\mu_{\mathbf{f}}}_{X^b|S^b}(x^b | f^b(s^r)). (4.53) \end{aligned}$$

Proof. First observe that due to (4.3) we have

$$\begin{aligned} &P_{X_t^a, X_t^b, Y_t, S_t^r | S_{[t-1]}^r, S_{[t+1,n]}^r}(x^a, x^b, y, s^r | \mu_{\mathbf{p}}, \mu_{\mathbf{f}}) \\ &\quad = P_{Y_t | S_t^r, X_t^a, X_t^b}(y | s^r, x^a, x^b) P_{X_t^a, X_t^b, S_t^r | S_{[t-1]}^r, S_{[t+1,n]}^r}(x^a, x^b, s^r | \mu_{\mathbf{p}}, \mu_{\mathbf{f}}). \qquad (4.54) \end{aligned}$$

Let us now consider the second term in (4.54). We have

$$\begin{aligned} &P_{X_t^a, X_t^b, S_t^r | S_{[t-1]}^r, S_{[t+1,n]}^r}(x^a, x^b, s^r | \mu_{\mathbf{p}}, \mu_{\mathbf{f}}) \\ &= \sum_{w_a \in \mathcal{W}_a} \sum_{w_b \in \mathcal{W}_b} P_{\mathbf{W}, X_t^a, X_t^b, S_t^r | S_{[t-1]}^r, S_{[t+1,n]}^r}(\mathbf{w}, x^a, x^b, s^r | \mu_{\mathbf{p}}, \mu_{\mathbf{f}}) \\ &\overset{(i)}{=} \sum_{w_a \in \mathcal{W}_a} \sum_{w_b \in \mathcal{W}_b} 1_{\left\{x^l = \phi^{(l)}\left(w_l, f^l(s^r, \mu_{\mathbf{p}}, \mu_{\mathbf{f}})\right),\ l=a,b\right\}} P_{W_a, W_b, S_t^r | S_{[t-1]}^r, S_{[t+1,n]}^r}(w_a, w_b, s^r | \mu_{\mathbf{p}}, \mu_{\mathbf{f}}) \end{aligned}$$

$$\stackrel{(ii)}{=} \sum_{w_a \in \mathcal{W}_a} \sum_{w_b \in \mathcal{W}_b} 1_{\left\{x^l = \phi^{(l)}\left(w_l, f^l(s^r, \mu_{\mathbf{p}}, \mu_{\mathbf{f}})\right),\ l=a,b\right\}} \frac{1}{|\mathcal{W}_a|} \frac{1}{|\mathcal{W}_b|} P_{S_t^r}(s^r)$$
$$= P_{S_t^r}(s^r) \sum_{w_a \in \mathcal{W}_a} \frac{1}{|\mathcal{W}_a|} 1_{\left\{x^a = \phi^{(a)}(w_a, f^a(s^r, \mu_{\mathbf{p}}, \mu_{\mathbf{f}}))\right\}} \sum_{w_b \in \mathcal{W}_b} \frac{1}{|\mathcal{W}_b|} 1_{\left\{x^b = \phi^{(b)}\left(w_b, f^b(s^r, \mu_{\mathbf{p}}, \mu_{\mathbf{f}})\right)\right\}}$$
$$\stackrel{(iii)}{=} \frac{1}{|\mathcal{W}_a|} \sum_{w_a \in \Upsilon^a_{\mu_{\mathbf{p}}, \mu_{\mathbf{f}}}(x^a, f^a(s^r))} \frac{1}{|\mathcal{W}_b|} \sum_{w_b \in \Upsilon^b_{\mu_{\mathbf{p}}, \mu_{\mathbf{f}}}(x^b, f^b(s^r))}$$
$$\stackrel{(iv)}{=} P_{S_t^r}(s^r) \pi^{\mu_{\mathbf{p}}, \mu_{\mathbf{f}}}_{X^a|S^a}(x^a|f^a(s^r)) \pi^{\mu_{\mathbf{p}}, \mu_{\mathbf{f}}}_{X^b|S^b}(x^b|f^b(s^r)) \tag{4.55}$$

where (i) follows since $X_t^i = \phi^{(i)}(W_i, f^i(S^r_{[n]}))$, $i = \{a, b\}$, (ii) is valid since W_a and W_b are independent of $S^r_{[n]}$ and state process being i.i.d. and (iii) follows due to (4.51) and (iv) follows due to (4.52). Substituting (4.55) in (4.54) completes the proof. □

We can now complete the proof of Theorem 4.2.3. With Lemma 4.2.3, it is shown that any achievable rate pair can be approximated by the convex combinations of rate conditions given in (4.37)-(4.39) which are indexed by $(\mu_{\mathbf{p}}, \mu_{\mathbf{f}})$ and satisfy (4.40) for joint state-input-output distributions. Hence, since $\lim_{\epsilon \to 0} \eta(\epsilon) = 0$, any achievable rate pair belongs to $\overline{co}\left(\bigcup_{\bar{\pi}} \mathcal{R}^{nc}(\bar{\pi}) \right)$. □

Consider now the setup in Section 4.2 in order to observe that for the non-causal case the optimality of Shannon strategies is not guaranteed. Recall that, we have

$$I(\mathbf{W}; Y_{[n]}, S_{[n]}) \leq \sum_{t=1}^{n} \left[H(Y_t|S_{[n]}, Y_{[t-1]}) - H(Y_t|\mathbf{W}, S_{[n]}, Y_{[t-1]}, \mathbf{T}_t) \right] \tag{4.56}$$

where $\mathbf{T}_t := (T_t^a, T_t^b)$. Consider now the right hand side of (4.56) and observe that

$$P_{Y_t|\mathbf{W}, S_{[n]}, Y_{[t-1]}, T_t^a, T_t^b}(y_t|\mathbf{w}, s_{[n]}, y_{[t-1]}, t_t^a, t_t^b)$$
$$= \sum_{s_t^a, s_t^b} P_{Y_t|S_t, S_t^a, S_t^b, T_t^a, T_t^b}(y_t|s_t, s_t^a, s_t^b, t_t^a, t_t^b) P_{S_t^a, S_t^b|Y_{[t-1]}, S_t}(s_t^a, s_t^b|y_{[t-1]}, s_t),$$

and therefore, the past channel outputs cannot be eliminated.

4.2.3 Asymmetric Noisy CSIT with Delays

Consider the problem defined in Section 4.2.1 where the two encoders have accesses to asymmetrically delayed, where delays are $d_a \geq 1$ and $d_b \geq 1$, respectively, and noisy versions of the state information S_t at each time $t \geq 1$, modeled by $S^a_{t-d_a} \in \mathcal{S}_a$, $S^b_{t-d_b} \in \mathcal{S}_b$, respectively. The rest of the channel model is identical and hence, (4.1), (4.2) and (4.3) are valid throughout this section. We also assume that S_t is fully available at the receiver. A code can be defined as in Definition 4.2.1, except now

$$\phi^{(a)}_t : \mathcal{S}_a^{t-d_a} \times \mathcal{W}_a \to \mathcal{X}_a, \;\; t = 1, 2, ...n;$$

$$\phi^{(b)}_t : \mathcal{S}_b^{t-d_b} \times \mathcal{W}_b \to \mathcal{X}_b, \;\; t = 1, 2, ...n.$$ [1]

Let $\mathcal{C}^{ad}$ denotes the capacity region of the delayed setup.

In the main result of this section the team policies are composed of probability distributions on the channel inputs rather than Shannon strategies.

Definition 4.2.4. *A memoryless stationary (in time) team policy is a family*

$$\tilde{\Pi} = \left\{ \tilde{\pi} = (\pi_{X^a}(\cdot), \pi_{X^b}(\cdot)) \in \mathcal{P}(\mathcal{X}^a) \times \mathcal{P}(\mathcal{X}^b) \right\}. \tag{4.57}$$

For every memoryless stationary team policy $\tilde{\pi}$, $\mathcal{R}^{ad}(\tilde{\pi})$ denotes the region of all rate pairs $R = (R_a, R_b)$ satisfying

$$R_a < I(X^a; Y | X^b, S) \tag{4.58}$$

$$R_b < I(X^b; Y | X^a, S) \tag{4.59}$$

$$R_a + R_b < I(X^a, X^b; Y | S) \tag{4.60}$$

where S, X^a, X^b and Y are random variables taking values in $\mathcal{S}$, $\mathcal{X}^a$, $\mathcal{X}^b$ and $\mathcal{Y}$,

[1]Obviously, when $d_l \geq t$, $l = a, b$ then $X^a_t = \phi^{(a)}_t(W_a)$ and $X^b_t = \phi^{(b)}_t(W_b)$.

respectively and whose joint probability distribution factorizes as

$$P_{S,X^a,X^b,Y}(s,x^a,x^b,y) = P_S(s)P_{Y|X^a,X^b,S}(y|x^a,x^b,s)\pi_{X^a}(x^a)\pi_{X^b}(x^b). \tag{4.61}$$

Let $\overline{co}\Big(\bigcup_{\tilde{\pi}} \mathcal{R}^{ad}(\tilde{\pi})\Big)$ denotes the closure of the convex hull of the rate regions $\mathcal{R}^{ad}(\tilde{\pi})$ given by (4.58)-(4.60) associated to all possible memoryless stationary team polices as defined in (4.57).

Theorem 4.2.4. $\mathcal{C}^{ad} = \overline{co}\Big(\bigcup_{\tilde{\pi}} \mathcal{R}^{ad}(\tilde{\pi})\Big)$.

Achievability can be shown via random coding arguments.

Converse Proof of Theorem 4.2.4. In the proof, we will use the fact that the delayed setup can be modeled by taking the last d_a, d_b entries of causal setup as empty. Recall that α_μ is defined in (4.41).

Lemma 4.2.5. *Assume that a rate pair $R = (R_a, R_b)$, with block length $n \geq 1$ and a constant $\epsilon \in (0, 1/2)$, is achievable. Then,*

$$\begin{aligned}
R_a &\leq \sum_{\mu \in \mathcal{S}^{(n)}} \alpha_\mu I(X_t^a; Y_t | X_t^b, S_t, S_{[t-1]} = \mu) + \eta(\epsilon) && (4.62)\\
R_b &\leq \sum_{\mu \in \mathcal{S}^{(n)}} \alpha_\mu I(X_t^b; Y_t | X_t^a, S_t, S_{[t-1]} = \mu) + \eta(\epsilon) && (4.63)\\
R_a + R_b &\leq \sum_{\mu \in \mathcal{S}^{(n)}} \alpha_\mu I(X_t^a, X_t^b; Y_t | S_t, S_{[t-1]} = \mu) + \eta(\epsilon). && (4.64)
\end{aligned}$$

Proof. For the sum-rate, observe that the derivation in (4.20) can be performed to verify (4.64), as for $d_i \geq 1$, $T_t^i = X_t^i$ by taking $S^i_{[t-d_i+1,t-1]} = \emptyset$, $i = \{a, b\}$.

Let us now consider encoder a. We have

$$R_a \leq \frac{1}{n}\log(|\mathcal{W}_a|) \leq \frac{1}{1-\epsilon}\frac{1}{n}\left(I(W_a; Y_{[n]}, S_{[n]}) + H(\epsilon)\right). \tag{4.65}$$

Furthermore,

$$I(W_a; Y_{[n]}, S_{[n]}) \overset{(i)}{\leq} I(W_a; Y_{[n]}, S_{[n]} | W_b, S^b_{[n]})$$

$$
\begin{aligned}
&= \sum_{t=1}^{n} \big[H(Y_t, S_t | S_{[t-1]}, Y_{[t-1]}, W_b, S^b_{[n]}) - \\
&\qquad\qquad H(Y_t, S_t | S_{[t-1]}, Y_{[t-1]}, \mathbf{W}, S^b_{[n]}) \big] \\
&\overset{(ii)}{=} \sum_{t=1}^{n} \big[H(Y_t | S_{[t]}, Y_{[t-1]}, W_b, S^b_{[n]}) - H(Y_t | S_{[t]}, Y_{[t-1]}, \mathbf{W}, S^b_{[n]}) \big] \\
&\overset{(iii)}{=} \sum_{t=1}^{n} \big[H(Y_t | S_{[t]}, Y_{[t-1]}, W_b, S^b_{[n]}, X^b_{[n]}) - \\
&\qquad\qquad H(Y_t | S_{[t]}, Y_{[t-1]}, \mathbf{W}, S^b_{[n]}, X^b_{[n]}) \big] \\
&\overset{(iv)}{\leq} \sum_{t=1}^{n} \big[H(Y_t | S_{[t]}, X^b_t) - H(Y_t | S_{[t]}, Y_{[t-1]}, \mathbf{W}, S^b_{[n]}, X^b_{[n]}, X^a_{[n]}) \big] \\
&\overset{(v)}{=} \sum_{t=1}^{n} \big[H(Y_t | S_{[t]}, X^b_t) - H(Y_t | S_{[t]}, X^b_t, X^a_t) \big] \\
&= \sum_{t=1}^{n} I(X^a_t; Y_t | X^b_t, S_{[t]}) \qquad (4.66)
\end{aligned}
$$

where (i) is due to (4.2) and conditioning reduces entropy, (ii) is valid since

$$
\begin{aligned}
P_{S_t|S^b_t}(s_t|s^b_t) &= P_{S_t|Y_{[t-1]},S_{[t-1]},W_a,W_b,S^b_{[n]}}(s_t|y_{[t-1]}, s_{[t-1]}, w_a, w_b, s^b_{[n]}) \\
&= P_{S_t|Y_{[t-1]},S_{[t-1]},W_b,S^b_{[n]}}(s_t|y_{[t-1]}, s_{[t-1]}, w_b, s^b_{[n]}) \qquad (4.67)
\end{aligned}
$$

where the second equality is due to (4.2), (iii) is valid since $X^b_t = \phi^{(b)}_t\left(W_b, S^b_{[t-d_b]}\right)$, (iv) is valid since conditioning reduces entropy and finally, (v) is valid by (4.3).

Now, recall that $\chi(\epsilon) = \frac{H(\epsilon)}{n(1-\epsilon)}$ and, combining (4.65) and (4.66) gives

$$
R_a \leq \frac{1}{n} \sum_{t=1}^{n} I(X^a_t; Y_t | X^b_t, S_{[t]}) + \eta(\epsilon). \qquad (4.68)
$$

Furthermore,

$$
I(X^a_t; Y_t | X^b_t, S_{[t]}) = n \sum_{\mu \in \mathcal{S}^{t-1}} \alpha_\mu I(X^a_t; Y_t | X^b_t, S_t, S_{[t-1]} = \mu), \qquad (4.69)
$$

and substituting the above into (4.68) yields (4.62).

Finally, for encoder b, (4.63) can be verified by following the similar steps of encoder a. $\square$

Now since, for any $t \geq 1$, conditional mutual information terms given in (4.62)-(4.64) are functions of $P_{X_t^a,X_t^b,Y_t,S_t|S_{[t-1]}}(x^a,x^b,y,s|\mu)$, in order to complete the proof of the converse, we need to show that this term factorizes as in (4.61).

Lemma 4.2.6. *For every $1 \leq t \leq n$ and $\mu \in \mathcal{S}^{t-1}$, the following holds*

$$P_{X_t^a,X_t^b,Y_t,S_t|S_{[t-1]}}(x^a,x^b,y,s|\mu) = P_S(s)P_{Y|S,X^a,X^b}(y|s,x^a,x^b)\pi^{\mu}_{X^a}(x^a)\pi^{\mu}_{X^b}(x^b). \quad (4.70)$$

Note that one of the crucial step in verifying the product form for the causal setup, see (4.18) and (4.19), is the independence of Shannon strategies of the current state. This also holds in the delayed setup. Therefore, let

$$\Upsilon^i_{\mu_\mathbf{i}}(x^i) := \{w_i : \phi_t^{(i)}(w_i, s^i_{[t-d_i]} = \mu_\mathbf{i}) = x^i\}, \;\; i = a, b \quad (4.71)$$

and

$$\pi^{\mu_\mathbf{i}}_{X^i}(x^i) := \sum_{w_i \in \Upsilon^i_{\mu_\mathbf{i}}(x^i)} \frac{1}{|\mathcal{W}_i|}, \;\; \pi^{\mu}_{X^i}(x^i) := \sum_{\mu_\mathbf{i}} \pi^{\mu_\mathbf{i}}_{X^i}(x^i) P_{S^i_{[t-d_i]}|S_{[t-1]}}(\mu_\mathbf{i}|\mu), \;\; i = a, b.$$

Hence, (4.70) can be shown following the same steps in Lemma 4.2.2. □

We can now complete the converse proof of Theorem 4.2.4. With Lemma 4.2.5 it is shown that any achievable rate pair can be approximated by the convex combinations of rate conditions which are indexed by $\mu \in \mathcal{S}^{(n)}$ and satisfy (4.61) for joint state-input-output distributions. Hence, any achievable pair $(R_a, R_b) \in \overline{co}\big(\bigcup_{\tilde{\pi}} \mathcal{R}^{ad}(\tilde{\pi})\big)$.

Remark 4.2.3 (Strictly Causal CSIT)**.** *When $d_a = d_b = 1$, Theorem 4.2.4 is the capacity region of the setup with strictly causal CSITs. This case was considered in the literature, e.g., see [LS13b], [LSY13], [LS13a] and [ZPS11], where it is shown that strictly causal side information is helpful. Theorem 4.2.4 verifies that since the full CSI is available at the receiver and since the decoder does not need to access the* *current CSI at the encoders, there exists no loss of optimality if the past information at the encoders are ignored.*

Remark 4.2.4 (CSI at Only One Encoder)**.** *One other conclusion of Theorem 4.2.4 is that in a situation where one of the encoders, say a, does not have an access to the state information (i.e., d_a is large) then, there exists no loss of optimality if the the past information at the other encoder is ignored.*

4.2.4 Degraded Message Set with Noisy CSIT

Assume a common message is provided to both encoders and one of the encoders has its own private message. Assume further that the encoder with the private message has causal noisy CSI, whereas the encoder with the common message only observes noisy state information with delay $d_a \geq 1$. Let the common and the private messages be W_a and W_b, respectively, and $S^a_{[t-d_a]}$, $d_a \geq 1$, and $S^b_{[t]}$ denote the CSI at encoder a, b, respectively, where (S_t, S^a_t, S^b_t) satisfies (4.1) and (4.2). Hence, $X^a_t = \phi^{(a)}_t(W_a, S^a_{[t-d_a]})$ and $X^b_t = \phi^{(b)}_t(W_a, W_b, S^b_{[t]})$; see Fig. 4.2. Let $\mathcal{C}^{dm}$ denote the capacity region for this channel. Recall that $\mathcal{T}_b = \mathcal{X}_b^{|\mathcal{S}_b|}$.

Definition 4.2.5. *A memoryless stationary (in time) team policy is a family*

$$\hat{\Pi} = \left\{\hat{\pi} = \left(\pi_{X^a,T^b}(\cdot,\cdot)\right) \in \mathcal{P}(\mathcal{X}^a \times \mathcal{T}^b)\right\} \tag{4.72}$$

of probability distributions on $(\mathcal{X}_a, \mathcal{T}_b)$.

Let for every $\hat{\pi}$, $\mathcal{R}^{dm}(\hat{\pi})$ denote the region of all rate pairs $R = (R_a, R_b)$ satisfying

$$R_b \ < \ I(T^b; Y|X^a, S) \tag{4.73}$$

$$R_a + R_b \ < \ I(X^a, T^b; Y|S) \tag{4.74}$$

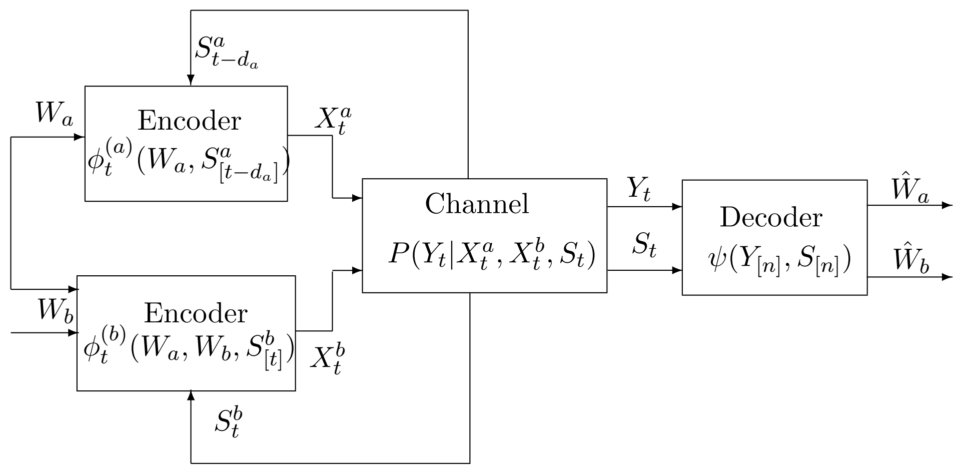

Figure 4.2: The multiple-access channel with degraded message set and with causal noisy state information.

where S, X^a, T^b and Y are random variables taking values in $\mathcal{S}$, $\mathcal{X}_a$, $\mathcal{T}_b$ and $\mathcal{Y}$, respectively and whose joint probability distribution factorizes as

$$P_{S,X^a,T^b,Y}(s,x^a,t^b,y) = P_S(s)P_{Y|X^a,T^b,S}(y|x^a,t^b,s)\pi_{X^a,T^b}(x^a,t^b). \tag{4.75}$$

Let $\overline{co}\left(\bigcup_{\hat{\pi}} \mathcal{R}^{dm}(\hat{\pi})\right)$ denotes the closure of the convex hull of the rate regions $\mathcal{R}^{dm}(\hat{\pi})$ given by (4.73) and (4.74) associated to all possible memoryless stationary team polices as defined in (4.72).

Theorem 4.2.5. $\mathcal{C}^{dm} = \overline{co}\left(\bigcup_{\hat{\pi}} \mathcal{R}^{dm}(\hat{\pi})\right)$.

Achievability. Fix $(R_a, R_b) \in \mathcal{R}^{dm}(\hat{\pi})$.

Codebook Generation Fix $\pi_{X^a}(x^a)$ and $\pi_{T^b|X^a}(t^b|x^a)$. For each $w_a \in \{1, \cdots, 2^{nR_a}\}$, randomly generate $x^a_{[n],w_a}$, each according to $\prod_{i=1}^n \pi_{X^a_i}(x^a_{i,w_a})$. Reveal this codebook to encoder b and, for each $w_a \in \{1, \cdots, 2^{nR_a}\}$ and $w_b \in \{1, \cdots, 2^{nR_b}\}$, encoder b randomly generates $t^b_{[n],w_b,w_a}$, each according to $\prod_{i=1}^n \pi_{T^b_i|X^a_i}(t^b_{i,w_b}|x^a_{i,w_a})$. These codeword pairs form the codebook, which is revealed to the decoder.

