An Information-theoretic approach to integrated sensing and communication

Mehrasa Ahmadipour

INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2022IPPAT041

Thèse de doctorat

An Information-Theoretic Approach to Integrated Sensing and Communication

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à Télécom Paris

École doctorale n^{0}626 Institut Polytechnique de Paris (ED IP Paris)
Spécialité de doctorat: Réseaux, Information et Communications

Thèse présentée et soutenue à Palaiseau, le 10 November 2022, par

MEHRASA AHMADIPOUR

Composition du Jury :

M. Aslan Tchamkerten Professeur, Télécom Paris	Président
M. Tobias J Oechtering Professeur, KTH	Rapporteur
Mme Meryem Benammar Associate Professeure, ISAE SUPAERO	Examinatrice
M. Giuseppe Caire Professeur, Technical University of Berlin	Examinateur
M. Malcolm Egan Chargé de Recherche, INRIA	Examinateur
M. Shlomo Shamai Professeur, Technion	Examinateur
Mme Mari Kobayashi Ingénieure, Apple	Membre invitée
Mme Michèle Wigger Professeure,Télécom Paris	Directrice de thèse

To

Fearless women in Iran who are battling for liberty

And

My mother, Homa, My father, Fariborz

Acknowledgements

I am deeply indebted to my supervisor Prof. Michèle Wigger for her support and advice. Working and learning under her guidance was a great privilege, and I could not have undertaken this journey without her.

I would like to express my gratitude to Prof. Mari Kobayashi for the many discussions that enhanced my research experience, especially in the first half of my Ph.D. I likewise appreciate the opportunity to collaborate with Prof. Guissepe Caire.

In addition, I'm incredibly grateful to have Prof. Ofer Shayevits and Prof.Tobias Oechtering as my thesis reporters. They provided valuable suggestions and insightful comments, which enriched my insight.

I express my most profound appreciation to Prof. Shlomo Shamai for being an examiner on my jury with very challenging questions to augment my future research directions. In addition, I am appreciative to have Dr. Meryem Benammar and Dr. Malcolm Egan serving on my committee, and their questions made me think of new ideas for future work.

I am very grateful to my friends in Télécom, especially Akram, Mustapha, Yue, Abdelaziz, Robert, and others in the COMELEC department, for all our fun, discussions, and coffee breaks in the past years. Furthermore, I have been fortunate to have Homa, with whom I carry many pleasant memories and who supported me during these years.

My deepest gratitude goes to my family for all their love and devotion. This endeavor would not have been possible without their support.

I should mention that during these years of my Ph.D., I have always thought about my country, Iran, and the people fighting for freedom. There have been moments of tears for those we lost through these years. But also, there have emerged hopes for a brighter future for humanity.

Résumé

Les réseaux sans fil de la prochaine génération devraient prendre en charge plusieurs techniques de détection et de localisation précises. Les systèmes de transport intelligents, où les véhicules détectent en permanence les changements environnementaux et échangent des informations et des données de détection avec des véhicules déjà détectés, des stations de base ou des serveurs centraux, en sont des exemples importants. Une approche courante mais naïve pour aborder la détection et la communication consiste à séparer les deux tâches en systèmes indépendants et à répartir les ressources disponibles, telles que la largeur de bande et la puissance, entre les deux systèmes. Cependant, les coûts élevés du spectre et du matériel encouragent l'intégration des tâches de détection et de communication via une seule forme d'onde et une seule plate-forme matérielle. Ces systèmes intégrés sont plus compliqués à concevoir, notamment en raison du compromis inhérent qu'ils présentent entre les performances de détection et de communication.

Cette thèse s'appuie sur [1], qui a introduit le premier modèle informationnel théorique pour la détection et la communication intégrées (ISAC) et a caractérisé les limites fondamentales des performances de détection et de communication dans ce modèle. Le modèle de [1] considère un canal sans mémoire dépendant de l'état (SDMC) avec des signaux de rétroaction généralisée observés au niveau de l'émetteur (Tx), il mesure également les performances de communication en termes de taux de données fiables et les performances de détection en termes de distorsion moyenne. On remarque que la rétroaction généralisée est bien adaptée à la modélisation des systèmes ISAC, car elle peut décrire le comportement du canal et donc de l'environnement, elle capture la nature passive des signaux d'écho observés au Tx, et elle dépend également de la forme d'onde de transmission. Les résultats présentés dans [1], montrent que pour une configuration point à point (P2P) à émetteur unique (Tx) et récepteur unique (Rx), le compromis optimal entre les performances de communication et de détection est atteint par des constructions de code aléatoire standard telles qu'utilisées pour la communication de données traditionnelle, où les statistiques des entrées du canal, cependant, doivent être adaptées pour obtenir la performance de détection souhaitée.

Cette thèse se concentre sur la théorie de l'information ISAC sur les réseaux multi-Tx ou multi-Rx. Plus précisément, notre première contribution est de caractériser le compromis fondamental entre les taux de communication et

la distorsion de détection des canaux de diffusion (BC) mono-Tx et bi-Rx dépendants de l'état qui sont physiquement dégradés. Nous fournissons également des limites intérieures et extérieures sur les compromis taux-distorsion réalisables pour les canaux de diffusion généraux. Nos résultats montrent des compromis intéressants entre les performances de détection et de communication réalisables simultanément, ce qui implique que pour améliorer le taux de communication, il faut sacrifier les performances de détection et vice versa. Ce compromis est illustré à l'aide de divers exemples. De plus, nous décrivons une classe de réseaux de base où les performances de communication et de détection n'ont pas de compromis, et où les deux tâches peuvent être satisfaites simultanément avec leurs performances optimales.

Les canaux P2P et les BC sont tous deux des réseaux à émission unique, pour lesquels on peut montrer que la stratégie de détection optimale de l'émetteur est un simple estimateur symbole par symbole de l'état caché étant donné les entrées et sorties du canal au niveau de la borne de détection. L'optimalité de cet estimateur découle du fait que les canaux de rétroaction généralisés et la séquence d'état se comportent tous deux sans mémoire pour une séquence d'entrée fixe. Ce n'est pas nécessairement le cas dans les configurations où le terminal de détection n'est pas le seul à fournir des entrées aux canaux sans mémoire, par exemple dans les réseaux multi-Tx ou dans les réseaux où la détection est effectuée à la Rx. Dans ce cas, la perturbation effective pour la détection n'est pas nécessairement sans mémoire puisque les entrées des autres terminaux créent également des perturbations et peuvent avoir une mémoire. Dans la deuxième partie de cette thèse, où l'accent est mis sur les réseaux à deux Tx, des stratégies de détection plus complexes sont donc nécessaires. Plus précisément, nous nous concentrons sur les scénarios de communication par canal multi-accès (MAC) à deux Tx et par dispositif interactif à dispositif (D2D), c'est-à-dire le canal bidirectionnel à deux Tx. Pour ces réseaux, nous introduisons la théorie de l'information *collaborative sensing*, un concept qui a déjà reçu une attention significative dans les communautés de communication et de traitement du signal. Dans nos schémas de détection collaborative, les émetteurs compriment les informations sur leurs signaux de retour et transmettent ces informations aux autres émetteurs dans le but de les aider dans leurs performances de détection.

Plus précisément, pour le MAC, nous étendons naturellement le schéma de codage de Willem pour transmettre les informations d'état d'une Tx à l'autre sur le chemin de communication construit sur la liaison de rétroaction généralisée. Le schéma proposé peut être considéré comme un schéma de codage source-canal distinct dans le sens où chaque Tx compresse d'abord les sorties et les entrées obtenues pour extraire les informations d'état, puis transmet l'indice de compression à l'aide d'un code de canal pur à l'autre Tx. Les schémas ISAC collaboratifs que nous proposons permettent d'obtenir un meilleur rendement de détection qu'un schéma ISAC antérieur [2] où les Tx ne s'entraident pas pour améliorer le rendement de détection. Ainsi, nous obtenons généralement un meilleur compromis taux-distorsion que le schéma précédent.

Pour le scénario D2D, nous décrivons deux schémas ISAC collaboratifs. Le premier schéma est similaire dans son esprit au schéma MAC et est basé sur la séparation source-canal et le schéma de canal bidirectionnel de Han. Le second schéma repose sur une stratégie améliorée basée sur le codage conjoint source-canal (JSCC), plus précisément le codage hybride. Nous montrons des performances améliorées des deux schémas de détection collaborative D2D. Dans le scénario MAC et D2D, nos schémas ISAC sont strictement concaves dans les paires taux-distorsion et améliorent donc également les stratégies classiques de partage du temps ou des ressources.

Abstract

Next-generation wireless networks are expected to support several accurate sensing and localization techniques. Important examples are intelligent transport systems, where vehicles continuously sense environmental changes and exchange sensing information and data with already detected vehicles, base stations, or central servers. A common but naive approach to address sensing and communication is to separate the two tasks into independent systems and split the available resources, such as bandwidth and power, between the two systems. The high spectrum and hardware costs, however, encourage integrating the sensing and communications tasks via a single waveform and a single hardware platform. Such integrated systems are more complicated to design, in particular also because of the inherent tradeoff they bear between sensing and communication performances.

This thesis builds on [1], which introduced the first information-theoretic model for integrated sensing and communication (ISAC) and characterized the fundamental limits of sensing and communication performances in this model. The model in [1] considers on a state-dependent memoryless channel (SDMC) with generalized feedback signals observed at the transmitter (Tx), also, it measures communication performance in reliable data rates and sensing performance in terms of average distortion. Notice that generalized feedback is well-suited to model ISAC systems, because it can describe the behaviour of the channel and thus the environment, it captures the passive nature of the echo signals observed at the Tx, and it also depends on the transmit waveform. The results in [1], show that for a single-Tx single-Receiver (Rx) point-to-point (P2P) setup, the optimal tradeoff between communication and sensing performances is achieved by standard random code constructions as used for traditional data communication, where the statistics of the channel inputs, however, must be adapted to obtain the desired sensing performance.

This thesis focuses on information-theoretic ISAC over multi-Tx or multi-Rx networks. More specifically, our first contribution is to characterize the fundamental tradeoff between communication rates and sensing distortion of state-dependent single-Tx two-Rx broadcast channels (BC) that are physically degraded. We also provide inner and outer bounds on the achievable rate-distortion tradeoffs for general BCs. Our results show interesting tradeoffs between the simultaneously achievable sensing and communication performances, implying that to improve com-

munication rate, one has to sacrifice sensing performance and vice versa. This tradeoff is illustrated at the hand of various examples. Moreover, a class of BCs is described where communication and sensing performances experience no tradeoff, and both tasks can be satisfied with their optimal performances simultaneously.

P2P channels and BCs are both single-Tx networks, for which it can be shown that the Tx's optimal sensing strategy is a simple symbol-by-symbol estimator of the hidden state given the channel inputs and outputs at the sensing terminal. The optimality of this estimator stems from the fact that the generalized feedback channels and the state-sequence both behave in a memoryless manner for a fixed input sequence. This is not necessarily the case in setups where the sensing terminal is not the only terminal feeding inputs to the memoryless channels, for example, in multi-Tx networks or in networks where sensing is performed at the Rx. In this case, the effective disturbance for the sensing is not necessarily memoryless since the inputs from the other terminals also create disturbances and can have memory. In the second part of this thesis, where the focus is on two-Tx networks, more complicated sensing strategies are thus required. Specifically, we focus the two-Tx single-Rx multiaccess channel (MAC) and interactive device-to-device (D2D) communication scenarios, i.e., the two-Tx two-way channel. For these networks, we introduce information-theoretic *collaborative sensing*, a concept that has already received significant attention in the communication and signal processing communities. In our collaborative sensing schemes, Txs compress information about their feedback signals and convey this information to other Txs with the goal of helping them in their sensing performances.

Specifically, for the MAC, we naturally extend Willem's coding scheme to convey state-information from one Tx to the other over the communication path built over the generalized feedback link. The proposed scheme can be considered a separate source-channel coding scheme in the sense that each Tx first compresses the obtained outputs and inputs to extract state information, then transmits the compression index using a pure channel code to the other Tx. Our proposed collaborative ISAC schemes achieve a better sensing performance than a previous ISAC scheme [2] where Txs do not help each other to improve sensing performance. Thus, we generally achieve a better rate-distortion tradeoff than the previous scheme.

For D2D scenario, we describe two collaborative ISAC schemes. The first scheme is similar in spirit to the MAC scheme and is based on source-channel separation and Han's two-way channel scheme. The second scheme is based on an improved strategy based on joint source-channel coding (JSCC), specifically hybrid coding. We show enhanced performances of both D2D collaborative sensing schemes. In both the MAC and the D2D scenario, our ISAC schemes are strictly concave in the rate-distortion pairs and thus also improve over classical time- or resource-sharing strategies.

Contents

List of Figures

List of Abbreviations

BC Broadcast Channel

MAC Multiple Access Channel

CSI Channel State Information

MIMO Multiple-Input Multiple-output

Rx Receiver

Tx Transmitter

ISAC Integrated Sensing And Communication

Chapter 1

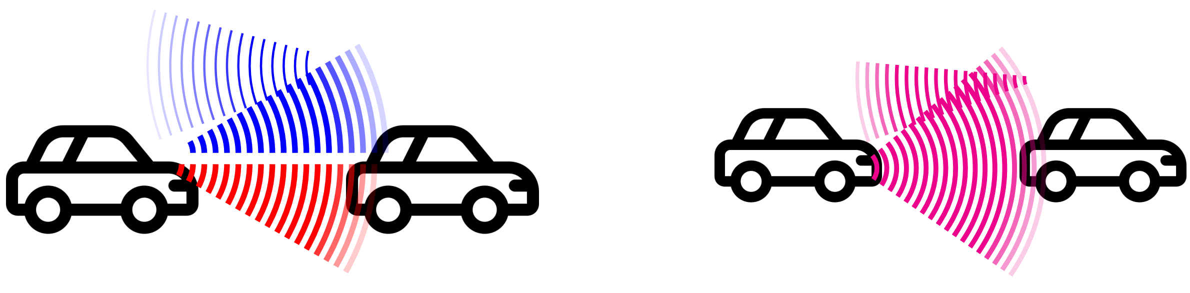
Introduction

1.1 Background and Motivation

Next-generation wireless networks are expected to support several autonomous and intelligent applications that rely heavily on accurate sensing and localization techniques, as this is one of the critical features of 6G wireless networks [4]. One example are intelligent transport systems, where vehicles are equipped to sense environmental changes and exchange information with other vehicles for various purposes.

The standard assumption of such a cooperative sensing and communication system is that a transmitter (Tx) wishes to convey a message to an already detected while a radio detection or radar receiver (Rx) that is co-located with this Tx wishes to estimate parameters of the environment.

A common but naive approach to address sensing and communication is to separate the two tasks in independent systems and to split the available resources such as bandwidth and power between the two. In the information-theoretic models that we present in the following chapters, such a system corresponds to *resource-sharing* between communication and sensing (see Figure 1.1a). The high costs of spectrum and hardware however encourages in-



(a) Resource-splitting

(b) Synergic wave-form

Figure 1.1: Resource-splitting vs. synergic wave-form.

tegrating the sensing and communication tasks via a single waveform and a single hardware platform [5, 6] (see Figure 1.1b). First attempts towards such integrated systems use a standard communication system and exploit the backscattered signal from this waveform for sensing purposes. The employed transmit waveform is either optimized for sensing or communication. The literature has widely studied such *integrated sensing and communication (ISAC)* schemes (see, e.g., [7, 8] and references therein).

In particular, several ISAC schemes have been proposed to optimize performance metrics capturing the tension between sensing and communication performances [9–14]. While these works provide system guidelines or propose waveforms suitable to specific scenarios, this thesis addresses the problem from a more theoretical angle. We aim to characterize the fundamental performance limits of ISAC systems under a given model for the parameters that need to be estimated and for the channel over which communication and sensing take place.

The present thesis is thus related to the fields of information theory and estimation and their interplay. The literature of related works in these fields is vast, and we shall mainly focus on the most related ones. We, however, would also like to mention the seminal work in [15], which connects the input-output mutual information and the minimum mean-square error (MMSE) achievable by optimal estimation of the input given the output.

The fundamental performance limits of ISAC systems were first considered in [1]. Specifically, [1] introduced an information-theoretic model for ISAC based on a memoryless state-dependent channel and generalized-feedback. The state-sequence is meant to model the evolution of the environment and thus captures the sensing parameter that the Tx wishes to estimate. The generalized feedback captures two underlying assumptions used in radar signal processing: generalized feedback captures the inherently passive nature of the backscattered signal observed at the Tx, which cannot be controlled but is determined by its surrounding environment, and it models the fact that the backscattered signal depends on the waveform employed by the Tx. It was proposed to use the classical average per-letter block-distortion to measure the Tx's sensing performance on the i.i.d. state-sequence.

The authors of [1], see also [16], characterized the exact *capacity-distortion tradeoff* of arbitrary single-Tx and single-Rx discrete memoryless channels (DMCs) with generalized feedback. The capacity-distortion tradeoff measures the inherent tradeoff between increasing data rate and reducing sensing distortion in such integrated systems. Interestingly, the results show that the optimal tradeoff is achieved by standard random code constructions as used for traditional data communication, where the statistics of the channel inputs (and thus of the codewords) however has to be adapted to meet the desired sensing performance. This observation is consistent with the signal-processing literature on the search for adequate channel input waveforms which allow to meet the desired sensing performance while still achieving high communication rates.

As mentioned, to achieve the optimal capacity-distortion tradeoff for single-Tx single-Rx DMCs, standard codes

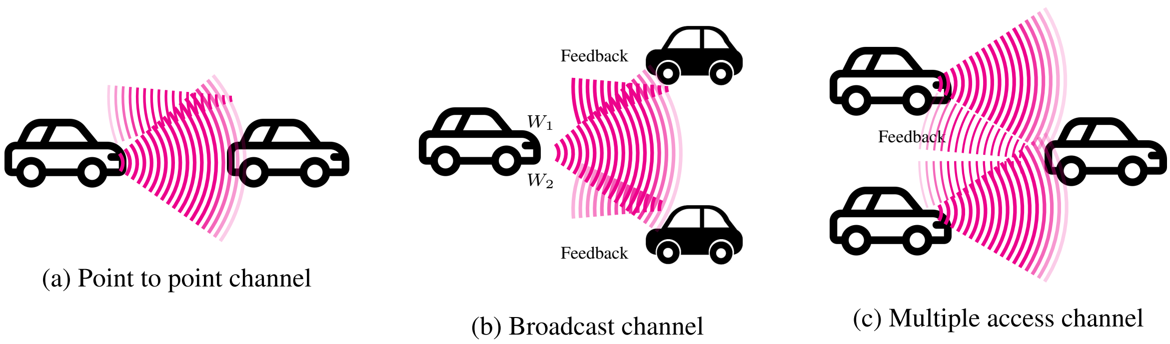


Figure 1.2: Different sensing and communication constellations

can be used for communication that ignore entirely the generalized feedback, i.e. the backscattered signals used for sensing. The reason is that any kind of feedback does not increase the standard capacity (i.e., the highest rate of communication) of memoryless single-Tx and single-Rx channels. The situation changes however significantly for multi-Tx or multi-Rx networks, where feedback can increase capacity even for memoryless channels. For example, feedback increases the capacity of memoryless single-Tx multi-Rx broadcast channels (BC) [17–19] because it enables the Tx to send common information that is useful to both Rxs at the same time (see e.g., [20, Section 17]). In a dual manner, in multi-Tx single-Rx multi-access channels (MAC), feedback allows the distributed Txs to cooperate in their transmissions by sending useful common or correlated information to the single Rx. In network ISAC setups, the generalized feedback thus improves both sensing and communication performances. As we already mentioned, the sensing and communication performances in single-Tx single-Rx point-to-point (P2P) setups depend on each other only through the common choice of the waveform. We show in this thesis that a similar observation also holds for memoryless BCs, i.e., the sensing performance is independent of the employed BC-feedback-code and only depends on the chosen waveform but not on other details of the code construction. It allows to base integrated coding and sensing systems on known BC-feedback code constructions such as [17–19] where one only has to adapt the statistics of the channel inputs to achieve the desired sensing performance. Based on the scheme in [17], we provide a general inner bound on the capacity-distortion region for general memoryless BCs with generalized feedback. We also provide a general outer bound on the capacity-distortion region of memoryless BCs by extending a known converse technique that reveals the outputs at one of the Rxs to the other Rx. Inner and outer bounds coincide in special cases.

Completely characterizing the capacity-distortion tradeoff region for a general memoryless state-dependent BC seems extremely challenging since even the capacity region (without sensing) is unknown both in the case without and with feedback (see e.g., [17–19, 21]). Instead, we characterize the capacity-distortion region for the particular case of physically degraded BCs. Analogously to the single-user case, feedback does not enlarge the capacity of physically degraded BCs and is useful only for sensing but not for communication. We illustrate the merit of optimal ISAC schemes over time- or resource-sharing schemes through various numerical examples.

The sensing situation becomes more interesting and challenging when the sensing terminal is not the only terminal feeding the input to the channel. In this case, the effective disturbance for the sensing is not necessarily memoryless since the inputs from the other terminals also create disturbances and can have memory. That being so, a strategy that first attempts to guess the other Txs' codewords followed by a symbol-wise estimator based on the observations and the guessed codewords can lead to a smaller (and thus better) distortion. A similar phenomenon has also been observed in [22], where communication is over a DMC, and state estimation is performed at the Rx's side. In this case, the optimal sensing strategy is first to decode the Tx's codeword and then to apply an optimal symbol-by-symbol estimator and the observed channel outputs to this codeword.

A similar strategy was applied in the two-Tx single-Rx multi-access channel (MAC) ISAC scenario of [2] in which, through the generalized feedback, each Tx first decodes part of the data sent by the other Tx and then applies a symbol-by-symbol estimator to the decoded codeword as well as its own channel inputs and outputs. The ISAC scheme of [2] is based on Willems' scheme for the MAC with generalized feedback, where each Tx encodes its data into two super-positioned codewords, of which the other Tx decodes the lower data-layer. This data is then repeated by both Txs in the next block as part of a third lowest-layer codeword, allowing the two Txs to transmit data cooperatively.

Somewhat naturally, [2] suggests using this decoded lower layer also for sensing purposes in the sense that each Tx applies the symbol-by-symbol estimator not only to its inputs and outputs but also to this decoded codeword. In this thesis, we suggest using this decoded codeword not only to exchange data but also to exchange sensing information. The concept of exchanging sensing information for ISAC has been studied in the signal processing literature under the paradigm of *collaborative sensing*. In this sense, we also introduce the concept of collaborative sensing for ISAC to the information-theoretic literature, where we focus on the MAC and the related device-to-device (D2D) communication, i.e., the two-way channel.

On a more technical level, we extend Willem's coding scheme so as to convey also state-information from one Tx to the other over the communication path built over the generalized feedback link. The proposed scheme can be considered as a separate source-channel coding scheme in the sense that each Tx first compresses the obtained outputs and inputs so as to extract state information, and then transmits the compression index using a pure channel code (here Willems' coding scheme) to the other Tx. As a result, the proposed scheme obtains a better sensing performance than the previous ISAC scheme [2] without collaborative sensing, and thus a better distortion-capacity tradeoff. A related idea was previously used in [23, 24] for the state-dependent MAC, where the Txs compress and transmit their state information to the Rx. In their setup, the transmission of the state is beneficial over pure data transmission because it helps the Rx to decode the data.

We further present a collaborative sensing ISAC scheme for the two-way channel, which models device-to-device communication (D2D), based on source-channel separation and use Han's two-way channel scheme. We also propose an improved scheme that is based on joint source-channel coding (JSCC), more specifically on hybrid coding [25], and show enhanced performances of both collaborative sensing schemes. In both the MAC and the D2D scenario, performances of the proposed ISAC schemes improve over classical time- or resource-sharing strategies.

Various other information-theoretic works have recently-analyzed the fundamental limits of ISAC systems [26–29]. For example, [29] analyzed systems with secrecy constraints, while [26–28] studied channels that depend on a single fixed parameter and Txs or sensor nodes wish to estimate this parameter based on backscattered signals. Their model is thus suited for scenarios where the estimation parameters change at a much slower time scale compared to the channel symbol period, while in [27] sensing (parameter estimation) is performed at the Tx, in [26] it is performed at a sensor that is close but not collocated with the Tx. The study in [28] analyzes the detection-error exponents of open-loop and close-loop coding strategies.

Related are also the information-theoretic the capacity-distortion tradeoff studies in which differ from our ISAC setup in that a state is estimated at the Rx and not at the Tx. [22, 30–33], but where the Rx estimates the state and not the Tx as in our ISAC setup.

1.2 Contribution and Organization of the Thesis

This thesis contains the following technical contributions:

1. Chapter 2 provides an introduction to sensing and communication as well as to the first approaches to integrating sensing and communication schemes. It also describes non-overlapped resource-sharing strategies for comparison purposes in the following chapters.

2. Chapter 3 review over related work. Specifically:

   - It presents the exact capacity-distortion-cost tradeoff for memoryless single-Tx single-Rx systems where the Tx wishes to estimate the state. It also proves the optimality of a symbol-by-symbol estimator for this setup. Moreover, we review results on slowly-varying channels.

   - It reviews the results of ISAC where the Rx estimates the state.

3. Chapter 4 presents the following results on single-Tx two-Rx broadcast channel (BC) systems::

- It presents inner and outer bounds on the capacity-distortion region for general BC channels. The inner bound is based on [17] and can be achieved using a block-Markov strategy that combines Marton coding with a lossy version of Gray-Wyner coding with side-information.
- We also characterize the capacity-distortion tradeoff region of physically degraded state-dependent memoryless BCs.
- The proposed inner and outer bounds match for physically degraded BCs, thus characterizing the exact capacity-distortion tradeoff region.

4. Chapter 5 introduces collaborative sensing for ISAC over the two-Tx single-Rx multiaccess channel (MAC) systems. It presents an improved inner bound on the fundamental rate-distortions tradeoff over two-Tx MACs based on a scheme where each Tx codes over the generalized feedback to improve the state estimation at the other Tx. We also show that it is of general nature and in particular can model scenarios with partial or perfect channel state information at the Rx as well as scenarios where the Txs wish to reconstruct functions or distorted versions of the actual state that is governing the channel.

5. Chapter 6 proposes two collaborative-sensing ISAC D2D schemes. The first is based on a separate source-channel coding approach and and the second on an improved JSCC approach using hybrid coding. In both schemes, the transmitted codeword carries not only data but also compression information that the other terminal can exploit for sensing. While the separation-based scheme employs Wyner-Ziv compression to account for the side-information at the other Tx, the JSCC based scheme uses implicitly binning as in standard hybrid coding.

1.3 Notation

We use calligraphic letters to denote sets, e.g., $\mathcal{X}$. The sets of real and nonnegative real numbers, however, are denoted by $\mathbb{R}$ and $\mathbb{R}_0^+$. Random variables are denoted by uppercase letters, e.g., X, and their realizations by lowercase letters, e.g., x. For vectors, we use boldface notation, i.e., lower case boldface letters such as $\boldsymbol{x}$ for deterministic vectors. We use $[1:X]$ to denote the set $\{1,\cdots,X\}$. We use X^n for the tuple of random variables $(X_1,\cdots,X_n)$. We abbreviate *independent and identically distributed* as *i.i.d.* and *probability mass function* as *pmf*. Logarithms are taken with respect to base 2.

We use the shorthand notations “Rx" for “Receiver" and “Tx" for “Transmitter". The set of all integers is denoted by $\mathbb{Z}$, the set of positive integers by $\mathbb{Z}^+$ and the set of real numbers by $\mathbb{R}$. For other sets we use calligraphic letters, e.g., $\mathcal{X}$. Random variables are denoted by uppercase letters, e.g., X, and their realizations by lowercase letters, e.g., x. For vectors we use boldface notation, i.e., upper case boldface letters such as $\mathbf{X}$ for random vectors and lower case boldface letters such as $\mathbf{x}$ for deterministic vectors.) Matrices are depicted with sans serif font, e.g., H. We write X^n for the tuple of random variables $(X_1, \ldots, X_n)$ and $\mathbf{X}^n$ for the tuple of random vectors $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$. We denote the entropy function by $H(\cdot)$, and the mutual information function by $I(\cdot)$. We use $\perp$ to indicate independence between random variables. Moreover, $\mathbb{1}\{\cdot\}$ denotes the indicator function. For vectors we use boldface notation, i.e., lower case boldface letters such as $\boldsymbol{x}$ for deterministic vectors. For positive integers n, we use $[1:n]$ to denote the set $\{1, \cdots, n\}$. We define $a * b \triangleq a\bar{b} + \bar{a}b$.

Chapter 2

Prelude: Sensing and Communication

The first part of this chapter commences with an introduction to radar systems, wireless communication and revisits some early studies related to the co-existence of radar and communication systems. It also describes co-existence solutions that we later use for comparison.

2.1 The Radar System

Radar is a system that utilizes radio waves to learn about positions, motions, or the mere presence of target objects in an environment through the analysis of backscattered signals. In fact, a radar terminal radiates a waveform that propagates through space until it reaches a target, where it is reflected in a way that depends on the properties of the target. The radar terminal collects and analyzes the backscattered signals so as to gain information about these properties. In this sense, the target almost acts like a passive transmitter and the radar terminal as a receiver. In the radar system, if the presence and position of a target are already known, the transmitter steers all the energy of transmit waveform towards the target, so as to obtain more information through the backscattered waveform. Thus the radar uses Line of Sight (LoS) techniques. Traditional radar systems mainly operate on the 24-79GHz band.

Sensing tasks can be roughly classified into three categories, detection, estimation, and recognition, which are all based on collecting signals/data concerning the sensed objects. Detection refers to making decisions on an object's *state* given some observations, such as the presence/absence of the target or other events related to the target. We can model the detection problem as a binary or multi-hypothesis testing problem. In the binary hypothesis testing problem as an example, we select from two hypotheses; the alternative hypothesis H_1 and the null hypothesis H_0. Detection metrics are the probability that H_1 holds but the detector chooses H_0 (miss-detection probability), and the probability that H_0 holds but the detector chooses H_1 (false-alarm probability).

Estimation performance can for example be measured by mean squared error (MSE). When the expectation of estimation is $E[\hat{S}] = S$ we call the estimator unbiased. In this case, MSE is equal to variance of S. The Cramér–Rao bound (CRB) expresses a lower bound on the variance of unbiased estimators of a deterministic (fixed, but unknown) parameter, the variance of any such estimator is at least as high as the inverse of the Fisher information.

Proposition 1. *Suppose S is an unknown deterministic r.v. which we want to estimate from m independent observations on Y; each from a distribution according to a given pdf $f(y, s)$. The variance of any unbiased estimator $\hat{S}$, defined as the inverse of the Fisher Information (FI), is lower bounded by*

$$var(\hat{S}) \geq \frac{1}{I(S)}, \tag{2.1}$$

where the Fisher information $I(S)$ is defined by

$$I(S) := m\mathbb{E}_s[\big(\frac{\partial \log(f(y,s))}{\partial s}\big)^2]. \tag{2.2}$$

This is the case when the estimated parameter is an scalar unbiased [34].

FI is the expectation of the curvature (negative second derivative) of the likelihood function concerning the valid parameter, which measures the "sharpness" or the estimator's accuracy. Estimation refers to extracting valuable parameters of the sensed object from observations. For example, distance/velocity/angle/quantity/size of targets are some possible parameters that a sensor desires to estimate.

One practical example in the wireless channel is the Gaussian channel, as we shall see in 3.30. However, MMSE is used to indicate the error of sensing in this Gaussian example in real-world models, and there are some difficulties in computing the closed form of it, which leads us to depend on bounds.

2.2 Wireless Communication System

Wireless communication systems mainly operate on the 2.4 GHz band. A transmitter wishes to transfer either data bits or source samples (for example, samples of an audio or video file) to a distant receiver. The data or source information is coded onto a transmitted waveform and the receiver collects and analyzes the propagated waveform to produce a guess of the transmitted information. The following performance metrics are usually considered for a communication system:

- *Energy or spectral efficiency* measure how many bits of information are communicated using a given energy budget or a given bandwidth.

- For data transmission, *bit-error rate (BER)*, *symbol-error rate (SER)*, or *frame-error rate (FER)* are used to measure the robustness of the communication. In fact, due to the disturbances in the communication channel, the data bits guessed at the receiver can be faulty.

- For source communication, robustness is either measured in bit-error rate or more often in distortion such as the average mean-squared error.

To outline the main differences between radar and communication systems, glance at Table 2.1.

Communication	Sensing
2.4 GHz	24-79 GHz
Data/Source Transmission	Estimation/Detection
Bit/Signal/Frame Rate Distortion	MMSE-CRB Detection/False Aarm prob.
All propagation paths	LoS

Table 2.1: Communication vs Sensing.

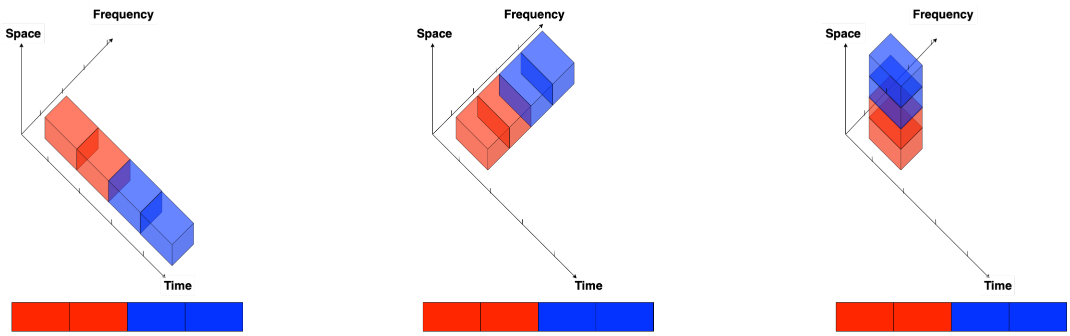
2.3 First Ideas of ISAC Systems

In this section, we introduce some basic ISAC ideas. The early work [35] modulates the communication bits on the missile range radar pulse interval. Interference rejection and robustness in multipath fading environments, inherent properties of spread spectrum systems, also make chirp (signals used in radar application) signaling very active for the expanding wireless communications market. Another approach in [36] as early as 1962 is based on chirp signals proposed for both analog and digital communication [37] but are also commonly used in radar applications. These works can be categorized as the first steps towards *Integrated Sensing and Communication (ISAC)*. In the survey [6], there is a fair string of evolution to pre-ISAC systems where a category of solutions is revisited. Some straightforward solutions are called *Non-Overlapped Resource Allocation*. In our information-theoretic model used throughout this thesis, such a system corresponds to time- or resource-sharing between communication and sensing; we shall call this *Basic time-sharing (TS)*, and with a minor modification, we will introduce *Improved time-sharing*.

2.4 Non-Overlapped Resource Allocation

A common but naive approach to address sensing and communication is to separate the two tasks into independent systems and split the available resources, such as bandwidth and power, between them so that they do not interfere. Time-division ISAC can be conveniently implemented into the existing commercial systems by splitting the transmission duration into radar and radio cycles, for example [38]. For radar sensing, frequency-modulated continuous waveform (FMCW) with up-and-down- chirp modulations were used, while various different modulation schemes (e.g., BPSK, PPM) can be used for communication.

In an OFDM system, frequency-division ISAC can implemented by allocating different communication and sensing tasks to different subcarriers as a function of the channel conditions and power budget of the Tx [39]. ISAC with non-overlapped resources can also be implemented over orthogonal spatial resources e.g., different antenna groups [40]. Thus, non-overlapping resource allocation can be performed in time, frequency, or spatial domains, as illustrated also in Figure 2.1.



(a) Time-sharing (b) Frequency-sharing (c) Spatial-sharing

Figure 2.1: The red cubes demonstrate the communication waveform, and the blue cubes demonstrate the sensing waveform.

This thesis considers two baseline schemes: the *Basic TS* scheme and the *Improved TS* scheme. The Basic TS scheme corresponds to the non-overlapped resource allocation approach, and splits its resources (time, bandwidth or spatial dimensions) between the following two modes:

- *Sensing mode*

  The system aims to design a suitable waveform to attain the minimum possible distortion. In our model, the waveform is translated into an input distribution; thus, the input pmf P_X is chosen to minimize the distortion, and hence the minimum distortion is achieved. Communication rate is zero.

- *Communication mode*

The system is designed to transfer as much reliable data as possible. Therefore, the input distribution is chosen to maximize the rate and communicates rate equals channel capacity. The estimator is set to a constant value regardless of the feedback and the input signals. The mode thus suffers from a large distortion.

The improved time-sharing *Improved TS* scheme still performs a sort of non-overlapped resource allocation, but resources are not exclusively dedicated to only sensing or only communication. It is simply that one of these tasks is prioritized. The second baseline scheme is called *Improved TS scheme* and can simultaneously perform the communication and sensing tasks. This scheme time-shares between the following modes.

- *Sensing mode with communication*

  The input pmf P_X is chosen to achieve minimum distortion. At the same time, the transmitter is also equipped with a communication encoder. It uses this input pmf to simultaneously transmit data at the rate given by the input-output mutual information of the system.

- *Communication mode with sensing*

  The input distribution is chosen to maximize the communication rate. i.e. achieve the capacity of the channel. The transmitter is however also equipped with a radar estimation device that optimally guesses the state-sequence based on the transmitted and backscattered signals.

We will see in the next chapter that all non-extreme operating points of the Basic and Improved TS schemes are typically suboptimal compared to an optimal ISAC scheme.

Chapter 3

An Information-Theoretic Model and Results

In this chapter, we review the single-Tx single-Rx as a P2P channel [1]. Notice that the model in [1] was the first information-theoretic model of an ISAC system. However, we slightly generalize the model and results in [1] by relaxing the assumption that the channel states are necessarily revealed to the Rx. As we shall explain, our model allows for any kind of channel state information (CSI) at the Rx. The model also includes as special cases scenarios where the Tx is interested in sensing a state that is related but not necessarily equal to the channel state.

The capacity-distortion-cost tradeoff is characterized, which allows us to quantify the merit of an optimal ISAC scheme over the baseline schemes introduced in the previous chapter. The results show that without loss in optimality, the communication scheme can ignore the generalized feedback signals, which are only used for state sensing. Communication and sensing performances depend on each other through the choice of the common waveform. The results further show a tradeoff between the simultaneously achievable sensing and communication performances arises in most situations except for “matched" situations where the same waveform simultaneously achieves capacity and minimum distortion. A Blahut-Arimoto type algorithm is used to evaluate the capacity-distortion-cost tradeoff numerically.

3.1 System Model

We describe a slight generalization of the first information-theoretic ISAC model of [1]. Consider the P2P communication scenario depicted in Fig. 3.1, where a Tx wishes to communicate a message to a Rx over a memoryless state-dependent channel and simultaneously estimate the state from generalized feedback.

The channel output at the Rx Y_i and the feedback signal Z_i at a given time i are generated according to its stationary channel law $P_{YZ|XS}(\cdot,\cdot|x_i,s_i)$ given the time-i channel input $X_i = x_i$ and state realization $S_i = s_i$,

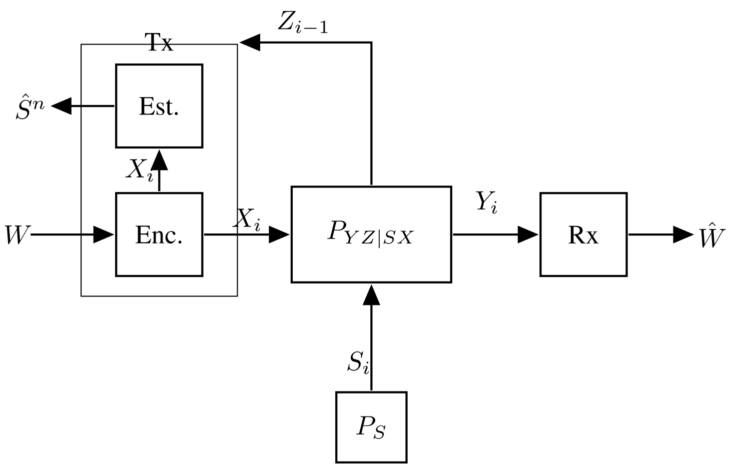


Figure 3.1: Joint sensing and communication model.

irrespective of the past inputs, outputs and state signals. Except for some Gaussian examples, assume that the channel states S_i, inputs X_i, outputs Y_i, and feedback signals Z_i take value in finite sets $\mathcal{S}$, $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, respectively. The state sequence $\{S_i\}_{i\geq 1}$ is assumed i.i.d. according to a given state distribution $P_S(\cdot)$.

A $(2^{n\mathsf{R}}, n)$ code for the state-dependent memoryless channel (SDMC) consists of

1. a discrete message set $\mathcal{W}$ of size $|\mathcal{W}| \geq 2^{n\mathsf{R}}$;

2. a sequence of encoding functions $\phi_i \colon \mathcal{W} \times \mathcal{Z}^{i-1} \to \mathcal{X}$, for $i = 1, 2, \ldots, n$;

3. a decoding function $g \colon \mathcal{Y}^n \to \mathcal{W}$;

4. a state estimator $h \colon \mathcal{X}^n \times \mathcal{Z}^n \to \hat{\mathcal{S}}^n$, where $\hat{\mathcal{S}}$ denotes a given finite reconstruction alphabet.

For a given code, the random message W is uniformly distributed over the message set $\mathcal{W}$ and the inputs are obtained as $X_i = \phi_i(W, Z^{i-1})$, for $i = 1, \ldots, n$. The corresponding channel outputs Y_i and Z_i at time i are obtained from the state S_i and the input X_i according to the SDMC transition law $P_{YZ|SX}$. Let $\hat{S}^n := (\hat{S}_1, \cdots, \hat{S}_n) = h(X^n, Z^n)$ denote the state estimate at the Tx and $\hat{W} = g(Y^n)$ the decoded message at the receiver.

The quality of the state estimates is measured by the expected average per-block distortion

$$\Delta^{(n)} := \mathbb{E}[d(S^n, \hat{S}^n)] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[d(S_i, \hat{S}_i)], \tag{3.1}$$

where $d : \mathcal{S} \times \hat{\mathcal{S}} \mapsto \mathbb{R}_0^+$ is a given bounded *distortion function*:

$$\max_{(s,\hat{s})\in\mathcal{S}\times\hat{\mathcal{S}}} d(s, \hat{s}) < \infty. \tag{3.2}$$

In practical communication systems, we typically impose an expected cost constraint on the channel inputs such as an average or peak power constraint. These cost constraints can often be expressed as

$$\mathbb{E}[b(X^n)] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[b(X_i)], \tag{3.3}$$

for some given cost functions $b\colon \mathcal{X} \mapsto \mathbb{R}_0^+$.

Remark 1 (Equivalent Model)**.** *Consider the related model with a triple of sequences $(\tilde{S}^n, \tilde{S}_T^n, \tilde{S}_R^n)$ that are i.i.d. according to a joint pmf $P_{\tilde{S},\tilde{S}_T,\tilde{S}_R}$ and where the Tx estimates $\tilde{S}_T^n$, the Rx observes $\tilde{S}_R^n$, and the two terminals observe the generalized feedback signals and channel outputs produced by the SDMC $P_{\tilde{Y}Z|X\tilde{S}}$. See Figure 3.2.This seemingly more general model is equivalent to our model if we set $Y = (\tilde{Y}, \tilde{S}_R)$, $S = \tilde{S}_T$ and*

$$P_{YZ|XS}((\tilde{y}, \tilde{s}_R), z \mid x, s) = \sum_{\tilde{s}\in\mathcal{S}} \frac{P_{\tilde{Y}Z|X\tilde{S}}(\tilde{y}, z|x, \tilde{s})}{P_{\tilde{S}\tilde{S}_T\tilde{S}_R}(\tilde{s}, \tilde{s}_T, \tilde{s}_R)} \tag{3.4}$$

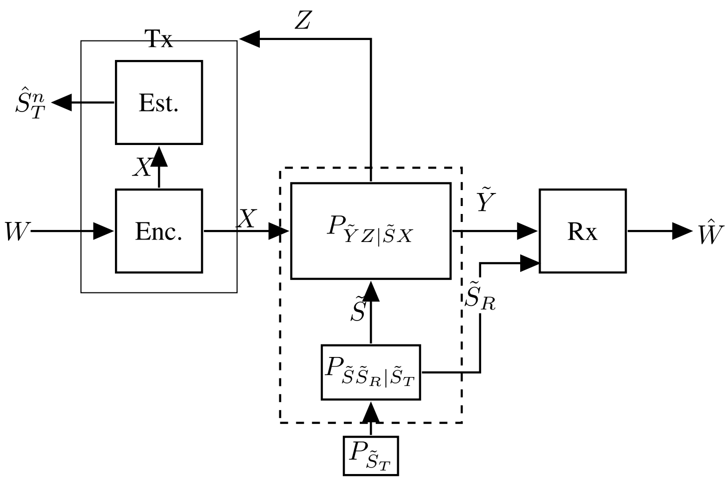


Figure 3.2: An equivalent model to the model in [1]

3.2 Capacity-Distortion-Cost Tradeoff

We start by defining the desired performance for communication and distortion, followed by the definition and properties of the capacity-distortion-cost function, and the main theorem of [1].

Definition 1. *A rate-distortion-cost tuple* $(\mathsf{R}, \mathsf{D}, \mathsf{B})$ *is said achievable if there exists a sequence (in n) of* $(2^{n\mathsf{R}}, n)$

codes that simultaneously satisfy

$$\lim_{n\to\infty} P_e^{(n)} = 0, \tag{3.5a}$$

$$\overline{\lim_{n\to\infty}}\, \Delta^{(n)} \leq \mathsf{D}, \tag{3.5b}$$

$$\overline{\lim_{n\to\infty}}\, \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[b(X_i)] \leq \mathsf{B}, \tag{3.5c}$$

for $P_e^{(n)} := \Pr\left(\hat{W} \neq W\right)$.

The capacity-distortion-cost tradeoff $\mathcal{C}(\mathsf{D},\mathsf{B})$ *is the largest rate* R *such that the rate-distortion-cost tuple* $(\mathsf{R},\mathsf{D},\mathsf{B})$ *is achievable.*

In order to characterize useful properties of the capacity-distortion-cost function, we define the following sets:

$$\mathcal{P}_{\mathsf{B}} = \left\{ P_X \,\middle|\, \sum_{x\in\mathcal{X}} P_X(x) b(x) \leq \mathsf{B} \right\}, \tag{3.6a}$$

$$\mathcal{P}_{\mathsf{D}} = \left\{ P_X \,\middle|\, \sum_{x\in\mathcal{X}} P_X(x) c(x) \leq \mathsf{D} \right\}. \tag{3.6b}$$

The minimum distortion for a given cost B is given by

$$\mathsf{D}_{\min}(\mathsf{B}) := \min_{P_X\in\mathcal{P}_{\mathsf{B}}} \sum_{x\in\mathcal{X}} P_X(x) c(x). \tag{3.7}$$

Definition 2. *Define the* information-theoretic tradeoff function $\mathcal{C}_{\text{inf}} : [\mathsf{D}_{\min}(\mathsf{B}),\infty) \times [0,\infty) \to \mathbb{R}_0^+$ *as*

$$\mathcal{C}_{\text{inf}}(\mathsf{D},\mathsf{B}) := \max_{P_X\in\mathcal{P}_{\mathsf{D}}\cap\mathcal{P}_{\mathsf{B}}} I(X;Y), \tag{3.8}$$

where $(X,S,Y,Z) \sim P_X P_S P_{YZ|SX}$ *and the maximum is over all* P_X *satisfying both the distortion and cost constraints* (3.6b) *and* (3.6a).

Lemma 1. *Given a SDMC* $P_{YZ|SX}$ *with state-distribution* P_S*, the capacity-distortion-cost tradeoff function* $\mathcal{C}_{\text{inf}}(\mathsf{D},\mathsf{B})$ *has the following properties.*

i) $\mathcal{C}_{\text{inf}}(\mathsf{D},\mathsf{B})$ *is non-decreasing and concave in* $\mathsf{D} \geq \mathsf{D}_{\min}(\mathsf{B})$ *and* $\mathsf{B} \geq 0$.

ii) $\mathcal{C}_{\text{inf}}(\mathsf{D},\mathsf{B})$ *saturates at channel capacity:*

$$\mathcal{C}_{\text{inf}}(\mathsf{D},\mathsf{B}) = \mathcal{C}_{\text{NoEst}}(\mathsf{B}), \quad \forall \mathsf{D} \geq \mathsf{D}_{\max}(\mathsf{B}), \tag{3.9}$$

where $\mathcal{C}_{\text{NoEst}}(\mathsf{B}) := \max_{P_X \in \mathcal{P}_{\mathsf{B}}} I(X;Y|S)$ *denotes the classical channel capacity of the SDMC for a given cost* B*, and* $\mathsf{D}_{\max}(\mathsf{B})$ *denotes the corresponding distortion*

$$\mathsf{D}_{\max}(\mathsf{B}) := \sum_{x \in \mathcal{X}} P_{X_{\max}}(x) c(x), \tag{3.10}$$

for $P_{X_{\max}} := \operatorname{argmax}_{P_X \in \mathcal{P}_{\mathsf{B}}} I(X;Y)$.

Proof: The proof is a straightforward extension of [33, Corollary 1] to the case of two cost functions and the state dependent channel. The nondecreasing property follows immediately from the definition in (3.8) because we have $\mathcal{P}_{\mathsf{D}_1} \subseteq \mathcal{P}_{\mathsf{D}_2}$ and $\mathcal{P}_{\mathsf{B}_1} \subseteq \mathcal{P}_{\mathsf{B}_2}$ for any $\mathsf{D}_1 \leq \mathsf{D}_2$ and $\mathsf{B}_1 \leq \mathsf{B}_2$.

In order to verify the concavity of $\mathcal{C}_{\text{inf}}(\mathsf{D}, \mathsf{B})$ with respect to (D, B), we consider time-sharing between two input distributions, denoted by $P_X^{(1)}$ and $P_X^{(2)}$, that achieve $\mathcal{C}_{\text{inf}}(\mathsf{D}_1, \mathsf{B}_1)$ and $\mathcal{C}_{\text{inf}}(\mathsf{D}_2, \mathsf{B}_2)$, respectively. To make the dependency of the mutual information with respect to the input distribution more explicit, we adapt the following notation: for any pmf P_X over the input alphabet $\mathcal{X}$, let $\mathcal{I}(P_X, P_{Y|XS}) := I(X;Y)$ for $(S, X, Y) \sim P_S P_X P_{Y|XS}$.

For any $\theta \in (0, 1)$, we have:

$$\begin{aligned}
\theta \mathcal{C}_{\text{inf}}(\mathsf{D}_1, \mathsf{B}_1) + (1-\theta)\mathcal{C}_{\text{inf}}(\mathsf{D}_2, \mathsf{B}_2) &\overset{(a)}{=} \theta \mathcal{I}\left(P_X^{(1)}, P_{Y|XS}\right) + (1-\theta)\mathcal{I}\left(P_X^{(2)}, P_{Y|XS}\right) \\
&\overset{(b)}{\leq} \mathcal{I}\left(\theta P_X^{(1)} + (1-\theta) P_X^{(2)}, P_{Y|XS}\right) \\
&\overset{(c)}{=} \mathcal{C}_{\text{inf}}\left(\theta \mathsf{D}_1 + (1-\theta)\mathsf{D}_2, \theta \mathsf{B}_1 + (1-\theta)\mathsf{B}_2\right).
\end{aligned} \tag{3.11}$$

where (a) follows by definition, (b) follows from the concavity of the mutual information functional with respect to the input distribution, (c) follows by the linearity of the constraints and because for any $k = 1, 2$ the pmf $P_X^{(k)}$ has expected cost no larger than B_k and expected distortion no larger than D_k.

This establishes the concavity of $\mathcal{C}_{\text{inf}}(\mathsf{D}, \mathsf{B})$. ■

3.3 Main Results

The main result of this chapter is an exact characterization of $\mathcal{C}(\mathsf{D}, \mathsf{B})$. We begin by describing the optimal estimator h, which is independent of the choice of encoding and decoding functions, and operates on a symbol-by-symbol basis, i.e., it computes estimate $\hat{S}_i$ only in function of X_i and Z_i but not of the other inputs and feedback signals.

Lemma 2. *Define the function*

$$\hat{s}^*(x,z) := \arg\min_{s'\in\hat{\mathcal{S}}} \sum_{s\in\mathcal{S}} P_{S|XZ}(s|x,z)d(s,s'), \tag{3.12}$$

where ties can be broken arbitrarily and

$$P_{S|XZ}(s|x,z) = \frac{P_S(s)P_{Z|SX}(z|s,x)}{\sum_{\tilde{s}\in\mathcal{S}} P_S(\tilde{s})P_{Z|SX}(z|\tilde{s},x)}. \tag{3.13}$$

Irrespective of the choice of encoding and decoding functions, distortion $\Delta^{(n)}$ in (3.5b) *is minimized by the estimator*

$$h^*(x^n,z^n) := (\hat{s}^*(x_1,z_1), \hat{s}^*(x_2,z_2), \ldots, \hat{s}^*(x_n,z_n)). \tag{3.14}$$

Notice that the function $\hat{s}(\cdot,\cdot)$ only depends on the SDMC channel law $P_{YZ|SX}$ and the state distribution P_S.

Proof: See Appendix A.1.1. ■

Lemma 2 implies that one can focus without loss in optimality on a symbol-by-symbol deterministic estimator. Based on (3.12), we define the estimation cost $c(x)$ of the optimal estimator

$$c(x) := \mathsf{E}[d(S,\hat{s}^*(X,Z))|X=x]\,. \tag{3.15}$$

We now are prepared to state the main theorem.

Theorem 1. *The capacity-distortion-cost tradeoff of a SDMC $P_{YZ|SX}$ with state-distribution P_S is:*

$$\mathcal{C}(\mathsf{D},\mathsf{B}) = \mathcal{C}_{\text{inf}}(\mathsf{D},\mathsf{B}), \quad \mathsf{D} \geq \mathsf{D}_{\min}(\mathsf{B}), \ \ \mathsf{B} \geq 0. \tag{3.16}$$

Proof: The proof of Theorem 1 is similar to the proof of the classic capacity-cost function [41], except that one also has to account for the sensing performance. Both in the converse proof and the achievability proof, this can be accomplished by evaluating the performance of the optimal (per-symbol) estimator $\hat{s}^*(\cdot,\cdot)$ in Lemma 2. In particular, a standard random coding argument can be used to prove achievability of Theorem 1. ■

Capacity of a memoryless channel is known to be achieved with i.i.d. inputs. Also because of the memoryless nature of the optimal estimator $h(\cdot,\cdot)$ in Lemma 2, this observation extends to the joint sensing and communication setup.