Encoding Define the encoding functions as follows: $x^a_i(w_a) = \phi^a_i(w_a, s^a_{[i-d_a]})$ and

$x_i^b(w_b) = \phi_i^b(w_b, w_a, s_{[i]}^b) = t_{i,w_b,w_a}^b(s_i^b)$ where x_{i,w_a}^a and t_{i,w_b,w_a}^b denote the ith component of $x_{[n],w_a}^a$ and $t_{[n],w_b,w_a}^b$, respectively. Therefore, to send the messages w_a and w_b, transmit the corresponding $x_{[n],w_a}^a$ and $t_{[n],w_b,w_a}^b$, respectively.

Decoding After receiving $(y_{[n]}, s_{[n]})$, the decoder looks for the only (w_a, w_b) pair such that $(x_{[n],w_a}^a, t_{[n],w_b}^b, y_{[n]}, s_{[n]})$ are jointly $\epsilon-$typical and declares this pair as its estimate $(\hat{w}_a, \hat{w}_b)$.

Error Analysis Let $E_{\alpha,\beta} \stackrel{\triangle}{=} \{(X_{[n],\alpha}^a, T_{[n],\beta,\alpha}^b, Y_{[n]}, S_{[n]}) \in A_\epsilon^n\}$, $\alpha \in \{1, \cdots, 2^{nR_a}\}$ and $\beta \in \{1, \cdots, 2^{nR_b}\}$ and assume that $(w_a, w_b) = (1, 1)$ was sent. Then

$$\begin{aligned} P_e^n &= P\big(E_{1,1}^c \bigcup_{(\alpha,\beta)\neq(1,1)} E_{\alpha,\beta}\big) \\ &\leq P(E_{1,1}^c) + \sum_{\alpha=1,\beta\neq 1} P(E_{\alpha,\beta}) + \sum_{\alpha\neq 1,\beta=1} P(E_{\alpha,\beta}) + \sum_{\alpha\neq 1,\beta\neq 1} P(E_{\alpha,\beta}). \end{aligned} \tag{4.76}$$

Since $\{Y_i, S_i, X_i^a, T_i^b\}_{i=1}^\infty$ is an i.i.d. sequence hence, $P(E_{1,1}^c) \to 0$ for $n \to \infty$. Next, let us consider the second term

$$\begin{aligned} &\sum_{\alpha=1,\beta\neq 1} P(E_{\alpha=1,\beta\neq 1}) \\ &= \sum_{\alpha=1,\beta\neq 1} P((X_{[n],1}^a, T_{[n],\beta}^b, Y_{[n]}, S_{[n]}) \in \mathrm{A}_\epsilon^n) \\ &\stackrel{(i)}{=} \sum_{\alpha=1,\beta\neq 1} \sum_{(x_{[n]}^a, t_{[n]}^b, y_{[n]}, s_{[n]}) \in \mathrm{A}_\epsilon^n} P_{T_{[n]}^b|X_{[n]}^a}(t_{[n]}^b|x_{[n]}^a) P_{X_{[n]}^a,Y_{[n]},S_{[n]}}(x_{[n]}^a, y_{[n]}, s_{[n]}) \\ &\leq \sum_{\alpha=1,\beta\neq 1} |A_\epsilon^n| 2^{-n[H(T^b|X^a)-\epsilon]} 2^{-n[H(X^a,Y,S)-\epsilon]} \\ &\leq 2^{nR_b} 2^{-n[H(T^b|X^a)+H(X^a,Y,S)-H(X^a,T^b,Y,S)-3\epsilon]} \\ &\stackrel{(ii)}{=} 2^{n[R_b - I(T^b;Y|S,X^a)-3\epsilon]} \end{aligned} \tag{4.77}$$

where (i) due to $T_{[n],\beta}^b$ is independent of $(Y_{[n]}, S_{[n]})$ given $X_{[n],1}^a$ and (ii) is due to

$$\begin{aligned} &H(T^b|X^a) + H(X^a, Y, S) - H(X^a, T^b, Y, S) \\ &\quad = H(T^b|X^a) + H(X^a, Y, S) - H(Y|X^a, T^b, S) - H(X^a, T^b, S) \end{aligned}$$

$$= H(X^a, Y, S) - H(Y|X^a, T^b, S) - H(X^a, S)$$
$$= I(T^b; Y|S, X^a)$$

where the second equality follows since T^b and S are independent given X^a. Finally,

$$\begin{aligned}
\sum_{\alpha\neq 1,\beta\neq 1} P(E_{\alpha\neq 1,\beta\neq 1}) &= \sum_{\alpha\neq 1,\beta\neq 1} P((X^a_{[n],\alpha}, T^b_{[n],\beta}, Y_{[n]}, S_{[n]}) \in \mathrm{A}^n_\epsilon) \\
&\overset{(iii)}{=} \sum_{\alpha\neq 1,\beta\neq 1} \sum_{(x^a_{[n]}, t^b_{[n]}, y_{[n]}, s_{[n]})\in \mathrm{A}^n_\epsilon} P_{T^b_{[n]}, X^a_{[n]}}(t^b_{[n]}, x^a_{[n]}) P_{Y_{[n]}, S_{[n]}}(y_{[n]}, s_{[n]}) \\
&\leq \sum_{\alpha\neq 1,\beta\neq 1} |A^n_\epsilon| 2^{-n[H(T^b, X^a)-\epsilon]} 2^{-n[H(Y,S)-\epsilon]} \\
&\leq 2^{n(R_a+R_b)} 2^{-n[H(T^b,X^a)+H(Y,S)-H(X^a,T^b,Y,S)-3\epsilon]} \\
&\overset{(iv)}{=} 2^{n[R_a+R_b-I(X^a,T^b;Y|S)-3\epsilon]} \qquad (4.78)
\end{aligned}$$

where (iii) holds since for $\alpha, \beta \neq 1$, $(T^b_{[n],\beta}, X^a_{[n],\alpha})$ is independent of $(Y_{[n]}, S_{[n]})$ and (iv) follows since

$$\begin{aligned}
&H(T^b, X^a) + H(Y, S) - H(X^a, T^b, Y, S) \\
&\quad = H(T^b, X^a) + H(Y, S) - H(Y|X^a, S, T^b) - H(X^a, S, T^b) \\
&\quad = H(T^b, X^a) + H(Y, S) - H(Y|X^a, S, T^b) - H(X^a, T^b) - H(S) \\
&\quad = I(X^a, T^b; Y|S),
\end{aligned}$$

and the rate conditions of the $\mathcal{R}_C(\hat{\pi})$ imply that each term tends in (4.76) tends to zero as $n \to \infty$. Finally, observe that the analysis for the error event $\sum_{\alpha\neq 1,\beta=1} P(E_{\alpha,\beta})$ is identical to the case of $\sum_{\alpha\neq 1,\beta\neq 1} P(E_{\alpha,\beta})$ which induces the same sum-rate constraint.

□

To prove the converse, first note that the main motivation in indexing mutual information terms by the past CSI, is to get a product form on the team policies. In the cooperative setup, we do not require a product form and therefore, the convex combination argument is not essential. However, we herein keep this indexing (see (4.75)) to avoid the use of a time sharing auxiliary random variable.

Converse Proof. First observe that, since $X_t^b = \phi_t^{(b)}\left(W_a, W_b, S_{[t-1]}^b, S_t^b\right)$, we have

$$T_t^b = \phi_t^{(b)}\left(W_a, W_b, S_{[t-1]}^b\right) \in \mathcal{X}_b{}^{|\mathcal{S}_b|}. \tag{4.79}$$

Lemma 4.2.7. *Let $T_t^b \in \mathcal{T}_b$ be the Shannon strategy induced by $\phi_t^{(b)}$ as shown in (4.79). Assume that a rate pair $R = (R_a, R_b)$, with block length $n \geq 1$ and a constant $\epsilon \in (0, 1/2)$, is achievable. Then,*

$$R_b \leq \sum_{\mu \in \mathcal{S}^{(n)}} \alpha_\mu I(T_t^b; Y_t | X_t^a, S_t, S_{[t-1]} = \mu) + \eta(\epsilon) \tag{4.80}$$

$$R_a + R_b \leq \sum_{\mu \in \mathcal{S}^{(n)}} \alpha_\mu I(X_t^a, T_t^b; Y_t | S_t, S_{[t-1]} = \mu) + \eta(\epsilon) \tag{4.81}$$

where α_μ and $\eta(\epsilon)$ are defined in (4.13).

Proof. Let us first consider the sum-rate condition. Since,

$$\begin{aligned} I(\mathbf{W}; Y_{[n]}, S_{[n]}) &\leq \sum_{t=1}^{n} \left[H(Y_t|S_{[t]}) - H(Y_t|\mathbf{W}, S_{[t]}, Y_{[t-1]}, X_t^a, T_t^b)\right] \\ &\overset{(i)}{=} \sum_{t=1}^{n} \left[H(Y_t|S_{[t]}) - H(Y_t|S_{[t]}, X_t^a, T_t^b)\right] \\ &= \sum_{t=1}^{n} I(X_t^a, T_t^b; Y_t|S_{[t]}), \end{aligned} \tag{4.82}$$

where (i) can be shown in a similar way as (4.21), we have,

$$\begin{aligned} R_a + R_b &\leq \frac{1}{n}\sum_{t=1}^{n} I(X_t^a, T_t^b; Y_t|S_{[t]}) + \eta(\epsilon) \\ I(X_t^a, T_t^b; Y_t|S_{[t]}) &= n \sum_{\mu \in \mathcal{S}^{t-1}} \alpha_\mu I(X_t^a, T_t^b; Y_t|S_t, S_{[t-1]} = \mu). \end{aligned} \tag{4.83}$$

Substituting the above into (4.83) yields (4.81). Let us now consider encoder b. With

Fano's inequality and standard steps, we get

$$R_b \leq \frac{1}{n}\log(|\mathcal{W}_b|) \leq \frac{1}{1-\epsilon}\frac{1}{n}\left(I(W_b;Y_{[n]},S_{[n]})+H(\epsilon)\right). \tag{4.84}$$

Following similar reasonings as in (4.66) we get,

$$\begin{aligned}
I(W_b;Y_{[n]},S_{[n]}) &\leq I(W_b;Y_{[n]},S_{[n]}|W_a,S^a_{[n]}) \\
&= \sum_{t=1}^{n}\big[H(Y_t|S_{[t]},Y_{[t-1]},W_a,S^a_{[n]}) \\
&\qquad\qquad -H(Y_t|S_{[t]},Y_{[t-1]},W_a,W_b,S^a_{[n]})\big] \\
&= \sum_{t=1}^{n}\big[H(Y_t|S_{[t]},Y_{[t-1]},W_a,S^a_{[n]},X^a_{[n]}) \\
&\qquad\qquad -H(Y_t|S_{[t]},Y_{[t-1]},W_a,W_b,S^a_{[n]},X^a_{[n]})\big] \\
&\leq \sum_{t=1}^{n}\big[H(Y_t|S_{[t]},X^a_t) \\
&\qquad\qquad -H(Y_t|S_{[t]},Y_{[t-1]},W_a,W_b,S^a_{[n]},X^a_{[n]},T^b_t)\big] \\
&\overset{(i)}{=} \sum_{t=1}^{n}\big[H(Y_t|S_{[t]},X^a_t)-H(Y_t|S_{[t]},X^a_t,T^b_t)\big] \\
&= \sum_{t=1}^{n} I(T^b_t;Y_t|X^a_t,S_{[t]})
\end{aligned} \tag{4.85}$$

where (i) is valid since

$$\begin{aligned}
&P_{Y_t|S_{[t]},Y_{[t-1]},\mathbf{W},S^a_{[n]},X^a_{[n]},T^b_t}(y_t|s_{[t]},y_{[t-1]},\mathbf{w},s^a_{[n]},x^a_{[n]},t^b_t) \\
&\overset{(ii)}{=} \sum_{s^b_t\in\mathcal{S}_b} P_{Y_t|S_t,S^b_t,X^a_t,T^b_t}(y_t|s_t,s^b_t,x^a_t,t^b_t) \\
&\qquad\qquad P_{S^b_t|S_{[t]},Y_{[t-1]},\mathbf{W},S^a_{[n]},X^a_{[n]},T^b_t}(s^b_t|s_{[t]},y_{[t-1]},\mathbf{w},s^a_{[n]},x^a_{[n]},t^b_t) \\
&\overset{(iii)}{=} \sum_{s^b_t\in\mathcal{S}_b} P_{Y_t|S_t,S^b_t,X^a_t,T^b_t}(y_t|s_t,s^b_t,x^a_t,t^b_t)P_{S^b_t|S_t}(s^b_t|s_t) \\
&= P_{Y_t|S_t,X^a_t,T^b_t}(y_t|s_t,x^a_t,t^b_t).
\end{aligned} \tag{4.86}$$

where (ii) is due to (4.3) and (iii) is due to (4.1) and (4.2). Following (4.21), we verify (4.80). □

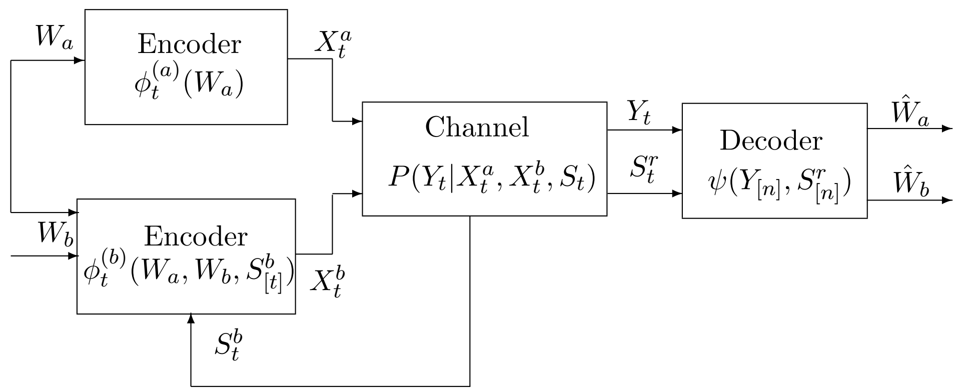

Figure 4.3: The multiple-access channel with degraded message set and with noisy state information at the receiver.

We now need to show that the distribution $P_{X_t^a, T_t^b, Y_t, S_t | S_{[t-1]}}(x^a, t^b, y, s|\mu)$ factorizes as in (4.75). Let first $\pi^\mu_{X^a, T^b}(x^a, t^b) := P_{X_t^a, T_t^b | S_{[t-1]}}(x^a, t^b|\mu)$ and observe that

$$\begin{aligned} & P_{X_t^a, T_t^b, Y_t, S_t | S_{[t-1]}}(x^a, t^b, y, s|\mu) \\ &= \sum_{s_t^b \in \mathcal{S}^b} P_{Y_t|X_t^a, X_t^b, S_t}(y|x^a, t^b(s_t^b), s) P_{S_t^b|S_t}(s_t^b|s_t) P_{S_t}(s) P_{X_t^a, T_t^b|S_{[t-1]}}(x^a, t^b|\mu) \\ &= \pi^\mu_{X^a, T^b}(x^a, t^b) P_{S_t}(s) P_{Y_t|X_t^a, T_t^b, S_t}(y|x^a, t^b, s) \end{aligned} \tag{4.87}$$

where the equalities are verified by (4.3), by (4.1) and by the fact that (X_t^a, T_t^b) is independent of S_t.

We can now complete the converse proof of Theorem 4.2.5. With Lemma 4.2.7 it is shown that any achievable rate pair can be approximated by the convex combinations of rate conditions which are indexed by $\mu \in \mathcal{S}^{(n)}$ and satisfy (4.75) for joint state-input-output distributions. Hence, any achievable pair $(R_a, R_b) \in \overline{co}\big(\bigcup_{\hat{\pi}} \mathcal{R}^{dm}(\hat{\pi})\big)$.

□

Remark 4.2.5. *Theorem 4.2.5 shows that when the common message encoder does* *not have access to the current noisy CSI (since the delay $d_a \geq 1$), by enlarging the optimization space of the other encoder, via Shannon strategies, the past CSI can be ignored without loss of optimality if the decoder is provided with complete CSI.*

Note that in the degraded message set scenario a product form on the pair (X^a, T^b) is not required (see (4.75)). In connection with this observation, let us consider the following noisy CSIR setup.

Let the encoder with the private message causally observe noisy state information, whereas the encoder with the common message has no CSI, i.e., $X_t^a = \phi_t^{(a)}(W_a)$ and $X_t^b = \phi_t^{(b)}(W_a, W_b, S_{[t]}^b)$, and the decoder also has access to noisy CSI at time t, $S_t^r \in \mathcal{S}_r$; see Fig. 4.3, where,

$$P_{S_{[n]}, S_{[n]}^r, S_{[n]}^b, \mathbf{W}}(s_{[n]}, s_{[n]}^r, s_{[n]}^b, \mathbf{w}) = \prod_{t=1}^{n} \frac{1}{|\mathcal{W}_a|} \frac{1}{|\mathcal{W}_b|} P_{S_t, S_t^r, S_t^b}(s_t, s_t^r, s_t^b) \tag{4.88}$$

and let $\bar{\mathcal{C}}^{dm}$ denote the capacity region for this setup.

Let for every memoryless stationary team policy $\hat{\pi}$ defined in (4.72), $\bar{\mathcal{R}}^{dm}(\hat{\pi})$ denote the region of all rate pairs $R = (R_a, R_b)$ satisfying,

$$R_b < I(T^b; Y | X^a, S^r) \tag{4.89}$$

$$R_a + R_b < I(X^a, T^b; Y | S^r) \tag{4.90}$$

where S^r, X^a, T^b and Y are random variables taking values in $\mathcal{S}_r$, $\mathcal{X}_a$, $\mathcal{T}_b$ and $\mathcal{Y}$, respectively and whose joint probability distribution factorizes as

$$P_{S^r, X^a, T^b, Y}(s^r, x^a, t^b, y) = P_{S^r}(s^r) P_{Y|X^a, T^b, S^r}(y|x^a, t^b, s^r) \pi_{X^a, T^b}(x^a, t^b). \tag{4.91}$$

Let $\overline{co}\Big(\bigcup_{\hat{\pi}} \bar{\mathcal{R}}^{dm}(\hat{\pi})\Big)$ denotes the closure of the convex hull of the rate regions $\bar{\mathcal{R}}^{dm}(\hat{\pi})$ given by (4.89) and (4.90) associated to all possible $\hat{\pi}$ as defined in (4.72).

Theorem 4.2.6. $\bar{\mathcal{C}}^{dm} = \overline{co}\Big(\bigcup_{\hat{\pi}} \bar{\mathcal{R}}^{dm}(\hat{\pi})\Big)$.

Proof. The achievability proof is identical to that of Theorem 4.2.5. The converse proof is also similar and therefore, we only provide a sketch. In particular, observe the following lines of equations for the converse proof of the condition on R_b:

$$\begin{aligned}
I(W_b;Y_{[n]},S^r_{[n]}) &\leq I(W_b;Y_{[n]},S^r_{[n]}|W_a)\\
&= \sum_{t=1}^{n}\big[H(Y_t,S^r_t|S^r_{[t-1]},Y_{[t-1]},W_a)\\
&\qquad\qquad -H(Y_t,S^r_t|S^r_{[t-1]},Y_{[t-1]},W_a,W_b)\big]\\
&\stackrel{(i)}{=} \sum_{t=1}^{n}\big[H(Y_t|S^r_{[t]},Y_{[t-1]},W_a)-H(Y_t|S^r_{[t]},Y_{[t-1]},W_a,W_b)\big]\\
&= \sum_{t=1}^{n}\big[H(Y_t|S^r_{[t]},Y_{[t-1]},W_a,X^a_t)\\
&\qquad\qquad -H(Y_t|S^r_{[t]},Y_{[t-1]},W_a,W_b,X^a_t)\big]\\
&\stackrel{(ii)}{\leq} \sum_{t=1}^{n}\big[H(Y_t|S^r_{[t]},X^a_t)-H(Y_t|S^r_{[t]},Y_{[t-1]},W_a,W_b,X^a_t,T^b_t)\big]\\
&\stackrel{(iii)}{=} \sum_{t=1}^{n}\big[H(Y_t|S^r_{[t]},X^a_t)-H(Y_t|S^r_{[t]},X^a_t,T^b_t)\big]\\
&= \sum_{t=1}^{n}I(T^b_t;Y_t|X^a_t,S^r_{[t]}) \qquad (4.92)
\end{aligned}$$

where (i) follows since state is i.i.d., where T^b_t is the Shannon strategy induced by encoder b at time t as shown in (4.79), and (ii) is valid since conditioning does not increase entropy, and (iii) is valid since

$$\begin{aligned}
&P_{Y_t|S^r_{[t]},Y_{[t-1]},W_a,W_b,X^a_t,T^b_t}(y_t|s^r_{[t]},y_{[t-1]},w_a,w_b,x^a_t,t^b_t)\\
&\stackrel{(iv)}{=} \sum_{s_t\in\mathcal{S},s^b_t\in\mathcal{S}_b} P_{Y_t|S_t,S^b_t,X^a_t,T^b_t}(y_t|s_t,s^b_t,x^a_t,t^b_t)\\
&\qquad\qquad P_{S^b_t,S_t|S^r_{[t]},Y_{[t-1]},W_a,W_b,X^a_t,T^b_t}(s^b_t,s_t|s^r_{[t]},y_{[t-1]},w_a,w_b,x^a_t,t^b_t)\\
&\stackrel{(v)}{=} \sum_{s_t\in\mathcal{S},s^b_t\in\mathcal{S}_b} P_{Y_t|S_t,S^b_t,X^a_t,T^b_t}(y_t|s_t,s^b_t,x^a_t,t^b_t)P_{S^b_t,S_t}(s^b_t,s_t|s^r_t)\\
&= P_{Y_t|S^r_t,X^a_t,T^b_t}(y_t|s^r_t,x^a_t,t^b_t) \qquad (4.93)
\end{aligned}$$

where (iv) is due to (4.3) and (v) holds due to (4.88). Hence,

$$R_b \leq \sum_{\mu_{\mathbf{r}} \in \mathcal{S}_r^{(n)}} \alpha_{\mu_{\mathbf{r}}} I(T_t^b; Y_t | X_t^a, S_t^r, S_{[t-1]}^r = \mu_{\mathbf{r}}) + \eta(\epsilon) \quad (4.94)$$

$$R_a + R_b \leq \sum_{\mu_{\mathbf{r}} \in \mathcal{S}_r^{(n)}} \alpha_{\mu_{\mathbf{r}}} I(X_t^a, T_t^b; Y_t | S_t^r, S_{[t-1]}^r = \mu_{\mathbf{r}}) + \eta(\epsilon) \quad (4.95)$$

where $\alpha_{\mu_{\mathbf{r}}} := \frac{1}{n} P_{S_{[t-1]}^r}(\mu_{\mathbf{r}})$ and $\eta(\epsilon)$ is given in (4.13). We now need to show that the joint distribution $P_{X_t^a, T_t^b, Y_t, S_t^r | S_{[t-1]}^r}(x^a, t^b, y, s^r | \mu_{\mathbf{r}})$ satisfies (4.91). Let $\pi_{X^a, T^b}^{\mu_{\mathbf{r}}}(x^a, t^b) := P_{X_t^a, T_t^b | S_{[t-1]}^r}(x^a, t^b | \mu_{\mathbf{r}})$ and observe that

$$\begin{aligned} & P_{X_t^a, T_t^b, Y_t, S_t^r | S_{[t-1]}^r}(x^a, t^b, y, s^r | \mu_{\mathbf{r}}) \\ &= \sum_{s_t^b \in \mathcal{S}^b} \sum_{s_t \in \mathcal{S}} P_{Y_t | X_t^a, X_t^b, S_t}(y | x^a, t^b(s_t^b), s) P_{S_t^b, S_t, S_r}(s_t^b, s_t, s^r) P_{X_t^a, T_t^b | S_{[t-1]}^r}(x^a, t^b | \mu_{\mathbf{r}}) \\ &= \pi_{X^a, T^b}^{\mu}(x^a, t^b) P_{S_t^r}(s^r) P_{Y_t | X_t^a, T_t^b, S_t^r}(y | x^a, t^b, s^r) \end{aligned} \quad (4.96)$$

where the first equality is verified by (4.3) and by the fact that (X_t^a, T_t^b) is independent of (S_t, S_t^b, S_t^r). □

Remark 4.2.6. *It should be observed that unlike Theorem 4.2.5 and results in the previous sections, for the validity of Theorem 4.2.6, it is not required to have a Markov condition on* $P_{S_t, S_t^b, S_t^r}(s_t, s_t^b, s_t^r)$*. Furthermore, the result also holds with no CSIR, i.e.,* $\mathcal{S}_r = \emptyset$ *is allowed, and in this case Theorem 4.2.6 is as an extension of [SBSV08, Theorem 4] to a noisy setup.*

Remark 4.2.7. *For the validity of converse proof of Theorem 4.2.6 it is crucial that* X_t^a *only depends on* W_a*. To be more explicit, let us assume* $\mathcal{S}_r = \emptyset$ *and consider the following steps of the converse*

$$\begin{aligned} I(W_b; Y_{[n]}) &\leq \sum_{t=1}^{n} H(Y_t | Y_{[t-1]}, X_{[n]}^a) - H(Y_t | Y_{[t-1]}, W_a, W_b, X_{[n]}^a, T_t^b) \\ &= \sum_{t=1}^{n} H(Y_t | Y_{[t-1]}, X_{[n]}^a) - H(Y_t | Y_{[t-1]}, X_t^a, T_t^b). \end{aligned} \quad (4.97)$$

Since S_t is not available to the decoder, the above equality is valid if $X^a_{[n]}$ does not provide any information about S_t. Hence, in other words, whether CSITs are noisy or not, if there is no CSI or noisy CSI at the decoder, the arguments above would fail if the uninformed encoder observes some degree of CSI, i.e., $d_a < \infty$ so that $X^a_{[n]}$ carry some information about (S_t, S^b_t, S^r_t).