Combining Lemma 1 and Theorem 1, one can conclude that the rate-distortion tradeoff function $\mathcal{C}(\mathsf{D},\mathsf{B})$ is nondecreasing and concave in $\mathsf{D} \geq \mathsf{D}_{\min}(\mathsf{B})$ and $\mathsf{B} \geq 0$, and for any $\mathsf{B} \geq 0$ it saturates at the channel capacity $\mathcal{C}_{\text{NoEst}}(\mathsf{B})$. For many channels, given $\mathsf{B} \geq 0$, the tradeoff $\mathcal{C}(\mathsf{D}, \mathsf{B})$ is strictly increasing in D until it reaches $\mathcal{C}_{\text{NoEst}}(\mathsf{B})$. However, for SDMBCs and costs $\mathsf{B} \geq 0$ where the capacity-achieving input distribution $P_{X_{\max}} := \operatorname{argmax}_{P_X \in \mathcal{P}_\mathsf{B}} I(X; Y \mid S)$ also achieves minimum distortion $\mathsf{D}_{\min}(\mathsf{B})$ in (3.7), the capacity-distortion tradeoff is constant $\mathcal{C}(\mathsf{D}, \mathsf{B}) = \mathcal{C}_{\text{NoEst}}(\mathsf{B})$, irrespective of the allowed distortion D.

Remark 2. *As mentioned in Remark 1, our model includes the scenario with imperfect CSI at Rx as a special case. Theorem 1 is easily adapted to this special case where the Rx observes sequence S_R^n for (S^n, S_R^n) i.i.d. according to an arbitrary distribution P_{SS_R}, as*

$$\mathcal{C}^{\text{imp}}(\mathsf{D}, \mathsf{B}) = \max_{P_X \in \mathcal{P}_\mathsf{D} \cap \mathcal{P}_\mathsf{B}} I(X; Y S_R) = \max_{P_X \in \mathcal{P}_\mathsf{D} \cap \mathcal{P}_\mathsf{B}} I(X; Y \mid S_R), \qquad \mathsf{D} \geq \mathsf{D}_{\min}(\mathsf{B}), \;\; \mathsf{B} \geq 0, \tag{3.17}$$

where $(X, S_R, Y, Z) \sim \sum_S P_X P_{SS_R} P_{YZ|SS_RX}$ and the definitions of the sets $\mathcal{P}_\mathsf{B}$ and $\mathcal{P}_\mathsf{D}$ are kept as in (3.6a) *and* (3.6b)*, same as the definition of the function $c(x)$ in* (3.15)*.*
Notice that the symbolwise estimator in (3.14) *remains optimal also in this related setup.*

Before going through the examples, it is beneficial to see some naive solutions, with the help of which we can arrive at the conclusion that the ISAC scheme performs better than them.

3.4 Baseline Schemes

We review the previously introduced TS schemes with more technical details. Recall that the Basic TS scheme time-shares between the following modes:

- *Sensing mode without communication (achieves rate-distortion pair $(0, \mathsf{D}_{\min}(\mathsf{B}))$)*

  The input pmf P_X is chosen to minimize the distortion:

  $$P_{X_{\min}} := \operatorname*{argmin}_{P_X \in \mathcal{P}_\mathcal{B}} \sum_x P_X(x) c(x), \tag{3.18}$$

  and thus the minimum distortion $\mathsf{D}_{\min}(\mathsf{B})$ defined in (3.7) is achieved. Communication rate is zero.

- *Communication mode without sensing (achieves $(\mathcal{C}_{\text{NoEst}}(\mathsf{B}), \mathsf{D}_{\text{trivial}}(\mathsf{B}))$)*

The input pmf P_X is chosen to maximize the rate:

$$P_{X_{\max}} = \operatorname*{argmax}_{P_X \in \mathcal{P}_{\mathcal{B}}} I(X;Y), \tag{3.19}$$

and this mode thus communicates at a rate equal to the channel capacity $\mathcal{C}_{\text{NoEst}}(\mathsf{B})$. The estimator is set to a constant value regardless of the feedback and the input signals. The mode thus achieves distortion

$$\mathsf{D}_{\text{trivial}}(\mathsf{B}) := \min_{s' \in \hat{\mathcal{S}}} \sum_{s \in \mathcal{S}} P_S(s) d(s, s'). \tag{3.20}$$

The Improved TS scheme time-shares between the following two modes.

- <u>*Sensing mode with communication*</u> *(achieves* $(\mathsf{R}_{\min}(\mathsf{B}), \mathsf{D}_{\min}(\mathsf{B}))$*)*

  The input pmf P_X is choosen according to (3.18) to achieve the minimum distortion. The chosen pmf $P_{X_{\min}}$ can achieve the following communication rate:

$$\mathsf{R}_{\min} := I(X_{\min}; Y), \qquad \text{for } X_{\min} \sim P_{X_{\min}}. \tag{3.21}$$

- <u>*Communication mode with sensing*</u> *(achieves* $(\mathcal{C}_{\text{NoEst}}(\mathsf{B}), \mathsf{D}_{\max}(\mathsf{B}))$*)*

  The input pmf $P_{X_{\max}}$ is chosen as in (3.19) to maximize the communication rate. The mode thus communicates at the capacity $\mathcal{C}_{\text{NoEst}}(\mathsf{B})$ of the channel. Sensing is performed by means of the optimal estimator in (3.12). The mode thus achieves distortion

$$\mathsf{D}_{\max} := \sum_{x \in \mathcal{X}} P_{X_{\max}}(x) c(x), \quad \text{for } X_{\max} \sim P_{X_{\max}}. \tag{3.22}$$

It is worth noticing that for any cost $\mathsf{B} \geq 0$, the two operating points of the two modes in the Improved TS scheme, $(\mathsf{R}_{\min}(\mathsf{B}), \mathsf{D}_{\min}(\mathsf{B}))$ and $(\mathcal{C}_{\text{NoEst}}(\mathsf{B}), \mathsf{D}_{\max}(\mathsf{B}))$, also lie on the capacity-distortion-cost tradeoff curve $\mathcal{C}(\mathsf{D}, \mathsf{B})$ presented in Theorem 1. These two points are extreme operating points of optimal ISAC schemes. As we will see at hand of the following examples, all other operating points of the Improved TS scheme are typically suboptimal compared to an optimal ISAC scheme.

3.5 Examples

To evaluate Theorem 1 without unnecessary complication, we assume perfect CSI at Rx. The capacity-cost-distortion tradeoff thus is given by (3.17) for $S_R = S$.

3.5.1 Binary Channel with Multiplicative Bernoulli State

Consider a channel $Y = SX$ with binary alphabets $\mathcal{X} = \mathcal{S} = \mathcal{Y} = \{0, 1\}$ and where the state S is Bernoulli-q, for $q \in (0, 1)$. The feedback to to the Tx is perfect $Y = Z$ and Hamming distortion measure $d(s, \hat{s}) = s \oplus \hat{s}$ is used for sensing. No cost constraint is imposed.

The following corollary specializes Theorem 1 to this example.

Corollary 1. *The capacity-distortion tradeoff of a binary channel with multiplicative Bernoulli state is given by*

$$\mathcal{C}(\mathsf{D}) = qH_{\mathsf{b}}\left(\frac{\mathsf{D}}{\min\{q, 1-q\}}\right), \tag{3.23}$$

where $H_{\mathsf{b}}(p)$ denotes the binary entropy function. In other words, the curve $\mathcal{C}(\mathsf{D})$ is parameterized as

$$\{(\mathcal{C} = qH_{\mathsf{b}}(p),\ \mathsf{D} = p\min\{q, 1-q\}) \colon p \in [0, 1/2]\}. \tag{3.24}$$

Proof: Since Y is deterministic given (S, X), and it equals 0 whenever $S = 0$, so following equalities are valid:

$$I(X; Y, S) = I(X; Y \mid S) = P_S(0)H(Y \mid S = 0) + P_S(1)H(Y \mid S = 1) = P_S(1)H(X). \tag{3.25}$$

Setting $p := P_X(0)$, following is obtained

$$I(X; Y \mid S) = qH_{\mathsf{b}}(p). \tag{3.26}$$

To calculate the distortion, notice the optimal estimator $\hat{s}^*(\cdot, \cdot)$ in Lemma 2 sets

$$\hat{s}^*(x, z) = \begin{cases} z, & \text{if } x = 1 \\ \operatorname{argmax}_{s\in\{0,1\}} P_S(s), & \text{if } x = 0. \end{cases} \tag{3.27}$$

In fact, whenever $x = 1$ the Tx acquires full state knowledge because $z = y = s$. In this case $c(x = 1) = 0$. For

$x = 0$, the Tx does not receive any useful information about the state and hence uses the best constant estimator, irrespective of the feedback z. In this case,

$$c(x=0) = \mathsf{E}\left[d\Big(S, \operatorname*{argmax}_{s\in\{0,1\}} P_S(s)\Big)\Big| X=0\right] = \min_{s\in\{0,1\}} P_S(s) = \min\{q, 1-q\}, \tag{3.28}$$

where the independence of S and X is used. The expected distortion of the optimal estimator thus evaluates to:

$$\mathsf{D} = \sum_x P_X(x)c(x) = P_X(0)c(0) = p\min\{q, 1-q\}. \tag{3.29}$$

■

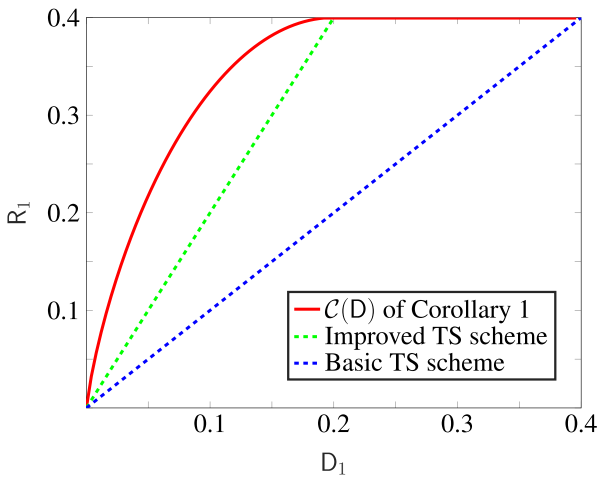


Figure 3.3: Capacity-distortion tradeoff of the binary channel with multiplicative Bernoulli state of parameter $q = 0.4$.

The capacity-distortion tradeoff of Corollary 1 is illustrated in Fig. 3.3 for state parameter $q = 0.4$. The figure also presents the performances of the two baseline TS schemes, which perform significantly worse than the optimal ISAC scheme. In the following, we derive the performance of the TS schemes. This example with a derivation of the parameters of the TS schemes is concluded.

The capacity-achieving input distribution for this channel is easily found as $P_{X_{\max}}(0) = P_{X_{\max}}(1) = 1/2$, and by (3.26) and (3.29) one can find $\mathcal{C}_{\text{NoEst}} = q$ and $\mathsf{D}_{\max} = \min\{q, 1-q\}/2$. Minimum distortion $\mathsf{D}_{\min} = 0$ is achieved by always sending $X = 1$, i.e., $P_{X_{\min}}(1) = 1$ and $P_{X_{\min}}(0) = 0$, in which case $D_{\min} = 0$ and $R_{\min} = 0$, see also (3.26) and (3.29). The Improved TS scheme thus achieves all pairs on the line connecting the two points $(0, 0)$ with $(q, \min\{q, 1-q\}/2)$. To determine the performance of the basic TS scheme, recall that the best constant estimator (that does not consider the feedback) which is $\hat{s}_{\text{const}} = \operatorname{argmax}_{s\in\{0,1\}} P_S(s)$, which allows to conclude that $\mathsf{D}_{\text{trivial}} = \min\{q, 1-q\}$. The basic TS scheme thus achieves all rate-distortion pairs on the line connecting the points $(0, 0)$ and $(q, \min\{q, 1-q\})$.

3.5.2 Real Gaussian Channel with Rayleigh Fading

In this example, the real Gaussian channel with Rayleigh fading is considered:

$$Y_i = S_i X_i + N_i, \tag{3.30}$$

where X_i is the channel input satisfying $\overline{\lim}_{n\to\infty} \frac{1}{n}\sum_i \mathsf{E}\left[|X_i|^2\right] \le \mathsf{B} = 10\text{dB}$, and both sequences $\{N_i\}$ and $\{S_i\}$ are independent of each other and i.i.d. Gaussian with zero mean and unit variance. The Tx observes the noisy feedback

$$Z_i = Y_i + N_{\text{fb},i}, \tag{3.31}$$

where $\{N_{\text{fb},i}\}$ are i.i.d. zero-mean Gaussian of variance $\sigma_{\text{fb}}^2 \ge 0$. Consider the quadratic distortion measure $d(s, \hat{s}) = (s - \hat{s})^2$.

First, the two operating points achieved by the Improved TS baseline scheme are characterized. The capacity of this channel is achieved with a Gaussian input $X_{\max} \sim \mathcal{N}(0, \mathsf{B})$, and thus the communication mode with sensing achieves the rate-distortion pair

$$\mathcal{C}_{\text{NoEst}}(\mathsf{B}) = \frac{1}{2}\mathsf{E}\left[\log(1 + |S|^2\mathsf{B})\right] = 1.213, \tag{3.32}$$

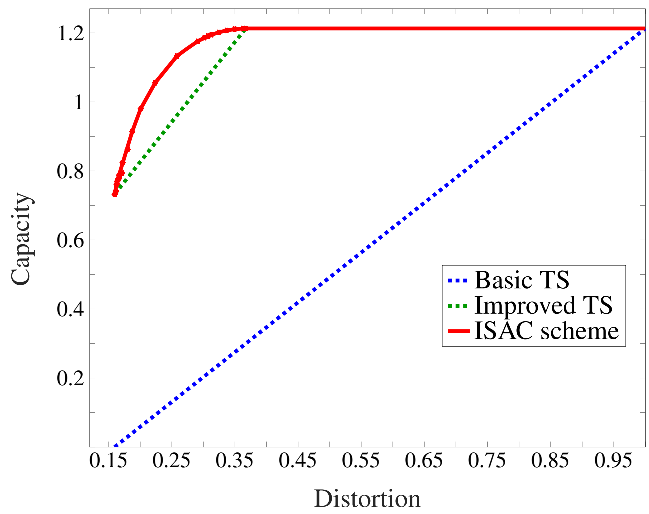


Figure 3.4: Capacity-distortion tradeoff of fading AWGN channel $\mathsf{B} = 10$ dB and $\sigma_{\text{fb}}^2 = 1$.

$$\mathsf{D}_{\max}(\mathsf{B}) = \mathsf{E}\left[\frac{(1+\sigma_{\text{fb}}^2)}{1+|X_{\max}|^2+\sigma_{\text{fb}}^2}\right] = 0.367, \tag{3.33}$$

where they have set $\sigma_{\text{fb}}^2 = 1$ and $P = 10$dB to obtain the numerical values. Minimum distortion $\mathsf{D}_{\min}$ is achieved by 2-ary pulse amplitude modulation (PAM), and thus the sensing mode with communication achieves rate-distortion pair

$$\mathsf{R}_{\min}(\mathsf{B}) = 0.733, \quad \mathsf{D}_{\min}(\mathsf{B}) = \frac{1+\sigma_{\text{fb}}^2}{1+P+\sigma_{\text{fb}}^2} = 0.166, \tag{3.34}$$

where the numerical value again corresponds to $\sigma_{\text{fb}} = 1$ and $B = 10$dB. Next, they characterize the performance of the basic TS baseline scheme. The best constant estimator for this channel is $\hat{s} = 0$, and the communication mode without sensing achieves rate-distortion pair $(\mathcal{C}_{\text{NoEst}}(\mathsf{B}), \mathsf{D}_{\text{trivial}}(\mathsf{B}) = 1)$. The sensing mode without communication achieves rate-distortion pair $(0, \mathsf{D}_{\min}(\mathsf{B}))$.

In Fig. 3.4, the rate-distortion tradeoff achieved by these two TS baseline schemes is compared with a numerical approximation of the capacity-distortion-cost tradeoff $\mathcal{C}(\mathsf{D}, \mathsf{B})$ of this channel. As previously explained, $\mathcal{C}(\mathsf{D}, \mathsf{B})$ also passes through the two end points $(\mathsf{R}_{\min}(\mathsf{B}), \mathsf{D}_{\min}(\mathsf{B}))$ and $(\mathcal{C}_{\text{NoEst}}(\mathsf{B}), \mathsf{D}_{\max}(\mathsf{B}))$ of the Improved TS scheme. To obtain a numerical approximation of the points on $\mathcal{C}(\mathsf{D}, \mathsf{B})$ in between these two operating points they use the Blahut-Arimoto type [1]. Specifically, the input alphabet is quantized to a $M = 16$-ary PAM constellation

$$\mathcal{X}_{\text{q}} := \{(2m-1-M)\kappa, m = 1, \cdots, M\}, \tag{3.35}$$

where $\kappa := \sqrt{3P/(M^2-1)}$. The Gaussian noise N is quantized with a centered equally-spaced 50-points alphabet, and the state S is quantized by applying an equally-spaced 8000-points quantizer on the Chi-square distributed random variable S^2. Denoting the quantized input, noise, and state by X_{q}, N_{q}, and S_{q}, they keep the multiplicative-state, additive-noise channel model to generate the channel outputs used to run Algorithm 1 to obtain the numerical approximations:

$$Y_{\text{q}} = S_{\text{q}} X_{\text{q}} + N_{\text{q}}. \tag{3.36}$$

3.6 Related Information-Theoretic Works on P2P ISAC Systems

In this section, we first review related information-theoretic works. We start in Subsection 3.6.1 with a model that is similar to the one in Section 3.1, but where the state does not consist of a sequence but only of a single parameter. In the following Subsection 3.6.2, we again consider i.i.d. state-sequences, but state estimation is performed at the Rx while the Tx may have access to some side-information.

3.6.1 Slowly-Varying State

Reconsider the model introduced in Section3.1, where now the state varies slower than the blocklength of communication [26–28] and is thus modeled as a single parameter

$$S_1 = S, \tag{3.37}$$

$$S_2 = \cdots = S_n = S_1, \tag{3.38}$$

where S is distributed according to P_S. In this model, the Tx produces a single estimate of the state as

$$\hat{S} = h(X^n, Z^n). \tag{3.39}$$

The model is depicted in Figure 3.5, where notice the switch from the Estimator to the Encoder, which indicates whether this latter can produce its channel input X_i also as a function of the past feedback outputs $Z_1, \ldots, Z_{i-1}$. Both scenarios, with closed and open switch, have been considered in the literature. Closed-switch systems are sometimes referred to as adaptive or closed-loop coding and mono-static radar, and open-switch systems as non-adapative or open-loop coding or bi-stating radar. In our previous model in Section 3.1 the switch was closed.

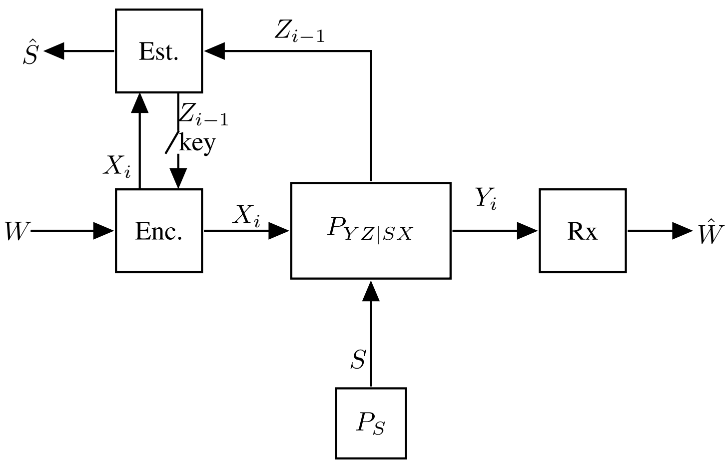


Figure 3.5: The slow-varying state channel.

The performance of the system is measured in terms of the asymptotic rate of reliable communication and asymptotic detection-error exponent which we discuss shortly.

Definition 3 (Definition 2, [28]). *For any $s \in \mathcal{S}$, define the state-dependent error probabilities as*

$$P_{c,s} \triangleq \max_{w} \Pr[g(Y^n) \neq w \mid W = w, S = s], \tag{3.40}$$

$$P_{d,s} \triangleq \frac{1}{|\mathcal{W}|} \sum_{\mathcal{W}} \Pr[h(Z^n, X^n) \neq s \mid W = w, S = s], \tag{3.41}$$

A rate-detection-error exponent (R, D) *is achievable if for any* $\epsilon > 0$*, there exists a sufficiently large* n *and* 2^{nR} *code so that for any* $s \in \mathcal{S}$*:*

$$P^n_{c,s} \leq \epsilon, \tag{3.42}$$

$$E^n_{d,s} \geq \mathsf{D} - \epsilon, \tag{3.43}$$

where the state-dependent error exponent is defined as $E^n_{d,s} \triangleq -\frac{1}{n} \log P^n_{d,s}$.

Theorem 2 (Theorem 5, [28]). *If the switch in Figure 3.5 is closed, then the set of all rate-detection-error exponent pairs contains all pairs* (R, D) *satisfying*

$$\mathsf{R} \leq I(X;Y), \tag{3.44}$$

$$\mathsf{D} \leq \psi_s(p_x), \tag{3.45}$$

for some P_X*, and where*

$$\psi_s(P_X) = \min_{s' \neq s} \max_{l \in [0,1]} - \sum_x P_X(x) \log \big(\sum_z P_{Z|XS}(z|x,s)^l P_{Z|XS}(z|x,s')^{1-l} \big). \tag{3.46}$$

Theorem 3 (Theorem 3, [28]). *If the switch in Figure 3.5 is open, then the set of all rate-detection-error exponent pairs is given by*

$$\mathsf{R} \leq \min_{s \in \mathcal{S}} I(X;Y), \tag{3.47}$$

$$\mathsf{D} \leq \phi(p_x), \tag{3.48}$$

where

$$\phi(P_X) = \min_{s \in \mathcal{S}} \min_{s' \neq s} \max_{l \in [0,1]} - \sum_x P_X(x) \log \big(\sum_z P_{Z|XS}(z|x,s)(z)^l P_{Z|XS}(z|x,s')^{1-l} \big). \tag{3.49}$$

The open loop system with a binary state $|\mathcal{S}| = 2$ was independently also solved in [26,27], but for setups where only the feedback channel depends on the state but not the forward channel, i.e., where

$$P_{YZ|XS} = P_{Y|X} P_{Z|XS}. \tag{3.50}$$

The work in [26] also considered the openloop system with Gaussian channels.

Reference [28] also discusses the benefits of adaptive over non-adaptive coding, i.e., they prove that the set of (R, D) pairs described in Theorem 2 is strictly larger than the set in Theorem 3. Notice that this property depends on the assumption of the constant state sequence $S_1 = \ldots = S_n$. In fact, as we had discussed previously, under our i.i.d. state-assumption adaptive and non-adaptive coding schemes perform equally well.

3.6.2 Sensing at the Rx

Consider a model similar to Section 3.1 where a Tx wishes to communicate over a P2P SDMC to a Rx, which now also wishes to estimate the channel state as we illustrate in Figure 3.6. The encoder might have a side-information on the state-sequence S^n which is now again assumed i.i.d. according to a given P_S. Depending on what side-information is available at encoder, we define an $(2^{n\mathsf{R}}, n)$ code to consist of

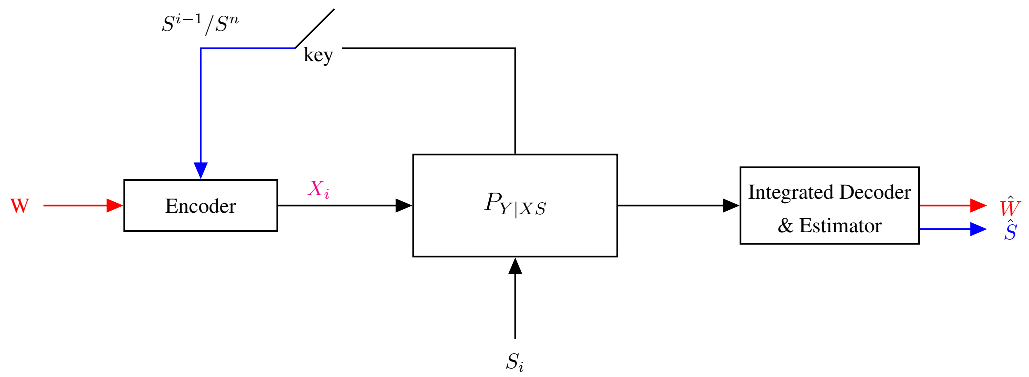


Figure 3.6: Joint communication and Rx channel estimation.

1. a discrete message set $\mathcal{W}$ of size $|\mathcal{W}| \geq 2^{n\mathsf{R}}$;

2. sequence of encoding functions $\{\phi_i\}$ that

   - in case of no CSI [33,42] is of the form

$$\phi_i \colon \mathcal{W} \to \mathcal{X}, \quad \text{for } i = 1, 2, \ldots, n, \tag{3.51}$$

- in case of state communication [22] is of the form

$$\phi_i \colon \mathcal{S}^{i-1} \to \mathcal{X}, \quad \text{for } i = 1, 2, \dots, n, \tag{3.52}$$

- in the case of strictly causal state communication [22] is of the form

$$\phi_i \colon \mathcal{W} \times \mathcal{S}^{i-1} \to \mathcal{X}, \quad \text{for } i = 1, 2, \dots, n, \tag{3.53}$$

- in the case of causal state and message communication [22], is of the form

$$\phi_i \colon \mathcal{W} \times \mathcal{S}^{i} \to \mathcal{X}, \quad \text{for } i = 1, 2, \dots, n, \tag{3.54}$$

- in the case of noncausal CSI and message communication [43] is of the form

$$\phi_i \colon \mathcal{W} \times \mathcal{S}^{n} \to \mathcal{X}, \quad \text{for } i = 1, 2, \dots, n \tag{3.55}$$

3. a decoding function $g \colon \mathcal{Y}^n \to \mathcal{W}$;
4. a state estimator $h \colon \mathcal{Y}^n \to \hat{\mathcal{S}}^n$, where $\hat{\mathcal{S}}$ denotes a given finite reconstruction alphabet.

For a given code and for $i = 1, \dots, n$, the random message W is assumed uniform over the message set $\mathcal{W}$ and the inputs are obtained as $X_i = \phi_i(W, Z)$ where Z denotes the encoder side-information (CSI), i.e., $Z = \emptyset$, $Z = S^{i-1}$, $Z = S^i$ or $Z = S^n$.

In [22, 33, 42–44], the quality of the state estimation is measured by the expected average per-block distortion as in (3.1) which we recall here

$$\Delta^{(n)} := \mathbb{E}[d(S^n, \hat{S}^n)] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[d(S_i, \hat{S}_i)]. \tag{3.56}$$

An achievable pair (R, D) is defined similar to Definition 1 but where the cost constraint (3.5c) is only present in the Gaussian case. We shall later explain a different sensing performance introduced in [30], which is measured in terms of the probability of a list decoding error and modify the achievability-definition accordingly, see Definition 4.

The Expected Average per-Block Distortion

The models introduced in [22, 33, 42–44] are mainly differ with respect to the side-information available at the encoder, see (3.51)-(3.55).

We first commence with the model introduced in [33, 42] where there is no state-information at the Tx and the encoding function is as in (3.51). This model is the dual to the model introduced in Section 3.1 and the main theorem of [42] is similar to Theorem 1. The optimal estimator at the Rx is symbol-by-symbol based on the observed sequence of outputs and the decoded codeword.

Theorem 4 (Theorem 1, [33,42])**.** *When the Tx has no CSI, the capacity-distortion function is given by*

$$\mathcal{C}_{\text{No-CSI}} := \max_{P_X \in \mathcal{P}_{\mathsf{D}}} I(X;Y), \tag{3.57}$$

where

$$\mathcal{P}_{\mathsf{D}} = \left\{ P_X \,\middle|\, \sum_{x \in \mathcal{X}} P_X(x) E[d(S, \hat{s}(X,Y))] \le \mathsf{D} \right\}, \tag{3.58}$$

$s^*(\cdot,\cdot)$ *is the optimal estimator introduced in* (3.12)*, with Z replaced by Y.*

As an example in [33], they looked at the Gaussian channel where the distortion is measured in MSE.

Corollary 2. *The tradeoff region for a SDGC (state-dependent Gaussian channel) with no CSI at Tx is the union of all pairs* $(\mathsf{R}^{\text{Gaus.}}_{\text{No-CSI}}, \mathsf{D}^{\text{Gaus.}}_{\text{No-CSI}})$ *satisfying*

$$\mathsf{R}^{\text{Gaus.}}_{\text{No-CSI}} = \log\left(1 + \frac{P}{N+Q}\right), \tag{3.59}$$

$$\mathsf{D}^{\text{Gaus.}}_{\text{No-CSI}} \ge \frac{QN}{Q+N}. \tag{3.60}$$

Also, similar results are provided for the MAC in [42].

Consider next the problem with strictly causal and causal CSI at the Tx, i.e., where the encoding function is presented in (3.53). These setups were considered in [22].

We denote its capacity-distortion functions by $\mathcal{C}_{Str-Caus}$.

Theorem 5 (Theorem 2, [22])**.** *The capacity–distortion function for strictly causal state communication is*

$$\mathcal{C}_{\text{Str-caus.}} = \max_{P_X P_{U|XS}} \Big(I(U,X;Y) - I(U,X;S) \Big), \tag{3.61}$$

where the estimator is reconstruction functions $\hat{s}(u,x,y)$ such that $\mathbb{E}[d(S,\hat{s}(u,x,y))] \le \mathsf{D}$.

Moreover, by choosing $V = X$ and $U = \emptyset$ the result also reproduces the case in [33]. As a special case, Theorem 5 is simplified to following theorem where there is no message to transmit.

Theorem 6 (Theorem 3, [22]). *The capacity–distortion function for causal state communication is*

$$\mathcal{C}_{\text{Caus.}}(\mathsf{D}) = \max \big(I(U,V;Y) - I(U,V;S)\big) \tag{3.62}$$

where the maximum is over all conditional pmfs $P_V P_{U|V,S}$ and functions $x(v,s)$ and reconstruction functions $\hat{s}(u,v,y)$ such that $\mathbb{E}[d(S,\hat{s}(u,v,y))] \leq \mathsf{D}$.

Note that by setting $V = X$ in this theorem, the capacity–distortion function for strictly causal communication in 3.61 is recovered. It has been proved that rate splitting between information transmission and state communication at Tx achieves optimal tradeoff.

The following theorem gives upper and lower bounds on the capacity–distortion function when the encoder has access to noncausal state at Tx as in (3.55).

Theorem 7 (Theorems 1 and 2, [43]). *The capacity–distortion function for non-causal CSI is bounded by*

$$\max_{p_{u|s}p_{x|u,s}} \Big(I(U;Y) - I(U;S)\Big) \leq \mathcal{C}_{\text{Non-caus.}} \leq \max_{p_{u|s}p_{x|u,s}} \Big(I(X,S;Y) - I(U,Y;S)\Big), \tag{3.63}$$

where the maximum is over all conditional pmfs $P_{U|S}P_{X|U,S}$ and reconstruction functions $\hat{s}(u,y)$ such that $\mathbb{E}[d(S,\hat{s}(u,y))] \leq D$.

The work [44] considers the special case of a Gaussian channel $Y = X + S + Z$ where the state S and the noise Z are independent zero-mean Gaussian sequences of variances Q and N, respectively, and MSE is used as a distortion metric.

Theorem 8 (Theorem 2, [44]). *The tradeoff region for a SDGC $Y^n = X^n(W,S^n) + S^n + Z^n$ with state information noncausally at Tx is given by the closure of the convex hull of all $(\mathsf{R}^{\text{Gaus.}}, \mathsf{D}^{\text{Gaus.}})$ pairs satisfying*

$$\mathsf{R}^{\text{Gaus.}} \leq \frac{1}{2}\log\left(1 + \frac{rP}{N}\right), \tag{3.64}$$

$$\mathsf{D}^{\text{Gaus.}} \geq Q\frac{rP+N}{(\sqrt{Q}+\sqrt{(1-r)P})^2 + rP + N}. \tag{3.65}$$

The Tx shares the input power using parameter r and sends $X = \sqrt{\frac{(1-r)P}{Q}}S$.

List Decoding Estimation

Both the causal and non-causal CSI cases were considered in [30], under a sensing measure given by the probability of list decoding error $P_{e,s}^{(n)}$ instead of expected average per-block distortion. The Rx thus not only produces a message guess but also a list $\mathcal{L}_n(Y^n)$ of possible state-sequences. If we denote by $P_{e,s}^n$ the probability that S^n is not in $\mathcal{L}_n(Y^n)$,

$$P_{e,s}^n \triangleq \Pr[S^n \notin \mathcal{L}_n(Y^n)], \tag{3.66}$$

[30] requires that this probability of error vanishes as $n \to \infty$. The list size is bounded as

$$|\mathcal{L}_n| \leq 2^{n\mathsf{R}_D}, \tag{3.67}$$

for a given rate R_D.

Definition 4 ([30]). *A pair $(\mathsf{R}, \mathsf{R}_D)$ is said to be achievable under the list-error-probability criterion if there exists a sequence of $(2^{nR}, 2^{nR_D}, n)-$ codes with $P_e^{(n)} \to 0$ and $P_{e,s}^{(n)} \to 0$ as $n \to \infty$.*

In the following theorem, the Tx has access to the state sequence causally.

Theorem 9 (Theorem 2, [30]). *The tradeoff region for a SDMC with state information causally known at the Tx is the union of all $(\mathsf{R}^{\text{Caus.}}, \mathsf{R}_D^{\text{Caus.}})$ pairs satisfying*

$$\mathsf{R}^{\text{Caus.}} = \max I(U;Y), \tag{3.68}$$

$$\mathsf{R}_D^{\text{Caus.}} \leq H(S), \tag{3.69}$$

$$\mathsf{R}^{\text{Caus.}} + \mathsf{R}_D^{\text{Caus.}} \leq I(X,S;Y), \tag{3.70}$$

for some $P_S P_U P_{X|US} P_{Y|XS}$.

Theorem 10 (Theorem 1, [30]). *The tradeoff region for a SDC with state information non-causally available at Tx is the union of all pairs $(\mathsf{R}^{\text{Non-caus.}}, \mathsf{R}_D^{\text{Non-caus.}})$ satisfying*

$$\mathsf{R}^{\text{Non-caus.}} \leq I(U;Y) - I(U;S), \tag{3.71}$$

$$\mathsf{R}_D^{\text{Non-caus.}} \leq H(S), \tag{3.72}$$

$$\mathsf{R}^{\text{Non-caus.}} + \mathsf{R}_D^{\text{Non-caus.}} \leq I(X,S;Y), \tag{3.73}$$

for some $P_S P_{UX|S} P_{Y|XS}$.

Remark 3. *The set of achievable pairs with causal CSI in Theorem 9 is generally smaller than the set of achievable pairs with noncausal CSI in Theorem 10. This can be seen by noting that in Theorem 9 the random variable U cannot depend on S.*

3.7 Conclusion

We revisited the first information-theoretic model and results on ISAC by Kobayashi et al. [1] in which a full characterization of the capacity-distortion tradeoff for memoryless state-dependent P2P channels was provided. Through several illustrative examples, we reviewed that the optimal ISAC scheme offers non-negligible gain compared to the basic time-sharing scheme that performs either sensing or communication, as well as compared to the improved time-sharing scheme that integrates both tasks into a single system but chooses the common waveform to *prioritize* one of the tasks. The results also showed that optimal sensing depends only on the employed waveform but not on the underlying coding scheme for the single-Tx systems studied in this chapter.

We also reviewed related works where either the state is constant over the entire blocklength and sensing performance is thus measured either in discrimination error probability or estimation error. Completely new sensing tools are required in this case, and interestingly, closed-loop strategies can now outperform open-loop strategies, unlike in the i.i.d. case studied in [1]. We also reviewed a different line of related works that assumes that sensing is performed at the receiver. Most of these works again assume i.i.d. states and use average distortion to measure sensing performance as in [1].

In this thesis, we formulate the tradeoff between communication and sensing in terms of the capacity-distortion tradeoff, as suggested by [33] and already previously used in related works. Though possibly inaccurate, it allows for an information-theoretic treatment of the problem. The signal-processing and communications works typically employ a dual formulation for this tradeoff allowing them to use tools from their communities.

Chapter 4

The Broadcast Channel

In this chapter, we consider ISAC over a single-Tx two-Rx broadcast channels (BC) by building on a simple single-Tx two-Rx communication model with a discrete memoryless channel and i.i.d. state sequences. The Tx observes strictly causal *generalized feedback signals*, used for state sensing. For the P2P channel studied in the previous chapter, feedback was only useful for state sensing but not to increase capacity of the communication channel. As we will see, this is not the case in memoryless Broadcast Channels.

As in the main model of the previous chapters, the Tx observes generalized feedback signals, which now in the BC setup is known to increase [17–19]. In fact, it enables the Tx to transmit some common information that is useful to both Rxs at the same time (see, e.g., [20, Section 17]).
ln our ISAC over BC setup, the generalized feedback thus improves both sensing and communication performances. Nevertheless, like in the P2P setup, the two performances only depend on each other through the common choice of the waveform. In other words, we show that the optimal state-sensing is independent of the employed BC-feedback-code and only depends on the chosen waveform but not on other details of the code construction. This allows to base joint coding and sensing systems on known BC-feedback code constructions such as [17–19].

4.1 System Model

Consider the two-Rx BC scenario depicted in Fig. 4.1. The model comprises a two-dimensional memoryless state sequence $\{(S_{1,i}, S_{2,i})\}_{i\geq 1}$ whose samples at any given time i are distributed according to a given joint law $P_{S_1S_2}$ over the state alphabets $\mathcal{S}_1 \times \mathcal{S}_2$. The Tx communicates with both Rxs over a state-dependent memoryless broadcast channel (SDMBC), where given time-i input $X_i = x$ and state realizations $S_{1,i} = s_1$ and $S_{2,i} = s_2$, the time-i outputs $Y_{1,i}$ and $Y_{2,i}$ observed at the Rxs and the Tx's feedback signal Z_i are distributed according to the stationary

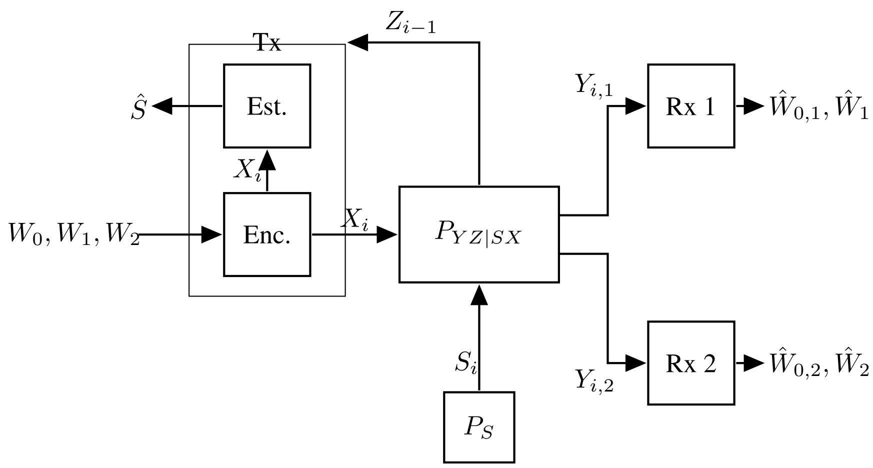


Figure 4.1: State-dependent broadcast channel with generalized feedback and state-estimator at the Tx.

channel transition law $P_{Y_1Y_2Z|S_1S_2X}(\cdot,\cdot,\cdot|s_1,s_2,x)$. We again assume that all alphabets $\mathcal{X},\mathcal{Y}_1,\mathcal{Y}_2,\mathcal{Z},\mathcal{S}_1,\mathcal{S}_2$ are finite.

The goal of the Tx is to convey a common message W_0 to both Rxs and individual messages W_1 and W_2 to Rxs 1 and 2, respectively, while estimating the states sequences $\{S_{1,i}\}$ and $\{S_{2,i}\}$ within some target distortions. For simplicity, the input cost constraint is omitted.

A $(2^{n\mathsf{R}_0}, 2^{n\mathsf{R}_1}, 2^{n\mathsf{R}_2}, n)$ code for an SDMBC thus consists of

1. three message sets $\mathcal{W}_0 = [1 : 2^{n\mathsf{R}_0}]$, $\mathcal{W}_1 = [1 : 2^{n\mathsf{R}_1}]$, and $\mathcal{W}_2 = [1 : 2^{n\mathsf{R}_2}]$;

2. a sequence of encoding functions $\phi_i \colon \mathcal{W}_0 \times \mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{Z}^{i-1} \to \mathcal{X}$, for $i = 1, 2, \ldots, n$;

3. for each $k = 1, 2$ a decoding function $g_k \colon \mathcal{Y}_k^n \to \mathcal{W}_0 \times \mathcal{W}_k$;

4. for each $k = 1, 2$ a state estimator $h_k \colon \mathcal{X}^n \times \mathcal{Z}^n \to \hat{\mathcal{S}}_k^n$, where $\hat{\mathcal{S}}_1$ and $\hat{\mathcal{S}}_2$ are given reconstruction alphabets.

For a given code, we let the random messages W_0, W_1, and W_2 be uniform over the message sets $\mathcal{W}_0$, $\mathcal{W}_1$, and $\mathcal{W}_2$ and the inputs $X_i = \phi_i(W_0, W_1, W_2, Z^{i-1})$, for $i = 1, \ldots, n$. The corresponding outputs $Y_{1,i}Y_{2,i}, Z_i$ at time i are obtained from the states $S_{1,i}$ and $S_{2,i}$ and the input X_i according to the SDMBC transition law $P_{Y_1Y_2Z|S_1S_2X}$. Further, for $k = 1, 2$ let $\hat{S}_k^n := (\hat{S}_{k,1}, \cdots, \hat{S}_{k,n}) = h_k(X^n, Z^n)$ be the Tx's estimates for state S_k^n and $(\hat{W}_{0,k}, \hat{W}_k) = g_k(Y_k^n)$ the messages decoded by Rx k. The quality of the state estimates $\hat{S}_k^n$ is again measured by bounded per-symbol distortion functions $d_k \colon \mathcal{S}_k \times \hat{\mathcal{S}}_k \mapsto [0, \infty)$, i.e., we assume

$$\max_{s_k \in \mathcal{S}_k, \hat{s}_k \in \hat{\mathcal{S}}_k} d_k(s_k, \hat{s}_k) < \infty, \quad k = 1, 2. \tag{4.1}$$

Our interest is in the two *expected average per-block distortions*

$$\Delta_k^{(n)} := \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[d_k(S_{k,i}, \hat{S}_{k,i})], \quad k = 1, 2, \tag{4.2}$$

and the joint probability of error

$$P_e^{(n)} := \Pr\Big((\hat{W}_{0,k}, \hat{W}_1) \neq (W_0, W_1) \quad \text{or} \quad (\hat{W}_{0,k}, \hat{W}_2) \neq (W_0, W_2)\Big). \tag{4.3}$$

Definition 5. *A rate-distortion tuple* $(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ *is achievable if there exists a sequence (in* n*) of* $(2^{n\mathsf{R}_0}, 2^{n\mathsf{R}_1}, 2^{n\mathsf{R}_2}, n)$ *codes that simultaneously satisfy*

$$\lim_{n\to\infty} P_e^{(n)} = 0 \tag{4.4a}$$

$$\overline{\lim_{n\to\infty}}\, \Delta_k^{(n)} \leq \mathsf{D}_k, \quad \text{for } k = 1, 2. \tag{4.4b}$$

Definition 6. *The capacity-distortion region* $\mathcal{CD}$ *is given by the closure of the union of all achievable rate-distortion tuples* $(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$.

4.2 Main Results

In this section, we present bounds on the capacity-distortion region $\mathcal{CD}$. As in the single-Rx case, one can easily determine the optimal estimator functions h_1 and h_2, which are independent of the encoding and decoding functions and operate on a symbol-by-symbol basis.

Lemma 3. *For each* $k = 1, 2$*, define the function*

$$\hat{s}_k^*(x, z) := \arg\min_{s' \in \hat{\mathcal{S}}_k} \sum_{s_k \in \mathcal{S}_k} P_{S_k|XZ}(s_k|x, z) d(s_k, s'), \tag{4.5}$$

where ties can be broken arbitrarily.

Irrespective of the choice of encoding and decoding functions, distortions $\Delta_1^{(n)}$ *and* $\Delta_2^{(n)}$ *are minimized by the estimators for* $k = 1, 2$

$$h_k^*(x^n, z^n) = (\hat{s}_k^*(x_1, z_1), \hat{s}_k^*(x_2, z_2), \ldots, \hat{s}_k^*(x_n, z_n)). \tag{4.6}$$

Proof: Analogously to proof in A.2.4. ■

Analogously to the definition in Equation (3.15) we can then define the optimal estimation cost for each input symbol $x \in \mathcal{X}$:

$$c_k(x) := \mathsf{E}[d_k(S_k, \hat{s}_k^*(X,Z))|X = x] , \qquad k = 1, 2. \tag{4.7}$$

Characterizing the capacity-distortion region is very challenging in general, because even the capacity regions of the SDMBC with and without feedback are unknown to date. We first present the exact capacity-distortion region for the class of physically degraded SDMBCs and then provide bounds for general SDMBCs. We shall also compare our results on the capacity-distortion regions to the performances achieved by simple TS baseline schemes, in analogy to the single-Rx setup.

4.2.1 Physically Degraded SDMBC: The Capacity-Distortion Region

This subsection characterizes the capacity-distortion region for *physically degraded SDMBCs* and evaluates it for two binary examples.

Definition 7. *An SDMBC $P_{Y_1Y_2Z|S_1S_2X}$ with state pmf $P_{S_1S_2}$ is called* physically degraded *if there are conditional laws $P_{Y_1|XS_1}$ and $P_{S_2Y_2|S_1Y_1}$ such that*

$$P_{Y_1Y_2|S_1S_2X}P_{S_1S_2} = P_{S_1}P_{Y_1|S_1X}P_{S_2Y_2|S_1Y_1}. \tag{4.8}$$

That means for any arbitrary input P_X, the tuple $(X, S_1, S_2, Y_1, Y_2) \sim P_X P_{S_1S_2} P_{Y_1Y_2|S_1S_2X}$ satisfies the Markov chain

$$X -\!\!\circ\!\!- (S_1, Y_1) -\!\!\circ\!\!- (S_2, Y_2). \tag{4.9}$$

Theorem 11. *The capacity-distortion region $\mathcal{CD}$ of a physically degraded SDMBC is given by the closure of the set of all tuples $(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ for which there exists a joint law P_{UX} so that the tuple $(U, X, S_1, S_2, Y_1, Y_2, Z) \sim P_{UX}P_{S_1S_2}P_{Y_1Y_2Z|S_1S_2X}$ satisfies the two rate constraints*

$$\mathsf{R}_1 \leq I(X; Y_1 \mid U) \tag{4.10}$$

$$\mathsf{R}_0 + \mathsf{R}_2 \leq I(U; Y_2), \tag{4.11}$$

and the distortion constraints

$$\mathbb{E}[d_k(S_k, \hat{s}_k^*(X, Z))] \leq \mathsf{D}_k, \quad k = 1, 2. \tag{4.12}$$

Proof: The achievability can be proved by standard superposition coding and using the optimal estimators in Lemma 3. The converse also follows from standard steps and the details are provided in Appendix A.2.1. ■

As mentioned in the proof, data communication is performed by simple superposition coding that ignores the feedback. Thus, also for physically degraded BCs feedback only facilitates state sensing but is useless for communications.

Remark 4. *Similarly to the single-Rx case, an input cost-constraint as in* (3.5c) *can be added to our model. Theorem 11 remains valid in this case, if the choice of the input distribution P_X is limited to satisfy the cost constraint*

$$\sum_{x \in \mathcal{X}} P_X(x) b(x) \leq B. \tag{4.13}$$

The analogous remark also applies to the non-physically degraded BC ahead and the presented inner and outer bounds.

Remark 5. *Similarly to what we described in Remark 2, the result in Theorem 11 can be extended to the case with imperfect Rx state-informations $S^n_{R,1}$ and $S^n_{R,2}$. For $(S^n, S^n_{R,1}, S^n_{R,2})$ i.i.d. $\sim P_{SS_{R,1},S_{R,2}}$ it suffices to replace in the rate-constraints* (4.10) *and* (4.11) *of Theorem 11 the state S_1 by $S_{R,1}$ and the state S_2 by $S_{R,2}$. The analogous remark also applies to the non-physically degraded BC ahead and the presented inner and outer bounds.*

Before evaluating Theorem 11 for two examples, we present details for the basic and the improves TS scheme for this BC setup.

4.2.2 Baseline Schemes

We again have a basic TS scheme that performs either sensing or communication at a time, and an *improved TS baseline scheme* that is able to perform both functions simultaneously via a common waveform by prioritizing either sensing or communication. Analogously to the single-Rx setup, each of the two baseline schemes time-shares between a sensing mode and a communication mode. However, since we now have two distortions and three rates, the choice of the “optimal" pmf P_X for each mode is not necessarily unique, but rather a continuum, depending on which function of the two distortions or the three rates one wishes to optimize. For fixed input pmf, the difference between the communication mode *without sensing* (employed by the basic TS scheme) and the communication mode *with sensing* (employed by the improved TS scheme) lies in the choice of the estimators. In the former mode, the Tx applies the best *constant estimators* for the two state-sequences, irrespective of its inputs and feedback outputs. In the latter mode, it applies the optimal estimators in Lemma 3, which depend on the input and the feedback output.

Similarly, the difference between the communication modes *without and with sensing* is that in the former all rates are zero and in the latter the chosen input pmf P_X can be used for communication at positive rates.

Example 3: Binary BC with Multiplicative Bernoulli States

Consider the physically degraded SDMBC with binary input and output alphabets $\mathcal{X} = \mathcal{Y}_1 = \mathcal{Y}_2 = \{0,1\}$ and binary state alphabets $\mathcal{S}_1 = \mathcal{S}_2 = \{0,1\}$. The channel input-output relation is described by

$$Y_k = S_k X, \qquad k = 1, 2, \tag{4.14}$$

with the joint state pmf

$$P_{S_1 S_2}(s_1, s_2) = \begin{cases} 1-q, & \text{if } (s_1, s_2) = (0,0) \\ 0, & \text{if } (s_1, s_2) = (0,1) \\ q\gamma, & \text{if } (s_1, s_2) = (1,1) \\ q(1-\gamma) & \text{if } (s_1, s_2) = (1,0), \end{cases} \tag{4.15}$$

for $\gamma, q \in [0,1]$. Notice that S_2 is a degraded version of S_1, which together with the transition law (4.14) ensures the Markov chain $X -\!\!\circ\!\!- (S_1, Y_1) -\!\!\circ\!\!- (S_2, Y_2)$ and the physically degradedness of the SDMBC. We consider output feedback

$$Z = (Y_1, Y_2), \tag{4.16}$$

and set the common rate $\mathsf{R}_0 = 0$ for simplicity. We first introduce TS schemes, then we present the ISAC in Corollary 3 for the same example.

In this SDMBC, zero distortions $\mathsf{D}_1 = \mathsf{D}_2 = 0$ can be achieved by deterministically choosing $X = 1$ exactly as for the single-Rx case. This choice however cannot achieve any positive communication rates, i.e., $\mathsf{R}_1 = \mathsf{R}_2 = 0$. In the sensing mode with and without communication, we thus have:

$$(\mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2) = (0, 0, 0, 0). \tag{4.17}$$

The optimal input distribution for communication is $X_{\max} \sim \mathcal{B}(1/2)$, in which case all rate-pairs $(\mathsf{R}_1, \mathsf{R}_2)$ satisfying

$$\mathsf{R}_k \leq P_{S_k}(1), \qquad k = 1, 2, \tag{4.18}$$

are achievable. The input $X_{\max} \sim \mathcal{B}(1/2)$ simultaneously maximizes both communication rates $\mathsf{R}_1, \mathsf{R}_2$.

In the communication mode *without* sensing, the Tx applies the optimal constant estimator for each state, namely

$$\hat{s}_{\text{const},k} := \operatorname*{argmax}_{\hat{s} \in \{0,1\}} P_{S_k}(\hat{s}), \qquad k = 1, 2, \tag{4.19}$$

and thus achieves all tuples

$$(\mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2) = (qr, \gamma q(1-r), \mathsf{D}_{1,\max}, \mathsf{D}_{2,\max}) \tag{4.20}$$

where $\mathsf{D}_{1,\max} := \min\{q, 1-q\}$ and $\mathsf{D}_{2,\max} := \min\{\gamma q, 1-\gamma q\}$, and $r \in [0,1]$ denotes the time-sharing parameter between the two communication rates.

In the communication mode *with* sensing, the same input $X_{\max}$ is used. The Tx however applies the optimal estimator for $k = 1, 2$:

$$\hat{s}_k^*(x, y_1, y_2) = \begin{cases} y_k, & \text{if } x = 1 \\ \hat{s}_{\text{const},k}, & \text{if } x = 0, \end{cases} \tag{4.21}$$

and achieves the tuple

$$(\mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2) = \left(qr, \gamma q(1-r), \frac{\mathsf{D}_{1,\max}}{2}, \frac{\mathsf{D}_{2,\max}}{2}\right), \tag{4.22}$$

where r again denotes the time-sharing parameter between the two communication rates.

The basic and improved TS baseline schemes achieve the time-sharing lines between points (4.17) and (4.20) and points (4.17) and (4.22), respectively. The following corollary evaluates Theorem 11 to obtain the performance of the optimal co-design scheme.

Corollary 3. *The capacity-distortions region $\mathcal{CD}$ of the binary physically degraded SDMBC in* (4.14)–(4.16) *is the set of all tuples* $(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ *satisfying*

$$\mathsf{R}_0 + \mathsf{R}_1 \leq qH_{\mathsf{b}}(p)r, \tag{4.23a}$$

$$\mathsf{R}_0 + \mathsf{R}_2 \leq \gamma q H_{\mathsf{b}}(p)(1-r), \tag{4.23b}$$

$$\mathsf{D}_1 \geq p \min\{q, 1-q\}, \tag{4.23c}$$

$$\mathsf{D}_2 \geq p \min\{\gamma q, 1-\gamma q\}, \tag{4.23d}$$

for some choice of the parameters $r, p \in [0,1]$.

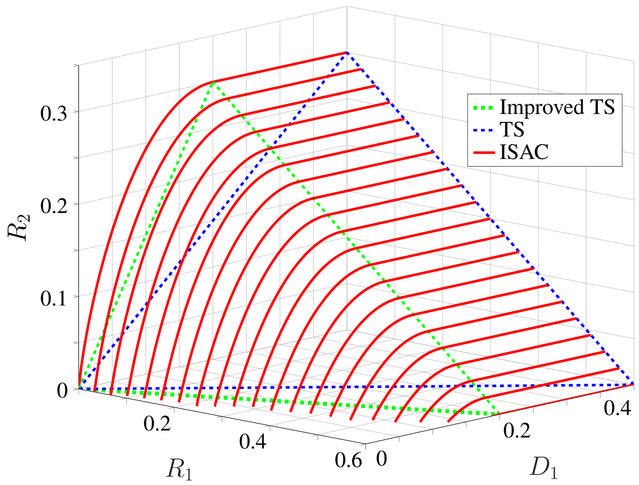


Figure 4.2: Boundary of the capacity-distortion region $\mathcal{CD}$ for Example 3 in Subsection 11.