It should be noted that for the setup given in [SBSV08, Theorem 4], Theorem 4.2.6 provides an equivalent characterization. Recall that in [SBSV08, Theorem 4] the informed encoder has full CSI, i.e., $X^b_t = \phi^{(b)}_t(W_a, W_b, S_{[t]})$, both the uniformed encoder and the decoder have no CSI and the capacity region, $\mathcal{C}_{AS}$, is given as the closure of all rate pairs (R_a, R_b) satisfying

$$R_b \;<\; I(U;Y|X^a) \tag{4.98}$$

$$R_b + R_a \;<\; I(U, X^a; Y) \tag{4.99}$$

for some joint measure on $\mathcal{S} \times \mathcal{X}_a \times \mathcal{X}_b \times \mathcal{Y} \times \mathcal{U}$ having the form

$$P_{Y|X^a,X^b,S}(y|x^a, x^b, s)P_{X^b|U,X^a,S}(x^b|u, x^a, s)P_S(s)P_{X^a,U}(x^a, u), \tag{4.100}$$

where $|\mathcal{U}| \leq |\mathcal{S}||\mathcal{X}_a||\mathcal{X}_b| + 1$. On the other hand, for this setup, Theorem 4.2.6 gives the capacity region, $\underline{\mathcal{C}}^{dm}$, as $\overline{co}\left(\bigcup_{\hat{\pi}} \mathcal{R}'_C(\hat{\pi})\right)$ where $\mathcal{R}'_C(\hat{\pi})$ denotes the region of all rate pairs $R = (R_a, R_b)$ satisfying

$$R_b \;<\; I(T;Y|X^a) \tag{4.101}$$

$$R_a + R_b \;<\; I(T, X^a; Y) \tag{4.102}$$

where $P_{Y,T,X^a,X^b,S}(y, t, x^a, x^b, s)$ factorizes as

$$P_{Y|X^a,X^b,S}(y|x^a, x^b, s)P_{X^b|S,T}(x^b|s, t)P_S(s)\hat{\pi}_{X^a,T}(x^a, t), \tag{4.103}$$

and $T : \mathcal{S} \to \mathcal{X}_b$.

Although the relation between an auxiliary variable and Shannon strategies is well understood for the single-user case (e.g., see [KSM07, Section 3.2]), it requires more attention in the multi-user case; in particular, note the difference between $|\mathcal{U}|$ and $|\mathcal{T}|$. Hence, we provide a proof for $\underline{\mathcal{C}}^{dm} = \mathcal{C}_{AS}$.

Theorem 4.2.7. $\underline{\mathcal{C}}^{dm} = \mathcal{C}_{AS}$.

Lemma 4.2.8. $\underline{\mathcal{C}}^{dm} \subseteq \mathcal{C}_{AS}$.

Proof. Recall that $T \in |\mathcal{T}| = |\mathcal{X}_b|^{|\mathcal{S}|}$ and $|\mathcal{U}| \leq |\mathcal{X}_a||\mathcal{X}_b||\mathcal{S}| + 1$. Hence, we have either $|\mathcal{U}| > |\mathcal{T}|$ or else. In the case where $|\mathcal{U}| < |\mathcal{T}|$, we note that $|\mathcal{U}|$ is limited to a finite set without loss of generality. Hence, we can always take $|\mathcal{U}|$ at least $|\mathcal{T}|$ such that it satisfies (4.98), (4.99) and (4.100). Then we can directly conclude that $\underline{\mathcal{C}}^{dm} \subseteq \mathcal{C}_{AS}$ since $P_{X^b|S,T}(x^b|s,t) = P_{X^b|S,T}(x^b|s,t,x^a) = 1_{\{x^b=t(s)\}}$ and this is a special case of $P_{X^b|U,X^a,S}(x^b|u,x^a,s)$. □

In order to prove the other direction, i.e., $\mathcal{C}_{AS} \subseteq \underline{\mathcal{C}}^{dm}$, let $\mathcal{C}_{AS}^E$ be the closure of all rate pairs (R_a, R_b) satisfying

$$R_b < I(U;Y|X^a) \tag{4.104}$$

$$R_b + R_a < I(U,X^a;Y) \tag{4.105}$$

for some joint measure on $\mathcal{S} \times \mathcal{X}_a \times \mathcal{X}_b \times \mathcal{Y} \times \mathcal{U}$ having the form

$$P_{Y|X^a,X^b,S}(y|x^a,x^b,s)1_{\{x^b=\mathbf{m}(s,x^a,u)\}}P_S(s)P_{X^a,U}(x^a,u), \tag{4.106}$$

for some $\mathbf{m} : \mathcal{U} \times \mathcal{X}_a \times \mathcal{S} \to \mathcal{X}_b$, where $|\mathcal{U}| \leq |\mathcal{S}||\mathcal{X}_a||\mathcal{X}_b| + 1$, and we first show that $\mathcal{C}_{AS} = \mathcal{C}_{AS}^E$, and following this, we show that $\mathcal{C}_{AS}^E \subseteq \underline{\mathcal{C}}^{dm}$.

Lemma 4.2.9. $\mathcal{C}_{AS} = \mathcal{C}_{AS}^E$.

Proof. It is obvious that $\mathcal{C}_{AS}^E \subseteq \mathcal{C}_{AS}$ and hence, we need to show that $\mathcal{C}_{AS} \subseteq \mathcal{C}_{AS}^E$. Let $\bar{P}_{X^b,X^a,U,S}(x^b,x^a,u,s)$ be a joint distribution in the form of (4.100), i.e.,

$$\bar{P}_{X^b,X^a,U,S}(x^b,x^a,u,s) = \bar{P}_{X^b|X^a,U,S}(x^b|x^a,u,s)P_S(s)\bar{P}_{X^a,U}(x^a,u). \tag{4.107}$$

Let $\bar{\mathbf{\Lambda}}$ denote a $|\mathcal{X}_a||\mathcal{U}||\mathcal{S}|$-by-$|\mathcal{X}_b|$ matrix where $\bar{\mathbf{\Lambda}}_{i,jkl} = \bar{P}_{X^b|X^a,U,S}(i|j,k,l)$, $1 \le i \le |\mathcal{X}_b|$, $1 \le j \le |\mathcal{X}_a|$, $1 \le k \le |\mathcal{U}|$ and $1 \le l \le |\mathcal{S}|$. Hence, $\bar{\mathbf{\Lambda}}$ is a $|\mathcal{X}_a||\mathcal{U}||\mathcal{S}|$-by-$|\mathcal{X}_b|$ row stochastic matrix, i.e., $\bar{\mathbf{\Lambda}}_{i,jkl} \ge 0,\ \forall i,j,k,l$ and $\sum_{i=1}^{|\mathcal{X}_b|} \bar{\mathbf{\Lambda}}_{i,jkl} = 1,\ \forall j,k,l$. Let $\mathbf{\Lambda}$ denote a $|\mathcal{X}_a||\mathcal{U}||\mathcal{S}|$-by-$|\mathcal{X}_b|$ binary stochastic matrix, that is a matrix with each row has exactly one non-zero element, which is 1. Observe now that any row stochastic matrix can be written as a convex combination of binary stochastic matrices (e.g., see [Hög77, Lemma 5] and [NFT07, Proposition IV.1]). Therefore, we have

$$\bar{\mathbf{\Lambda}} = \sum_{i=1}^{k} \lambda_i \mathbf{\Lambda}^{(i)},\ \ \sum_{i=1}^{k} \lambda_i = 1, \tag{4.108}$$

where $\mathbf{\Lambda}^{(i)}$ is a binary stochastic matrix and by [Hög77, Lemma 5], $k \le (|\mathcal{X}_a||\mathcal{U}||\mathcal{S}|)^2$.

Let, for the joint distribution $\bar{P}_{X^b,X^a,U,S}(x^b,x^a,u,s)$,

$$\bar{R}_b \ < \ I(U;Y|X^a)_{\bar{\mathbf{\Lambda}}}, \tag{4.109}$$

$$\bar{R}_a + \bar{R}_b \ < \ I(U,X^a;Y)_{\bar{\mathbf{\Lambda}}}. \tag{4.110}$$

Therefore, $(\bar{R}_a,\bar{R}_b) \in \mathcal{C}_{AS}$. Now, observe that for a fixed distribution $P_{X^a,U}(x^a,u)$, both $I(U,X^a;Y)$ and $I(U;Y|X^a)$ are convex in $P_{Y|X^a,U}(y|x^a,u)$ and hence, convex in $P_{X^b|X^a,U,S}(\cdot|x^a,u,s)$. This and (4.108) imply that

$$I(U;Y|X^a)_{\bar{\mathbf{\Lambda}}} \ \le \ \sum_{i=1}^{k} \lambda_i I(U;Y|X^a)_{\mathbf{\Lambda}^{(i)}}, \tag{4.111}$$

$$I(U,X^a;Y)_{\bar{\mathbf{\Lambda}}} \ \le \ \sum_{i=1}^{k} \lambda_i I(U,X^a;Y)_{\mathbf{\Lambda}^{(i)}}, \tag{4.112}$$

where $I(U;Y|X^a)_{\mathbf{\Lambda}^{(i)}}$ and $I(U,X^a;Y)_{\mathbf{\Lambda}^{(i)}}$ denote the mutual information terms induced by $\mathbf{\Lambda}^{(i)}$.

Now, let (R_a^i, R_b^i), $1 \le i \le k$, be such that

$$R_b^i \le I(U;Y|X^a)_{\mathbf{\Lambda}^{(i)}},$$

$$R_b^i + R_a^i \le I(U,X^a;Y)_{\mathbf{\Lambda}^{(i)}},$$

and hence, $(R_a^i, R_b^i) \in \mathcal{C}_{AS}^E$, $1 \le i \le k$. Let $(R_a^f, R_b^f) = \sum_{i=1}^k \lambda_i (R_a^i, R_b^i)$. Since a convex combination of achievable rates is also achievable, so $(R_a^f, R_b^f) \in \mathcal{C}_{AS}^E$. This observation and inequalities (4.109)-(4.112) complete the claim that $(\bar{R}_a, \bar{R}_b) \in \mathcal{C}_{AS}^E$.

□

Up to now, we have shown that $\underline{\mathcal{C}}^{dm} \subseteq \mathcal{C}_{AS}$ and $\mathcal{C}_{AS}^E = \mathcal{C}_{AS}$. In order to prove that $\underline{\mathcal{C}}^{dm} = \mathcal{C}_{AS}$, it remains to show that $\mathcal{C}_{AS}^E \subseteq \underline{\mathcal{C}}^{dm}$. Note that $\mathcal{C}_{AS}^E$ still depends on $P_{X^a,U}(x^a,u)$ in which $|\mathcal{U}|$ can be larger than $|\mathcal{T}|$. Hence, in the next lemma we basically show that for every $P_{X^a,U}(x^a,u)$, there exists a $\hat{\pi}_{T^a,U}(t^a,u)$ which induces the same rate constraints as induced by $P_{X^a,U}(x^a,u)$.

Lemma 4.2.10. $\mathcal{C}_{AS}^E \subseteq \underline{\mathcal{C}}^{dm}$.

Proof. Let us fix a joint distribution $P^*_{Y,X^a,X^b,U,S}(y,x^a,x^b,u,s)$ satisfying (4.106), i.e., $P^*_{Y,X^a,X^b,U,S}(y,x^a,x^b,u,s) = P^*_{Y|X^a,X^b,S}(y|x^a,x^b,s)1_{\{x^b=\mathbf{m}(s,x^a,u)\}}P_S(s)P^*_{X^a,U}(x^a,u)$.

Observe that for every $\mathbf{m}$ satisfying $x^b = \mathbf{m}(u,x^a,s)$, one can define

$$x^b = \mathbf{m}(u,x^a,s) = \bar{\mathbf{m}}(x^a,u)(s), \;\; \bar{\mathbf{m}}(x^a,u) \in \mathcal{T}, \tag{4.113}$$

where $\mathcal{T}$ is the set of all mappings from $\mathcal{S}$ to $\mathcal{X}_b$. Now, let

$$\left(I(U;Y|X^a)_{P^*_{Y,X^a,U}(y,x^a,u)}, I(U,X^a;Y)_{P^*_{Y,X^a,U}(y,x^a,u)} \right), \tag{4.114}$$

denote the mutual information pair induced by $P^*_{Y,X^a,U}(y,x^a,u)$. We have

$I(U,X^a;Y)_{P^*_{Y,X^a,U}(y,x^a,u)}$

$$
\begin{aligned}
&= \sum_{u\in\mathcal{U}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,U}(y,x^a,u)\log\frac{P^*_{Y,U,X^a}(y,u,x^a)}{P^*_Y(y)P^*_{U,X^a}(u,x^a)}\\
&= \sum_{t\in\mathcal{T}}\sum_{u\in\mathcal{U}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,U,T}(y,x^a,u,t)\log\frac{P^*_{Y,U,X^a}(y,u,x^a)}{P^*_Y(y)P^*_{U,X^a}(u,x^a)}\\
&\overset{(i)}{=} \sum_{t\in\mathcal{T}}\sum_{u\in\mathcal{U}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,U,T}(y,x^a,u,t)\log\frac{P^*_{Y,U,X^a,T}(y,u,x^a,t)}{P^*_Y(y)P^*_{U,X^a,T}(u,x^a,t)}\\
&\overset{(ii)}{=} \sum_{t\in\mathcal{T}}\sum_{u\in\mathcal{U}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,U,T}(y,x^a,u,t)\log\frac{P^*_{Y|X^a,T}(y|x^a,t)P^*_{U,T,X^a}(u,t,x^a)}{P^*_Y(y)P^*_{U,T,X^a}(u,t,x^a)}\\
&= \sum_{t\in\mathcal{T}}\sum_{u\in\mathcal{U}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,U,T}(y,x^a,u,t)\log\frac{P^*_{Y,X^a,T}(y,x^a,t)}{P^*_Y(y)P^*_{X^a,T}(x^a,t)}\\
&= \sum_{t\in\mathcal{T}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,T}(y,x^a,t)\log\frac{P^*_{Y,X^a,T}(y,x^a,t)}{P^*_Y(y)P^*_{X^a,t}(x^a,t)}\\
&= I(T,X^a;Y)_{P^*_{Y,X^a,T}(y,x^a,t)}, \qquad (4.115)
\end{aligned}
$$

where (i) is valid since $\bar{\mathbf{m}}(x^a,u)\in\mathcal{T}$, i.e., for each (x^a,u) there exists only one $t\in\mathcal{T}$ such that $P_{T|X^a,U}(t|x^a,u)=1$, (ii) is valid since

$$
\begin{aligned}
P^*_{Y|X^a,T,U}(y|x^a,t,u) &\overset{(iii)}{=} \sum_{s\in\mathcal{S}} P^*_{Y|X^a,T,U,S}(y|x^a,t,u,s)P_S(s)\\
&\overset{(iv)}{=} \sum_{s\in\mathcal{S}} P_{Y|X^a,T,S}(y|x^a,t,s)P_S(s)\\
&= \sum_{s\in\mathcal{S}} P^*_{Y,S|X^a,T}(y,s|x^a,t) \;=\; P^*_{Y|X^a,T}(y|x^a,t), \qquad (4.116)
\end{aligned}
$$

where (iii) is valid since S and (X^a,T,U) are independent and (iv) is valid due to (4.3). Similarly, we have

$$
\begin{aligned}
&I(U;Y|X^a)_{P^*_{Y,X^a,U}(y,x^a,u)}\\
&= \sum_{u\in\mathcal{U}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,U}(y,x^a,u)\log\frac{P^*_{Y,U|X^a}(y,u|x^a)}{P^*_{Y|X^a}(y|x^a)P^*_{U|X^a}(u|x^a)}\\
&= \sum_{u\in\mathcal{U}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,U}(y,x^a,u)\log\frac{P^*_{Y,U,X^a}(y,u,x^a)}{P^*_{Y|X^a}(y|x^a)P^*_{U,X^a}(u,x^a)}
\end{aligned}
$$

$$\begin{aligned}
&\overset{(v)}{=} \sum_{t\in\mathcal{T}}\sum_{u\in\mathcal{U}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,U,T}(y,x^a,u,t)\log\frac{P^*_{Y,U,X^a,T}(y,u,x^a,t)}{P^*_{Y|X^a}(y|x^a)P^*_{U,X^a,T}(u,x^a,t)} \\
&\overset{(vi)}{=} \sum_{t\in\mathcal{T}}\sum_{u\in\mathcal{U}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,U,T}(y,x^a,u,t)\log\frac{P^*_{Y|T,X^a}(y|t,x^a)P^*_{U,T,X^a}(u,t,x^a)}{P^*_{Y|X^a}(y|x^a)P^*_{U,T,X^a}(u,t,x^a)} \\
&= \sum_{t\in\mathcal{T}}\sum_{u\in\mathcal{U}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,U,T}(y,x^a,u,t)\log\frac{P^*_{Y,T|X^a}(y,t|x^a)}{P^*_{Y|X^a}(y|x^a)P^*_{T|X^a}(t|x^a)} \\
&= \sum_{t\in\mathcal{T}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,T}(y,x^a,t)\log\frac{P^*_{Y,T|X^a}(y,t|x^a)}{P^*_{Y|X^a}(y|x^a)P^*_{T|X^a}(t|x^a)} \\
&= I(T;Y|X^a)_{P^*_{Y,X^a,T}(y,x^a,t)}, \qquad (4.117)
\end{aligned}$$

where (v) and (vi) follows from the same reasonings of (i) and (ii), respectively. Now, let $R_b^{'} < I(U;Y|X^a)_{P^*_{Y,X^a,U}(y,x^a,u)}$ and $R_b^{'} + R_a^{'} < I(U,X^a;Y)_{P^*_{Y,X^a,U}(y,x^a,u)}$. Hence, $(R_a^{'}, R_b^{'}) \in \mathcal{C}^E_{AS}$. Observe now that for a distribution in the form of $P^*_{Y,X^a,T}(y,x^a,t)$, one can define $\hat{\pi}_{X^a,T}(x^a,t) = P^*_{X^a,T}(x^a,t)$. Therefore, since $\underline{\mathcal{C}}^{dm} = \overline{co}\Big(\bigcup_{\hat{\pi}}\mathcal{R}_C^{'}(\hat{\pi})\Big)$, and due to (4.115) and (4.117), $(R_a^{'}, R_b^{'}) \in \mathcal{C}^G_{FS}$, which completes the claim. □

Proof of Theorem 4.2.7. Follows from Lemmas 4.2.8 - 4.2.10. □

4.3 Examples

We present two examples. In the first example, we discuss the state dependent modulo-additive MAC with noisy CSIT and complete CSIR (as in Section 4.2.1) and show that the proposed inner and outer bounds are tight and yield the capacity region. In the second example we consider the problem defined in Section 4.2.2 where the channel is a binary multiplier MAC whose state is an interference sequence.

4.3.1 Modulo-additive MAC with Noisy CSIT

Recall that the results of Section 4.2.1 are given in terms of Shannon-strategies. Hence, their computation requires an optimization over an extended space of the input alphabet to a space of strategies and is often hard; in fact, very few explicit solutions exist even in the single-user case. In [EZ00], modulo-additive single-user channel with complete CSIT is considered and a closed-form solution for the capacity is derived. Based on this result, we now consider the modulo-additive state-dependent MAC with asymmetric noisy CSIT and show that for the sum-rate capacity, the optimal set of strategies has uniform distribution. This enable us to determine the entire capacity region by observing that under the uniform distribution both inner and outer bounds are tight.

To be more explicit, we consider a two-user state-dependent MAC in which the channel noise, defined by a process $\{Z_t\}_{t=1}^{\infty}$, is correlated with the state process. The channel is given by $Y = X^a \oplus X^b \oplus Z$ where $\mathcal{X}_a = \mathcal{X}_b = \mathcal{Y} = \mathcal{Z} = \{0, \cdots, q-1\}$ and Z, is conditionally independent of (X^a, X^b) given the state S and in the sequel addition (and subtraction) is understood to be performed mod-q. Assume further that we have the setup of Section 4.2.1. The following theorem is the main result of this example and can be though as an extension of [EZ00, Theorem 1] to a noisy multi-user setting.