Proof: We start by noticing that for this example $I(X;Y_1 \mid U,S_1) = qH(X|U)$ and $I(U;Y_2 \mid S_2) = q\gamma(H(X) - H(X \mid U))$. Setting $p := P_X(0)$ and $r := \frac{H(X|U)}{H(X)}$, directly leads to the desired rate constraints. The distortion constraints are obtained from the optimal estimators in (4.21). Following the same steps as in the single-Rx case, i.e. (3.28) and (3.29), we obtain

$$\mathsf{D}_k \geq p \min\{P_{S_k}(0), P_{S_k}(1)\}, \tag{4.24}$$

which concludes the proof. ■

Notice that above Corollary 3 reduces to Corollary 1 in the special case of $\mathsf{R}_0 = \mathsf{R}_2 = 0$ and $\mathsf{D}_2 = \infty$, i.e., when we ignore Rx 2. Fig. 4.2 shows in red color the boundary of the projection of the tradeoff region $\mathcal{CD}$ of this example onto the 3-dimensional plane $(\mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1)$, for parameters $\gamma = 0.5$ and $q = 0.6$. The tradeoff with D_2 is omitted for simplicity and because D_2 is a scaled version of D_1. The figure also shows the boundaries of the basic and improved TS baseline schemes. We again notice a significant gain for an optimal co-design scheme compared to the TS baseline schemes.

So far, there was no tradeoff between the two distortion constraints D_1 and D_2. This is different in the next example, which otherwise is very similar.

Example 4: Binary BC with Multiplicative Bernoulli States and Flipping Inputs

Reconsider the same state pmf $P_{S_1S_2}$ as in the previous example, but now an SDMBC with a transition law that flips the input for Rx 2:

$$Y_1 = S_1 X, \qquad Y_2 = S_2(1 - X). \tag{4.25}$$

As in the previous example we consider output feedback $Z = (Y_1, Y_2)$.

Corollary 4. *The capacity-distortion region $\mathcal{CD}$ of the binary SDMBC with flipping inputs in* (4.25) *and output feedback is the set of all tuples* $(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ *satisfying*

$$\mathsf{R}_1 \leq qH_\mathrm{b}(p)r, \tag{4.26a}$$

$$\mathsf{R}_0 + \mathsf{R}_2 \leq \gamma q H_\mathrm{b}(p)(1-r), \tag{4.26b}$$

$$\mathsf{D}_1 \geq p \min\{q(1-\gamma), (1-q)\}, \tag{4.26c}$$

$$\mathsf{D}_2 \geq (1-p)q \min\{\gamma, 1-\gamma\}, \tag{4.26d}$$

for some choice of the parameters $r, p \in [0,1]$.

The capacity-distortion region expression above captures the tradeoffs between the two rates through the parameter r, between the rates and the distortions through the parameter p, and between the two distortions through the parameter p.

Comparing above Corollary 4 to the previous Corollary 3, we remark the identical rate constraints and the relaxed distortion constraints for both D_1 and D_2 in Corollary 4. The reason is that the flipping input allows the Tx to perfectly estimate S_1 from (X, Y_1, Y_2) not only when $X = 1$ but also when $X = 0$ and $Y_2 = 1$ because they imply that $S_2 = 1$ and by (4.15) also $S_1 = 1$.

Proof: The proof is similar to the proof of Corollary 3, except for the description of the optimal estimators. To determine these optimal estimators, we remark that only four input-output relations are possible: $(x, y_1, y_2) \in \{(0,0,0), (0,0,1), (1,0,0), (1,1,0)\}$. Moreover, when $X = 1$, then $Y_1 = S_1$, and when $X = 0$, then $Y_2 = S_2$. In particular, when $X = 0$ and $Y_2 = 1$, then $S_2 = 1$ and also $S_1 = 1$, see (4.15). The optimal estimator for state S_1 thus is:

$$\hat{s}_1^*(x, y_1, y_2) = \begin{cases} y_1, & \text{if } x = 1 \\ 1, & \text{if } (x, y_2) = (0,1) \\ \arg\min\limits_s P_{S_1|S_2}(s|0), & \text{else,} \end{cases} \tag{4.27}$$

and $\hat{s}_1^*(X, Y_1, Y_2) = S_1$ unless $X = 0$, $Y_2 = 0$, and $S_1 \neq \arg\min_s P_{S_1|S_2}(s|0)$, which is equivalent to $(X = 0, S_2 = 0)$ and $S_1 \neq \arg\min_s P_{S_1|S_2}(s|0)$. This yields $c_1(1) = 0$ and because S_2 is independent of X:

$$c_1(0) = P_{S_2}(0) \min_s P_{S_1|S_2}(s|0). \tag{4.28}$$

Recalling $p = P_X(0)$, we readily obtain the distortion for state S_1:

$$\mathsf{D}_1 = p \min_s P_{S_1,S_2}(s,0) = p \min\{q(1-\gamma), 1-q\}. \tag{4.29}$$

The optimal estimator and the corresponding distortion for state S_2 can be obtained in a similar way. ■

4.2.3 General SDMBC: Bounds on the Capacity-Distortion Region

In the remainder of this section, we reconsider general SDMBCs, for which we present bounds on $\mathcal{CD}$. We start with a simple outer bound.

Theorem 12 (Outer Bound on $\mathcal{CD}$)**.** *If* $(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ *lies in* $\mathcal{CD}$ *for a given SDMBC* $P_{Y_1Y_2Z|S_1S_2X}$ *with state pmf* $P_{S_1S_2}$*, then there exist pmfs* $P_X, P_{U_1|X}, P_{U_2|X}$ *such that the random tuple* $(U_k, X, S_1, S_2, Y_1, Y_2, Z) \sim P_{U_k|X}P_XP_{S_1S_2}P_{Y_1Y_2Z|S_1S_2X}$ *satisfies the rate constraints*

$$\mathsf{R}_0 + \mathsf{R}_k \leq I(U_k; Y_k), \quad k = 1, 2, \tag{4.30a}$$

$$\mathsf{R}_0 + \mathsf{R}_1 + \mathsf{R}_2 \leq I(X; Y_1, Y_2), \tag{4.30b}$$

and the average distortion constraint

$$\mathbb{E}[d_k(S_k, \hat{s}_k^*(X,Z))] \leq \mathsf{D}_k, \quad k = 1, 2, \tag{4.31}$$

where the function $\hat{s}_k^*(\cdot,\cdot)$ *is defined in* (4.5).

Proof: See Appendix A.2.2. ■

Achievability results are easily obtained by combining existing achievability results for SDMBCs with generalized feedback with the optimal estimator in Lemma 3. We consider the block-Markov coding scheme in [17], which combined with the optimal estimator in Lemma 3 yields the following proposition.

Proposition 2 (Inner Bound on $\mathcal{CD}$)**.** *Consider an SDMBC* $P_{Y_1Y_2Z|S_1S_2X}$ *with state pmf* $P_{S_1S_2}$*. The capacity-distortion region* $\mathcal{CD}$ *includes all tuples* $(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ *that satisfy inequalities* (4.32) *on top of this page and the distortion constraints* (4.31). *where* $(U_0, U_1, U_2, X, S_1, S_2, Y_1, Y_2, Z, V_0, V_1, V_2) \sim P_{U_0U_1U_2X}P_{S_1S_2}P_{Y_1Y_2Z|S_1S_2X}P_{V_0V_1V_2|U_0U_1U_2Z}$*, for some choice of (conditional) pmfs* $P_{U_0U_1U_2X}$ *and* $P_{V_0V_1V_2|U_0U_1U_2Z}$.

$$\mathsf{R}_0 + \mathsf{R}_1 \leq I(U_0, U_1; Y_1, V_1) - I(U_0, U_1, U_2, Z; V_0, V_1 \mid Y_1) \tag{4.32a}$$

$$\mathsf{R}_0 + \mathsf{R}_2 \leq I(U_0, U_2; Y_2, V_2) - I(U_0, U_1, U_2, Z; V_0, V_2 \mid Y_2) \tag{4.32b}$$

$$\begin{aligned}\mathsf{R}_0 + \mathsf{R}_1 + \mathsf{R}_2 \leq\ & I(U_1; Y_1, V_1|U_0) + I(U_2; Y_2, V_2 \mid U_0) + \min_{k\in\{1,2\}} I(U_0; Y_k, V_k) - I(U_1; U_2 \mid U_0) \\ & -I(U_0, U_1, U_2, Z; V_1 \mid V_0, Y_1) - I(U_0, U_1, U_2, Z; V_2|V_0, Y_2) \\ & - \max_{k\in\{1,2\}} I(U_0, U_1, U_2, Z; V_0 \mid Y_k)\end{aligned} \tag{4.32c}$$

$$\begin{aligned}2\mathsf{R}_0 + \mathsf{R}_1 + \mathsf{R}_2 \leq\ & I(U_0, U_1; Y_1, V_1) + I(U_0, U_2; Y_2, V_2) - I(U_1; U_2 \mid U_0) \\ & -I(U_0, U_1, U_2, Z; V_0, V_1 \mid Y_1) - I(U_0, U_1, U_2, Z; V_0, V_2 \mid Y_2)\end{aligned} \tag{4.32d}$$

Proof: See [17] for the proof. Here, we bring the coding scheme for the BC with generalized feedback in [17] The scheme in [17] combines Marton's no-feedback scheme with LGW-SI (Lossy Gray-Wyner scheme with Side-information) scheme using a block-Markov framework. Transmission takes place over consecutive blocks, where the first B blocks are of length n and the last block is of length γn for $\gamma > 1$. We denote the input, output, feedback sequences in Block $b \in \{1, \cdots B\}$ by $X_{(b)}, Y_{1,(b)}, Y_{2,(b)}, Z_{(b)}$ respectively, and the input, output sequences in Block $B+1$ by $X'_{(B+1)}, Y'_{1,(B+1)}, Y_{2,(B+1)'}$. The messages to be sent are in a product form $W_k = (W_{k,(1)}, \cdots, W_{k,(B)})$, for $k \in \{0, 1, 2\}$, where each message of each block is uniformly distributed over a message set. At the end of block b , after observing the feedback $Z_{(b)}$ and with its block-information b $U_{0,(b)}, U_{1,(b)}, U_{2,(b)}$. the Tx applies lossy Gray-Wyner coding [45] according to the conditional law $P_{V_0V_1V_2|U_0U_1U_2Z}$ to generates the update information $V_{0,(b)}, V_{1,(b)}, V_{2,(b)}$. In the following block $b{+}1$, the Tx then uses Marton's no-feedback scheme to send the Messages $W_{0,(b+1)}, W_{1,(b+1)}, W_{2,(b+1)}$ together with update information $V_{0,(b)}, V_{1,(b)}, V_{2,(b)}$ (see also Fig 4.3). Decoding is performed backward, starts from the last block. Each Rx k first tries to decode update information of the last block $\bar{V}_{(B+1)} \triangleq (V_{0,(B+1)}, V_{k,(B+1)})$ (because there is no new information in the last block) from $Y_{k,(B+1)}$. Then, Rx k attempts to decode the new information in block B by using the decoded update information of block $B+1$, and this procedure in a backward manner. ∎

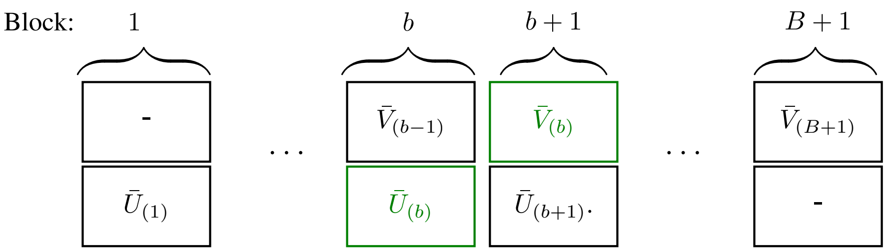


Figure 4.3: Backward decoding for block Markov coding used in Proposition 2

In the following subsections, we evaluate the proposed inner and outer bounds on $\mathcal{CD}$ for two examples and identify setups where the provided bounds allow us to conclude that no tradeoff between the achievable rates and distortions occurs.

4.2.4 No Rate-Distortion Tradeoff

Notice that for some SDMBCs there is no tradeoff between the achievable distortions and communication rates. In this case, for the BC, the capacity-distortion region is given by the Cartesian product between the SDMBC's capacity region:

$$\mathcal{C} := \{(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2) \colon \mathsf{D}_1 \geq 0,\ \mathsf{D}_2 \geq 0 \quad \text{s.t. } (\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2) \in \mathcal{CD}\}, \tag{4.33}$$

and its distortion region:

$$\mathcal{D} := \{(\mathsf{D}_1, \mathsf{D}_2) \colon \mathsf{R}_0 \geq 0,\ \mathsf{R}_1 \geq 0,\ \mathsf{R}_2 \geq 0 \text{ s.t. } (\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2) \in \mathcal{CD}\}. \tag{4.34}$$

Proposition 3 (No Rate-Distortion Tradeoff)**.** *Consider an SDMBC $P_{Y_1Y_2Z|S_1S_2X}$ with state pmf $P_{S_1S_2}$ for which there exist functions ψ_1 and ψ_2 with domain $\mathcal{X} \times \mathcal{Z}$ so that irrespective of the input distribution P_X the relations*

$$(S_k, \psi_k(Z, X)) \perp X, \tag{4.35}$$

$$S_k \multimap\!\!- \psi_k(Z, X) \multimap\!\!- (Z, X), \quad k = 1, 2, \tag{4.36}$$

hold for $(S_1, S_2, X_2, Z) \sim P_{S_1}P_{S_2}P_X P_{Z|XS_1,X_2}$. The capacity-distortion region of this SDMBC is the product of the capacity region and the distortion region:

$$\mathcal{CD} = \mathcal{C} \times \mathcal{D}. \tag{4.37}$$

Proof: The proof is provided in Appendix A.2.5. ■

Example 5: Erasure BC with Noisy Feedback

Our first example satisfies Conditions (4.35) and (4.36) in Proposition 3 for an appropriate choice of ψ_1 and ψ_2, and its capacity-distortion region is thus given by the product of the capacity region and the distortion region.

Let $(E_1, S_1, E_2, S_2) \sim P_{E_1S_1E_2S_2}$ over $\{0,1\}^4$ be given but arbitrary. Consider the state-dependent erasure BC

$$Y_k = \begin{cases} X & \text{if } S_k = 0, \\ ? & \text{if } S_k = 1, \end{cases}, \qquad k = 1, 2, \tag{4.38}$$

where the feedback signal $Z = (Z_1, Z_2)$ is given by

$$Z_k = \begin{cases} Y_k & \text{if } E_k = 0, \\ ? & \text{if } E_k = 1, \end{cases}, \qquad k = 1, 2. \tag{4.39}$$

Further consider Hamming distortion measures $d_k(s, \hat{s}) = s \oplus \hat{s}$, for $k = 1, 2$. For the choice

$$\psi_k(Z_k) = \begin{cases} 1, & \text{if} \quad Z_k = ?, \\ 0, & \text{else}, \end{cases} \tag{4.40}$$

the described SDMBC satisfies the conditions in Proposition 3, thus yielding the following corollary.

Corollary 5. *The capacity-distortion region of the state-dependent erasure BC with noisy feedback in* (4.38)–(4.39) *is the Cartesian product between the capacity region of the SDMBC and its distortion region:*

$$\mathcal{CD} = \mathcal{C} \times \mathcal{D}. \tag{4.41}$$

When $P_{E_1S_1E_2S_2} = P_{E_1S_1}P_{E_2S_2}$, then the distortion region is given by:

$$\mathcal{D} = \{(\mathsf{D}_1, \mathsf{D}_2) \colon \mathsf{D}_k \geq P_{E_kS_k}(1, 0)\}. \tag{4.42}$$

Proof. The state can perfectly be estimated ($S_k = 0$) with zero distortion if $(S_k, E_k) = (0, 0)$. Otherwise, the feedback is $Z_k = ?$ and provides no information. The optimal estimator is then given by the best constant estimator, which in this example is:

$$\hat{s}_{\text{const},k} = \mathbf{1}\{P_{S_k}(1) \geq P_{S_kE_k}(0, 1)\}. \tag{4.43}$$

This immediately yields the distortion constraint in (4.42). □

Notice that the capacity region $\mathcal{C}$ of the SDMBC (4.38) is unknown even with perfect feedback. In [46, 47], the capacity region of this SDMBC with perfect feedback was characterized when each Rx is informed about the state realizations at *both* Rxs.

4.2.5 Example 6: State-Dependent Dueck's BC with Multiplicative Bernoulli States

We consider a state-dependent version of Dueck's example in [48], which first served to show that feedback can increase capacity of a memoryless BC. Interestingly, despite its simplicity, the state-dependent extension of this example allows observing various kinds of tradeoffs between communication and sensing performances and also between performances at the various Rxs. For example, for specific choices of parameters, the problems of sensing and communication decompose (Corollary 6), and it is possible to simultaneously achieve the optimal sensing and communication performances. For other parameters a tradeoff arises. The present example also shows nicely that our presented co-design scheme can significantly outperform the two TS methods.

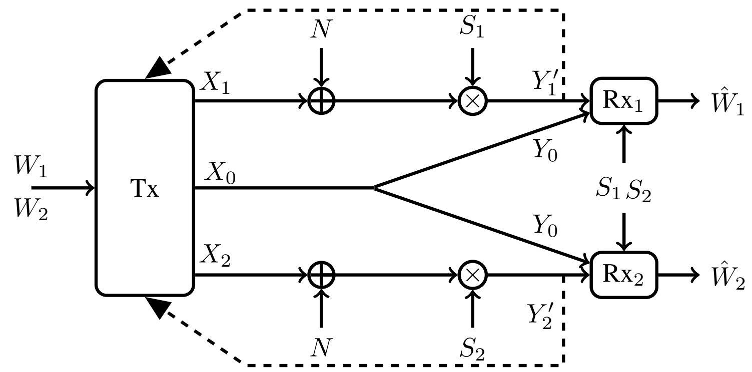


Figure 4.4: State-dependent Dueck Broadcast Channel.

Consider the state-dependent BC in Figure 4.4 with input $X = (X_0, X_1, X_2) \in \{0,1\}^3$, i.i.d. Bernoulli states $S_1, S_2 \sim \mathcal{B}(q)$, for $q \in [0,1]$, and outputs

$$Y_k = (X_0, Y_k', S_1, S_2), \qquad k = 1, 2, \tag{4.44}$$

where

$$Y_k' = S_k(X_k \oplus N), \qquad k = 1, 2, \tag{4.45a}$$

and the noise $N \sim \mathcal{B}(1/2)$ is independent of the inputs and the states. The feedback signal is

$$Z = (Y_1', Y_2'), \tag{4.46}$$

and for simplicity we again ignore the common rate R_0.

We notice that only X_1 and X_2 are corrupted by the state and the noise. Since X_0 is received without any state or noise, it is thus completely useless for sensing. In fact, the optimal estimator of Lemma 3 for $k = 1, 2$ is (see Appendix A.2.3)

$$\hat{s}_k^*(x_1, x_2, y_1', y_2') = \begin{cases} \mathbb{1}\{q \geq (1-q)\}, & y_k' = 0 y_{\bar{k}}' = 1, x_1 \neq x_2 \\ 0, & y_k' = 0, y_{\bar{k}}' = 1, x_1 = x_2 \\ 1, & y_k' = 1 \\ 0, & y_1 = y_2 = 0, x_1 \neq x_2 \\ \mathbb{1}\{q \geq (1-q)(2-q)\}, & y_1' = y_2' = 0, x_1 = x_2, \end{cases} \tag{4.47}$$

where we slightly abuse notation by omitting the argument x_0 for the estimator $\hat{s}_k^*$ because this latter does not depend on x_0.

For a given input pmf with probability $t := \Pr[X_1 \neq X_2]$, the expected distortion achieved by the optimal estimators in (4.47) is (see Appendix A.2.3):

$$\begin{aligned} \mathsf{E}\left[d_k(S_k, \hat{s}_k^*(X_1, X_2, Y_1', Y_2'))\right] &= \frac{1}{2} tq \left(\min\{q, 1-q\} + (1-q)\right) \\ &\quad + \frac{1}{2}(1-t)\min\{q, (1-q)(2-q)\} \end{aligned} \tag{4.48}$$

We observe different cases: i) for $q \in [0, 1/2]$, both minima are achieved by q; ii) for $q \in \left(1/2, 2-\sqrt{2}\right]$, the first and second minima are achieved by $1-q$ and q, respectively; iii) for $q \in \left(2-\sqrt{2}, 1\right]$, the first and second minimum are achieved by $(1-q)$ and $(1-q)(2-q)$, respectively. The distortion constraint (4.31) thus evaluates to:

$$\mathsf{D}_k \geq \begin{cases} q/2, & q \in [0, 1/2] \\ q(1-t(2q-1))/2, & q \in \left(1/2, 2-\sqrt{2}\right] \\ (1-q)(2-q+t(3q-2))/2, & q \in \left(2-\sqrt{2}, 1\right]. \end{cases} \tag{4.49}$$

We notice that for $q \in [0, 1/2]$, the distortion constraint is independent of t and thus of P_X, and the minimum expected distortions are $\mathsf{D}_{\min,1} = \mathsf{D}_{\min,2} = \frac{1}{2}q$. For $q \in \left(1/2, 2-\sqrt{2}\right]$, the minimum expected distortions are achieved for $t = 1$ and the same holds also for $q \in \left(2-\sqrt{2}, 2/3\right]$. For $q \in [2/3, 1]$, the distortions are minimized for $t=0$. We thus have $\mathsf{D}_{\min,1} = \mathsf{D}_{\min,2} = \mathsf{D}_{\min}$, where

$$\mathsf{D}_{\min} := \begin{cases} q/2, & q \in [0,1/2] \\ q(1-q), & q \in [1/2,2/3] \\ (1-q)(2-q)/2, & q \in [2/3,1]. \end{cases} \tag{4.50}$$

We obtain a characterization of the distortion region:

$$\mathcal{D} = \{(\mathsf{D}_1,\mathsf{D}_2)\colon \mathsf{D}_1 \geq \mathsf{D}_{\min},\ \mathsf{D}_2 \geq \mathsf{D}_{\min}\}. \tag{4.51}$$

The private-messages capacity region is:

$$\mathcal{C} = \{(\mathsf{R}_1,\mathsf{R}_2)\colon \mathsf{R}_1 \leq 1,\ \mathsf{R}_2 \leq 1,\ \text{and} \quad \mathsf{R}_1+\mathsf{R}_2 \leq 1+q^2\}. \tag{4.52}$$

The converse and achievability proofs are provided in Appendices A.2.3 and A.2.3, respectively.

Reconsider now the case where $q \in [0,1/2]$. As previously explained, the distortion is independent of the input distribution, and the capacity-distortion region $\mathcal{CD}$ degenerates to the product of the capacity and distortion regions:

Corollary 6 (No Rate-Distortion Tradeoff)**.** *For above state-dependent Dueck example with $q \in [0,1/2]$:*

$$\mathcal{CD} = \mathcal{C} \times \mathcal{D}. \tag{4.53}$$

For the general case, we only have bounds on the capacity-distortion region $\mathcal{CD}$. We first present our outer bound, which is based on Theorem 12 and proved in Appendix A.2.3.

Corollary 7 (Outer Bound)**.** *The capacity-distortion region $\mathcal{CD}$ (without common message) of Dueck's state-dependent BC is included in the set of tuples $(\mathsf{R}_1,\mathsf{R}_2,\mathsf{D}_1,\mathsf{D}_2)$ that for some choice of the parameters $t \in [0,1]$ satisfy the rate-constraints*

$$\mathsf{R}_k \leq 1, \qquad k=1,2, \tag{4.54}$$

$$\mathsf{R}_1+\mathsf{R}_2 \leq 1+q^2 H_{\mathrm{b}}(t) \tag{4.55}$$

and the distortion constraints in (4.49).

The inner bound is based on Proposition 2, see Appendix A.2.3. Together with the outer bound in Corollary 7 it

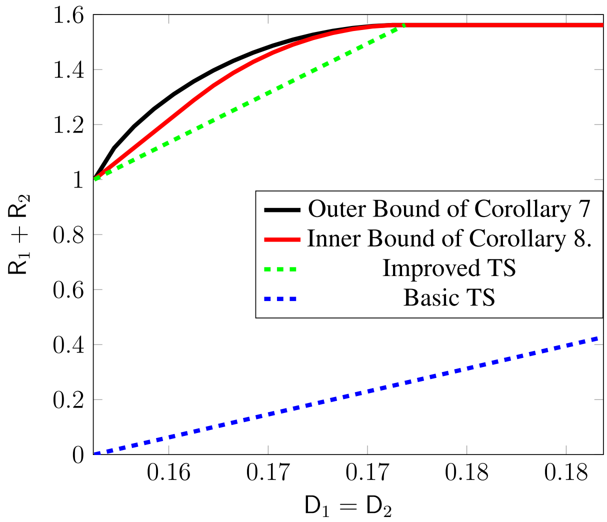


Figure 4.5: Sum-rate $\mathsf{R}_1 + \mathsf{R}_2$ vs. symmetric distortion $\mathsf{D}_1 = \mathsf{D}_2$ for the state-dependent Dueck BC with $q = 3/4$.

characterizes both the distortion region $\mathcal{D}$ and the capacity region $\mathcal{C}$ in (4.51) and (4.52).

Corollary 8 (Inner bound)**.** *The capacity-distortion region $\mathcal{CD}$ of the state-dependent Dueck BC includes all rate-distortion tuples* $(\mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ *that for some choice of $t \in [0,1]$ satisfy* (4.49) *and*

$$\mathsf{R}_k \leq 1, \quad k = 1, 2, \tag{4.56}$$

$$\mathsf{R}_1 + \mathsf{R}_2 \leq 1 + qH_\mathrm{b}(t) - q(1-q), \tag{4.57}$$

as well as the convex hull of all these tuples.

Fig. 4.5 shows our outer and inner bounds in Corollaries 7 and 8 for $q = 3/4$, where in the inner bound we consider the convex hull operation through convex combinations between values of $t > 0$ and $t = 0$. The figure also shows the performances of the basic and improved TS baseline schemes, whose modes we explain next. (Recall that the basic TS scheme time-shares the sensing mode without communication and the communication mode without sensing, and the improved TS scheme time-shares the sensing mode with communication and the communication mode with sensing.)

Sensing mode with and without communication:

In the sensing mode with communication, one can choose an arbitrary pmf for X_0, e.g., X_0 Bernoulli-$1/2$ because this input does not affect the sensing. From (4.50), the minimum distortions of $\mathsf{D}_{\min,1} = \mathsf{D}_{\min,2} = 5/32$ are achieved by setting $X_1 = X_2$ with probability 1. For $X_1 = X_2$ the sum-rate cannot exceed $\mathsf{R}_1 + \mathsf{R}_2 \leq 1$, because $I(X_0, X_1, X_2; Y_1, Y_2) = I(X_0, X_2; Y_1, Y_2) \leq H(X_0) + I(X_2; Y_1', Y_2'|X_0) \leq 1$ as Y_1' and Y_2' are corrupted by the

Bernoulli-$1/2$ noise N. On the other hand, any rate pair $(\mathsf{R}_1, \mathsf{R}_2)$ of sum-rate $\mathsf{R}_1 + \mathsf{R}_2 = 1$ is trivially achievable by communicating only with the noiseless X_0-input.

We conclude that the sensing mode with communication achieves the rate-distortion tuple $(\mathsf{R}_1, \mathsf{R}_1, \mathsf{D}_1, \mathsf{D}_2)$ satisfying

$$\mathsf{R}_1 + \mathsf{R}_2 \leq 1 \quad \text{and} \quad \mathsf{D}_k \geq 5/32, \quad k = 1, 2. \tag{4.58}$$

If the Tx cannot perform communication and sensing tasks simultaneously, the same minimum distortions are achieved but the rates are trivially zero.

$$\mathsf{R}_1 + \mathsf{R}_2 = 0 \quad \text{and} \quad \mathsf{D}_k \geq 5/32, \quad k = 1, 2. \tag{4.59}$$

Communication mode with and without sensing:

The optimal pmf P_X achieving the capacity region in (4.52) corresponds to i.i.d. Bernoulli-$1/2$ distributed X_0, X_1, X_2 (Appendix A.2.3). The corresponding sum rate is $\mathsf{R}_1 + \mathsf{R}_2 = 1 + q^2 = 25/16$. The minimum achievable distortions are thus obtained from (4.49) by setting $t = \Pr[X_1 \neq X_2] = 1/2$, i.e., $\mathsf{D}_{\max,1} = \mathsf{D}_{\max,2} = 11/64$. The best constant estimator is $\hat{S}_1 = \hat{S}_2 = 1$ because $3/4 = P_{S_k}(1) > P_{S_k}(0) = 1/4$, which achieves distortions $\mathsf{D}_{\text{trivial},1} = \mathsf{D}_{\text{trivial},2} = 1/4$. We can conclude that the communication mode with sensing achieves all rate-distortion tuples $(\mathsf{R}_1, \mathsf{R}_1, \mathsf{D}_1, \mathsf{D}_2)$ satisfying

$$\begin{aligned} \mathsf{R}_1 + \mathsf{R}_2 &\leq 25/16, \\ \mathsf{R}_k \leq 1 \quad &\text{and} \quad \mathsf{D}_k \geq 11/64 \quad k = 1, 2 \end{aligned} \tag{4.60}$$

and the communication mode without sensing achieves all rate-distortion tuples $(\mathsf{R}_1, \mathsf{R}_1, \mathsf{D}_1, \mathsf{D}_2)$ satisfying

$$\begin{aligned} \mathsf{R}_1 + \mathsf{R}_2 &\leq 25/16 \\ \mathsf{R}_k \leq 1 \quad &\text{and} \quad \mathsf{D}_k \geq 1/4, \quad k = 1, 2. \end{aligned} \tag{4.61}$$

4.3 Related Information-Theoretic Works on Multi-Hop ISAC Systems

In this section, we review related works to ISAC over Single-Tx Two-Rx networks. The first work is similar to our model, where the Tx estimates the state(s), but one of the Rxs is treated as an eavesdropper. The other work describes the model in which the Rx estimates the state(s) over a state-dependent multi-hop channel.

4.3.1 Secure ISAC

In [29], secrecy aspects of ISAC systems are considered. The model consists of i.i.d states, perfect output feedback to the Tx, and each Rx knows its corresponding state, where part of the message is kept from the eavesdropper. The characterization of the secrecy-distortion region is exact for the physically-degraded BC when the legitimate Rx is stronger. The Tx sends two encoded messages W_1, W_2 over a channel to a legitimate Rx who observes Y_1^n (and knows S_1^n). At the same time, the eavesdropper who observes Y_2^n and knows S_2^n should learn only a vanishing amount of information about W_2, which is captured by the additional secrecy condition $I(W_2; Y_2^n \mid S_2^n) \leq \delta$ achievability is defined as:

Definition 8 (Definition 1, [29])**.** *A secrecy-distortion tuple (R_1, R_2, D_1, D_2) is achievable if it is possible to find a sequence (in the blocklength n) of encoding and decoding functions satisfying for any $\delta > 0$:*

$$\frac{1}{n}\log|\mathcal{W}_k| \geq R_k - \delta \tag{4.62}$$

$$\Pr[W \neq \hat{W}] \leq \delta \tag{4.63}$$

$$I(W_2; Y_2^n \mid S_2^n) \leq \delta \tag{4.64}$$

$$E[d(S_k, \hat{S}_k)] \leq D_k + \delta \tag{4.65}$$

for $k \in \{1, 2\}$.

Theorem 13 (Theorem 1, [29])**.** *For a physically-degraded BC and with the partial secrecy requirement* (4.64)*, the set of achievable quadruples (R_1, R_2, D_1, D_2) is given by the convex hull of all quadruples satisfying*

$$R_1 \leq I(V; Y_1 \mid S_1) \tag{4.66}$$

$$R_2 \leq \min\{H(Y_1, S_1 \mid Y_2, S_2) - H(S_1 \mid Y_2, S_2, V), \\ I(V; Y_1 \mid S_1) - R_1\} \tag{4.67}$$

$$D_k \geq E[d(S_k, \hat{S}_k))], \quad \text{for} \quad k \in \{1, 2\}, \tag{4.68}$$

for some $P_{VXY_1Y_2S_1S_2} = P_X P_{V|X} P_{S_1S_2} P_{Y_1|S_1X} P_{Y_2|S_2Y_1}$, *and it suffices to use the deterministic per-letter estimator in Lemma* (3) *with replacing the feedback* Z_1 *by the perfect output feedback* Y_1, Y_2.

They also characterized secrecy-distortion characterization for reversely physically degraded BC which is slightly modified compared to Theorem 13. Output statistics of random binning (OSRB) is used in the achievability proofs based on [49]. Also, they presented inner and outer bounds for the general model where there is no degradedness condition on BC.

4.3.2 Sensing at Rx

The model in [50] is an extension of the P2P ISAC system that had been introduced in [22, 33]. In [50], a multihop channel is considered where the CSIs are unavailable to neither the Tx nor the Rx. Similar to [22, 33], the Rx estimates CSIs while a reliable information transmission is happening.
Consider a memoryless two-hop channel where the states are estimated at the destination. A $(2^{nR}, n)-$code consists of:

- Encoding functions: a source encoding function $f_{1n} : \mathcal{W} \to \mathcal{X}_1^n$, and a sequence of relay encoding functions $f_{2,i} : \mathcal{Y}_{1,i} \to \mathcal{X}_{2,i}, i = 1, \cdots, n$,
- Decoding functions: a message decoding function $g_n : \mathcal{Y}_2^n \to \mathcal{W}$, and two state estimation functions $h_{1n} : \mathcal{Y}_2^n \to \mathcal{S}_1^n$ and $h_{2n} : \mathcal{Y}_2^n \to \mathcal{S}_2^n$.

The average probability of decoding error is as in (4.3) in which $W_0 = \emptyset$ and the quality of channel states estimation is measured by bounded per-symbol distortion functions.

Definition 9 ([50]). *If there exists a sequence of* $(2^{nR}, n)$ *codes satisfying* (4.4)*, the rate-distortion tuple* (R, D_1, D_2) *is achievable and supremum of it is defined as capacity-distortion* $C(D_1, D_2)$.

Theorem 14 (Theorem 1, [50]). *The capacity-distortion function for a two-hop channel with independent states is*

$$C(D_1, D_2) = \max_{p_{x_1}p_{x_2}} \min\{I(X_1; Y_1), \quad I(X_2; Y_2) - \min_{p_{\hat{s}_1|x_1y_1}} I(Y_1; \hat{S}_1 \mid X_1)\}, \tag{4.69}$$

such that following distortion constraints are satisfied

$$E_{X_1Y_1,S_1,\hat{S}_1}[d_1(S_1, \hat{S}_1)] \le D_1, \tag{4.70}$$

$$E_{X_2Y_2}[d_2(S_2, \hat{S}_2)] \le D_2, \tag{4.71}$$

where $S_1 -\!\!\circ\!\!- (X_1, Y_1) -\!\!\circ\!\!- \hat{S}_1$ *form a Markov chain and* $\hat{S}_2$ *is the optimal one-shot estimator based on* (X_2, Y_2)*, that is, for each* $x_2 \in \mathcal{X}_2$*, the estimator* $h : X_2 \times Y_2 \to S_2$ *is chosen to minimize* $E[d_2(S_2, h_2(X_2, Y_2)) \mid X_2 = x_2]$.

The result shows that a decode-(indirectly)-compress-and-forward strategy achieves the capacity-distortion function.

4.4 Conclusion

We fully characterized the capacity-distortion tradeoff for physically-degraded BCs. We presented inner and outer bounds on the capacity-distortion region for general BCs. Through several illustrative examples, we demonstrated that the optimal ISAC scheme offers non-negligible gain compared to the time-sharing schemes. Interestingly, there were ideal situations where the capacity was achieved without compromising the sensing performance. Our results also revealed that the optimal sensing depends only on the employed waveform but not on the underlying coding scheme for the single-Tx two-Rx systems.

According to Chapters 3 and 4, we conclude that whenever we have a single Tx, we can accommodate the communication schemes to serve ISAC merely by being cautious about the restrictions on the input distribution to allow for the desired sensing performance. However, this is not the case with more Txs, when Txs can collaborate, both for communication and sensing through generalized feedback signals.

Chapter 5

The Multiple Access Channel

This chapter considers information-theoretic models for ISAC over multi-access channels (MAC) and device-to-device (D2D) communication. The models are general and include as special cases scenarios with and without perfect or imperfect state-information at the MAC Rx as well as causal state-information at the D2D terminals. For both setups, we propose collaborative sensing ISAC schemes where terminals not only convey data to the other terminals but also state-information that they extract from their previous observations. This state-information can be exploited at the other terminals to improve their sensing performances. Indeed, as we show through examples, our schemes improve over previous non-collaborative schemes in terms of their achievable rate-distortion tradeoffs. For D2D we propose two schemes, one where compression of state information is separated from channel coding and one where it is integrated via a hybrid coding approach.

5.1 System Model

Consider the two-Tx single-Rx MAC scenario in Fig. 5.1. The model consists of a two-dimensional memoryless state sequence $\{(S_{1,i}, S_{2,i})\}_{i\geq 1}$ whose samples at any given time i are distributed according to a given joint law $P_{S_1S_2}$ over the state alphabets $\mathcal{S}_1 \times \mathcal{S}_2$. Given that at time-$i$ Tx 1 sends input $X_{1,i} = x_1$ and Tx 2 input $X_{2,i} = x_2$ and given state realizations $S_{1,i} = s_1$ and $S_{2,i} = s_2$, the Rx's time-i output Y_i and the Txs' feedback signals $Z_{1,i}$ and $Z_{2,i}$ are distributed according to the stationary channel transition law $P_{YZ_1Z_2|S_1S_2X_1X_2}(\cdot, \cdot, \cdot|s_1, s_2, x_1, x_2)$. Input and output alphabets $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, \mathcal{Z}_1, \mathcal{Z}_2, \mathcal{S}_1, \mathcal{S}_2$ are assumed finite. A $(2^{n\mathsf{R}_1}, 2^{n\mathsf{R}_2}, n)$ code consists of

1. two message sets $\mathcal{W}_1 = [1 : 2^{n\mathsf{R}_1}]$ and $\mathcal{W}_2 = [1 : 2^{n\mathsf{R}_2}]$;

2. a sequence of encoding functions $\Omega_{k,i} \colon \mathcal{W}_k \times \mathcal{Z}_k^{i-1} \to \mathcal{X}_k$, for $i = 1, 2, \ldots, n$ and $k = 1, 2$;

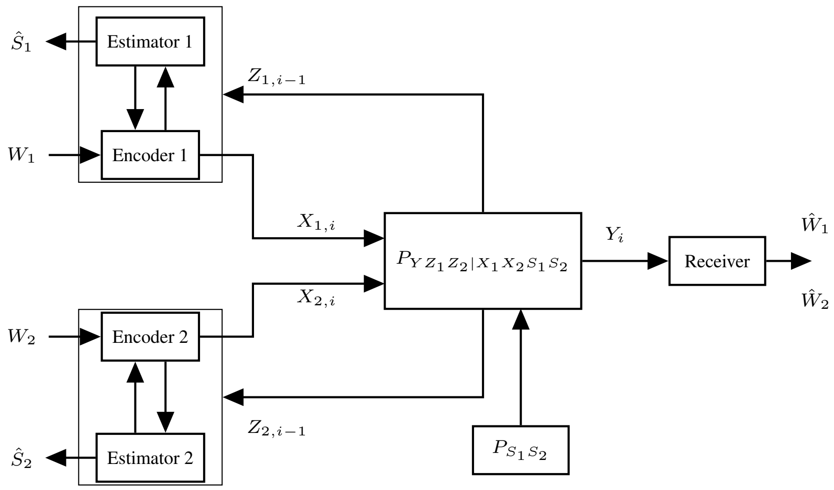


Figure 5.1: State-dependent discrete memoryless multiaccess channel with sensing at the transmitters.

3. a decoding function $g\colon \mathcal{Y}^n \to \mathcal{W}_1 \times \mathcal{W}_2$;

4. for each $k = 1, 2$ a state estimator $\phi_k\colon \mathcal{X}_k^n \times \mathcal{Z}_k^n \to \hat{\mathcal{S}}_k^n$, where $\hat{\mathcal{S}}_1$ and $\hat{\mathcal{S}}_2$ are given reconstruction alphabets.

For a given code, let the random message W_k, for $k = 1, 2$, be uniform over the message set $\mathcal{W}_k$ and the inputs $X_{k,i} = \phi_{k,i}(W_k, Z_k^{i-1})$, for $i = 1, \ldots, n$. The Txs' state estimates are obtained as $\hat{S}_k^n := (\hat{S}_{k,1}, \cdots, \hat{S}_{k,n}) = \phi_k(X_k^n, Z_k^n)$ and the Rx's guess of the messages as $(\hat{W}_1, \hat{W}_2) = g(Y^n)$. We shall measure the quality of the state estimates $\hat{S}_k^n$ by bounded per-symbol distortion functions $d_k\colon \mathcal{S}_k \times \hat{\mathcal{S}}_k \mapsto [0, \infty)$, and consider *expected average block distortions*

$$\Delta_k^{(n)} := \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[d_k(S_{k,i}, \hat{S}_{k,i})], \quad k = 1, 2. \tag{5.1}$$

The probability of decoding error is defined as:

$$P_e^{(n)} := \Pr\Big(\hat{W}_1 \neq W_1 \quad \text{or} \quad \hat{W}_2 \neq W_2\Big). \tag{5.2}$$

Definition 10. *A rate-distortion tuple* $(\mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ *is achievable if there exists a sequence (in* n*) of* $(2^{n\mathsf{R}_1}, 2^{n\mathsf{R}_2}, n)$ *codes that simultaneously satisfy*

$$\lim_{n\to\infty} P_e^{(n)} = 0 \tag{5.3a}$$

$$\overline{\lim_{n\to\infty}}\, \Delta_k^{(n)} \leq \mathsf{D}_k, \quad \text{for } k = 1, 2. \tag{5.3b}$$

Definition 11. *The capacity-distortion region* $\mathcal{CD}$ *is the closure of the set of all achievable tuples* $(\mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$.

Remark 6 (On the States)**.** *Notice that the general law* $P_{S_1S_2}$ *governing the states* S_1^n *and* S_2^n *allows to model* *various types of situations including scenarios where the state sequences are highly correlated (even identical) or scenarios where the state-sequences are independent.*

Our model also includes a scenario where the channel is governed by an internal i.i.d. state sequence S^n of pmf P_S and the states S_1^n, S_2^n are related to S^n over an independent memoryless channel $P_{S_1S_2|S}$. For example, the states S_1^n and S_2^n can be imperfect or noisy versions of the actual state sequence S^n. To see that this scenario can be included in our model, notice that since no terminal observes S^n nor attempts to reconstruct S^n, both the distortions and the error probabilities only depend on the conditional law

$$P_{YZ_1Z_2|X_1X_2S_1S_2}(y,z_1,z_2|x_1,x_2,s_1,s_2) = \sum_s P_{YZ_1Z_2|X_1X_2S}(y,z_1,z_2|x_1,x_2,s)\frac{P_S(s)P_{S_1S_2|S}(s_1,s_2|s)}{P_{S_1S_2}(s_1,s_2)}, \tag{5.4}$$

where $P_{S_1S_2}(s_1,s_2) = \sum_s P_S(s)P_{S_1S_2|S}(s_1,s_2|s)$ denotes the joint pmf of the two states. Computing the channel law in (5.4) *and plugging it into our results in the next section, thus immediately also provides results for the described setup where the actual state is S^n and the states S_1^n and S_2^n are noisy versions thereof.*

Remark 7 (State-Information)**.** *Our model also includes scenarios with perfect or imperfect state-information at the Rx. In fact, considering our model with an output*

$$Y = (T, Y') \tag{5.5}$$

where Y' denotes the actual MAC output and T the Rx's imperfect channel state-information about the states S_1^n and S_2^n. Notice that in our model, the Rx observes the state-information T^n only in a causal manner. Causality is however irrelevant here since the Rx only has to decode the messages at the end of the entire transmission. Therefore, plugging the choice (5.5) *into our results for T the Rx state-information and Y' the actual MAC output, our results in the following section directly lead to results for this related setup with Rx state-information.*

5.2 A Collaborative ISAC Scheme

Before describing our collaborative ISAC scheme for the MAC, we review literature on the MAC and in particular Willem's scheme for the MAC with generalized feedback, which acts as a building block on our scheme.

While the capacity region of the MAC without feedback was determined in [51], a computable single-letter expression for the capacity region is only known in special cases such as the two-user Gaussian MAC with perfect feedback [52] or a class of semi-deterministic MACs [53] with one-sided perfect feedback. Kramer had obtained a multi-letter characterization of the capacity region of the MAC with feedback in [54]. For most channels it seems however impossible to evaluate and compute Kramer's region. Various computable inner and outer bounds have been obtained on the MAC with feedback, the one most relevant to our work is Willems's inner bound [55]. Like previous inner bounds, it is based on the idea that each Tx decodes part of the data sent by the other Tx, which allows the two Txs to cooperatively resend these data parts in the next block using a more efficient coding scheme.

5.2.1 Willems' Coding Scheme with Generalized Feedback and the ISAC extension

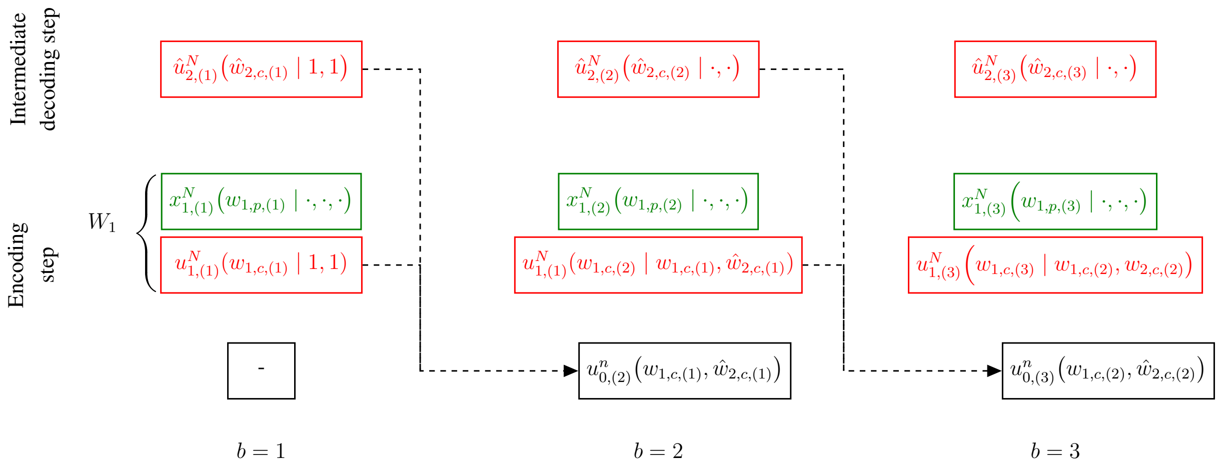


Figure 5.2: Operations at Tx 1 in Willems' scheme during the first three blocks. After each block b Tx 1 decodes message $W_{2,c,(b)}$ based on its generalized feedback output $Z^N_{1,(b)}$. The decoded message is then retransmitted in block $b+1$ jointly with $W_{1,c,(b)}$.

Willems' scheme splits the blocklength n into $B+1$ blocks of length $N = n/(B+1)$ each. Accordingly, throughout, we let $X^N_{1,(b)}, X^N_{2,(b)}, S^N_{1,(b)}, S^N_{2,(b)}, Z^N_{1,(b)}, Z^N_{2,(b)}, Y^N_{(b)}$ denote the block-b inputs, states and outputs, e.g., $S^N_{1,(b)} := (S_{1(b-1)N+1}, \dots, S_{1,bN})$. We also represent the two messages W_1 and W_2 in a one-to-one way as the $2B$-length tuples

$$W_k = (W_{k,c,(1)}, \dots, W_{k,c,(B)}, W_{k,p,(1)}, \dots, W_{k,p,(B)}), \qquad k \in \{1,2\}, \tag{5.6}$$

where all pairs $(W_{k,c,(b)}, W_{k,p,(b)})$ are independent and uniformly distributed over $\left[2^{N\bar{R}_{k,c}}\right] \times \left[2^{N\bar{R}_{k,p}}\right]$ for $\bar{R}_{k,c} \triangleq \frac{B+1}{B} R_{k,c}$ and $\bar{R}_{k,p} \triangleq \frac{B+1}{B} R_{k,p}$ and $R_{k,c} + R_{k,p} = R_k$.

An independent superposition code is constructed for each block b (see also Figure 5.2):

- A lowest-level code $\mathcal{C}_{0,(b)}$ consisting of $2^{N\bar{R}_{1,c}} \cdot 2^{N\bar{R}_{2,c}}$ codewords $u_{0,(b)}(w_{1,c}, w_{2,c})$ is constructed by drawing all entries i.i.d. according to a auxiliary pmf P_{U_0}.

- On each lowest-level codeword $u_{0,(b)}(w_{1,c}, w_{2,c})$ we superposition two codebooks $\{u^N_{k,(b)}(w'_{k,c} \mid w_{1,c}, w_{2,c})\}$, for $k \in \{1,2\}$ and $w'_{k,c} \in [2^{NR_{k,c}}]$, by drawing the i-th entry of each codeword according to $P_{U_k|U_0}(\cdot \mid u_0)$ where u_0 denotes the i-th entry of $u_0^N(w_{1,c}, w_{2,c})$.

- On each second-layer codeword $u_{k,(b)}^N(w'_{k,c}|w_{1,c}, w_{2,c})$, we superposition a codebook $\{x_{k,(b)}^N(w'_{k,p}|w'_{k,c}, w_{1,c}, w_{2,c})\}$, for $k \in \{1,2\}$ and $w'_{k,p} \in [2^{NR_{k,p}}]$, by drawing the i-th entry of each codeword according to $P_{X_k|U_0U_k}(\cdot \mid u_0, u_k)$ where u_k denotes the i-th entry of $u_{k,(b)}^N(w'_{k,c} \mid w_{1,c}, w_{2,c})$.

As depicted in Figure 5.2, in Willems' scheme, Tx 1 sends the following block-b channel inputs

$$x_{1,(b)}^N = x_{1,(b)}^N\left(W_{1,p,(b)}\middle|W_{1,c,(b)}, W_{1,c,(b-1)}, \hat{W}_{2,c,(b-1)}\right), \qquad b \in \{1, \ldots, B+1\}, \tag{5.7}$$

where $\hat{W}_{2,c,(b-1)}$ denotes the message part that Tx 1 decodes after reception of the block-$(b-1)$ generalized feedback signal $Z_{1,(b-1)}^N$, e.g., through a joint typicality decoding rule. Also, we set throughout $W_{k,c,(0)} = \hat{W}_{k,c,(0)} = W_{k,p,(B+1)} = 1$, for $k \in \{1,2\}$.

Decoding at the Rx is performed backwards, starting with the last block $B+1$ based on which the Rx decodes the pair of common messages $(W_{1,c,(B)}, W_{2,c,(B)})$ using for example a joint-typicality decoder. It then uses knowledge of these common messages and the outputs in block B to decode the block-B private messages $(W_{1,p,(B)}, W_{2,p,(B)})$ and the block $(B-1)$ common messages $(W_{1,c,(B-1)}, W_{2,c,(B-1)})$, etc. The backward decoding procedure is also depicted in Figure 5.3.

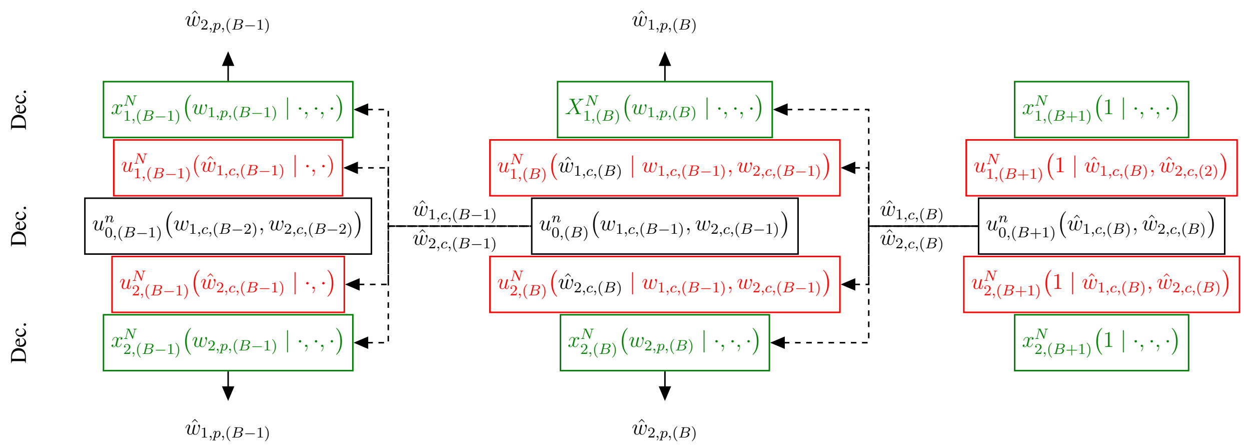


Figure 5.3: Backward decoding procedure at the Rx in Willems' scheme. The pair of common messages $(W_{1,c,(b-1)}, W_{2,c,(b-1)})$ and private messages $(W_{1,p,(b)}, W_{2,p,(b)})$ are jointly decoded based on the block-b outputs $Y_{(b)}^N$ and using the previously decoded $(\hat{W}_{1,c,(b)}, \hat{W}_{2,c,(b)})$.

As Willems showed, his scheme can achieve the following rate-region.

Theorem 15 (Willems' Achievable Region [55]). *Any nonnegative rate-pair (R_1, R_2) is achievable over the MAC with generalized feedback if it satisfies the following inequalities*

$$R_k \leq I(X_k; Y \mid X_{\bar{k}} U_k U_0) + I(U_k; Z_{\bar{k}} \mid X_{\bar{k}} U_0), \qquad k \in \{1, 2\}, \tag{5.8}$$

$$R_1 + R_2 \leq I(X_1 X_2; Y), \tag{5.9}$$

$$R_1 + R_2 \leq I(X_1 X_2; Y \mid U_0 U_1 U_2) + I(U_1; Z_2 \mid X_2 U_0) + I(U_2; Z_1 \mid X_1 U_0), \tag{5.10}$$

for some choice of pmfs $P_{U_0}, P_{U_1|U_0}, P_{U_2|U_0}, P_{X_1|U_0U_1}, P_{X_2|U_0U_2}$, and where above mutual informations are calculated according to the pmf $P_{U_0} P_{U_1|U_0} P_{U_2|U_0} P_{X_1|U_0U_1} P_{X_2|U_0U_2} P_{S_1S_2}\ P_{YZ_1Z_2|S_1S_2X_1X_2}$.