Theorem 4.3.1. *The capacity region of the modulo-additive state-dependent MAC defined above is given by the closure of the rate pairs* (R_a, R_b) *satisfying*

$$R_a < \log q - H_{\min}$$

$$R_b < \log q - H_{\min}$$

$$R_a + R_b < \log q - H_{\min} \tag{4.118}$$

where $H_{\min} := \min_{t^a,t^b} H(Z + t^a(S^a) + t^b(S^b)|S)$.

Proof. First, recall the rate condition given in Theorem 4.2.2;

$$R_a + R_b \leq H(Y|S) - H(Y|T^a, T^b, S). \tag{4.119}$$

We first determine the optimal distributions of t^a, t^b, the distributions achieving the sum-rate capacity, and then show that these distributions yield the same inner bound. Let us first consider $H(Y|T^a, T^b, S)$. Clearly, $P_{Y|X^a,X^b,S}(y|x^a, x^b, s) = P_{Z|S}(y - x^a - x^b|s)$ and $H(Y|T^a, T^b, S) \geq \min_{t^a,t^b} H(Y|T^a = t^a, T^b = t^b, S)$. Observe that

$$\begin{aligned} P_{Y|T^a,T^b,S}(y|t^a, t^b, s) &= \sum_{s^a,s^b} P_{Y|T^a,T^b,S^a,S^b,S}(y|t^a, t^b, s^a, s^b, s) P_{S^a,S^b|S}(s^a, s^b|s) \\ &= \sum_{s^a,s^b} P_{Z|S}(Z = y - t^a(s^a) - t^b(s^b)|s) P_{S^a,S^b|S}(s^a, s^b|s) \\ &= P_{Z+t^a(S^a)+t^b(S^b)|S}(y|s). \end{aligned} \tag{4.120}$$

where the second step is valid since Z is conditionally independent of (S^a, S^b) given S. Therefore, $H(Y|T^a = t^a, T^b = t^b, S) = H(Z + t^a(S^a) + t^b(S^b)|S)$. Let (t^{a*}, t^{b*}) be two mappings from $\mathcal{S}_a$ to $\mathcal{X}_a$ and $\mathcal{S}_b$ to $\mathcal{X}_b$, respectively, for which $H(Y|T^a = t^{a*}, T^b = t^{b*}, S) = H_{\min}$. Now recall that, by Corollary 4.2.1, we have

$$\begin{aligned} \mathcal{C}_{\Sigma}^{FS} &= \sup_{\pi_{T^a}(t^a)\pi_{T^b}(t^b)} \left[H(Y|S) - H(Y|T^a, T^b, S)\right] \\ &\leq \sup_{\pi_{T^a}(t^a)\pi_{T^b}(t^b)} H(Y|S) - H_{\min}, \end{aligned} \tag{4.121}$$

and we now determine the policies $\{\pi_{T^a}(t^a),\ t^a \in \mathcal{T}_a\}$ and $\{\pi_{T^b}(t^b),\ t^b \in \mathcal{T}_b\}$ achieving the supremum above. Let us first define the following class of strategies

$$\mathcal{T}_a^* := \{t_\tau^a\}, \text{ where } t_\tau^a(s^a) = t^{a*}(s^a) + \tau,\ \ \tau = 1, \cdots, q \tag{4.122}$$

$$\mathcal{T}_b^* := \{t_\tau^b\}, \text{ where } t_\tau^b(s^b) = t^{b*}(s^b) - \tau,\ \ \tau = 1, \cdots, q. \tag{4.123}$$

It should be noted that $H(Y|T^a = t^{a*}, T^b = t^{b*}, S) = H(Y|T^a = t^a_\tau, T^b = t^b_\tau, S)$ since $H(Y|T^a = t^a, T^b = t^b, S) = H(Z + t^a(S^a) + t^b(S^b)|S)$. Note that $H(Y|S) \leq \log|\mathcal{Y}| = \log q$, but if we choose T^a and T^b uniformly distributed within $\mathcal{T}_a^*$ and $\mathcal{T}_b^*$, respectively (with zero mass on strategies not in $\mathcal{T}_a^*$ and $\mathcal{T}_b^*$), we would get

$$\begin{aligned}
P_{Y|S}(y|s) &\overset{(i)}{=} \sum_{s^a,s^b} \sum_{t^a\in\mathcal{T}_a^*} \sum_{t^b\in\mathcal{T}_b^*} P_{Y|T^b,T^b,S^a,S^b,S}(y|t^a,t^b,s^a,s^b,s)\frac{1}{q^2}P_{S^a,S^b|S}(s^a,s^b|s) \\
&= \sum_{s^a,s^b} P_{S^a,S^b|S}(s^a,s^b|s)\frac{1}{q^2} \sum_{t^a\in\mathcal{T}_a^*} \sum_{t^b\in\mathcal{T}_b^*} P_{Z|S}(y - t^a(s^a) - t^b(s^b)|s) \\
&\overset{(ii)}{=} \sum_{s^a,s^b} P_{S^a,S^b|S}(s^a,s^b|s)\frac{1}{q^2} \sum_{t^a\in\mathcal{T}_a^*} 1 \\
&\overset{(iii)}{=} \frac{1}{q}
\end{aligned} \tag{4.124}$$

where (i) valid since T^a and T^b are uniformly distributed, (ii) is due to (4.123) (i.e., follows from the fact that $t^b \in \mathcal{T}_b^*$ traces all possible values of Z) and finally, (iii) is valid since $|\mathcal{T}_a^*| = q$. Therefore, we get that $\mathcal{C}_{FS}^{\sum} = \log q - H_{\min}$ which is achieved by

$$\pi_{T^a}(t^a) = \frac{1}{q},\ \ \forall t^a \in \mathcal{T}_a^*,\ \ \pi_{T^b}(t^b) = \frac{1}{q},\ \ \forall t^b \in \mathcal{T}_b^*. \tag{4.125}$$

Let us now consider the inner bound. In particular, we need to show that the sets of policies in (4.125) give $H(Y|T^a, S) = H(Y|T^b, S) = \log q$. Consider $H(Y|T^a, S)$ and observe that

$$\begin{aligned}
P_{Y|T^a,S}(y|t^a, s) &\overset{(iv)}{=} \sum_{s^a,s^b} \sum_{t^b\in\mathcal{T}_b^*} P_{Y|T^b,T^b,S^a,S^b,S}(y|t^a,t^b,s^a,s^b,s)\frac{1}{q}P_{S^a,S^b|S}(s^a,s^b|s) \\
&= \sum_{s^a,s^b} P_{S^a,S^b|S}(s^a,s^b|s)\frac{1}{q} \sum_{t^b\in\mathcal{T}_b^*} P_{Z|S}(y - t^a(s^a) - t^b(s^b)|s) \\
&\overset{(v)}{=} \sum_{s^a,s^b} P_{S^a,S^b|S}(s^a,s^b|s)\frac{1}{q} \\
&= \frac{1}{q}
\end{aligned} \tag{4.126}$$

where (iv) is valid since T^b is uniformly distributed and (v) is due to (4.123) (i.e., follows from the fact that $t^b \in \mathcal{T}_b^*$ traces all possible values of Z). Thus, $H(Y|T^a, S) = \log q$. It can be shown similarly that under (4.125) $H(Y|T^b, S) = \log q$. □

Finally, it is easy to see that when there is no side information at the encoders and at the decoder the capacity region of modulo-additive state-dependent MAC is given by the closure of rate pairs (R_a, R_b) where

$$\begin{aligned} R_a &\leq \log q - H(Z) \\ R_b &\leq \log q - H(Z) \\ R_a + R_b &\leq \log q - H(Z). \end{aligned} \tag{4.127}$$

Observe that we have

$$\begin{aligned} H(Z + t^a(S^a) + t^b(S^b)|S) &\leq H(Z|S) + H(t^a(S^a) + t^b(S^b)|S) \\ H_{\min} = \min_{t^a, t^b} H(Z + t^a(S^a) + t^b(S^b)|S) &\leq \min_{t^a, t^b} \left[H(Z|S) + H(t^a(S^a) + t^b(S^b)|S)\right] \\ &\overset{(vi)}{=} H(Z|S) \\ &\overset{(vii)}{<} H(Z) \end{aligned}$$

where (vi) can be achieved with any deterministic mapping and (vii) is valid since Z and S (and hence S) are correlated. Therefore, availability of state information strictly increases, by an amount of at least $I(S; Z)$, the capacity region of the modulo-additive state-dependent MAC.

4.3.2 Binary MAC with Interference

Consider the binary multiplier MAC with state process interfering the output, namely $Y = X^a X^b \oplus S$ where $\mathcal{X}_a = \mathcal{X}_b = \mathcal{Y} = \mathcal{S} = \{0, 1\}$. Assume further that the communication setup is given as in Section 4.2.2 with $S^r = S \oplus Z^r$ where $Z^r \sim$

Ber(p_r) is Bernoulli with $P(Z^r = 1) = p_r$. We now show that the capacity region, with both causal and non-causal coding, of this channel is given by the closure of (R_a, R_b) where $R_a < 1 - H(S|S^r)$, $R_b < 1 - H(S|S^r)$ and $R_a + R_b < 1 - H(S|S^r)$.

First recall the capacity region given in Theorem 4.2.3 and observe that

$$H(Y|S^r, X^a, X^b) = H(X^a X^b \oplus S|S^r, X^a, X^b) = H(S|S^r, X^a, X^b) = H(S|S^r).$$

Hence, input distributions do not effect $H(Y|S^r, X^a, X^b)$. Clearly, $H(Y|S^r) \leq 1$, $H(Y|S^r, X^a) \leq 1$ and $H(Y|S^r, X^b) \leq 1$ and we now show that equalities can be achieved. More explicitly, we have the following optimizing distributions which can be obtained using standard inequalities

$$\begin{aligned} \underset{\pi_{X^l|S^l}(x^l|f(s^r)),\ l=a,b}{\text{argmax}} H(Y|S^r) &= \big\{\pi_{X^a|S^a}(0|f^a(0)) = \pi_{X^a|S^a}(0|f^a(1)) = 0.5, \\ &\qquad \pi_{X^b|S^b}(0|f^b(0)) = \pi_{X^b|S^b}(0|f^b(1)) = 0.5\big\} \quad (4.128) \\ \underset{\pi_{X^l|S^l}(x^l|f(s^r)),\ l=a,b}{\text{argmax}} H(Y|S^r, X^a) &= \big\{\pi_{X^a|S^a}(0|f^a(0)) = \pi_{X^a|S^a}(0|f^a(1)) = 0, \\ &\qquad \pi_{X^b|S^b}(0|f^b(0)) = \pi_{X^b|S^b}(0|f^b(1)) = 0.5\big\} \quad (4.129) \\ \underset{\pi_{X^l|S^l}(x^l|f(s^r)),\ l=a,b}{\text{argmax}} H(Y|S^r, X^b) &= \big\{\pi_{X^b|S^b}(0|f^b(0)) = \pi_{X^b|S^b}(0|f^b(1)) = 0, \\ &\qquad \pi_{X^a|S^a}(0|f^b(0)) = \pi_{X^b|S^b}(0|f^b(1)) = 0.5\big\} \quad (4.130) \end{aligned}$$

and in the rest, let us show that these yield the equalities in the conditional entropies. Let us start with R_a, i.e., $H(Y|S^r, X^b)$. Note that

$$H(Y|S^r, X^b) = \sum_{s^r} \sum_{x^b} P_{S^r}(s^r)\pi_{X^b|S^b}(x^b|f^b(s^r))H(Y|S^r = s^r, X^b = x^b). \quad (4.131)$$

Substituting (4.130) in (4.131) gives

$$H(Y|S^r, X^b) =$$

$$P_{S^r}(0)H(X^a \oplus S|X^b = 1, S^r = 0) + P_{S^r}(1)H(X^a \oplus S|X^b = 1, S^r = 1).$$

We next show that under (4.130) $H(X^a \oplus S|X^b = 1, S^r = 0) = 1$, for which it is enough to show that $P_{X^a \oplus S|X^b,S^r}(0|1,0) = 0.5$. We have

$$
\begin{aligned}
& P_{X^a \oplus S|X^b,S^r}(0|1,0) \\
&= \sum_{s\in\{0,1\}} \sum_{x^a\in\{0,1\}} P_{X^a\oplus S|S,X^a,X^b,S^r}(0|s,x^a,1,0) P_{S|S^r}(s|0)\pi_{X^a|S^a}(x^a|f^a(0)) \quad (4.132) \\
&= P_{S|S^r}(0|1)\left[0.5P_{X^a\oplus S|S,X^a,X^b,S^r}(0|0,0,1,0) + 0.5P_{X^a\oplus S|S,X^a,X^b,S^r}(0|0,1,1,0)\right] \\
&\quad + P_{S|S^r}(1|1)\left[0.5P_{X^a\oplus S|S,X^a,X^b,S^r}(0|1,0,1,0) + 0.5P_{X^a\oplus S|S,X^a,X^b,S^r}(0|1,1,1,0)\right] \\
&= 0.5
\end{aligned}
$$

where (4.132) is due to the Markov condition $S \to S^r \to (X^a, X^b)$ and (4.36). We can similarly show that $P_{X^a\oplus S|X^b,S^r}(0|1,1) = 0.5$ and hence, $H(X^a \oplus S|X^b = 1, S^r = 1) = 1$. Therefore, $H(Y|S^r, X^b) = 1$. Since the above derivation is symmetric, under (4.129) $H(Y|X^a, S^r) = 1$.

It now remains to show that with (4.128), $H(Y|S^r)$ is equal to one. It should be observed that

$$
\begin{aligned}
& P_{X^aX^b\oplus S|S^r}(\cdot|s^r) \\
&\overset{(i)}{=} \sum_{x^a,x^b,s} P_{X^aX^b\oplus S|X^a,X^b,S}(\cdot|x^a,x^b,s)\pi_{X^a|S^a}(x^a|f^a(s^r))\pi_{X^b|S^b}(x^b|f^b(s^r))P_{S|S^r}(s|s^r) \\
&\overset{(ii)}{=} 0.25 \sum_{s\in\{0,1\}} P_{S|S^r}(s|s^r) \sum_{x^a,x^b\{0,1\}} P_{X^aX^b\oplus S|X^a,X^b,S}(\cdot|x^a,x^b,s) \\
&= 0.5
\end{aligned}
$$

where (i) is due to is due to the Markov condition $S \to S^r \to (X^a, X^b)$ and (4.36), (ii) is due to (4.128) and the last step is valid since for given s, there are only two pairs of (x^a, x^b) for which $P_{X^aX^b\oplus S|X^a,X^b,S}(\cdot|x^a,x^b,s) = 1$ (and zero for the other two). Hence, $H(Y|S^r) = 1$.

Finally, it can be easily shown that the capacity region of $Y = X^aX^b \oplus S$ without

CSIT and CSIR is given by the closure of (R_a, R_b) where $R_a < 1-H(S)$, $R_b < 1-H(S)$ and $R_a + R_b < 1 - H(S)$. Therefore, availability of noisy CSI at the encoders (both causal and non-causal) and at the decoder increases the capacity region by an amount of $I(S; S^r)$.

4.4 Conclusion and Remarks

We have considered several scenarios for the memoryless state-dependent MAC with an i.i.d. state process, asymmetric noisy CSI at the encoders and complete and noisy CSI at the receiver. When the encoders have access to causal noisy CSI, single-letter inner and outer bounds, which are tight for the sum-rate capacity, are obtained. In order to reduce the space of optimization, from Shannon strategies to channel inputs, we consider the case where CSITs are asymmetric deterministic functions of noisy CSIR. The causal setup of this problem is considered in [CY11] and a single-letter characterization for capacity region is provided. Hence, we considered the non-causal setup and showed that the causal and non-causal capacity regions are identical.

When the decoder does not need to access the current CSI at the encoder, which matches with the delayed scenario, we observe that a single-letter characterization of the capacity region can be obtained. We further discuss a degraded message set scenario and show that when the common message encoder does not have an access to the current noisy CSI, due to delay, it is possible to obtain a single-letter expression for the capacity region. Since a product form is not required in this case, we observed that as long as the common message encoder does not have access to CSI, then in any noisy setup (the cases where no CSIR or noisy CSIR) it is possible to obtain the capacity region.

Chapter 5

Multiple Access Channel without Receiver Side Information

In this chapter, we generalize the sum-rate capacity result presented in the previous section. In particular, it is shown that when the state processes are asymmetric, Shannon strategies are optimal if the decoder is provided with some information which makes the CSITs conditionally independent. With this result at hand, the next step is to investigate what the minimum rate required to transmit such information to the receiver is when there is no CSIR. By using the lossless CEO approach [GP79] and adopting the recent proof technique of [LS13b, Theorem 1], we characterize the rate required to transmit this information to the receiver. Therefore, we demonstrate how far the Shannon strategies are away from the optimality when there is no CSIR. Recall once again that when there is no CSIR, Shannon strategies are suboptimal; see [LS13b] for a particular example.

The rest of the chapter is organized as follows. In Section 5.1 we formally state the problem, in Section 5.2 we present the main result on the sum-rate capacity and give the converse proof sketch. In Section 5.3 we consider the case where there is no CSIR and provide an inner bound to the capacity region and in Section 5.4 we present concluding remarks.

5.1 Problem Setup

Consider a two-user memoryless state dependent MAC, with two encoders, a, b, and two independent message sources W_a and W_b which are uniformly distributed in the finite sets $\mathcal{W}_a$ and $\mathcal{W}_b$, respectively. The channel inputs from the encoders are $X^a \in \mathcal{X}_a$ and $X^b \in \mathcal{X}_b$, respectively, and the channel output is $Y \in \mathcal{Y}$. The channel state process is modeled as a sequence $\{S_t\}_{t=1}^{\infty}$ of i.i.d. random variables in some finite space $\mathcal{S}$. The two encoders have access to causal possibly correlated versions of the state information S_t at each time $t \geq 1$, modeled by $S_t^a \in \mathcal{S}_a$, $S_t^b \in \mathcal{S}_b$, respectively. We also assume that $\{(S_t, S_t^a, S_t^b)\}_{t=1}^{\infty}$ is a sequence of i.i.d. triples and independent from (W_a, W_b). Therefore, we have that for any $n \geq 1$ (4.2) is satisfied. The channel inputs at time t, i.e., X_t^a and X_t^b, are functions of $(W_a, S_{[t]}^a)$ and $(W_b, S_{[t]}^b)$, respectively. Let $\mathbf{W} := (W_a, W_b)$ and $\mathbf{X_t} := (X_t^a, X_t^b)$, respectively. Then, the laws governing n-sequences of state, input and output letters are given by

$$P_{Y_{[n]}|\mathbf{W},\mathbf{X}_{[n]},S_{[n]},S_{[n]}^a,S_{[n]}^b}(y_{[n]}|\mathbf{w},\mathbf{x}_{[n]},s_{[n]},s_{[n]}^a,s_{[n]}^b) = \prod_{t=1}^{n} P_{Y_t|X_t^a,X_t^b,S_t}(y_t|x_t^a,x_t^b,s_t), \quad (5.1)$$

where the channel's transition probability distribution, $P_{Y_t|X_t^a,X_t^b,S_t}(y_t|x_t^a,x_t^b,s_t)$, is given a priori.

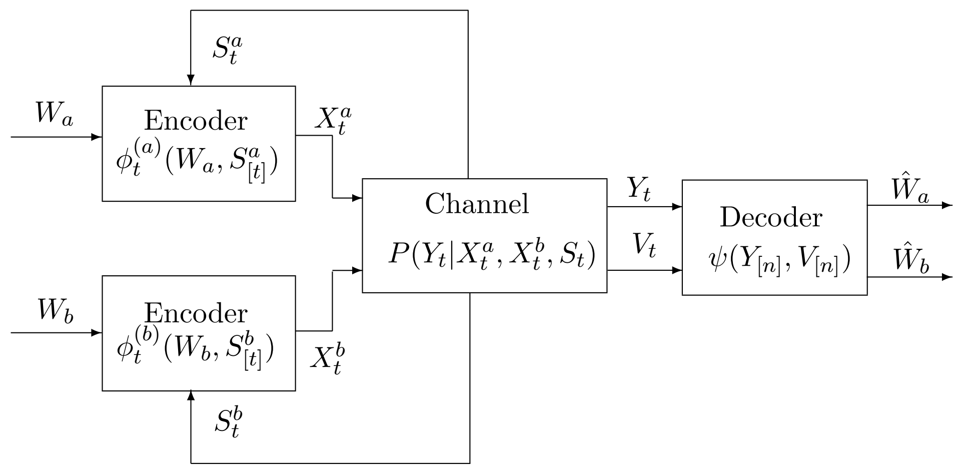

Figure 5.1: The multiple-access channel with asymmetric causal noisy state feedback.

5.2 Common Information at the Receiver

We first determine the sum-rate capacity when some information is available at the receiver. Let $\{(V_t)\}_{t=1}^{\infty}$, $V_t \in \mathcal{V}$, be an external sequence of random variables that the decoder observes (see Fig. 5.1) and satisfies the following

$$P_{V_{[n]},W_a,W_b}(v_{[n]}, w_a, w_b) = \frac{1}{|\mathcal{W}_a|}\frac{1}{|\mathcal{W}_b|}\prod_{i=1}^{n} P_{V_i}(v_i) \tag{5.2}$$

$$P_{S_{[t]},S_{[t]}^a,S_{[t]}^b,V_{[t]}}(s_{[t]}, s_{[t]}^a, s_{[t]}^b, v_{[t]}) = P(s_{[t]}^a|v_{[t]})P(s_{[t]}^b|v_{[t]})P(s_{[t]}, v_{[t]}), \quad 1 \le t \le n. \tag{5.3}$$

Remark 5.2.1. *Note that when $\{S_t^a\}_{t=1}^{\infty}$ and $\{S_t^b\}_{t=1}^{\infty}$ are independent the process $\{V_t\}_{t=1}^{\infty}$ can be taken as a deterministic (or null) sequence. A more general example is as follows: Let $\{Z_t\}_{t=1}^{\infty}$, $\{Z_t^a\}_{t=1}^{\infty}$ and $\{Z_t^b\}_{t=1}^{\infty}$ be three noise processes which are i.i.d. and independent of each other and where $\{Z_t\}_{t=1}^{\infty}$ is independent of $\{S_t\}_{t=1}^{\infty}$ and $\{Z_t^a\}_{t=1}^{\infty}$ and $\{Z_t^b\}_{t=1}^{\infty}$ are independent of $\{V_t\}_{t=1}^{\infty}$. Let $V_t = S_t + Z_t$, $S_t^a = V_t + Z_t^a$, $S_t^b = V_t + Z_t^b$. In this case, equation (5.3) holds and V_t is only a noisy version of S_t. These two examples demonstrate that the availability of the process $\{V_t\}_{t=1}^{\infty}$ at the receiver is in general a less restrictive scenario than the availability of complete CSI* *at the receiver.*

Definition 5.2.1. *An $(n, 2^{nR_a}, 2^{nR_b})$ code with block length n and rate pair (R_a, R_b) for the state dependent MAC with causal noisy state information consists of*

(1) A sequence of mappings for each encoder

$$\phi_t^{(a)} : \mathcal{S}_a^t \times \mathcal{W}_a \to \mathcal{X}_a, \ \ t = 1, 2, ...n;$$

$$\phi_t^{(b)} : \mathcal{S}_b^t \times \mathcal{W}_b \to \mathcal{X}_b, \ \ t = 1, 2, ...n.$$

2) An associated decoding function

$$\psi : \mathcal{V}^n \times \mathcal{Y}^n \to \mathcal{W}_a \times \mathcal{W}_b.$$

The system's probability of error, $P_e^{(n)}$, an achievable rate pair, the capacity region and the sum-rate capacity is defined in a similar manner of Chapter 4. Let $\mathcal{C}$ and $\mathcal{C}^{\Sigma}$ denote the capacity region and the sum-rate capacity, respectively. Let, as before, the set of all possible functions from $\mathcal{S}_a$ to $\mathcal{X}_a$ and $\mathcal{S}_b$ to $\mathcal{X}_b$ be denoted by $\mathcal{T}_a := {\mathcal{X}_a}^{|\mathcal{S}_a|}$ and $\mathcal{T}_b := {\mathcal{X}_b}^{|\mathcal{S}_b|}$, respectively. Recall also that $\mathcal{T}_a$-valued and $\mathcal{T}_b$-valued random vectors are called as Shannon strategies.