Kobayashi et al. [2] extended Willems' scheme to a ISAC scenario by adding a state estimator at the two Txs. Specifically, for any block b each Tx k applies the symbol-per-symbol estimation

$$\hat{s}^N_{k,(b)} = \tilde{\phi}^{*\otimes N}_k \left(x^N_{k,(b)}, z^N_{k,(b)}, u^N_{\bar{k},(b)} \left(W_{\bar{k},c,(b)} | W_{k,c,(b-1)}, \hat{W}_{\bar{k},c,(b-1)} \right) \right), \qquad b \in \{1, \dots, B\}, \tag{5.11}$$

where $\tilde{\phi}^*_k$ denotes the optimal estimator of S_k based on the tuple $(X_k, Z_k, U_{\bar{k}})$:

$$\tilde{\phi}^*_k(x_k, z_k, u_{\bar{k}}) := \arg \min_{s'_k \in \hat{\mathcal{S}}_k} \sum_{s_k \in \mathcal{S}_k} P_{S_k|X_k Z_k U_{\bar{k}}}(s_k | x_k, z_k, u_{\bar{k}})\ d_k(s_k, s'_k). \tag{5.12}$$

Thus, any of the two Txs bases its state-estimation not only on its inputs and outputs of a given block but also on the codeword that it decoded from the other Tx.

For the last block $B+1$, Tx k can produce any trivial estimate, e.g., $\hat{s}^N_{k,(B+1)}$ because its influence on the average distortion vanishes as the number of blocks grows, $B \to \infty$.

Combining the described state-estimation with Willems' scheme, the following rate-distortion region can be shown to be achievable.

Theorem 16. *[Kobayashi et al.'s ISAC region [2]] A rate-distortion tuple (R_1, R_2, D_1, D_2) is achievable if it satisfies*

$$R_k \leq I(X_k; Y \mid X_{\bar{k}} U_k U_0) + I(U_k; Z_{\bar{k}} \mid X_{\bar{k}} U_0), \qquad k \in \{1, 2\}, \tag{5.13}$$

$$R_1 + R_2 \leq I(X_1 X_2; Y), \tag{5.14}$$

$$R_1 + R_2 \leq I(X_1 X_2; Y \mid U_0 U_1 U_2) + I(U_1; Z_2 \mid X_2 U_0) + I(U_2; Z_1 \mid X_1 U_0), \tag{5.15}$$

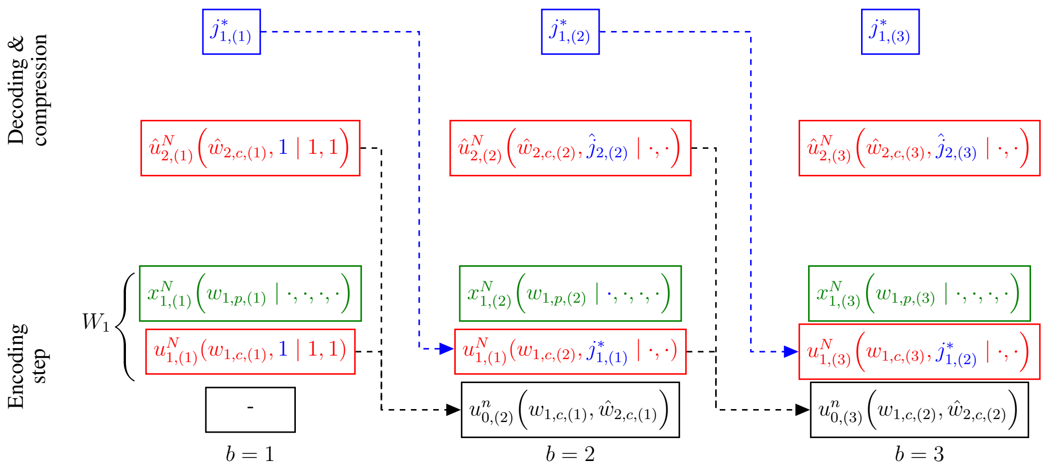


Figure 5.4: Our proposed scheme at Tx 1 during the firts three blocks

and

$$\mathrm{E}\Big[d_k(S_k, \tilde{\phi}_k^*(X_k, Z_k, U_{\bar{k}}, \text{const}))\Big] \leq D_k, \quad k=1,2, \tag{5.16}$$

for some choice of pmfs $P_{U_0}, P_{U_1|U_0}, P_{U_2|U_0}, P_{X_1|U_1U_0}, P_{X_2|U_2U_0}$.

5.2.2 Our Proposed Collaborative ISAC Scheme

We present our collaborative ISAC scheme. It extends the scheme in [2] in that the second-layer codeword of Willems' code construction is not only used to transmit data but also compression information useful for state sensing. The compression information is generated at one of the Txs and mainly intended for the other Tx to improve its sensing performance. In our scheme, the Rx however also decodes this information and uses it to improve its decoding performance.

Code construction

Choose pmfs $P_{U_0}, P_{U_1|U_0}P_{U_2|U_0}, P_{X_1|U_1U_0}, P_{X_2|U_2U_0}$, and define the pmf

$$P_{U_0U_1U_2X_1X_2S_1S_2YZ_1Z_2V_1V_2} = P_{U_0}P_{U_1|U_0}P_{U_2|U_0}P_{X_1|U_1U_0}P_{X_2|U_2U_0}P_{S_1S_2}P_{YZ_1Z_2|X_1X_2S_1S_2} P_{V_1|X_1Z_1U_2U_0}P_{V_1|X_1Z_1U_2U_0}. \tag{5.17}$$

Employ Willems' three-level superposition code construction for the given choice of pmfs, except that each second-layer codeword is indexed by a pair of indices. We thus denote the second-layer codewords by

$$\Big(u_{0,(b-1)}^N\Big(W_{1,c,(b-2)},\hat{W}_{2,c,(b-2)}^{(1)}\Big),\ u_{1,(b-1)}^N\Big(W_{1,c,(b-1)},J_{1,(b-2)}^*\ \Big|\ W_{1,c,(b-2)},\hat{W}_{2,c,(b-2)}^{(1)}\Big)$$
$$u_{2,(b-1)}^N\Big(\hat{w}_2,\hat{j}_2\ \Big|\ W_{1,c,(b-2)},\hat{W}_{2,c,(b-2)}^{(1)}\Big),$$
$$x_{1,(b-1)}^N\Big(W_{1,p,(b-1)}\ \Big|\ W_{1,c,(b-1)},J_{1,(b-2)}^*,W_{1,c,(b-2)},\hat{W}_{2,c,(b-2)}^{(1)}\Big),$$
$$v_{1,(b-1)}^N\Big(j_1^*\ \Big|\ J_{1,(b-2)}^*,W_{1,c,(b-1)},\hat{w}_2,\hat{j}_2,W_{1,c,(b-2)},\hat{W}_{2,c,(b-2)}^{(1)}\Big),Z_{1,(b-1)}^N\Big)\in\mathcal{T}_\epsilon^N(P_{U_0U_1U_2X_1V_1Z_1}) \tag{5.19}$$

$u_{k,(b)}^N(w'_{1,c},j_1 \mid w_{1,c},w_{2,c})$ and $u_{2,(b)}^N(w'_{2,c},j_2 \mid w_{1,c},w_{2,c})$ and accordingly the corresponding third-layer codewords by $x_{1,(b)}^N(w'_{1,p}|w'_{1,c},j_1,w_{1,c},w_{2,c})$ and $x_{1,(b)}^N(w'_{2,p}|w'_{2,c},j_2,w_{1,c},w_{2,c})$, where the indices j_1 and j_2 take value in the sets $[2^{nR'_1}]$ and $[2^{nR'_2}]$ for some positive auxiliary rates $R_{1,v}$ and $R_{2,v}$.

We further construct a compression codebook for each block and each of the two Txs, For each $b\in\{1,\ldots,B\}$ and each sixtuple $(w_{1,c},w_{2,c},w'_{1,c},j_1w'_{2,c},j_2)\in[2^{NR_{1,c}}]\times[2^{NR_{2,c}}]\times[2^{NR_{1,c}}]\times[2^{NR_{1,v}}]\times[2^{NR_{2,c}}]\times[2^{NR_{2,v}}]$ we generate a sequence $v_{1,(b)}^N(j'_1 \mid w'_{1,c},j_1,w'_{2,c},j_2,w_{1,c},w_{2,c})$ for each $j'_1\in[2^{NR_{1,v}}]$ and a sequence $v_{2,(b)}^N(j'_2 \mid w'_{1,c},j_1,w'_{2,c},j_2,w_{1,c},w_{2,c})$ for each $j'_2\in[2^{NR_{2,v}}]$. The sequences $v_{1,(b)}^N(j'_1 \mid w'_{1,c},j_1,w'_{2,c},j_2,w_{1,c},w_{2,c})$ and $v_{2,(b)}^N(j'_2 \mid w'_{1,c},j_1,w'_{2,c},j_2,w_{1,c},w_{2,c})$ are obtained by drawing their i-th entries according to $P_{V_1|U_0U_1U_2}(\cdot \mid u_0,u_1,u_2)$ and $P_{V_2|U_0U_1U_2}(\cdot \mid u_0,u_1,u_2)$, respectively, for u_0,u_1,u_2 denoting the i-th entries of the sequences $u_{0,(b)}^N(w_{1,c},w_{2,c})$, $u_{1,(b)}^N(w'_{1,c},j_1 \mid w_{1,c},w_{2,c})$, and $u_{2,(b)}^N(w'_{2,c},j_2 \mid w_{1,c},w_{2,c})$.

Operations at the Txs

In each block b, Tx k sends the block-b sequence

$$X_{k,(b)}^N=x_{k,(b)}^N\left(W_{k,p,(b)}|W_{k,c,(b)},J_{k,(b-1)}^*,W_{k,c,(b-1)},\hat{W}_{\bar{k},c,(b-1)}^{(k)}\right), \tag{5.18}$$

where Tx k generates the indices $J_{k,(b-1)}^*$ and $\hat{W}_{\bar{k},c,(b-1)}$ during a joint decoding and compression step at the end of block $b-1$ as follows. (For convenience we again set $W_{k,p,(B+1)}=W_{k,c,(0)}=\hat{W}_{\bar{k},c,(0)}^k=J_{k,(B+1)}^{\bar{k}}=1$.)

After receiving the generalized feedback signal $Z_{k,(b-1)}^N$, Tx k looks for a triple of indices j_k^*, $\hat{w}_{\bar{k}}$, and $\hat{j}_{\bar{k}}$ satisfying the joint typicality check (5.19), and if $b>2$ also the typicality check (5.20), which are displayed on top of the page. It randomly picks one of these triples and sets

$$J_{1,(b-1)}^*=j_1^*,\qquad \hat{W}_{2,(b-1)}^{(1)}=\hat{w}_2,\qquad \hat{J}_{2,(b-2)}^{(1)}=\hat{j}_2. \tag{5.21}$$

$$\begin{aligned}&\Big(u_{0,(b-2)}^N\Big(W_{1,c,(b-3)},\hat{W}_{2,c,(b-3)}^{(1)}\Big),\ u_{1,(b-2)}^N\Big(W_{1,c,(b-2)},J_{1,(b-2)}^*\ \Big|\ W_{1,c,(b-3)},\hat{W}_{2,c,(b-3)}^{(1)}\Big),\\&\quad u_{2,(b-2)}^N\Big(\hat{W}_{2,c,(b-2)}^{(1)},\hat{J}_{2,(b-3)}^{(1)}\ \Big|\ W_{1,c,(b-3)},\hat{W}_{2,c,(b-3)}^{(1)}\Big),\\&\quad x_{1,(b-2)}^N\Big(W_{1,p,(b-2)}\ \Big|\ W_{1,c,(b-2)},J_{1,(b-3)}^*,W_{1,c,(b-3)},\hat{W}_{2,c,(b-3)}^{(1)}\Big),\\&\quad v_{2,(b-2)}^N\Big(\hat{j}_2\ \Big|\ W_{1,c,(b-2)},J_{1,(b-3)}^*,\hat{W}_{2,c,(b-2)}^{(1)},\hat{J}_{2,(b-3)}^{(1)},W_{1,c,(b-3)},\hat{W}_{2,c,(b-3)}^{(1)}\Big),\\&\qquad\qquad\qquad\qquad Z_{1,(b-2)}^N\Big)\in\mathcal{T}_\epsilon^N(P_{U_0U_1U_2X_1V_2Z_1}).\end{aligned}\tag{5.20}$$

Tx k also produces the block-b state estimate

$$\begin{aligned}\hat{s}_{k,(b)}^n=\phi_k^{*\otimes N}\Big(&x_{k,(b)}^N\left(W_{k,p,(b)}|W_{k,c,(b)},J_{k,(b-1)},W_{k,c,(b-1)},\hat{W}_{\bar{k},c,(b-1)}^{(k)}\right),\\&z_{k,(b)}^N,u_{\bar{k},(b)}^N\left(W_{k,c,(b)}\mid J_{k,(b-1)},W_{k,c,(b-1)},\hat{W}_{\bar{k},c,(b-1)}^{(k)}\right),\\&v_{\bar{k},(b)}^N\left(J_{k,(b)}\mid W_{k,c,(b)},J_{k,(b-1)},W_{k,c,(b-1)},\hat{W}_{\bar{k},c,(b-1)}^{(k)}\right)\Big)\end{aligned}\tag{5.22}$$

where

$$\phi_k^*(x_k,z_k,u_{\bar{k}},v_{\bar{k}}):=\arg\min_{s_k'\in\hat{\mathcal{S}}_k}\sum_{s_k\in\mathcal{S}_k}P_{S_k|X_kZ_kU_{\bar{k}}V_{\bar{k}}}(s_k|x_k,z_k,u_{\bar{k}},v_{\bar{k}})\ d_k(s_k,s_k').\tag{5.23}$$

Without loss in performance as $B\to\infty$, the estimate in the last block $B+1$ can again be set to a dummy sequence.

Decoding at the Rx

Decoding at the Rx is similar to Willems' scheme and uses backward decoding. The difference is that the Rx in block b not only decodes the message tuple $(W_{1,p,(b)},W_{2,p,(b)},W_{1,c,(b-1)},W_{2,c,(b-1)})$ but also the compression indices $J_{1,(b-1)}^*$ and $J_{2,(b-2)}^*$. Specifically, in a generic block $b\in\{2,\dots,B\}$, the Rx looks for a unique sixtuple $(w_{1,p},w_{2,p},w_{1,c},w_{2,c},j_1,j_2)\in[2^{NR_{1,p}}]\times[2^{NR_{2,p}}]\times[2^{NR_{1,c}}]\times[2^{NR_{2,c}}]\times[2^{NR_{1,v}}]\times[2^{NR_{2,v}}]$ satisfying

$$\begin{aligned}\Bigg(&u_{0,b}^N(w_{1,c},w_{2,c}),\ u_{1,(b)}^N\Big(\hat{W}_{1,c,(b)},j_1\ \Big|\ w_{1,c},w_{2,c}\Big),\ u_{2,(b)}^N\Big(\hat{W}_{2,c,(b)},j_2\ \Big|\ w_{1,c},w_{2,c}\Big),\\&x_{1,(b)}^N\Big(w_{1,p}\ \Big|\ \hat{W}_{1,c,(b)},j_1,w_{1,c},w_{2,c}\Big),\ x_{2,(b)}^N\Big(w_{2,p}\ \Big|\ \hat{W}_{2,c,(b)},j_2,w_{1,c},w_{2,c}\Big),\\&\quad v_{1,(b)}^N\left(\hat{J}_{1,(b)}\ \Big|\ \hat{W}_{1,c,(b)},j_1,\hat{W}_{2,c,(b)},j_2,w_{1,c},w_{2,c}\right),\\&\quad v_{2,(b)}^N\left(\hat{J}_{2,(b)}\ \Big|\ \hat{W}_{1,c,(b)},j_1,\hat{W}_{2,c,(b)},j_2,w_{1,c},w_{2,c}\right),\ Y_{(b)}^N\Bigg)\in\mathcal{T}_{2\epsilon}(P_{U_0U_1U_2X_1X_2Y}).\end{aligned}\tag{5.24}$$

If such a unique sixtuple exists, it sets $\hat{W}_{1,c,(b-1)} = w_{1,c}$, $\hat{W}_{1,p,(b)} = w_{1,p}$, $\hat{W}_{2,c,(b-1)} = w_{2,c}$, $\hat{W}_{2,p,(b)} = w_{2,p}$, $\hat{J}_{1,(b-1)} = j_1$, and $\hat{J}_{2,(b-1)} = j_2$. Otherwise it declares an error.

The Rx finally declares the messages $\hat{W}_1$ and $\hat{W}_2$ that correspond to the produced guesses $\{(\hat{W}_{k,p,(b)}, \hat{W}_{k,c,(b)})\}$.

In Appendix A.3.1 we show that as $N \to \infty$ and $B \to \infty$, the described scheme achieves vanishing probabilities of error, the compressions are successful with probability 1, and the asymptotic expected distortions are bounded by D_1 and D_2 whenever B is sufficiently large and

$$R_{k,v} > I(V_k; X_k Z_k \mid \underline{U}) \tag{5.25a}$$

$$R_{\bar{k},v} + R_{k,c} < I(U_k V_{\bar{k}}; X_{\bar{k}} Z_{\bar{k}} \mid U_0 U_{\bar{k}}) \tag{5.25b}$$

$$R_{1,v} + R_{2,v} + R_{k,c} < I(U_k V_{\bar{k}}; X_{\bar{k}} Z_{\bar{k}} \mid U_0 U_{\bar{k}}) + I(V_k; X_{\bar{k}} Z_{\bar{k}} \mid \underline{U}) \tag{5.25c}$$

$$R_{k,p} < I(X_k; Y V_1 V_2 \mid \underline{U} X_{\bar{k}}) \tag{5.25d}$$

$$R_{k,v} + R_{k,p} < I(X_k; Y \mid U_0 X_{\bar{k}}) + I(V_2; X_1 X_2 Y V_1 \mid \underline{U}) + I(V_1; X_1 X_2 Y \mid \underline{U}) \tag{5.25e}$$

$$\begin{aligned} R_{k,v} + R_{k,p} + R_{\bar{k},p} < {} & I(X_1 X_2; Y \mid U_0 U_{\bar{k}}) + I(V_2; X_1 X_2 Y V_1 \mid \underline{U}) \\ & + I(V_1; X_1 X_2 Y \mid \underline{U}) \end{aligned} \tag{5.25f}$$

$$R_{1,p} + R_{2,p} < I(X_1 X_2; Y V_1 V_2 \mid \underline{U}) \tag{5.25g}$$

$$R_{1,v} + R_{1,p} + R_{2,v} + R_{2,p} < I(X_1 X_2; Y \mid U_0) + I(V_1; X_1 X_2 Y \mid \underline{U}) + I(V_2; X_1 X_2 Y V_1 \mid \underline{U}) \tag{5.25h}$$

$$R_{1,v} + R_1 + R_{2,v} + R_2 < I(X_1 X_2; Y) + I(V_1; X_1 X_2 Y \mid \underline{U}) + I(V_2; X_1 X_2 Y V_1 \mid \underline{U}), \tag{5.25i}$$

where $\underline{U} \triangleq (U_0, U_1, U_2)$ and

$$\mathrm{E}[d_k(S_k, \phi_k^*(X_k, Z_k, U_{\bar{k}}, V_{\bar{k}})] \leq D_k, \quad k = 1, 2, \tag{5.25j}$$

for ϕ_k^* defined in (5.23).

Using the Fourier-Motzkin Elimination (FME) algorithm it can be shown, see Appendix A.3.2, that such a choice of rates is possible under the rate-constraints (5.26).

Theorem 17. *The capacity-distortion region $\mathcal{CD}$ includes any rate-distortion tuple (R_1, R_2, D_1, D_2) that for some choice of pmfs $P_{U_0}, P_{U_1|U_0}, P_{U_2|U_0}, P_{X_1|U_0U_1}, P_{X_2|U_0U_2}, P_{V_1|U_0U_2X_1Z_1}, P_{V_2|U_0U_1X_2Z_2}$ satisfies Inequalities* (5.26) *on top of the previous page (where $\underline{U} := (U_0, U_1, U_2)$) as well as the distortion constraints* (5.25j).

Proof. See Appendix A.3.1. □

$$
\begin{aligned}
R_k \leq\ & I(U_k; X_{\bar{k}} Z_{\bar{k}} \mid U_0 U_{\bar{k}}) + I(V_k; X_{\bar{k}} Z_{\bar{k}} \mid \underline{U}) - I(V_k; X_k Z_k \mid \underline{U}) + \min\{ \\
& I(X_k; Y \mid U_0 X_{\bar{k}}) + I(V_k; X_1 X_2 Y \mid \underline{U}) + I(V_{\bar{k}}; X_1 X_2 Y V_k \mid \underline{U}) \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad -I(V_k; X_k Z_k \mid \underline{U}), \\
& I(X_1 X_2; Y \mid U_0 U_k) + I(V_k; X_1 X_2 Y \mid \underline{U}) + I(V_{\bar{k}}; X_1 X_2 Y V_k \mid \underline{U}) \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad -I(V_{\bar{k}}; X_{\bar{k}} Z_{\bar{k}} \mid \underline{U}), \\
& I(X_1 X_2; Y \mid U_0) + I(V_k; X_1 X_2 Y \mid \underline{U}) + I(V_{\bar{k}}; X_1 X_2 Y V_k \mid \underline{U}) \\
& \quad -I(V_k; X_k Z_k \mid \underline{U}) - I(V_{\bar{k}}; X_{\bar{k}} Z_{\bar{k}} \mid \underline{U}),\ I(X_k; Y V_1 V_2 \mid \underline{U} X_{\bar{k}})\},\ k = 1, 2,
\end{aligned} \tag{5.26a}
$$

$$
\begin{aligned}
R_1 + R_2 \leq\ & I(U_2; X_1 Z_1 \mid U_0 U_1) + I(V_2; X_1 Z_1 \mid \underline{U}) - I(V_2; X_2 Z_2 \mid \underline{U}) \\
& + I(U_1; X_2 Z_2 \mid U_0 U_2) + I(V_1; X_2 Z_2 \mid \underline{U}) - I(V_1; X_1 Z_1 \mid \underline{U}) + \min\{ \\
& I(X_1 X_2; Y \mid U_0 U_2) + I(V_1; X_1 X_2 Y \mid \underline{U}) + I(V_2; X_1 X_2 Y V_1 \mid \underline{U}) - I(V_1; X_1 Z_1 \mid \underline{U}), \\
& I(X_1 X_2; Y \mid U_0 U_1) + I(V_1; X_1 X_2 Y \mid \underline{U}) + I(V_2; X_1 X_2 Y V_1 \mid \underline{U}) - I(V_2; X_2 Z_2 \mid \underline{U}), \\
& I(X_1 X_2; Y \mid U_0) + I(V_1; X_1 X_2 Y \mid \underline{U}) + I(V_2; X_1 X_2 Y V_1 \mid \underline{U}) \\
& \qquad\qquad\qquad\qquad\qquad\qquad -I(V_1; X_1 Z_1 \mid \underline{U}) - I(V_2; X_2 Z_2 \mid \underline{U}), \\
& I(X_1 X_2; Y V_1 V_2 \mid \underline{U})\}
\end{aligned} \tag{5.26b}
$$

$$
\begin{aligned}
R_1 + R_2 \leq\ & I(X_1 X_2; Y) + I(V_1; X_1 X_2 Y \mid \underline{U}) - I(V_1; X_1 Z_1 \mid \underline{U}) \\
& \qquad\qquad\qquad + I(V_2; X_1 X_2 Y V_1 \mid \underline{U}) - I(V_2; X_2 Z_2 \mid \underline{U})
\end{aligned} \tag{5.26c}
$$

and for $k = 1, 2$

$$
I(U_k; X_{\bar{k}} Z_{\bar{k}} \mid U_0 U_{\bar{k}}) + I(V_k; X_{\bar{k}} Z_{\bar{k}} \mid \underline{U}) \geq I(V_k; X_k Z_k \mid \underline{U}), \tag{5.26d}
$$

$$
\begin{aligned}
I(X_1 X_2; Y \mid U_0) + I(V_1; X_1 X_2 Y \mid \underline{U}) + I(V_2; X_1 X_2 Y V_1 \mid \underline{U}) \geq\ & I(V_1; X_1 Z_1 \mid \underline{U}) \\
& + I(V_2; X_2 Z_2 \mid \underline{U})
\end{aligned} \tag{5.26e}
$$

$$
I(X_k; Y \mid U_0 X_{\bar{k}}) + I(V_1; X_1 X_2 Y \mid \underline{U}) + I(V_2; X_1 X_2 Y V_1 \mid \underline{U}) \geq I(V_k; X_k Z_k \mid \underline{U}). \tag{5.26f}
$$

Notice that Theorem 17 recovers the previous achievable region in Theorem 16 through the choice $V_1 = V_2 =$constants, which removes the collaborative sensing between the two Txs.

Remark 8 (Wyner-Ziv Coding)**.** *In our scheme, no binning as in Wyner-Ziv coding is used for the compression of the V_1- and V_2-codewords. Instead, decoder side-information is taken into account through the additional typicality check* (5.20) *and by including the V_1- and V_2-codewords in the typicality check* (5.24)*. These strategies are known as implicit binning and allow multiple decoders to exploit different levels of side-information, see [56].*

5.2.3 Examples

The following two examples show the improvement of Theorem 17 over Theorem 16.

Example 1. *Consider a MAC with binary input, output, and state alphabets $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y} = \mathcal{S}_2 = \{0, 1\}$. State*

$S_2 \sim Ber(p_s)$, while $S_1 = 0$ is a constant. The channel input-output relation is described by

$$Y = S_2 X_2, \qquad (Z_1, Z_2) = (S_2, X_1). \tag{5.27}$$

For this channel, the following tuple

$$(\mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2) = (0, 0, 0, 0), \tag{5.28}$$

lies in the achievable region of Theorem 17 through the choice $V_1 = Z_1 = S_2$ and $(\hat{S}_2 = V_1, \hat{S}_1 = 0)$. Distortion $\mathsf{D}_2 = 0$ is however not achievable in Theorem 16 because S_2 is independent of $(U_1, U_2, U_0, X_1, X_2)$ and thus of (X_2, U_1, Z_2), and the optimal estimator is the trivial estimator $\hat{S}_2 = \psi_2^*(X_2, Z_2, U_1) = \mathbb{1}\{p_s > 1/2)$ which achieves distortion $\mathsf{D}_2 = \min\{1 - p_s, p_s\}$.

Example 2. *Consider binary noise, states and channel inputs $B_0, B_k, S_k, X_k \in \{0, 1\}$. The noise to the Rx B_0 is Bernoulli-t_0, and B_k, the noise on the feedback to Tx k, is Bernoulli-t_k. All noises are independent and also independent of the states S_1, S_2, which are i.i.d. Bernoulli-p_s. We can then des are described the channel as*

$$Y' = S_1 X_1 + S_2 X_2 + B_0, \qquad Y = (Y', S_1, S_2), \tag{5.29}$$

$$Z_1 = S_1 X_1 + S_2 X_2 + B_1, \qquad Z_2 = S_1 X_1 + S_2 X_2 + B_2. \tag{5.30}$$

In this example the Rx thus has perfect channel state-information, see also Remark 7. Hamming distance is considered as a distortion measure: $d(s, \hat{s}) = s \oplus \hat{s}$.