Definition 5.2.2. *A memoryless stationary (in time) team policy is a family*

$$\Pi = \{\pi = (\pi_{T^a}(\cdot), \pi_{T^b}(\cdot)) \in \mathcal{P}(\mathcal{T}_a) \times \mathcal{P}(\mathcal{T}_b)\} \tag{5.4}$$

of probability distribution pairs on $(\mathcal{T}_a, \mathcal{T}_b)$.

For every memoryless stationary team policy π, let $\mathcal{R}(\pi)$ denote the region of all rate pairs $R = (R_a, R_b)$ satisfying

$$R_a \ < \ I(T^a; Y | T^b, V) \tag{5.5}$$

$$R_b < I(T^b; Y|T^a, V) \tag{5.6}$$

$$R_a + R_b < I(T^a, T^b; Y|V) \tag{5.7}$$

where V, T^a, T^b and Y are random variables taking values in $\mathcal{V}$, $\mathcal{T}_a$, $\mathcal{T}_b$ and $\mathcal{Y}$, respectively, and whose joint probability distribution factorizes as

$$P_{V,T^a,T^b,Y}(v, t^a, t^b, y) = P_V(v) P_{Y|T^a,T^b,V}(y|t^a, t^b, v)\pi_{T^a}(t^a)\pi_{T^b}(t^b). \tag{5.8}$$

Let $\mathcal{C}^{in} := \overline{co}\left(\bigcup_\pi \mathcal{R}(\pi)\right)$ denote the closure of the convex hull of the rate regions $\mathcal{R}(\pi)$ given by (5.5)-(5.7) associated to all possible memoryless stationary team polices as defined in (5.4).

Theorem 5.2.1 (Inner Bound to $\mathcal{C}$)**.** $\mathcal{C}^{in} \subseteq \mathcal{C}$.

Note that $\{Y_i, V_i, T_i^a, T_i^b\}_{i=1}^\infty$ is an independent sequence and hence, the proof is identical to Theorem 4.2.1. Let

$$\mathcal{C}^o := \left\{(R_a, R_b) \in \mathbb{R}^+ \times \mathbb{R}^+ : R_a + R_b \leq \sup_{\pi_{T^a}(t^a)\pi_{T^b}(t^b)} I(T^a, T^b; Y|V)\right\},$$

where $\mathbb{R}^+$ is the set of positive reals.

Theorem 5.2.2 (Outer Bound to $\mathcal{C}$)**.** $\mathcal{C} \subseteq \mathcal{C}^o$.

Proof of Theorem 5.2.2. As the proof is similar to the proof of Theorem 4.2.2, we herein provide the sketch. We need to show that all achievable rates satisfy

$$R_a + R_b \leq \sup_{\pi_{T^a}(t^a)\pi_{T^b}(t^b)} I(T^a, T^b; Y|V),$$

i.e., a converse for the sum-rate capacity. Let $\alpha_\mu := \frac{1}{n} P_{V_{[t-1]}}(\mu)$ and $\eta(\epsilon) := \frac{\epsilon}{1-\epsilon}\log|\mathcal{Y}| + \frac{H(\epsilon)}{1-\epsilon}$. Observe that $\lim_{\epsilon\to 0}\eta(\epsilon) = 0$ and $\sum_{\mu\in\mathcal{V}^{(n)}} \alpha_\mu = 1$ where $\mathcal{V}^{(n)}$ is the sets of all $\mathcal{V}$-strings of length less than n. Recall further that, for all $t \geq 1$, $X_t^a = \phi_t^{(a)}\left(W_a, S_{[t]}^a\right) = \phi_t^{(a)}\left(W_a, S_{[t-1]}^a, S_t^a\right)$ and $X_t^b = \phi_t^{(b)}\left(W_b, S_{[t]}^b\right) = \phi_t^{(b)}\left(W_b, S_{[t-1]}^b, S_t^b\right)$. Then, we can define the Shannon strategies $T_t^a \in \mathcal{T}_a$ and $T_t^b \in \mathcal{T}_b$ by putting, for every $s_a \in \mathcal{S}_a$ and $s_b \in \mathcal{S}_b$,

$$\begin{aligned} T_t^a(s_a) &:= \phi_t^{(a)}\left(W_a, S_{[t-1]}^a, s_a\right) \\ T_t^b(s_b) &:= \phi_t^{(b)}\left(W_b, S_{[t-1]}^b, s_b\right). \end{aligned} \tag{5.9}$$

Lemma 5.2.1. *Let $T_t^a \in \mathcal{T}_a$ and $T_t^b \in \mathcal{T}_b$ be the Shannon strategies induced by $\phi_t^{(a)}$ and $\phi_t^{(b)}$, respectively, as shown in (5.9). Assume that a rate pair $R = (R_a, R_b)$, with block length $n \geq 1$ and a constant $\epsilon \in (0, 1/2)$, is achievable. Then,*

$$R_a + R_b \leq \sum_{\mu \in \mathcal{V}^{(n)}} \alpha_\mu I(T_t^a, T_t^b; Y_t | V_t, V_{[t-1]} = \mu) + \eta(\epsilon). \tag{5.10}$$

The main idea in the converse proof that we provide in this section is to show that there is no loss of optimality if we ignore the past CSI at the encoders given that the decoder is provided with the process $\{V_t\}$. The following steps show that memoryless stationary team policies are as good as any policy that the encoders can apply. To show this, observe that, for any $t \geq 1$, $I(T_t^a, T_t^b; Y_t | V_t, V_{[t-1]} = \mu)$ is a function of the joint conditional distribution of V_t, inputs T_t^a, T_t^b and output Y_t given the past realization $(V_{[t-1]} = \mu)$. Hence, to complete the proof of the outer bound, we need to show that $P_{T_t^a, T_t^b, Y_t, V_t | V_{[t-1]}}(t^a, t^b, y, v | \mu)$ factorizes as in (5.8). This is done in the lemma below. In particular, it is crucial to observe that having $V_{[t-1]}$ at the decoder is enough to provide a product form on T^a and T^b. Let

$$\begin{aligned} \Upsilon_{\mu_{\mathbf{a}}}^a(t^a) &:= \{w_a : \phi_t^{(a)}(w_a, s_{[t-1]}^a = \mu_{\mathbf{a}}) = t^a\} \\ \Upsilon_{\mu_{\mathbf{b}}}^b(t^b) &:= \{w_b : \phi_t^{(b)}(w_b, s_{[t-1]}^b = \mu_{\mathbf{b}}) = t^b\} \\ \pi_{T^a}^{\mu_{\mathbf{a}}}(t^a) &:= \sum_{w_a \in \Upsilon_{\mu_{\mathbf{a}}}^a(t^a)} \frac{1}{|\mathcal{W}_a|} \\ \pi_{T^b}^{\mu_{\mathbf{b}}}(t^b) &:= \sum_{w_b \in \Upsilon_{\mu_{\mathbf{b}}}^b(t^b)} \frac{1}{|\mathcal{W}_b|} \end{aligned} \tag{5.11}$$

$$\begin{aligned}\pi^{\mu}_{T^a}(t^a) &:= \sum_{\mu_{\mathbf{a}}} \pi^{\mu_{\mathbf{a}}}_{T^a}(t^a) P_{S^a_{[t-1]}|V_{[t-1]}}(\mu_{\mathbf{a}}|\mu) \\ \pi^{\mu}_{T^b}(t^b) &:= \sum_{\mu_{\mathbf{b}}} \pi^{\mu_{\mathbf{b}}}_{T^b}(t^b) P_{S^b_{[t-1]}|V_{[t-1]}}(\mu_{\mathbf{b}}|\mu),\end{aligned} \tag{5.12}$$

where $\mu_{\mathbf{a}}$ and $\mu_{\mathbf{b}}$ denote particular realizations of $S^a_{[t-1]}$ and $S^b_{[t-1]}$, respectively.

Lemma 5.2.2. *For every $1 \le t \le n$ and $\mu \in \mathcal{V}^{t-1}$, the following holds*

$$P_{T^a_t,T^b_t,Y_t,V_t|V_{[t-1]}}(t^a,t^b,y,v|\mu) = P_V(v)P_{Y|V,T^a,T^b}(y|v,t^a,t^b)\pi^{\mu}_{T^a}(t^a)\pi^{\mu}_{T^b}(t^b). \tag{5.13}$$

Proof follows from (4.2), (5.3), (5.9), (5.11) and (5.12).

We now complete the proof of Theorem 5.2.2. With Lemma 5.2.1 it is shown that the sum of any achievable rate pair can be approximated by the convex combinations of rate conditions given in (5.7) which are indexed by $\mu \in \mathcal{V}^{(n)}$ and satisfy (5.8). More explicitly, we have

$$\begin{aligned}R_a + R_b &\le \sum_{\mu \in \mathcal{V}^{(n)}} \alpha_\mu I(T^a_t, T^b_t; Y_t | V_t, V_{[t-1]} = \mu) + \eta(\epsilon) \\ &= \sum_{\mu \in \mathcal{V}^{(n)}} \alpha_\mu I(T^a_t, T^b_t; Y_t | V_t)_{\pi^{\mu}_{T^a}(t^a)\pi^{\mu}_{T^b}(t^b)} + \eta(\epsilon) \\ &\le \sup_{\left(\pi_{T^a}(t^a)\pi_{T^b}(t^b)\in\Pi\right)} I(T^a_t, T^b_t; Y_t|V_t) + \eta(\epsilon).\end{aligned}$$

Hence, since $\lim_{\epsilon\to 0}\eta(\epsilon) = 0$, any achievable pair satisfies

$$R_a + R_b \le \sup_{\pi_{T^a}(t^a)\pi_{T^b}(t^b)} I(T^a, T^b; Y|V).$$

□

As a consequence of Theorems 5.2.1 and 5.2.2, we have the following theorem.

Theorem 5.2.3.

$$\mathcal{C}^{\Sigma} = \sup_{\pi_{T^a}(t^a)\pi_{T^b}(t^b)} I(T^a, T^b; Y|V). \tag{5.14}$$

Proof of Theorem 5.2.3. We need to show that $\exists\ (R_a, R_b) \in \mathcal{C}^{in}$ achieving (5.14). Let us fix $\pi_{T^a}(t^a)\pi_{T^b}(t^b)$ and consider the rate constraints given in $\mathcal{C}^{in}$

$$I(T^a;Y|T^b,V) = H(T^a) - H(T^a|T^b,Y,V) \tag{5.15}$$

$$I(T^b;Y|T^a,V) = H(T^b) - H(T^b|T^a,Y,V) \tag{5.16}$$

and

$$I(T^a,T^b;Y|V) = H(T^a) + H(T^b) - H(T^a|T^b,Y,V) - H(T^b|Y,V), \tag{5.17}$$

where (5.15), (5.16) and (5.17) are valid since T^a and T^b are independent of each other and independent of V. Observe now that for any $\pi_{T^a}(t^a)\pi_{T^b}(t^b)$, $I(T^a;Y|T^b,V) + I(T^b;Y|T^a,V) \geq I(T^a,T^b;Y|S)$ since $H(T^b|Y,V) \geq H(T^b|T^a,Y,V)$. Therefore, the sum-rate constraint in $\mathcal{C}^{in}$ is always active and hence, there exists $(R_a, R_b) \in \mathcal{C}^{in}$ achieving (5.14). □

Remark 5.2.2. *Let $S_a = (S_p^a, S_c)$ and $S_b = (S_p^b, S_c)$, where S_p^a an S_p^b are independent. This scenario is considered in common information problems, e.g. [Wyn75]), and for the optimality of Shannon strategies for this setup, it is sufficient to take $V = S_c$. We also remark that the common information between the given set of random variables is a well studied subject in information theory and there are several notions of common information which are characterized by defining different measures on common randomness. The interested reader is referred to [EGK11, Section 14.2.2] for a more detailed discussion.*

Remark 5.2.3. *Let*

$$\begin{aligned}\mathcal{O} &:= \inf_{\{V:P_{S^a,S^b,V,S}(s^a,s^b,v,s)=P_{S^a|V}(s^a|v)P_{S^b|V}(s^b|v)P_{V,S}(v,s)\}} \sup_{\pi_{T^a}(t^a)\pi_{T^b}(t^b)} I(T^a,T^b;Y|V)\\ \mathcal{C}^{out} &:= \big\{(R_a,R_b) \in \mathbb{R}^+ \times \mathbb{R}^+ : R_a + R_b \leq \mathcal{O}\big\}. \end{aligned} \tag{5.18}$$

Due to Theorem 5.2.3, we have

$$\mathcal{C} \subseteq \mathcal{C}^{out}. \tag{5.19}$$

5.3 An Achievable Region without CSIR

So far, we have shown that for the sum-rate capacity, when CSITs are correlated, the Shannon strategies are optimal provided that the decoder is provided with a random variable under which the CSITs are conditionally independent. It is now natural to consider the situation when there is no CSIR and try to convey this random variable to the receiver in order to examine how far Shannon strategies are away from optimality. More explicitly, consider the process $\{(S_t^a, S_t^b, S_t)\}$ and let us assume there exists a process $\{V_t\}$ such that

$$P_{S_t^a, S_t^b, S_t, V_t}(s_t^a, s_t^b, s_t, v_t) = P_{S_t^a|V_t}(s_t^a|v_t) P_{S_t^b|V_t}(s_t^b|v_t) P_{S_t, V_t}(s_t, v_t). \tag{5.20}$$

The question of interest is if S_t^a and S_t^b are provided to the encoders such as in the previous section and the decoder has no CSI, what is the rate required to convey V_t to the receiver.

It should first be recalled that a similar problem appears in the source coding setting, which is known as the lossless CEO problem, and the tight characterization of the achievable rates is known under the Markov condition (5.20) [GP79, Theorem 1,2]; it can be summarized as follows. Let $\{(S_t^a, S_t^b, V_t)\}_{t=1}^n$ denote i.i.d. copies of a triplet of random variables such that

$$P_{S_t^a, S_t^b, V_t}(s^a, s^b, v) = P_{S_t^a|V_t}(s^a|v) P_{S_t^b|V_t}(s^b|v) P_{V_t}(v).$$

Let there be two encoders-a,b which observe S^a, S^b, respectively and the decoder is interested in a lossless reconstruction of V. Let $\mathcal{R}^{ceo}$ be the region of all achievable rate pairs. Let $(\tilde{S}^a, \tilde{S}^b)$, taking values in finite sets, be a pair of random variables with

$$P_{S^a,S^b,V,\tilde{S}^a,\tilde{S}^b}(s^a, s^b, v, \tilde{s}^a, \tilde{s}^b) = P_{\tilde{S}^a|S^a}(\tilde{s}^a|s^a)P_{\tilde{S}^b|S^b}(\tilde{s}^b|s^b)P_{S^a,S^b,V}(s^a, s^b, v). \quad (5.21)$$

Let $\mathcal{P}(\lambda)$ be the set of all random variables $\mathcal{A} = (S^a, S^b, V, \tilde{S}^a, \tilde{S}^b)$ satisfying (5.21) and $H(V|\tilde{S}^a, \tilde{S}^b) \leq \lambda$. For $\mathcal{A} \in \mathcal{P}(\lambda)$, let $\mathcal{R}^{gp}(\mathcal{A})$ be the region of rate pairs (R_a, R_b) where

$$R_a > I(S^a; \tilde{S}^a|\tilde{S}^b) \quad (5.22)$$

$$R_b > I(S^b; \tilde{S}^b|\tilde{S}^a) \quad (5.23)$$

$$R_a + R_b > I(S^a, S^b; \tilde{S}^a, \tilde{S}^b). \quad (5.24)$$

Theorem 5.3.1. *[GP79, Theorem 1,2]* $\mathcal{R}^{ceo} = \bar{co}\left(\bigcup_{\mathcal{A}\in\mathcal{P}(0)} \mathcal{R}^{gp}(\mathcal{A})\right)$ *where* $\mathcal{P}(0)$ *consists of* $\mathcal{A}$ *satisfying (5.21) with* $H(V|\tilde{S}^a, \tilde{S}^b) = 0$.

Note that $\mathcal{R}^{ceo}$ is non-empty if and only if $H(V|\tilde{S}^a, \tilde{S}^b) = 0$ is satisfied.

In [LS13b] state-dependent MAC with two independent states each known to one encoder causally and strictly causally is considered. The authors propose a novel coding scheme which combines distributed Wyner-Ziv source coding with side information and channel coding via a block Markov scheme with backward decoding. This inner bound is attained by using two compression indexes one for each of the state information at one encoder and sending both of them to the receiver. In order to convey V to the receiver, we use the technique in this inner bound and the rate region of the lossless CEO approach and obtain an inner bound.

Before we present this bound, note that only the decoding function in Definition 5.2.1 is changed to $\psi : \mathcal{Y}^n \to \mathcal{W}_a \times \mathcal{W}_b$ while the other definitions such as error probability, achievable rate pair and capacity region are unchanged. Let $\mathcal{C}^{ns}$ denote the capacity region for this scenario.

Let $\mathbf{V} := (V, V^a, V^b)$, $\mathbf{S} := (S, S^a, S^b)$, $\mathbf{X} := (X^a, X^b)$ and $\mathbf{v}$, $\mathbf{s}$, $\mathbf{x}$ denote their particular realizations, respectively. Let $\mathcal{P}^{cr}(\lambda)$ be the collection of all random variables

$$(V, V^a, V^b, T^a, T^b, Y, S, S^a, S^b, X^a, X^b), \text{ where } V \in \mathcal{V}, \;\; T^i \in \mathcal{T}_i, \;\; V^i \in \mathcal{V}_i, \;\; i = a, b$$

whose joint distribution can be written as

$$P_{\mathbf{V},T^a,T^b,\mathbf{S},\mathbf{X}}(\mathbf{v}, t^a, t^b, y, \mathbf{s}, \mathbf{x}) = P_{Y|\mathbf{X},S}(y|\mathbf{x}, s) 1_{\{x^a = t^a(s^a)\}} 1_{\{x^b = t^b(s^b)\}} P_{T^a}(t^a) P_{T^b}(t^b)$$
$$P_{V^a|S^a}(v^a|s^a) P_{V^b|S^b}(v^b|s^b) P_{S^a|V}(s^a|v) P_{S^b|V}(s^b|v) P_{V,S}(v, s) \qquad (5.25)$$

and satisfy $H(V|V^a, V^b) \leq \lambda$. Let $\mathcal{R}^{cr}$ be the convex hull of the collection of all (R_a, R_b) satisfying

$$R_a \leq I(T^a; Y|T^b, V^a, V^b, V) - I(V^a; S^a|V^b, Y) \qquad (5.26)$$
$$R_b \leq I(T^b; Y|T^a, V^a, V^b, V) - I(V^b; S^b|V^a, Y) \qquad (5.27)$$
$$R_a + R_b \leq I(T^a, T^b; Y|V^a, V^b, V) - I(V^a, V^b; S^a, S^b|Y) \qquad (5.28)$$

for some $(V, V^a, V^b, T^a, T^b, Y, S, S^a, S^b, X^a, X^b) \in \mathcal{P}^{cr}(0)$.

Theorem 5.3.2. $\mathcal{R}^{cr} \in \mathcal{C}^{ns}$.

Remark 5.3.1. *To the best of our knowledge, Theorem 5.2.3 covers the results in the literature for the sum-rate capacity of state dependent MACs. Note that when S_a and S_b are independent, we can simply take (V_a, V_b) deterministic and hence, the minus terms in (5.28) are zero. Now, the supremum of the sum-rate constraint gives the sum-rate capacity.*

The proof of Theorem 5.3.2 is an extension of the idea presented in [LS13b] and it introduces the random variable V, following the lossless CEO approach, which is a function of the auxiliary random variables (V^a, V^b). To summarize the idea, note that the minus terms in (5.26)- (5.28) correspond to the rates required to transmit (V^a, V^b) in a distributed Wyner-Ziv network [Gas04]. In [LS13b], S^a and S^b are independent and therefore so are V^a and V^b. However, it should be noted that the rate region of the distributed Wyner-Ziv network is obtained when the pair (S^a, S^b) is correlated. Therefore, since the Markov condition (5.25) implies that

$$I(V^a, V^b; S^a, S^b|Y) = \\ I(V^a; S^a) + I(V^b; S^b) + I(V^a; V^b|Y) - I(V^b; V^a, Y) - I(V^a; V^b, Y)$$

we can write (5.26)-(5.28) as

$$\begin{aligned}
R_a &\leq I(T^a; Y|T^b, V^a, V^b) - R_{a'} \\
R_b &\leq I(T^b; Y|T^a, V^a, V^b) - R_{b'} \\
R_a + R_b &\leq I(T^a, T^b; Y|V^a, V^b) - R_{a'} - R_{b'} \\
R_{a'} &\geq I(V^a; S^a) - I(V^a; V^b, Y) \\
R_{b'} &\geq I(V^b; S^b) - I(V^b; V^a, Y) \\
R_{a'} + R_{b'} &\geq I(V^a; S^a) + I(V^b; S^b) + I(V^a; V^b|Y) \\
&\quad -I(V^b; V^a, Y) - I(V^a; V^b, Y) \qquad (5.29)
\end{aligned}$$

where the last three equations correspond to the rate pairs given in [Gas04, Theorem 2] for distributed coding of correlated sources.

Throughout the proof, we keep the notation of [LS13b] and typical sequences and typical sets are used as defined in [CK81] (see Definition 2.2.3).

Before we prove Theorem 5.3.2, following [LS13b, Lemma 1], we need to get two intermediate results. First note that, for some distribution of the form (5.25) and

(R_a, R_b) satisfying (5.26)-(5.28) imply that

$$\begin{aligned} R_a + R_b &\leq I(T^a, T^b; Y|V^a, V^b, V) - I(V^a, V^b; S^a, S^b|Y) \\ &= I(T^a, T^b; Y, V^a, V^b, V) - I(V^a, V^b; S^a, S^b|Y) \\ &= I(T^a, T^b; Y) + I(T^a, T^b; V^a, V^b, V|Y) - I(V^a, V^b, V; S^a, S^b|Y) \\ &\leq I(T^a, T^b; Y) \end{aligned} \tag{5.30}$$

where the last line follows since $(V^a, V^b) \leftrightarrow (S^a, S^b) \leftrightarrow (Y, T^a, T^b)$. Note that if $I(T^a, T^b; Y) = 0$, then (5.30) imply that both R_a and R_b must be zero and hence achievable. Therefore, we can consider the case where $I(T^a, T^b; Y) > 0$. Note that this implies $I(T^a; Y|T^b) > 0$ or $I(T^b; Y|T^a) > 0$ since by (5.25), $I(T^b; Y|T^a) \geq I(T^b; Y)$ and therefore, $I(T^a, T^b; Y) \leq I(T^b; Y|T^a) + I(T^a; Y|T^b)$. In the rest of the proof, without loss of generality (WLOG), we assume

$$I(T^a; Y|T^b) > 0. \tag{5.31}$$

Furthermore, we have the following lemma.

Lemma 5.3.1. *If under a distribution of the form (5.25), (R_a, R_b) satisfy (5.26)-(5.28) and $I(T^b; Y, S^a|T^a) = 0$, then R_b must be zero and R_a can not exceed $I(T^a; Y)$. In this case (R_a, R_b) is achievable by using the MAC as a single user channel from X^a to Y.*

Proof of Lemma 5.3.1. Under the conditions of the lemma we have

$$\begin{aligned} R_b &\leq I(T^b; Y|T^a, V^a, V^b, V) - I(V^b; S^b|V^a, Y) \\ &= I(T^b; Y|T^a, V^a, V^b) - I(V^b; S^b|V^a, Y) \\ &\leq I(T^b; Y, S^a|T^a, V^a, V^b) - I(V^b; S^b|V^a, Y, S^a) \qquad (5.32) \\ &= I(T^b; Y, S^a|T^a, V^b) - I(V^b; S^b|Y, S^a). \qquad (5.33) \end{aligned}$$

Note that (5.32) is valid since $I(T^b;Y|T^a,V^a,V^b,V) \leq I(T^b;Y,S^a|T^a,V^a,V^b,V)$ and

$$\begin{aligned} I(V^b;S^b|V^a,Y) &= H(V^b|V^a,Y) - H(V^b|S^b,V^a,Y) \\ &\geq H(V^b|V^a,Y,S^a) - H(V^b|S^b,V^a,Y) \\ &\stackrel{(i)}{=} H(V^b|V^a,Y,S^a) - H(V^b|S^b,V^a,Y,S^a) \\ &= I(V^b;S^b|V^a,Y,S^a) \end{aligned}$$

where (i) follows since $V^b \to S^b \to (Y,S^a)$ which is implied by (5.25). We now verify (5.33) in two steps. First note that

$$\begin{aligned} I(V^b;S^b|V^a,Y,S^a) &= H(V^b|V^a,Y,S^a) - H(V^b|V^a,Y,S^a,S^b) \\ &= H(V^b|Y,S^a) - H(V^b|S^b) \end{aligned}$$

where the second equlaity is valid due to (5.25) which also implies $V^b \to (Y,S^a) \to V^a$. Furthermore, note that

$$\begin{aligned} I(T^b;Y,S^a|T^a,V^a,V^b) &= H(T^b|T^a,V^a,V^b) - H(T^b|Y,S^a,T^a,V^a,V^b) \\ &\stackrel{(ii)}{=} H(T^b|T^a,V^b) - H(T^b|Y,S^a,T^a,V^a,V^b) \\ &\stackrel{(iii)}{=} H(T^b|T^a,V^b) - H(T^b|Y,S^a,T^a,V^b) \qquad (5.34) \end{aligned}$$

where under (5.25), (ii) holds since T^b is independent of (V^a,V^b) and (iii) holds since $T^b \leftrightarrow (Y,S^a,T^a,V^b) \leftrightarrow V^a$ which can be observed as

$$\begin{aligned} &P_{T^b|Y,S^a,T^a,V^a,V^b}(t^b|y,s^a,t^a,v^a,v^b) \\ &= \frac{\sum_{s\in\mathcal{S}}\sum_{s^b\in\mathcal{S}_b} P_{\mathbf{S},T^b,Y,T^a,V^a,V^b}(\mathbf{s},t^b,y,t^a,v^a,v^b)}{\sum_{t^b\in\mathcal{T}_b}\sum_{s\in\mathcal{S}}\sum_{s^b\in\mathcal{S}_b} P_{\mathbf{S},T^b,Y,T^a,V^a,V^b}(s,s^b,s^a,t^b,y,t^a,v^a,v^b)} \\ &\stackrel{(iv)}{=} \frac{\sum_{s\in\mathcal{S}}\sum_{s^b\in\mathcal{S}_b} P_{Y|\mathbf{S},T^a,T^b}(y|\mathbf{s},t^a,t^b)P_{V^a|S^a}(v^a|s^a)P_{\mathbf{S},T^b,T^a,V^b}(\mathbf{s},t^b,t^a,v^b)}{\sum_{t^b\in\mathcal{T}_b}\sum_{s\in\mathcal{S}}\sum_{s^b\in\mathcal{S}_b} P_{Y|\mathbf{S},T^a,T^b}(y|\mathbf{s},t^a,t^b)P_{V^a|S^a}(v^a|s^a)P_{\mathbf{S},T^b,T^a,V^b}(\mathbf{s},t^b,t^a,v^b)} \\ &= \frac{P_{V^a|S^a}(v^a|s^a)\sum_{s\in\mathcal{S}}\sum_{s^b\in\mathcal{S}_b} P_{Y|\mathbf{S},T^a,T^b}(y|\mathbf{s},t^a,t^b)P_{\mathbf{S},T^b,T^a,V^b}(\mathbf{s},t^b,t^a,v^b)}{P_{V^a|S^a}(v^a|s^a)\sum_{t^b\in\mathcal{T}_b}\sum_{s\in\mathcal{S}}\sum_{s^b\in\mathcal{S}_b} P_{Y|\mathbf{S},T^a,T^b}(y|\mathbf{s},t^a,t^b)P_{\mathbf{S},T^b,T^a,V^b}(\mathbf{s},t^b,t^a,v^b)} \\ &= P(t^b|y,s^a,t^a,v^b) \end{aligned}$$

where (iv) is valid due to (4.3) and (5.25). We now show that R_b must be zero. Starting from (5.33), we have

$$\begin{aligned} R_b &\leq I(T^b;Y,S^a|T^a,V^b) - I(V^b;S^b|Y,S^a) \\ &= I(T^b;Y,S^a,V^b|T^a) - I(V^b;S^b|Y,S^a) \\ &= I(T^b;Y,S^a|T^a) + I(T^b;V^b|T^a,Y,S^a) - I(V^b;S^b|Y,S^a) \\ &\stackrel{(v)}{=} H(V^b|T^a,Y,S^a) - H(V^b|T^a,Y,S^a,T^b) - H(V^b|Y,S^a) + H(V^b|S^b,Y,S^a) \\ &= H(V^b|T^a,Y,S^a) - H(V^b|T^a,Y,S^a,T^b) - H(V^b|Y,S^a) + H(V^b|S^b) \\ &\leq H(V^b|S^b) - H(V^b|T^a,Y,S^a,T^b) \\ &\leq 0 \end{aligned} \tag{5.35}$$

where the second equality holds since under (5.25) T^b is independent of (T^a, V^b), (v) is due to the assumption that $I(T^b;Y,S^a|T^a) = 0$ and the others are due to (5.25) and the chain rule; in particular, the last step is due to the Markov condition $V^b \to S^b \to (T^a, Y, S^a, T^b)$. Note now that $R_a \leq I(T^a;Y)$, since $R_b = 0$ and

$$\begin{aligned} R_a &= R_a + R_b \\ &\stackrel{(vi)}{\leq} I(T^a,T^b;Y) \\ &\leq I(T^a;Y) + I(T^b;Y,S^a|T^a) \\ &= I(T^a;Y). \end{aligned}$$

where (vi) is due to (5.30). □

Therefore, in the proof of the main theorem, WLOG we assume (5.31) holds and

$$I(T^b;Y,S^a|T^a) > 0. \tag{5.36}$$

Proof of Theorem 5.3.2. Let (R_a, R_b) satisfy (5.26)-(5.28). The communication is performed in $B+3$ blocks where the first B blocks have blocklengths n and the last three blocks have blocklengths n_1, n_2 and n_3, respectively. The sole purpose of the last three blocks is to convey to the receiver the description of the states in block B. Let us now specify these lengths:

$$n_1 = \frac{nR_{a'}}{\mu_1} \tag{5.37}$$

$$n_2 = \frac{nR_{b'}}{\mu_2} \tag{5.38}$$

$$n_3 = \frac{n_2\left(H(S^a)+\delta\right)}{\mu_1} = \frac{nR_{b'}\left(H(S^a)+\delta\right)}{\mu_1\mu_2} \tag{5.39}$$

where $I(T^a; Y|T^b) > \mu_1 > 0$ for some $\mu_1 > 0$; it is crucial to have positive μ_1 which is satisfied by 5.31, and where $I(T^b; Y, S^a|T^a) > \mu_2 > 0$ for some μ_2, see (5.36).

Codebook Generation Pick $R_{S^a}, R_{S^b}, R_{a'}, R_{b'}$ such that $R_{a'} \leq R_{S^a}, R_{b'} \leq R_{S^b}$.

1) Generate $2^{nR_{S^a}}$ sequences $v^a_{[n]}[j_a]$, $j_a \in \{1, \cdots, 2^{nR_{S^a}}\}$ i.i.d. according to $P_{V^a}(\cdot)$. Randomly partition the indices $\{j_a : 1 \leq j_a \leq 2^{nR_{S^a}}\}$ into $2^{nR_{a'}}$ bins. Denote by $k_a(j_a)$ the bin j_a belongs to and by $\alpha_a(k_a)$ the content of the bin with number k_a.

2) Generate $2^{n(R_a+R_{a'})}$ vectors $t^a_{[n]}[w_a, k_a]$, $w_a \in \{1, \cdots, 2^{nR_a}\}$, $k_a \in \{1, \cdots, 2^{nR_{a'}}\}$ i.i.d. according to $P_{T^a}(\cdot)$.

3) Generate $2^{nR_{S^b}}$ sequences $v^b_{[n]}[j_b]$, $j_b \in \{1, \cdots, 2^{nR_{S^b}}\}$ i.i.d. according to $P_{V^b}(\cdot)$. Randomly partition the indices $\{j_b : 1 \leq j_b \leq 2^{nR_{S^b}}\}$ into $2^{nR_{b'}}$ bins. Denote by $k_b(j_b)$ the bin j_b belongs to and by $\alpha_b(k_b)$ the content of the bin with number k_b.

4) Generate $2^{n(R_b+R_{b'})}$ vectors $t^b_{[n]}[w_b, k_b]$, $w_b \in \{1, \cdots, 2^{nR_b}\}$, $k_b \in \{1, \cdots, 2^{nR_{b'}}\}$ i.i.d. according to $P_{T^b}(\cdot)$.

Repeat the above steps independently B times with the same distribution and rates. The codebook for the last three blocks are generated as follows:

Block B+1

1) Generate one length-n_1 codeword $t^b_{[n_1],B+1}$ i.i.d. according to $P_{T^b}(\cdot)$.
2) Generate $2^{nR_{a'}}$ length-n_1 codewords $t^a_{[n_1],B+1}[\tilde{k}_a]$, $\tilde{k}_a \in \{1, \cdots, 2^{nR_{a'}}\}$ i.i.d. according to $P_{T^a}(\cdot)$.

Block B+2

1) Generate one length-n_2 codeword $t^a_{[n_2],B+2}$ i.i.d. according to $P_{T^a}(\cdot)$.
2) Generate $2^{nR_{b'}}$ length-n_2 codewords $t^b_{[n_2],B+2}[\tilde{k}_b]$, $\tilde{k}_b \in \{1, \cdots, 2^{nR_{b'}}\}$ i.i.d. according to $P_{T^b}(\cdot)$.

Block B+3

1) Generate $2^{n_2(H(S^a)+\tilde{\delta})}$ length-n_3 codewords $t^a_{[n_3],B+3}[\bar{k}_a]$ i.i.d. according to $P_{T^a}(\cdot)$, $\bar{k}_a \in \{1, \cdots, 2^{n_2(H(S^a)+\tilde{\delta})}\}$.
2) Generate one length-n_3 codeword $t^b_{[n_3],B+3}$ i.i.d. according to $P_{T^b}(\cdot)$.

The codebook is revealed to the encoders and the decoder. The encoding operation depends on the block number. Let $\tau \in [1 : B]$ denote the block number.

Encoding Let $w_{a,\tau} \in [1 : 2^{nR_a}]$ and $w_{b,\tau} \in [1 : 2^{nR_b}]$ denote the messages to be sent in block τ.

Block 1. Encoder a, b sends $t^a_{[n],1}[w_{a,1}, 1]$ and $t^b_{[n],1}[w_{b,1}, 1]$, respectively.

Block τ*,* $\tau \in [2:B]$. Encoder q knows $s^q_{[n],\tau-1}$, $q \in \{a,b\}$, and looks for the sequences $v^q_{[n],\tau-1}$ (where $v^q_{[n],\tau-1}[\cdot]$ denote the sequences generated in block $\tau-1$) to find the first index $j_q \in \{1,\cdots,2^{nR_{S^q}}\}$, $q \in \{a,b\}$, such that

$$\left(v^q_{[n],\tau-1}[j_q], s^q_{[n],\tau-1}\right) \in \mathcal{T}_{V^q,S^q}. \tag{5.40}$$

Denote these indices by $j_{a,\tau-1}$ and $j_{b,\tau-1}$, respectively. If a vector $v^q_{[n],\tau-1}[j_{q,\tau-1}]$ satisfying (5.40) does not exist, then the encoder picks a default index, say $j_{q,\tau-1}=1$. Denote by $k_{q,\tau}$ the bin number to which $j_{q,\tau-1}$ belongs. Then

$$\begin{aligned} &\text{Encoder } a \text{ sends: } t^a_{[n],\tau}[w_{a,\tau}, k_{a,\tau}] \\ &\text{Encoder } b \text{ sends: } t^b_{[n],\tau}[w_{b,\tau}, k_{b,\tau}]. \end{aligned} \tag{5.41}$$

Block B+1. User a knows $s^a_{[n],B}$ and inspects the sequences $v^a_{[n],B}$, i.e., $v_{[n]}$ sequences generated in block B, and selects the first index $j_a \in [1:2^{nR_{S^a}}]$ such that

$$(v^a_{[n],B}[j_a], s^a_{[n],B}) \in \mathcal{T}_{V^a,S^a}. \tag{5.42}$$

Denote this index by $j_{a,B}$. If an index can not be found, then let $j_{a,B}=1$. Denote by $k_{a,B+1}$ the bin number to which $j_{a,B}$ belongs. Then

$$\begin{aligned} &\text{Encoder } a \text{ sends: } t^a_{[n_1],B+1}[k_{a,B+1}] \\ &\text{Encoder } b \text{ sends: } t^b_{[n_1],B+1}. \end{aligned} \tag{5.43}$$

Block B+2. User b knows $s^b_{[n],B}$ and inspects the sequences $v^b_{[n],B}$ and selects the first index $j_b \in [1:2^{nR_{S^b}}]$ such that

$$(v^b_{[n],B}[j_b], s^b_{[n],B}) \in \mathcal{T}_{V^b,S^b}. \tag{5.44}$$

Denote this index by $j_{b,B}$. If an index can not be found, then let $j_{b,B}=1$.

Denote by $k_{b,B+1}$ the bin number to which $j_{b,B}$ belongs. Then

$$\begin{aligned} \text{Encoder } a \text{ sends: } & t^a_{[n_2],B+2} \\ \text{Encoder } b \text{ sends: } & t^b_{[n_2],B+2}[k_{b,B+1}]. \end{aligned} \tag{5.45}$$

Block B+3. User a transmits an almost lossless description of $s^a_{[n_2],B+2}$, for example, using [CT06, Section 7.13] for transmitting a source over a noisy channel. Hence, if $k_{a,B+3}$ is the index of $s^a_{[n_2],B+2}$ in the set of all $\tilde{\delta}$-typical sequences, then

$$\begin{aligned} \text{Encoder } a \text{ sends: } & t^a_{[n_3],B+3}[k_{a,B+3}] \\ \text{Encoder } b \text{ sends: } & t^b_{[n_3],B+3}. \end{aligned} \tag{5.46}$$

Decoding Let $y_{[n],\tau}$ denote the channel output in block τ. Decoding starts at block $B+3$.

Block B+3. Decoder looks for an index $\hat{k}_{a,B+3}$ such that

$$\left(t^a_{[n_3],B+3}[\hat{k}_{a,B+3}], t^b_{[n_3],B+3}, y_{[n_3],B+3}\right) \in \mathcal{T}_{T^a,T^b,Y}. \tag{5.47}$$

If $\hat{k}_{a,B+3}$ does not exist or is not unique an error is declared. Otherwise, it sets $\hat{s}^a_{[n_2],B+2}$ to be the sequence whose index is $\hat{k}_{a,B+3}$ in the set of all $\tilde{\delta}$-typical sequences.

Block B+2. Decoder has $y_{[n_2],B+2}$ and also $\hat{s}^a_{[n_2],B+2}$. Decoder looks for an index $\hat{k}_{b,B+2}$ such that

$$\left(t^a_{[n_2],B+2}, t^b_{[n_2],B+2}[\hat{k}_{b,B+2}], \hat{s}^a_{[n_2],B+2}, y_{[n_2],B+2}\right) \in \mathcal{T}_{T^a,T^b,S^a,Y}. \tag{5.48}$$

If an index satisfying (5.48) does not exist or is not unique an error is declared.

Block B+1. Decoder looks for an index $\hat{k}_{a,B+1}$ such that

$$\left(t^a_{[n_1],B+1}[\hat{k}_{a,B+1}], t^b_{[n_1],B+1}, y_{[n_1],B+1}\right) \in \mathcal{T}_{T^a,T^b,Y}. \tag{5.49}$$

If an index satisfying (5.49) does not exist or is not unique an error is declared.

Block τ, $\tau \in [2:B]$. Decoder has $\hat{k}_{a,\tau+1}$, $\hat{k}_{b,\tau+1}$ and $y_{[n],\tau}$ and it looks for $v^a_{[n]}[\hat{j}_a] \in \alpha_a(\hat{k}_{a,\tau+1})$, and $v^b_{[n]}[\hat{j}_b] \in \alpha_b(\hat{k}_{b,\tau+1})$ such that

$$\left(v^a_{[n]}[\hat{j}_a], v^b_{[n]}[\hat{j}_b], y_{[n],\tau}\right) \in \mathcal{T}_{V^a,V^b,Y}. \tag{5.50}$$

If such a pair does not exist, or is not unique, an error is declared. Note that $(v^a_{[n]}[\hat{j}_a], v^b_{[n]}[\hat{j}_b])$ consists of the compressed state sequences in Block τ. Using this, decoder tries to estimate the messages $(w_{a,\tau}, w_{b,\tau})$ as well as $(k_{a,\tau}, k_{b,\tau})$, which are the bin numbers of states in block $\tau - 1$. Specifically, the decoder looks for

$$\left(t^a_{[n],\tau}[\hat{w}_{a,\tau}, \hat{k}_{a,\tau}], t^b_{[n],\tau}[\hat{w}_{b,\tau}, \hat{k}_{b,\tau}], v^a_{[n]}[\hat{j}_a], v^b_{[n]}[\hat{j}_b], y_{[n],\tau}\right) \in \mathcal{T}_{T^a,T^b,V^a,V^b,Y}. \tag{5.51}$$

If such indices do not exist or not unique an error is declared.

Block 1. Works as in the blocks $2, \cdots, B$ except $\hat{k}_{a,1} = \hat{k}_{b,1} = 1$.

Let $(\hat{w}_{a,\tau}, \hat{w}_{b,\tau})$, $\tau \in \{1, \cdots, B\}$ denote the decoder outputs.

Error Analysis Let us first verify that by decoding the last three blocks we obtain $(k_{a,B+1}, k_{b,B+1})$ with high probability.

Block B+3. Let $\mathcal{E}^{B+3}_{k'} := \left\{\left(t^a_{[n_3]}[k'], t^b_{[n_3]}, y_{[n_3]}\right) \in \mathcal{T}_{T^a,T^b,Y}\right\}$. Hence, the probability of error in block $B+3$, P^e_{B+3}, satisfies

$$\begin{aligned} P^e_{B+3} &\leq \Pr\left(\mathcal{E}^{B+3,c}_{k_{a,B+3}} \bigcup_{k_{a,B+3} \neq \hat{k}_{a,B+3}} \mathcal{E}^{B+3}_{\hat{k}_{a,B+3}}\right) \\ &\leq \Pr\left(\mathcal{E}^{B+3,c}_{k_{a,B+3}}\right) + \sum_{k_{a,B+3} \neq \hat{k}_{a,B+3}} \Pr\left(\mathcal{E}^{B+3}_{\hat{k}_{a,B+3}}\right). \end{aligned}$$

Note that $\Pr\left(\mathcal{E}^{B+3,c}_{k_{a,B+3}}\right) \to 0$ as $n_3 \to \infty$, which is satisfied when $n \to \infty$ (see

5.39). For the other term, following standard arguments, we have

$$\begin{aligned}
&\sum_{k_{a,B+3}\neq\hat{k}_{a,B+3}} \Pr\left(\mathcal{E}^{B+3}_{\hat{k}_{a,B+3}}\right)\\
&\quad= \sum_{k_{a,B+3}\neq\hat{k}_{a,B+3}} P\left(\left(t^a_{[n_3]}[\hat{k}_{a,B+3}], t^b_{[n_3]}, y_{[n_3]}\right)\in\mathcal{T}_{T^a,T^b,Y}\right)\\
&\quad= \sum_{k_{a,B+3}\neq\hat{k}_{a,B+3}} \sum_{\left(t^a_{[n_3]}[\hat{k}_{a,B+3}], t^b_{[n_3]}, y_{[n_3]}\right)\in\mathcal{T}_{T^a,T^b,Y}} P\left(t^a_{[n_3]}[\hat{k}_{a,B+3}]\right) P\left(t^b_{[n_3]}, y_{[n_3]}\right)\\
&\quad\leq 2^{n_2\left(H(S^a)+\tilde{\delta}\right)} 2^{-n_3\left(I(T^a;Y|T^b)\right)}\\
&\quad= 2^{\frac{nR_{b'}}{\mu_2}\left(H(S^a)+\tilde{\delta}\right)} 2^{-\frac{nR_{b'}(H(S^a)+\delta)}{\mu_1\mu_2}\left(I(T^a;Y|T^b)\right)} \qquad (5.52)
\end{aligned}$$

where (5.52) is due to (5.38) and (5.39). Therefore, $\sum_{k_{a,B+3}\neq\hat{k}_{a,B+3}} P\left(\mathcal{E}^{B+3}_{\hat{k}_{a,B+3}}\right)\to 0$ as $n\to\infty$ since $I(T^a;Y|T^b)>\mu_1>0$ (recall that $\mu_1>0$ is satisfied by 5.31) which guarantees that $k_{a,B+3}$ is decoded correctly.

Block B+2. The decoder has both $y_{[n_2],B+2}$ and $\hat{s}^a_{[n_2],B+2}$ and tries to estimate $k_{b,B+1}$. Let $\mathcal{E}^{B+2}_{k''} := \left\{\left(t^b_{[n_2]}[k''], t^a_{[n_2]}, y_{[n_2]}, \hat{s}^a_{[n_2],B+2}\right)\in\mathcal{T}_{T^a,T^b,Y,S^a}\right\}$. The probability of error in block $B+2$, P^e_{B+2}, satisfies

$$P^e_{B+2} \;\leq\; \Pr\left(\mathcal{E}^{B+2,c}_{k_{b,B+1}}\right) + \sum_{k_{b,B+1}\neq\hat{k}_{b,B+1}} \Pr\left(\mathcal{E}^{B+2}_{\hat{k}_{b,B+1}}\right).$$

Note that $\Pr\left(\mathcal{E}^{B+2,c}_{k_{b,B+1}}\right)\to 0$ as $n_2\to\infty$, which is satisfied when $n\to\infty$ (see 5.38). For the other term, we have

$$\begin{aligned}
&\sum_{k_{b,B+1}\neq\hat{k}_{b,B+1}} \Pr\left(\mathcal{E}^{B+2}_{\hat{k}_{b,B+1}}\right)\\
&\quad= \sum_{k_{b,B+1}\neq\hat{k}_{b,B+1}} P\left(\left(t^b_{[n_2]}[\hat{k}_{b,B+1}], t^a_{[n_2]}, y_{[n_2]}, \hat{s}^a_{[n_2],B+2}\right)\in\mathcal{T}_{T^a,T^b,Y,S^a}\right)\\
&\quad= \sum_{k_{b,B+1}\neq\hat{k}_{b,B+1}} \sum_{\left(t^b_{[n_2]}[\hat{k}_{b,B+1}], t^a_{[n_2]}, y_{[n_2]}, \hat{s}^a_{[n_2],B+2}\right)\in\mathcal{T}_{T^a,T^b,Y,S^a}} P\left(t^b_{[n_2]}[\hat{k}_{b,B+1}]\right)\\
&\qquad\qquad\times P\left(t^a_{[n_2]}, y_{[n_2]}, \hat{s}^a_{[n_2],B+2}\right)
\end{aligned}$$

$$\leq 2^{nR_{b'}} 2^{-n_2\left(I(T^b;Y,S^a|T^a)\right)}$$
$$= 2^{nR_{b'}} 2^{-\frac{nR_{b'}}{\mu_2}\left(I(T^b;Y,S^a|T^a)\right)} \tag{5.53}$$

where (5.53) is due to (5.38). Therefore, $\sum_{k_{b,B+1}\neq\hat{k}_{b,B+1}} P\left(\mathcal{E}^{B+2}_{\hat{k}_{a,B+2}}\right) \to 0$ as $n \to \infty$ since $I(T^b;Y,S^a|T^a) > \mu_2 > 0$ (recall that $\mu_2 > 0$ is satisfied by 5.36) which guarantees that $k_{b,B+1}$ is decoded correctly.

Block B+1. Similar to the decoding processes in blocks $B+3$, $B+2$, the decoder estimates $k_{a,B+1}$ with arbitrarily low error probability, as $n \to \infty$, by setting $n_1 = \frac{nR_{a'}}{\mu_1}$. Analysis is similar and hence, we skip the details.

For the error analysis in blocks $\tau \in [1:B]$, we use the fact that $k_{a,B+1}$ and $k_{b,B+1}$ is decoded with low error probability and WLOG assume that $(w_{a,\tau}, w_{b,\tau}) = (1,1),\ \forall\tau$ and specific sequence of pairs $(k_{a,[2:B+3]}, k_{b,[2:B+3]})$ is chosen.

Let $\mathbf{s}^q_{[B]} := \left(s^q_{[n],1}, \cdots, s^q_{[n],B}\right)$, $q \in \{a,b\}$, and define the following events:

$$\begin{aligned}
\mathcal{A}_{1,\tau}(s^a_{[n],\tau}, s^b_{[n],\tau}) \quad &:= \quad \Big\{\left(v^a_{[n]}[j_a], s^a_{[n],\tau}\right) \in \mathcal{T}_{V^a,S^a},\ \left(v^b_{[n]}[j_b], s^b_{[n],\tau}\right) \in \mathcal{T}_{V^b,S^b}, \\
&\qquad\qquad \text{for some } j_q \in \{1,\cdots,2^{nR_{S^q}}\}, q \in \{a,b\}\Big\} \\
\mathcal{A}_{2,1}(s^a_{[n],1}, s^b_{[n],1}) \quad &:= \quad \Big\{\left(t^a_{[n],1}[1,1], t^b_{[n],1}[1,1], s^a_{[n],1}, s^b_{[n],1}\right) \in \mathcal{T}_{T^a,T^b,S^a,S^b}\Big\} \\
\mathcal{A}_{3,\tau}(s^a_{[n],\tau}, s^b_{[n],\tau}) \quad &:= \quad \Big\{\left(t^a_{[n],\tau}[1,k_{a,\tau}], t^b_{[n],\tau}[1,k_{b,\tau}], s^a_{[n],\tau}, s^b_{[n],\tau}\right) \in \mathcal{T}_{T^a,T^b,S^a,S^b}\Big\} \\
\mathcal{B}_{2,\tau}(j_a, j_b) \quad &:= \quad \Big\{\left(v^a_{[n],\tau}[j_a], v^b_{[n],\tau}[j_b], y_{[n],\tau}\right) \in \mathcal{T}_{V^a,V^b,Y}\Big\} \\
\mathcal{B}_{3,\tau}(w_a,k_a,w_b,k_b,j_a,j_b) \quad &:= \\
&\Big\{\left(t^a_{[n],\tau}[w_a,k_a], t^b_{[n],\tau}[w_b,k_b], v^a_{[n],\tau}[j_a], v^b_{[n],\tau}[j_b], y_{[n],\tau}\right) \in \mathcal{T}_{T^a,T^b,V^a,V^b,Y}\Big\}
\end{aligned}$$

and

$$\mathcal{A} \quad := \quad \mathcal{A}_{2,1}(s^a_{[n],1}, s^b_{[n],1}) \bigcap_{\tau=1}^{B} \mathcal{A}_{1,\tau}(s^a_{[n],\tau}, s^b_{[n],\tau}) \bigcap_{\tau=2}^{B+1} \mathcal{A}_{3,\tau}(s^a_{[n],\tau}, s^b_{[n],\tau})$$

$$\beta_\tau := \mathcal{B}^c_{2,\tau}(j_{a,\tau}, j_{b,\tau}) \bigcup_{\substack{j_a\in\alpha_a(k_{a,\tau+1}),\ j_b\in\alpha_b(k_{b,\tau+1})\\ (j_a,j_b)\neq(j_{a,\tau},j_{b,\tau})}} \mathcal{B}_{2,\tau}(j_a, j_b) \tag{5.54}$$

$$\gamma_\tau := \mathcal{B}^c_{3,\tau}(1, k_{a,\tau}, 1, k_{b,\tau}, j_{a,\tau}, j_{b,\tau}) \bigcup_{(w_a,k_a,w_b,k_b)\neq(1,k_{a,\tau},1,k_{b,\tau})} \mathcal{B}_{3,\tau}(w_a, k_a, w_b, k_b, j_{a,\tau}, j_{b,\tau}). \tag{5.55}$$

The error event is $\mathcal{E} := \bigcup_{\tau=1}^{B} \{(\hat{w}_{a,\tau}, \hat{w}_{b,\tau}) \neq (w_{a,\tau}, w_{b,\tau})\}$ and hence, it is sufficient to show that

$$\lim_{n\to\infty} P(\mathcal{E}|\mathbf{s}^a_{[B]}, \mathbf{s}^b_{[B]}) = 0,\ \ \forall(\mathbf{s}^a_{[B]}, \mathbf{s}^b_{[B]}) \in \mathcal{T}^B_{S^a,S^b} \tag{5.56}$$

where $\mathcal{T}^B_{S^a,S^b}$ is the B product of $\mathcal{T}_{S^a,S^b}$. Note that

$$\mathcal{E} \subseteq \mathcal{A}^c \cup \beta_B \cup \gamma_B \cup \left\{\bigcup_{\tau=1}^{B-1}(\beta_\tau \cup \gamma_\tau)\right\}$$

and hence,

$$\begin{aligned} & P(\mathcal{E}|\mathbf{s}^a_{[B]}, \mathbf{s}^b_{[B]}) \\ &\leq P\left(\mathcal{A}^c \cup \beta_B \cup \gamma_B \cup \left\{\bigcup_{\tau=1}^{B-1}(\beta_\tau \cup \gamma_\tau)\right\} \middle| \mathbf{s}^a_{[B]}, \mathbf{s}^b_{[B]}\right) \\ &\leq P\left(\mathcal{A}^c \middle| \mathbf{s}^a_{[B]}, \mathbf{s}^b_{[B]}\right) + P\left(\beta_B \cup \gamma_B \cup \left\{\bigcup_{\tau=1}^{B-1}(\beta_\tau \cup \gamma_\tau)\right\} \middle| \mathcal{A}, \mathbf{s}^a_{[B]}, \mathbf{s}^b_{[B]}\right) \end{aligned} \tag{5.57}$$

where (5.57) is valid since for any events E_1, E_2

$$\begin{aligned} P(E_1 \cup E_2|\theta) &= P(E_1|\theta) + P(E_2 \cap E_1^c|\theta) \\ &\leq P(E_1|\theta) + P(E_2|E_1^c, \theta)P(E_1^c|\theta) \leq P(E_1|\theta) + P(E_2|E_1^c, \theta). \end{aligned}$$

For $\tau \in [1 : B]$, let $\gamma^c_{[\tau+1,B]} := \{\gamma^c_{\tau+1}, \cdots, \gamma^c_B\}$ and $\beta^c_{[\tau+1,B]} := \{\beta^c_{\tau+1}, \cdots, \beta^c_B\}$. Repeating the above steps recursively, we get

$$\begin{aligned} P(\mathcal{E}|\mathbf{s}^a_{[B]}, \mathbf{s}^b_{[B]}) &\leq P\left(\mathcal{A}^c|\mathbf{s}^a_{[B]}, \mathbf{s}^b_{[B]}\right) + P\left(\beta_B|\mathcal{A}, \mathbf{s}^a_{[B]}, \mathbf{s}^b_{[B]}\right) + P\left(\gamma_B|\mathcal{A}, \mathbf{s}^a_{[B]}, \mathbf{s}^b_{[B]}, \beta^c_B\right) \\ &+P\left(\bigcup_{\tau=1}^{B-1}(\beta_\tau \cup \gamma_\tau)|\mathcal{A}, \mathbf{s}^a_{[B]}, \mathbf{s}^b_{[B]}, \beta^c_B, \gamma^c_B\right) \end{aligned}$$

$$
\begin{aligned}
&\leq P\left(\mathcal{A}^c|\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]}\right)+P\left(\beta_B|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]}\right)+P\left(\gamma_B|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]},\beta^c_B\right)\\
&\quad+\sum_{\tau=1}^{B-1}\left[P\left(\beta_\tau|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]},\gamma^c_{[\tau+1,B]},\beta^c_{[\tau+1,B]}\right)+P\left(\gamma_\tau|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]},\gamma^c_{[\tau+1,B]},\beta^c_{[\tau,B]}\right)\right]\\
&\overset{(i)}{=} P\left(\mathcal{A}^c|\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]}\right)+P\left(\beta_B|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]}\right)+P\left(\gamma_B|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]},\beta^c_B\right)\\
&\quad+\sum_{\tau=1}^{B-1}\left[P\left(\beta_\tau|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]},\gamma^c_{\tau+1}\right)+P\left(\gamma_\tau|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]},\gamma^c_{\tau+1},\beta^c_\tau\right)\right]\\
&\overset{(ii)}{=} P\left(\mathcal{A}^c|\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]}\right)+P\left(\beta_B|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]}\right)+P\left(\gamma_B|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]},\beta^c_B\right)\\
&\quad+(B-1)\left[P\left(\beta_\tau|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]},\gamma^c_{\tau+1}\right)+P\left(\gamma_\tau|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]},\gamma^c_{\tau+1},\beta^c_\tau\right)\right] \qquad (5.58)
\end{aligned}
$$

for some fixed τ and (i) holds since $\beta_\tau \leftrightarrow (\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]},\gamma^c_{\tau+1}) \leftrightarrow (\gamma^c_{[\tau+2,B]},\beta^c_{[\tau+1,B]})$ and $\gamma_\tau \leftrightarrow (\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]},\gamma^c_{\tau+1},\beta^c_\tau) \leftrightarrow (\gamma^c_{[\tau+2,B]},\beta^c_{[\tau+1,B]})$ and (ii) holds because of the independent codebook construction across the blocks. It should be noted that conditioned on $\gamma^c_{\tau+1}$, the decoder has at hand the correct bin indices of the previous block, i.e., $(\hat{k}_{a,\tau+1},\hat{k}_{b,\tau+1}) = (k_{a,\tau+1},k_{b,\tau+1})$. Let us now consider each term in (5.58).

By the covering lemma 2.3.1,

$$
\lim_{n\to\infty} P\left(\mathcal{A}^c|\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]}\right)=0 \ \text{ whenever } \ R_{S^a}>I(V^a;S^a) \ \text{ and } \ R_{S^b}>I(V^b;S^b). \qquad (5.59)
$$

For the term $P\left(\beta_B|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]}\right)$, note that the decoder knows the correct bin indices $k_{a,B+1}$ and $k_{b,B+1}$. Since $P(\mathcal{B}^c_{2,B}(j_{a,B},j_{b,B}))\to 0$, we have,

$$
\begin{aligned}
&P\big(\beta_B|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]}\big)\leq\\
&\quad P\Big(\bigcup_{\substack{j_a\in\alpha_a(k_{a,B+1})\\ j_a\neq j_{a,B}}}\mathcal{B}_{2,B}(j_a,j_{b,B})\big|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]}\Big)+P\Big(\bigcup_{\substack{j_b\in\alpha_b(k_{b,B+1})\\ j_b\neq j_{b,B}}}\mathcal{B}_{2,B}(j_{a,B},j_b)\big|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]}\Big)\\
&\quad+P\Big(\bigcup_{\substack{j_a\in\alpha_a(k_{a,B+1}),\ j_b\in\alpha_b(k_{b,B+1})\\ j_a\neq j_{a,B},\ j_b\neq j_{b,B}}}\mathcal{B}_{2,B}(j_a,j_b)\big|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]}\Big). \qquad (5.60)
\end{aligned}
$$

By standard typicality arguments,

$$P\Big(\bigcup_{\substack{j_a\in\alpha_a(k_{a,B+1})\\ j_a\neq j_{a,B}}} \mathcal{B}_{2,B}(j_a,j_{b,B})\big|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]}\Big) \leq 2^{n(R_{S^a}-R_{a'})}2^{-n[I(V^a;V^b,Y)-\delta]} \quad (5.61)$$

for some $\delta > 0$, $\lim_{n\to\infty}\delta = 0$. Similarly,

$$P\Big(\bigcup_{\substack{j_b\in\alpha_b(k_{b,B+1})\\ j_b\neq j_{b,B}}} \mathcal{B}_{2,B}(j_{a,B},j_b)\big|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]}\Big) \leq 2^{n(R_{S^b}-R_{b'})}2^{-n[I(V^b;V^a,Y)-\delta]}. \quad (5.62)$$

For the last term in (5.60), we invoke [Gas04, Lemma 8] (see Appendix 2.3.2 for the statement of the lemma) and obtain

$$P\Big(\bigcup_{\substack{j_a\in\alpha_a(k_{a,B+1}),\ j_b\in\alpha_b(k_{b,B+1})\\ j_a\neq j_{a,B},\ j_b\neq j_{b,B}}} \mathcal{B}_{2,B}(j_a,j_{b,B})\big|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]}\Big) \leq$$
$$2^{n(R_{S^a}+R_{S^b}-R_{a'}-R_{b'})}2^{-n[I(V^a;Y,V^b)+I(V^b;Y,V^a)-I(V^a,V^b|Y)-\delta]}. \quad (5.63)$$

Hence, $\lim_{b\to\infty} P\left(\beta_B|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]}\right) = 0$ if

$$\begin{aligned} R_{S^a}-R_{a'} &< I(V^a;V^b,Y) \\ R_{S^b}-R_{b'} &< I(V^b;V^a,Y) \\ R_{S^a}+R_{S^b}-R_{a'}-R_{b'} &< I(V^a;Y,V^b)+I(V^b;Y,V^a)-I(V^a,V^b|Y). \end{aligned} \quad (5.64)$$

Consider now the term $P\left(\gamma_B|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]},\beta^c_B\right)$ where β^c_B guarantees that the decoder has at hand the compressed version of the state information at block B, $(v^a_{[n],B}[j_{a,B}],v^b_{[n],B}[j_{b,B}])$. Note that this pair is independent of Shannon strategies generated in block B. Recall now that

$$\gamma_B := \mathcal{B}^c_{3,B}(1,k_{a,B},1,k_{b,B},j_{a,B},j_{b,B}) \bigcup_{(w_a,k_a,w_b,k_b)\neq(1,k_{a,B},1,k_{b,B})} \mathcal{B}_{3,B}(w_a,k_a,w_b,k_b,j_{a,B},j_{b,B})$$

where $\lim_{n\to\infty} P\left(\mathcal{B}^c_{3,B}(1,k_{a,B},1,k_{b,B},j_{a,B},j_{b,B})\right) = 0$. We can decompose the union as

$$\bigcup_{\substack{(w_a,k_a,w_b,k_b)\neq\\(1,k_{a,B},1,k_{b,B})}} \mathcal{B}_{3,B}(w_a,k_a,w_b,k_b,j_{a,B},j_{b,B}) = \bigcup_{(w_a,k_a)\neq(1,k_{a,B})} \mathcal{B}_{3,B}(w_a,k_a,1,k_{b,B},j_{a,B},j_{b,B})$$

$$\bigcup_{(w_b,k_b)\neq(1,k_{b,B})} \mathcal{B}_{3,B}(1,k_{a,B},w_b,k_b,j_{a,B},j_{b,B}) \bigcup_{\substack{(w_b,k_b)\neq(1,k_{b,B})\\(w_a,k_a)\neq(1,k_{a,B})}} \mathcal{B}_{3,B}(w_a,k_a,w_b,k_b,j_{a,B},j_{b,B}). \tag{5.65}$$

For the first union in the right hand side of (5.65), we have

$$\begin{aligned}
&P\Big(\bigcup_{(w_a,k_a)\neq(1,k_{a,B})} \mathcal{B}_{3,B}(w_a,k_a,1,k_{b,B},j_{a,B},j_{b,B})\big|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]},\beta^c_B\Big) = \\
&P\Big(\bigcup_{(w_a,k_a)\neq(1,k_{a,B})} \big(t^a_{[n],B}[w_a,k_a],t^b_{[n],B}[1,k_{b,B}],v^a_{[n],B}[j_{a,B}],v^b_{[n],B}[j_{b,B}],y_{[n],B}\big) \\
&\qquad\qquad \in \mathcal{T}_{T^a,T^b,V^a,V^b,Y}\big|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]},\beta^c_B\Big) \\
&\leq \sum_{(w_a,k_a)\neq(1,k_{a,B})} \sum_{\substack{\left(t^a_{[n],B}[w_a,k_a],t^b_{[n],B}[1,k_{b,B}],v^a_{[n],B}[j_{a,B}],v^b_{[n],B}[j_{b,B}],y_{[n],B}\right)\\ \in\mathcal{T}_{T^a,T^b,V^a,V^b,Y}}} \\
&\quad\Big[P(t^a_{[n],B}[w_a,k_a]\big|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]},\beta^c_B) \\
&\qquad\qquad P(t^b_{[n],B}[1,k_{b,B}],v^a_{[n],B}[j_{a,B}],v^b_{[n],B}[j_{b,B}],y_{[n],B}\big|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]},\beta^c_B)\Big] \\
&\leq 2^{n(R_a+R_{a'})}2^{-n\left[-H(T^a,T^b,Y,V^a,V^b)+H(T^a)+H(T^b,Y,V^a,V^b)-\delta\right]} \\
&= 2^{n(R_a+R_{a'})}2^{-n\left[-I(T^a;Y|V^a,V^b,T^b)-\delta\right]}
\end{aligned} \tag{5.66}$$

for some $\delta > 0$, $\lim_{n\to 0}\delta = 0$. For each $\tau \in [1:B]$, we can obtain similar results for the terms in the right hand side of (5.65) and hence conclude that,

$$\lim_{n\to\infty} P\left(\gamma_B|\mathcal{A},\mathbf{s}^a_{[B]},\mathbf{s}^b_{[B]},\beta^c_B\right) = 0$$

if

$$\begin{aligned}
R_a + R_{a'} &< I(T^a;Y|V^a,V^b,T^b) \\
R_b + R_{b'} &< I(T^b;Y|V^a,V^b,T^a) \\
R_a + R_b + R_{a'} + R_{b'} &< I(T^a,T^b;Y|V^a,V^b).
\end{aligned} \tag{5.67}$$

We can similarly show that the other terms in (5.58), i.e., $P\left(\beta_\tau|\mathcal{A}, \mathbf{s}^a_{[B+1]}, \mathbf{s}^b_{[B+1]}, \gamma^c_{\tau+1}\right)$ and $P\left(\gamma_\tau|\mathcal{A}, \mathbf{s}^a_{[B+1]}, \mathbf{s}^b_{[B+1]}, \gamma^c_{\tau+1}, \beta^c_\tau\right)$, tends to zero with large blocklengths.

Recall that V is a function of (V^a, V^b), $V^a - S^a - (Y, V^b)$ and $V^b - S^b - (Y, V^a)$ and therefore, combining (5.59), (5.64) and (5.67) shows that (5.56) is satisfied whenever

$$\begin{aligned} R_a + R_{a'} &< I(T^a; Y|V^a, V^b, T^b, V) - I(V^a; S^a|V^b, Y) \\ R_b + R_{b'} &< I(T^b; Y|V^a, V^b, T^a) - I(V^b; S^b|V^a, Y) \\ R_a + R_b + R_{a'} + R_{b'} &< I(T^a, T^b; Y|V^a, V^b) - I(V^a, V^b; S^a, S^b|Y). \end{aligned} \tag{5.68}$$

In order to complete the proof, we finally need to show that the rate of the above scheme approaches (R_a, R_b) for large B. This follows from [LS13b, Equations (52),(53)]; we here give the details for the sake of completeness.

Recall that the total transmission time is $B + 3$ blocks where the first B blocks have length n and the rest have n_1, n_2 and n_3, respectively, where these lengths are given in (5.37)-(5.39). Therefore, the scheme has the effective rates, (R_a, R_b)

$$(R_a, R_b)\frac{nB}{nB + n_1 + n_2 + n_3} = (R_a, R_b)\frac{B}{B + \frac{R_{a'}}{\mu_1} + \frac{R_{b'}}{\mu_2} + \frac{R_{b'}(H(S^a)+\delta)}{\mu_1\mu_2}}$$

which approaches (R_a, R_b) for large B. □

Remark 5.3.2. *One conclusion of Theorem 5.3.2 and 5.2.3 is that the Shannon strategies are far from optimality by at most $I(V^a, V^b; S^a, S^b|Y)$ when there is no CSIR; note that the maximum sum-rate in the inner bound above is given as*

$$\max_{(V,V^a,V^b,T^a,T^b)\in\mathcal{P}^{cr}(0)} I(T^a, T^b; Y|V^a, V^b, V) - I(V^a, V^b; S^a, S^b|Y)$$

which can be shown via a similar proof as the one for Theorem 5.2.3, by showing that the sum-rate constraint is always active.

5.4 Conclusion and Remarks

In this chapter, we have investigated the memoryless state dependent MACs where each encoder is provided with asymmetric correlated CSI in a causal way. We first show that for the optimality of Shannon strategies for the sum-rate capacity, it is sufficient to have the decoder provided with a side information under which the correlated CSITs become conditionally independent. This observation generalizes the known results which either assume independent CSITs or full CSI at the receiver. We next consider the situation where there is no CSIR. We provide an inner bound, which is inspired from the lossless CEO approach and uses the technique provided in [LS13b], to demonstrate the rate required to convey the aforementioned side information to the receiver.

Chapter 6

Summary and Conclusion

6.1 Summary

The main purpose of this thesis is to investigate the influence, from the channel capacity perspective, of the channel output feedback and the channel side information (CSI) in the single and multiterminal communication systems. By the definition of channel capacity, one needs to work on large blocklengths and arguably, one of the most important objectives in channel coding problems is to obtain single letter expressions for the set of all achievable rates. Obtaining such expressions for the capacity regions in multiterminal setups is, in general, hard, and in the situations where the channel output feedback and the side information are also involved in the problem setup, it is arguably more challenging. Indeed, for many such problems single letter expressions are still unknown. The two of the essential ingredients of a channel coding problem when there is feedback or side information can be succinctly given by the facts that the problem has a stochastic dynamic and the information patterns at the decision makers, i.e., the encoder(s) and the decode(s), might have degraded characteristics. These two facts lead to alternative formulations of the problem in a stochastic control framework and many contributions have already been available for such channel coding problems. In a broad sense, the main contributions of this thesis are obtaining some complete solutions by using tools and ideas from stochastic control theory. Let us now recall these contributions.

The first part of this thesis has dealt with the channel output feedback problem for channels with memory. Motivated by the known fact that even for channels with memory, feedback might not increase capacity, we have obtained in Chapter 3 a larger class of channels with memory whose feedback and non-feedback capacities are identical. The main tool used to derive this result is the dynamic programming formulation of the optimization problem that the converse of feedback capacity problem reduces to.

The second part of this thesis has considered the state-dependent multiple access channels (MAC) where asymmetric noisy side information is provided to the encoders. In Chapter 4, by first assuming that there exists full CSI at the receiver, we have obtained single-letter expressions for the capacity regions under several scenarios. Achievability of these regions follows from standard arguments of joint typicality coding and hence, the original ingredient of these results is the team decision based converse coding approach. More explicitly, to obtain a single-letter expression for the capacity region, the presented converse coding approach shows that the past (and future in the non-causal setup) is irrelevant. This is obtained by showing that under any policy that one can achieve using an arbitrary decentralized coding policy, the same performance can be achieved by using memoryless stationary team policies. Depending on what information available to which decision makers as well as the statistical dependence between these information, we have further shown that in some situations it might not be required to have full CSI at the receiver. It is worth to stress that although the same results could be obtained via using the standard tools from information theory, such as introducing axillary random variables for time-sharing, we believe such a converse approach is insightful and can be used to obtain further results in multiuser information theory.

The state-dependent MAC has been an active research area and there exist many contributions based on the information classification (see the literature review in Chapter 1). Among these one can observe that in the situation when there is no CSI at the receiver, complete characterizations are available mainly under the assumption of private messages' degradation, i.e., one encoder knows the other's private message. Indeed, one of the most classical problems in this setup, a two-user MAC with causal (or non-causal) CSI available at the encoders and no CSI at the receiver, is still open. The associated single user problem has been solved by Shannon and the capacity is given in terms of the Shannon strategies which are shown to be suboptimal for the MAC problem. This fact motivated us to find the information required to be made available to the receiver so that a tight converse in terms of Shannon strategies is possible. Characterization of this information, which is shown in Chapter 5, is then used to determine the rate required to transmit this information to the receiver when the receiver has no CSI. This rate is determined by using the technique of a recently proposed inner bound as well as the result of the lossless CEO source coding problem. The obtained inner bound also demonstrates how far, at most, the Shannon strategies are away from optimality.

6.2 Suggestions for Future Research

In recent years a considerable amount of research has been devoted to understand the fundamental limits on information flow in networks and to design optimal coding techniques and protocols that achieve these limits [EGK11]. The theory in this direction has been referred to as Network Information Theory (NIT) and the results presented in this thesis contribute to this broad area. NIT aims to extend Shannon's point-to-point information theory results to networks such as broadcast, interference and relay networks. Hence, a potential future direction of research can be extending the results on the state-dependent MAC presented in this thesis to other types of multiterminal models.

Although for many NIT problems a complete solution is yet to be developed, there are many achievable regions with some of them having a tight converse. When the available results are examined, it is observed that in order to get a single-letter expression, the usual approach, which is also the critical step, is to correctly represent the time dependent variable via some auxiliary random variables. The single letter characterization then follows by applying some results from convex analysis such as the support lemma [CK81]. In classical Shannon theoretic results, such as DMC channel capacity, the expression is given in terms of physically represented parameters, such as the channel input and output, and hence, one critique that can be made against using auxiliary random variables is the lack of such a physical representation among the systems parameters. The team based approach presented in the second part of this thesis, on the other hand, follows a different approach by showing that the past or future is irrelevant. Therefore, it is worth to explore this approach for other multiterminal systems.

One should recall that the expression for the feedback capacity of channels with arbitrary memory is given in terms of the inf-information rates. We can define it more explicitly as follows.

Definition 6.2.1. *[VH93] The liminf in probability of a sequence of random variables* $\{X_t\}$ *is defined as the largest extended real number* α *such that* $\forall \epsilon > 0$

$$\lim_{t\to\infty} \mathbb{P}(X_t \leq \alpha - \epsilon) = 0.$$

Denote this number α *by* $\liminf_{inprob} X_t$.

Let

$$\begin{aligned} \overrightarrow{P}_{X_{[T]}|Y_{[T]}}(x_{[T]}|y_{[T]}) &\triangleq \prod_{i=1}^{T} P_{X_t|X_{[t-1]},Y_{[t-1]}}(x_t|x_{[t-1]}, y_{[t-1]}) \\ \overrightarrow{P}_{X_{[T]}|Y_{[T]}} P_{Y_{[T]}}(X_{[T]}, Y_{[T]}) &\triangleq \overrightarrow{P}_{X_{[T]}|Y_{[T]}}(x_{[T]}|y_{[T]}) P_{Y_{[T]}}(y_{[T]}). \end{aligned}$$

In [TM09] and [Tat00], it is shown that the feedback capacity of arbitrary channels is equal to

$$\begin{aligned} C_{FB} &= \sup_{\{\mathcal{D}_T\}_{T=1}^{\infty}} \underline{I}(X \to Y) \\ \mathcal{D}_T &= \{\{P_{X_t|X_{[t-1]},Y_{[t-1]}}(x_t|x_{[t-1]}, y_{[t-1]})\}_{t=1}^{T}\} \end{aligned}$$

where X and Y are channel input and output, respectively, and $\underline{I}(X \to Y) = \liminf_{inprob} \frac{1}{T} \overrightarrow{i}(X_{[T]}; Y_{[T]})$ where

$$\overrightarrow{i}(X_{[T]}; Y_{[T]}) \triangleq \log \frac{P_{X_{[T]},Y_{[T]}}(x_{[T]}, y_{[T]})}{\overrightarrow{P}_{X_{[T]}|Y_{[T]}} P_{Y_{[T]}}(x_{[T]}, y_{[T]})}.$$

It should be observed that the optimization problem given in C_{FB} is very difficult to solve. The usual way to deal with this is to assume some sort of ergodicity in the channel model [Tat00], more specifically, one makes suitable assumptions so that the channel is information stable. To that end the authors in [Tat00] examined the class of Markov channels and they show that the problem of feedback coding for Markov channels can be cast as a partially observed stochastic control problem. Hence, they can use the tools of dynamic programming to solve the mutual information optimization problem underlying the capacity problem [Tat00, Chapter 4].

Another possible future work can be extending the results presented in Chapter 3 to the case of arbitrary channels with memory. The first step required in this direction is to formulate C_{FB} as a stochastic control problem for arbitrary channels and use dynamic programming to solve the optimization problem.

Bibliography

[AF94] F. Alajaji and T. Fuja. Effect of feedback on the capacity of discrete additive channels with memory. In *Proc. IEEE Int. Symp. Information Theory*, Trondheim, Norway, June 1994.

[Ala95] F. Alajaji. Feedback does not increase capacity of discrete channels with additive noise. *IEEE Trans. Inform. Theory*, 41(1):546–549, Jan. 1995.

[Ala99] F. Alajaji. *Information Theory.* Queen's University, Lecture Notes, 1999.

[ALV+06] D.M. Arnold, H.A. Loeliger, P.O. Vontobel, A. Kavčić, and W. Zeng. Simulation-based computation of information rates for channels with memory. *IEEE Trans. Inform. Theory*, 52(8):3498–3507, Aug. 2006.

[AS12] A. Anastasopoulos and K. Sohn. An interpretation of the cover and leung capacity region for the mac with feedback through stochastic control. In *Proc. IEEE Int. Conf. Communications*, pages 2360 – 2364, Ottawa, Canada, June 2012.

[Ber01] D. P. Bertsekas. *Dynamic Programming and Optimal Control.* Athena Scientific, Belmont, MA, 2001.

[BSP12] U. Basher, A. Shirazi, and H. Permuter. Capacity region of finite state multiple-access channel with delayed state information at the transmitters. *IEEE Trans. Inform. Theory*, 58(6):3430–3452, June 2012.

[CA95] P.N. Chen and F. Alajaji. Strong converse, feedback capacity and hypothesis testing. *Journal of the Chinese Institute of Engineers*, 18:777–785, Nov. 1995.

[CB05] J. Chen and T. Berger. The capacity of finite-state Markov channels with feedback. *IEEE Trans. Inform. Theory*, 51(3):780–798, Mar. 2005.

[CK81] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Akademiai Kiado, 1981.

[CL81] M. T Cover and C. S. K Leung. An achievable rate region for the multiple-access channel with feedback. *IEEE Trans. Inform. Theory*, 27(3):292–298, May 1981.

[CS99] G. Caire and S. Shamai. On the capacity of some channels with channel state information. *IEEE Trans. Inform. Theory*, 45(6):2007–2019, Sep. 1999.

[CS05] Y. Cemal and Y. Steinberg. The multiple-access channel with partial state information at the encoders. *IEEE Trans. Inform. Theory*, 51(11):3992–4003, Nov. 2005.

[CT06] T. M. Cover and J. A. Thomas. *Elements of Information Theory Second Edition*. Wiley, New Jersey, 2006.

[CY11] G. Como and S. Yüksel. On the capacity of memoryless finite state multiple access channels with asymmetric state information at the encoders. *IEEE Trans. Inform. Theory*, 57(3):1267–1273, Mar. 2011.

[DG06] S. Diggavi and M. Grossglauser. On information transmission over a finite buffer channel. *IEEE Trans. Inform. Theory*, 52(3):1226–1237, Mar. 2006.

[DN02] A. Das and P. Narayan. Capacities of time-varying multiple access channels with side information. *IEEE Trans. Inform. Theory*, 48(1):4–25, Jan. 2002.

[EGK11] A. El Gamal and Y. H. Kim. *Network Information Theory.* Cambridge, New York, 2011.

[EM02] Y. Ephraim and N. Merhav. Hidden Markov processes. *IEEE Trans. Inform. Theory*, 48(6):1518–1569, June 2002.

[EZ00] U. Erez and R. Zamir. Noise prediction for channels with side ifnormation at the transmitter. *IEEE Trans. Inform. Theory*, 46(4):1610–1617, July 2000.

[Gal68] R. G. Galleger. *Information Theory and Reliable Communication.* Wiley, New York, 1968.

[Gas04] M. Gastpar. The Wyner-Ziv problem with multiple sources. *IEEE Trans. Inform. Theory*, 50(11):2762–2768, Nov. 2004.

[GP79] S. I. Gelfand and M. S. Pinsker. Coding of sources on the basis of observations with incomplete observations. *Prob. Pered. Inform.*, 15(2):45–57, July 1979.

[GP80] S. I. Gelfand and M. S. Pinsker. Coding for channels with random parameters. *Prob. of Cont. and Inform.*, 9:19–31, 1980.

[GV96] A. Goldsmith and P. Varaiya. Capacity, mutual information, and coding for finite-state Markov channels. *IEEE Trans. Inform. Theory*, 42(3):868–886, May 1996.

[GV97] A. Goldsmith and P. Varaiya. Capacity of fading channels with channel side information. *IEEE Trans. Inform. Theory*, 43(6):1986–1992, Nov. 1997.

[GW75] N. T. Gaarder and J. K. Wolf. The capacity region of a multiple-access discrete memoryless channel can increase with feedback. *IEEE Trans. Inform. Theory*, 21(1):100–102, Jan. 1975.

[Hög77] G. H. Högnäs. Random semigroup acts on a finite set. *Jour. Austral. Math. Soc.*, 23 (Series A):482–498, 1977.

[Jaf06] S. A. Jafar. Capacity with casual and non-casual side information-a unified view. *IEEE Trans. Inform. Theory*, 52(12):5468–5474, Dec. 2006.

[KL07] S. Kotagiri and J. Laneman. Multiaccess channels with state known to one encoder: A case of degraded message sets. In *Proc. IEEE Int. Symp. Information Theory*, pages 1566–1570, Nice, France, June 2007.

[KSM07] G. Keshet, Y. Steinberg, and N. Merhav. Channel coding in the presence of side information: Subject review. *Found. Trends Commun. Inf. Theory*, 4(6):445–586, 2007.

[KT74] N. V. Kuznetsov and B. S. Tsybakov. Coding in memories with defective cells. *Problemy peredachi informatsii*, 10(2):52–60, 1974.

[LS] A. Lapidoth and Y. Steinberg. A note on multiple-access channels with strictly-causal state information. *available at [arXiv:1106.0380v1].*

[LS13a] A. Lapidoth and Y. Steinberg. The multiple-access channel with causal side information: common state. *IEEE Trans. Inform. Theory*, 59(1):32–50, Jan. 2013.

[LS13b] A. Lapidoth and Y. Steinberg. The multiple-access channel with causal side information: double state. *IEEE Trans. Inform. Theory*, 59(3):1379–1393, Mar. 2013.

[LSY13] M. Li, O. Simeone, and A. Yener. Multiple access channels with states causally known at transmitters. *IEEE Trans. Inform. Theory*, 59(3):1394–1404, Mar. 2013.

[MD89] M. Mushkin and I. B. David. Capacity and coding for the gilbert-elliott channel. *IEEE Trans. Inform. Theory*, 35(6):1277–1290, Nov. 1989.

[MT09] A. Mahajan and D. Teneketsiz. Optimal performance of networked control systems with non-classical information structures. *SIAM Jour. Cont. and Opt.*, 48(3):1377–1404, May 2009.

[NFT07] U. Niesen, C. Fragouli, and D. Tuninetti. On capacity of line networks. *IEEE Trans. Inform. Theory*, 53(11):4039–4058, Nov. 2007.

[NT11] A. Nayyar and D. Teneketsiz. On the structure of real-time encoders and decoders in a multi-terminal communication system. *IEEE Trans. Inform. Theory*, 57(9):6196–6214, Sep. 2011.

[PA09] C. Pimentel and F. Alajaji. A discrete channel model for capturing memory and soft-decision information: A capacity study. In *Proc. IEEE Int. Conf. Communications*, Dresden, Germany, June 2009.

[PAM12] C. Pimentel, F. Alajaji, and P. Melo. A discrete queue-based model for capturing memory and soft-decision information in correlated fading channels. *IEEE Trans. Comm.*, 60(6):1702–1711, June 2012.

[PSSB11] H. Permuter, S. Shamai, and A. Somekh-Baruch. Message and state cooperation in multiple access channels. *IEEE Trans. Inform. Theory*, 57(10):6379–6396, Oct. 2011.

[PWC09] H. Permuter, T. Weissman, and J. Chen. Capacity region of the finite-state multiple access channel with and without feedback. *IEEE Trans. Inform. Theory*, 55:2455–2477, June 2009.

[PWG09] H. Permuter, T. Weissman, and A. J. Goldsmith. Finite state channels with time-invariant deterministic feedback. *IEEE Trans. Inform. Theory*, 55:644–662, Feb. 2009.

[Rez06] M. Rezaeian. Symmetric characterization of finite state Markov channels. In *Proc. IEEE Int. Symp. Information Theory*, Seattle, USA, July 2006.

[Sal92] M. Salehi. Capacity and coding for memories with real-time noisy defect information at encoder and decoder. *IEE-Proceedings-I*, 139:113–117, Apr. 1992.

[ŞAY09a] N. Şen, F. Alajaji, and S. Yüksel. Feedback capacity of a class of symmetric finite-state Markov channels. In *Proc. Forty-Seventh Aller. Conf. Comm. Cont. and Comp.*, Monticello,IL, Sep. 2009.

[ŞAY09b] N. Şen, F. Alajaji, and S. Yüksel. On the feedback capacity of a discrete non-binary noise channel with memory. In *Proc. Canadian Work. Information Theory*, Ottawa, Canada, May 2009.

[ŞAY11] N. Şen, F. Alajaji, and S. Yüksel. Feedback capacity of a class of symmetric finite-state Markov channels. *IEEE Trans. Inform. Theory*, 57(7):4110–4122, July 2011.

[ŞAYC] N. Şen, F. Alajaji, S. Yüksel, and G. Como. Memoryless multiple access channel with asymmetric noisy state information at the encoders. *submitted to IEEE Trans. Inform. Theory.*

[ŞAYC12] N. Şen, F. Alajaji, S. Yüksel, and G. Como. Multiple access channel with various degrees of asymmetric state information. In *Proc. IEEE Int. Symp. Information Theory*, pages 1697–1701, Cambridge, MA, July 2012.

[SBSV08] A. Somekh-Baruch, S. Shamai, and S. Verdú. Cooperative multiple-access encoding with states available at one transmitter. *IEEE Trans. Inform. Theory*, 54(10):4448–4469, Oct 2008.

[ŞCYA11] N. Şen, G. Como, S. Yüksel, and F. Alajaji. On the capacity of memoryless finite-state multiple access channels with asymmetric noisy state information at the encoders. In *Proc. Forty-Ninth Aller. Conf. Comm. Cont. and Comp.*, Monticello,IL, Sep. 2011.

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:479–423, Jul. and Oct. 1948.

[Sha56] C.E. Shannon. The zero-error capacity of a noisy channel. *IRE Trans. Inform. Theory*, 2:8–19, 1956.

[Sha58] C. E. Shannon. Channels with side information at the transmitter. *IBM J. Res. Develop.*, 2:289–293, 1958.

[SK05] S. Sigurjónsson and Y. H Kim. On multiple user channels with causal state information at the transmitters. In *Proc. IEEE Int. Symp. Information Theory*, pages 72–76, Adelaide, Australia, July 2005.

[SP09] B. Shrader and H. Permuter. Feedback capacity of the compound channel. *IEEE Trans. Inform. Theory*, 55(8):3629–3644, Aug. 2009.

[Tat00] S. C. Tatikonda. *Control Under Communication Constraints.* Ph.D Dissertion, M.I.T., Cambridge, MA, 2000.

[TM09] S. Tatikonda and S. Mitter. The capacity of channels with feedback. *IEEE Trans. Inform. Theory*, 55(1):329–343, Jan. 2009.

[VH93] S. Verdu and T.H. Han. Approximation theory of output statistics. *IEEE Trans. Inform. Theory*, 39:752–772, May 1993.

[VH94] S. Verdu and T.H. Han. A general formula for channel capacity. *IEEE Trans. Inform. Theory*, 40(4):11471157, July 1994.

[Vis99] H. Viswanathan. Capacity of Markov channels with receiver csi and delayed feedback. *IEEE Trans. Inform. Theory*, 45(2):761–771, Mar. 1999.

[Wit98] H. S. Witsenhausen. Equivalent stochastic control problems. *Mathemtics of Control, Signal and Systems*, 1:3–11, 1998.

[Wyn75] A. D. Wyner. The common information of two dependent random variables. *IEEE Trans. Inform. Theory*, 21(2):45–57, Mar. 1975.

[YB13] S. Yüksel and T. Başar. *Stochastic Networked Control Systems: Stabilization and Optimization under Information Constraints*. Birkhauser, Boston, MA, 2013.

[YKT05] S. Yang, A. Kavcic, and S. Tatikonda. Feedback capacity of finite-state machine channels. *IEEE Trans. Inform. Theory*, 51(3):799–810, Mar. 2005.

[YT07] S. Yüksel and S. Tatikonda. Capacity of Markov channels with partial state feedback. In *Proc. IEEE Int. Symp. Information Theory*, pages 1861–1865, Nice, France, June 2007.

[Yük13] S. Yüksel. On optimal causal coding of partially observed Markov sources in single and multi-terminal settings. *IEEE Trans. Inform. Theory*, 59(1):424–437, Jan. 2013.

[ZPS11] A. Zaidi, P. Piantanida, and S. Shamai. Multiple access channel with states known noncausally at one encoder and only strictly causally at the other encoder. In *Proc. IEEE Int. Symp. Information Theory*, St. Petersburg, Russia, July 2011.

[ZPSS] A. Zaidi, P. Piantanida, and S. Shamai (Shitz). Capacity region of multiple access channel with states known noncausally at one encoder and only strictly causally at the other encoder. *available at [arXiv:1201.3278]*.