We compare Theorems 16 and 17 on the following choices of random variables. Let

$$X_k = \underbrace{U_0 \oplus \Sigma}_{\triangleq U_1} \oplus \theta_k, \quad \text{for } k \in \{1, 2\} \tag{5.31}$$

where $U_0, \Sigma_1, \Sigma_2, \theta_1, \theta_2$ are all independent Bernoulli random variables of parameters p, q_1, q_2, r_1, r_2. For the evaluation of Theorem 17 we further choose the compression random variables

$$V_k = \begin{cases} \mathbb{1}\{Z_k = 1\} + 2 \cdot \mathbb{1}\{Z_k = 2\} & \text{if } E_k = 0 \\ \text{“?”} & \text{if } E_k = 1 \end{cases} \qquad \forall k = \{1, 2\} \tag{5.32}$$

for a binary E_k independent of $(S_1, S_2, B_0, B_1, B_2, U_0, U_1, U_2, \Sigma_1, \Sigma_2, \theta_1, \theta_2)$. For this choice, Tx k conveys information about Z_k to Tx $\bar{k}$, which helps this latter to better estimate its state $S_{\bar{k}}$. For instance, when $E_1 = 0$, Tx-2 *receives another noisy observation of the output which helps it to better estimate its state, because*

$$Y' = \begin{cases} 0 & \text{if } Z_2 \in \{0,1\}, V_1 = 0 \\ 1 & \text{if } V_1 = 1 \\ 2 & \text{if } Z_2 \in \{2,3\}, V_1 = 0 \end{cases} . \tag{5.33}$$

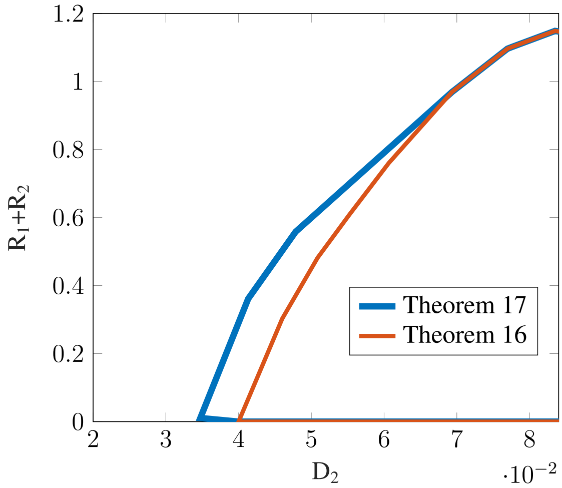


Figure 5.5: Sum-rate distortion tradeoff achieved by Theorems 16 and 17 in Example 2 for given channel parameters $p_s = 0.9$, $t_0 = 0.3$, $t_1 = 0.1$ and $t_2 = 0.1$.

Proof. First, we discuss the computation of distortion.

Calculation of Distortion

We evaluate the expected distortion of the estimators in (5.23), for a given input pmf $P_{X_1X_2}$. We consider the distortion on state S_2 where we notice different cases in the calculation:

- When Tx 2 sends $X_2 = 0$, the distortion calculation is the same as calculation in Corollary 16. When Tx 2 sends $X_2 = 0$, it vanishes the effect of S_2, and thus the distortion is a minima over two states
- When Tx 2 sends $X_2 = 1$, the distortion depends on different feedback observations as follows:
  - If Tx 2 observes $Z_2 = 0$, then it knows $S_1 = S_2 = 0$, thus, the distortion is zero. This observation also provides information about the noise $B_2 = 0$.
  - If Tx 2 observes $Z_2 = 3$, then Tx 2 knows $S_1 = S_2 = 1$, thus, the distortion is zero. Moreover it provides information regarding the noise $B_2 = 1$.

– The advantage of using the proposed auxiliary r.v V_1 in our scheme compared to Corollary 16 arises when Tx 2 observes $Z_2 = 1$ or 2 categorized as follows:
  * If Tx 2 observes $Z_2 = 1$, and
    · decodes $V_1 = 0$, which means $Z_1 = \{0, 3\}$. The zero-distortion estimation provides $S_2 = 0$ or,
    · decodes $V_1 = 1$ which means $Z_1 = 1$. The estimation stands with a non-zero distortion or,
    · decodes $V_1 = 2$, which means $Z_1 = 2$. The estimation stands with a non-zero distortion.
  * If Tx 2 observes $Z_2 = 2$ and
    · decodes $V_1 = 0$, it knows $Z_1 = 3$, which means $S_2 = 0$ and thus the distortion is zero or,
    · decodes $V_1 = 1$, which means $Z_1 = 1$. In this case the distortion is non-zero but less than in Corollary 16 because Z_1 is available at Tx 2 as side-information.
    · decodes $V_1 = 2$ which means $Z_1 = 2$. Analogous to previous case, the distortion is non-zero.

We obtain the following final expression for the distortion at Tx 2:

$$\begin{aligned} D_2^{(5.32)} &= P_{X_2}(0) \min\{p_s, \bar{p}_s\} \\ &+ P_{U_1X_2}(0,1) \cdot \min\Big\{\bar{p}_s \cdot p_s r_1 \bar{t}_1 \bar{t}_2 + (1 - p_s r_1) t_1 t_2, \quad p_s \cdot (1 - p_s r_1) \bar{t}_1 \bar{t}_2\Big\} \\ & && (5.34) \\ &+ P_{U_1X_2}(1,1) \cdot \min\Big\{\bar{p}_s \cdot p_s \bar{r}_1 \bar{t}_1 \bar{t}_2 + (1 - p_s \bar{r}_1) t_1 t_2, \quad p_s \cdot (1 - p_s \bar{r}_1) \bar{t}_1 \bar{t}_2\Big\} && (5.35) \end{aligned}$$

While in Corollary 16, the constraint (5.16) evaluates to

$$\begin{aligned} D_2^{\text{Corollary16}} &= P_{X_2}(0) \min\{p_s, \bar{p}_s\} + P_{U_1X_2}(0,1) \Bigg[\min\{\bar{p}_s(p_s r_1 \bar{t} + t(1 - p_s r_1)), \bar{t}(1 - p_s r_1) p_s\} && (5.36) \\ &\qquad + \min\{p_s(p_s r_1 \bar{t} + t(1 - p_s r_1))\} \Bigg] \\ &+ P_{U_1X_2}(1,1) \Bigg[\min\{\bar{p}_s(p_s \bar{r}_1 \bar{t} + t(1 - p_s \bar{r}_1)), \bar{t}(1 - p_s \bar{r}_1) p_s\} && (5.37) \\ &\qquad + \min\{p_s(p_s \bar{r}_1 \bar{t} + t(1 - p_s \bar{r}_1))\} \Bigg] \end{aligned}$$

The details of calculation is given in Appendix A.4.1 and the evaluation of the rate region is given in Appendix A.4.2.

□

To achieve Figure5.5, we optimize the final expressions over all $p, q_k, r_k, p_{e_k} \in [0, 1]$ for specific channel parameters $p_s = 0.9$, $t_0 = 0.3$, $t_1 = 0.1$ and $t_2 = 0.1$. Figure 5.5 shows the maximum sum-rate $R_1 + R_2$ in function of distortion D_2 achieved by Theorems 16 and 17, where recall that for the region in Theorem 16 we set $V_1 = V_2 = 0$. Notice that both curves are strictly concave and thus improve over classic time- and resource sharing strategies. The minimum distortions achieved by Theorems 16 and 16 are $D_{2,\min} = 0.035$ and $D_{2,\min} = 0.04$.

5.3 Conclusion

We considered integrated sensing and communication (ISAC) over multi-access channels (MAC) where different terminals help each other to improve sensing. We reviewed related communication schemes and proposed adaptations that fully integrate the collaborative sensing into information-theoretic data communication schemes. Through examples, we demonstrated the advantages of our collaborative sensing ISAC schemes compared to non-collaborative ISAC schemes with respect to the achieved rate-distortion regions.

Chapter 6

The Device-to-Device Channel

In this chapter, we consider the ISAC two-way channel, where two terminals exchange data over a common channel and based on their inputs and outputs also wish to estimate the state-sequences that govern the two-way channel.

6.1 System Model

Consider the two-terminal two-way communication scenario in Fig. 6.1. The model consists of a two-dimensional memoryless state sequence $\{(S_{1,i}, S_{2,i})\}_{i\geq 1}$ whose samples at any given time i are distributed according to a given joint law $P_{S_1S_2}$ over the state alphabets $\mathcal{S}_1 \times \mathcal{S}_2$. Given that at time-$i$ Tx 1 sends input $X_{1,i} = x_1$ and Tx 2 input $X_{2,i} = x_2$ and given state realizations $S_{1,i} = s_1$ and $S_{2,i} = s_2$, the Txs' time-i feedback signals $Z_{1,i}$ and $Z_{2,i}$ are distributed according to the stationary channel transition law $P_{Z_1Z_2|S_1S_2X_1X_2}(\cdot,\cdot|s_1, s_2, x_1, x_2)$. Input and output alphabets $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, \mathcal{Z}_1, \mathcal{Z}_2, \mathcal{S}_1, \mathcal{S}_2$ are assumed finite.

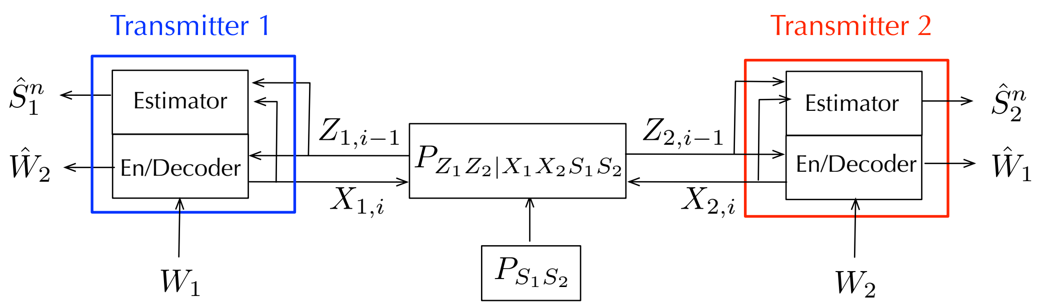


Figure 6.1: State-dependent discrete memoryless two-way channel with sensing at the terminals.

A $(2^{n\mathsf{R}_1}, 2^{n\mathsf{R}_2}, n)$ code consists of

1. two message sets $\mathcal{W}_1 = [1 : 2^{n\mathsf{R}_1}]$ and $\mathcal{W}_2 = [1 : 2^{n\mathsf{R}_2}]$;

2. sequences of encoding functions $\Omega_{k,i}\colon \mathcal{W}_k \times \mathcal{Z}_k^{i-1} \to \mathcal{X}_k$, for $i = 1, 2, \ldots, n$ and $k = 1, 2$;

3. decoding functions $g_k\colon \mathcal{Z}^n \to \mathcal{W}_k$, for $k = 1, 2$;

4. state estimators $\phi_k\colon \mathcal{X}_k^n \times \mathcal{Z}_k^n \to \hat{\mathcal{S}}_k^n$, for $k = 1, 2$, where $\hat{\mathcal{S}}_1$ and $\hat{\mathcal{S}}_2$ are given reconstruction alphabets.

For a given code, let the random message W_k, for $k = 1, 2$, be uniform over the message set $\mathcal{W}_k$ and consider the inputs $X_{k,i} = \phi_{k,i}(W_k, Z_k^{i-1})$, for $i = 1, \ldots, n$. Tx $k \in \{1, 2\}$ obtains its state estimate as $\hat{S}_k^n := (\hat{S}_{k,1}, \cdots, \hat{S}_{k,n}) = \phi_k(X_k^n, Z_k^n)$ and its message guess as $\hat{W}_{3-k} = g_k(Z_k^n, W_k)$.

We shall measure the quality of the state estimates $\hat{S}_k^n$ by bounded per-symbol distortion functions $d_k\colon \mathcal{S}_k \times \hat{\mathcal{S}}_k \mapsto [0, \infty)$, and consider *expected average block distortions*

$$\Delta_k^{(n)} := \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[d_k\left(S_{k,i}, \hat{S}_{k,i}\right)\right], \quad k = 1, 2. \tag{6.1}$$

The probability of decoding error is defined as:

$$P_e^{(n)} := \Pr\Big(\hat{W}_1 \neq W_1 \quad \text{or} \quad \hat{W}_2 \neq W_2\Big). \tag{6.2}$$

Definition 12. *A rate-distortion tuple* $(\mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ *is achievable if there exists a sequence (in* n*) of* $(2^{n\mathsf{R}_1}, 2^{n\mathsf{R}_2}, n)$ *codes that simultaneously satisfy*

$$\lim_{n\to\infty} P_e^{(n)} = 0 \tag{6.3a}$$

$$\overline{\lim_{n\to\infty}}\, \Delta_k^{(n)} \leq \mathsf{D}_k, \quad \text{for } k = 1, 2. \tag{6.3b}$$

Definition 13. *The capacity-distortion region $\mathcal{CD}$ is the closure of the set of all achievable tuples* $(\mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$.

Remark 9 (State-Information at the Terminals)**.** *Considering a two-way channel where*

$$Z_k = (S_{\bar{k}}, Z_k'), \qquad k \in \{1, 2\}, \tag{6.4}$$

for some output Z_k'. This models a situation where each terminal obtains strictly causal state-information about the other terminal's state. Inner bounds for this setup with strictly causal state-information can immediately be obtained from our results presented in the next section by plugging in the choice in (6.4)*. The same remark applies also to* *imperfect strictly-causal state-information in which case the output should be modeled as*

$$Z_k = (T_k, Z'_k), \qquad k \in \{1,2\}, \tag{6.5}$$

where Z'_k again models the actual channel output and T_k models the strictly causal imperfect state-information at Terminal k. Alternatively, T_k could even be related to the desired channel state S_k and not only to the other terminal's state $S_{\bar{k}}$. Plugging the choice (6.5) *into our results for an appropriate choice of T_k leads to results for this related setup with imperfect or generalized state-information at the terminals.*

In contrast, our model does not include causal or non-causal state-information. These are interesting extensions of our work, but left for future research. They would certainly require new tools such as dirty-paper coding [57].

6.2 A Collaborative ISAC Scheme

We first review Han's scheme for pure data communication over the two-way channel and then include the collaborative sensing idea in Han's scheme. Finally we integrate collaborative sensing and communication through joint source-channel coding (JSCC).

6.2.1 Han's Two-Way Coding Scheme

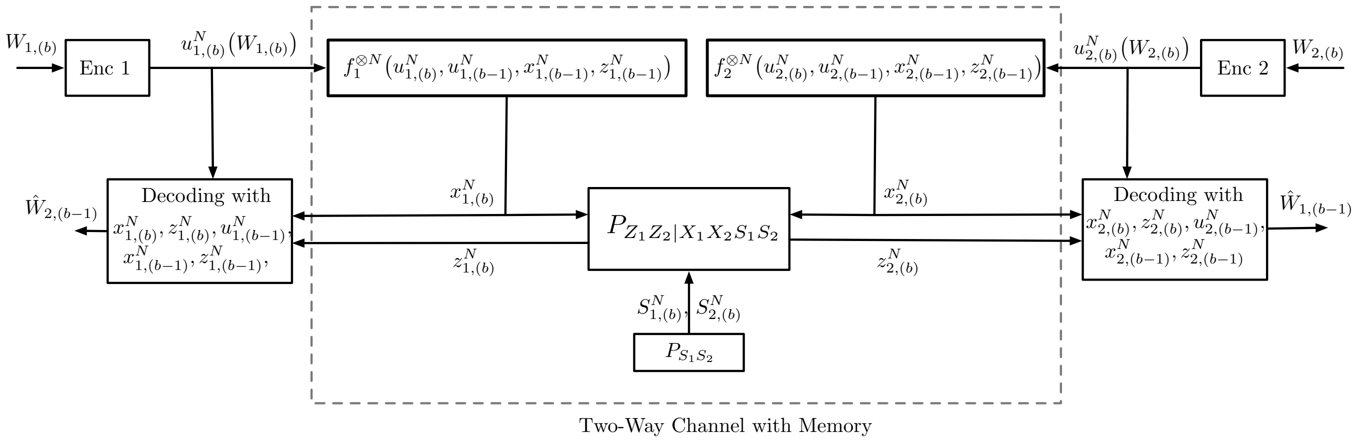


Figure 6.2: Han's coding scheme in a given block b. Encoders transform the discrete-memoryless two-way channel into a channel with memory so as to be able to correlate the inputs of the two terminals. Encoding is then performed through the independent codewords $u^N_{1,(b)}$ and $u^N_{2,(b)}$. Decoding of block-$(b-1)$ messages is performed based on the inputs/outputs in the two consecutive blocks $b-1$ and b.

The capacity of the two-way channel, and thus the optimal coding scheme is still open for general channels. The best known general scheme was proposed by Han [58]. The idea is to correlate the inputs of the two terminals in a stationary way so as to still allow for single-letter rate-expressions. An improved, more general scheme was proposed by Kramer [59]; it however leads to multi-letter expressions involving directed information terms and their evaluation thus seems unfeasible for most channels. We therefore base our scheme on Han's two-way coding scheme, which is depicted in Figure 6.2 and described in the following.

For convenience of notation, define

$$P_{Z_1Z_2|X_1X_2}(z_1,z_2|x_1,x_2) = \sum_{s_1\in\mathcal{S}_1, s_2\in\mathcal{S}_2} P_{S_1S_2}(s_1,s_2)P_{Z_1Z_2|X_1X_2S_1S_2}(z_1,z_2|x_1,x_2,s_1,s_2). \tag{6.6}$$

Han's scheme splits the blocklength n into $B+1$ blocks of length $N = n/(B+1)$ each. Accordingly, throughout, we let $X_{1,(b)}^N, X_{2,(b)}^N, S_{1,(b)}^N, S_{2,(b)}^N, Z_{1,(b)}^N, Z_{2,(b)}^N$ denote the block-b inputs, states and outputs, e.g., $S_{1,(b)}^N := (S_{1(b-1)N+1}, \ldots, S_{1,bN})$. We also represent the two messages W_1 and W_2 in a one-to-one way as the B-length tuples

$$W_k = (W_{k,(1)}, \ldots, W_{k,(B)}), \qquad k \in \{1,2\}, \tag{6.7}$$

where each $W_{k,(b)}$ is independent and uniformly distributed over $\left[2^{N\bar{R}_k}\right]$ for $\bar{R}_k \triangleq \frac{B+1}{B}R_k$.

Construct an independent code $\mathcal{C}_{k,(b)} = \left\{u_{k,(b)}^N(1), \ldots, u_{k,(b)}^N(2^{n\bar{R}_k})\right\}$ for each of the two terminals by picking entries i.i.d. according to some pmf P_{U_k}. As shown in Figure 6.2, Terminal k encodes Message $W_{k,(b)}$ by means of the codeword $u_{k,(b)}^N(W_{k,(b)})$ and sends the sequence

$$X_{k,(b)}^N = f_k^{\otimes N}\left(u_{k,(b)}^N(W_{k,(b)}),\ u_{k,(b-1)}^N(W_{k,(b-1)}),\ x_{k,(b-1)}^N,\ z_{k,(b-1)}^N\right) \tag{6.8}$$

over the channel during block b. Notice that by applying the function f_k to the block-b codeword symbols as well as to the symbols of the block-$(b-1)$ codeword $u_{k,(b-1)}^N(W_{k,(b-1)})$ and the block-$(b-1)$ channel inputs and outputs $x_{k,(b-1)}^N$ and $z_{k,(b-1)}^N$, the terminals introduce memory to the channel. An interesting point of view is to consider the transition of the codewords $u_{1,(b)}^N$ and $u_{2,(b)}^N$ to the channel outputs $z_{1,(b)}^N$ and $z_{2,(b)}^N$ as a virtual two-way channel with block-memory over which one can code and decode. Naturally, decoding of each message part $W_{k,(b)}$ is not based only on the signals in block (b) because other blocks depend on this message as well. In Han's scheme, decoding is over two consecutive blocks. Specifically, Terminal k decodes the block-b message $W_{\bar{k},(b)}$ using a joint-typicality decoder based on the block-b inputs, outputs, and own transmitted codewords $x_{k,(b)}^N, z_{k,(b)}^N$ and $u_{k,(b)}^N$, as well as on the block-$(b+1)$ inputs and outputs $x_{k,(b+1)}^N$ and $z_{k,(b+1)}^N$.

Notice that without any special care, the rate-region that is achievable with above scheme has to be described with a multi-letter expression because the joint pmf of the tuple $x_{1,(b+1)}^N, z_{1,(b+1)}^N, u_{1,(b)}^N, x_{1,(b)}^N, z_{1,(b)}^N$ that Terminal 1 uses to decode codeword $u^N_{2,(b)}(W_{k,(b)})$ varies with the block b. However, if one chooses a joint pmf $P_{U_1U_2X_1X_2Z_1Z_2}$ satisfying the stationarity condition

$$\begin{aligned}&P_{U_1U_2X_1X_2Z_1Z_2}(u_1,u_2,x_1,x_2,z_1,z_2)\\&\quad=\sum_{\tilde{u}_1,\tilde{u}_2,\tilde{x}_1,\tilde{x}_2,\tilde{z}_1,\tilde{z}_2} P_{Z_1Z_2|X_1X_2}(z_1,z_2|x_1,x_2)\mathbb{1}\{x_1=f_1(u_1,\tilde{u}_1,\tilde{x}_1,\tilde{z}_1)\}\\&\qquad\mathbb{1}\{x_2=f_2(u_2,\tilde{u}_2,\tilde{x}_2,\tilde{z}_2)\}\cdot P_{U_1}(u_1)P_{U_2}(u_2)P_{U_1U_2X_1X_2Z_1Z_2}(\tilde{u}_1,\tilde{u}_2,\tilde{x}_1,\tilde{x}_2,\tilde{z}_1,\tilde{z}_2),\end{aligned} \tag{6.9}$$

where P_{U_1} and P_{U_2} are the marginals of $P_{U_1U_2X_1X_2Z_1Z_2}$, then the pmf of the tuple of sequences $x^N_{1,(b+1)}, x_{2,(b+1)}, z^N_{1,(b+1)}, z^N_{2,(b+1)}, u^N_{1,(b)}, u^N_{2,(b)}, x^N_{1,(b)}, x^N_{2,(b)}, z^N_{1,(b)}, z^N_{2,(b)}$ is independent of the block index b. This allows to characterize the rate region achieved by the described coding scheme using a single-letter expression. All rate-pairs (R_1, R_2) are achievable that satisfy

$$R_1 \le I(U_1; X_2, Z_2, \tilde{U}_2, \tilde{X}_2, \tilde{Z}_2) \tag{6.10a}$$

$$R_2 \le I(U_2; X_1, Z_1, \tilde{U}_1, \tilde{X}_1, \tilde{Z}_1), \tag{6.10b}$$

where $(U_1, U_2, X_1, X_2, Z_1, Z_2, \tilde{U}_1, \tilde{U}_2, \tilde{X}_1, \tilde{X}_2, \tilde{Z}_1, \tilde{Z}_2)$ are distributed according to the pmf

$$\begin{aligned}&P_{U_1U_2X_1X_2Z_1Z_2\tilde{U}_1\tilde{U}_2\tilde{X}_1\tilde{X}_2\tilde{Z}_1\tilde{Z}_2}(u_1,u_2,x_1,x_2,z_1,z_2,\tilde{u}_1,\tilde{u}_2,\tilde{x}_1,\tilde{x}_2,\tilde{z}_1,\tilde{z}_2)\\&\quad=P_{Z_1Z_2|X_1X_2}(z_1,z_2|x_1,x_2)\mathbb{1}\{x_1=f_1(u_1,\tilde{u}_1,\tilde{x}_1,\tilde{z}_1)\}\mathbb{1}\{x_2=f_2(u_2,\tilde{u}_2,\tilde{x}_2,\tilde{z}_2)\}\\&\qquad\cdot P_{U_1}(u_1)P_{U_2}(u_2)P_{U_1U_2X_1X_2Z_1Z_2}(\tilde{u}_1,\tilde{u}_2,\tilde{x}_1,\tilde{x}_2,\tilde{z}_1,\tilde{z}_2).\end{aligned} \tag{6.11}$$

This recovers Han's theorem:

Theorem 18 (Han's Achievable Region for Two-Way Channels [58])**.** *Any nonnegative rate-pair* (R_1, R_2) *is achievable over the two-way channel if it satisfies Inequalities* (6.10) *for some choice of pmf* $P_{U_1U_2X_1X_2Z_1Z_2}$ *and functions* f_1 *and* f_2 *satisfying the stationarity condition* (6.9).

For certain cases above theorem can be simplified, and for certain channels the simplified region even coincides with capacity. The simplification is obtained by choosing the two functions f_1 and f_2 to simply produce the codewords $u^N_{1,(b-1)}$ and $u^N_{2,(b-1)}$ from the previous block[1] and ignore the other arguments. In this case, the set of rates that can be achieved coincides with the following inner bound that was first proposed by Shannon [60].

[1]The delay of a block introduced in this scheme is not crucial, it simply comes from the fact that Han's scheme decodes the block-$(b-1)$ codewords based on the block-b outputs. In this special case without adaptation, Han's scheme could be simplified by transmitting and decoding the codewords $u^N_{1,(b-1)}$ and $u^N_{2,(b-1)}$ directly in block $b-1$ without further delay.

Theorem 19 (Shannon's Inner Bound, [60]). *A pair of nonnegative pairs (R_1, R_2) is achievable if it satisfies*

$$R_1 \leq I(X_1; Z_2|X_2) \tag{6.12a}$$

$$R_2 \leq I(X_2; Z_1|X_2), \tag{6.12b}$$

for some input pmfs P_{X_1} and P_{X_2} and where $(X_1, X_2, Z_1, Z_2) \sim P_{X_1}P_{X_2}P_{Z_1Z_2|X_1X_2}$.

6.2.2 Collaborative Sensing and Communication based on Han's Two-Way Coding Scheme

We extend Han's coding scheme to include also collaborative sensing, that means each terminal compresses its block-b inputs and outputs so as to capture information about the other terminal's state and sends this state-information in the next-following block. In this first collaborative sensing and communication scheme that we present here, the sensing (compression) does not affect the communication (except possibly for the choice of the pmf $P_{U_1U_2X_1X_2Z_1Z_2}$). In fact, we again use Han's encodings and decodings as described in the previous subsection, except that the block-b

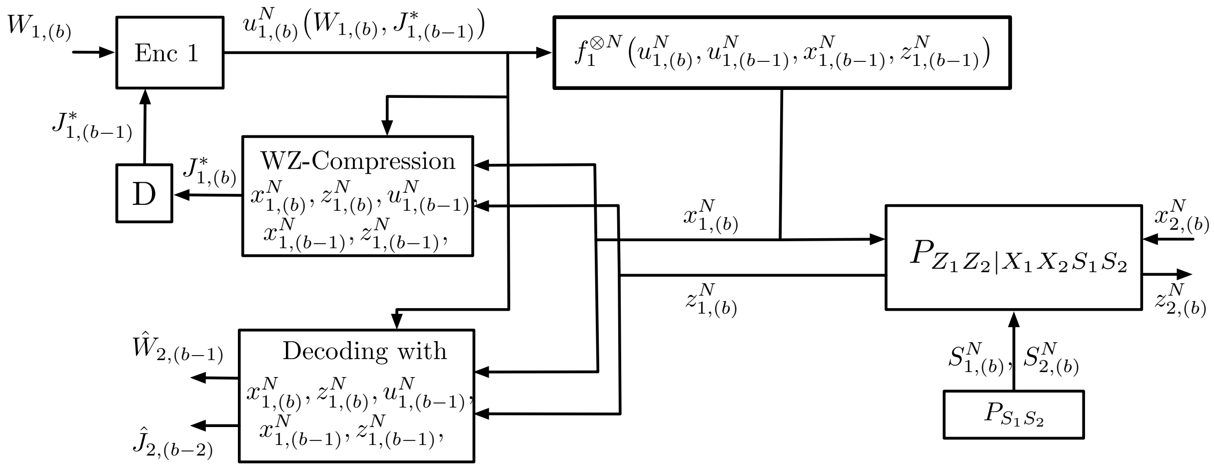


Figure 6.3: A first collaborative-sensing version of Han's coding scheme. The figure illustrates the encoding and decoding operations in a given block b at Terminal 1; Terminal 2 behaves analogously. To facilitate sensing at Terminal 2, Terminal 1 compresses its block-b channel inputs and outputs, together with its inputs, outputs, and codeword from the previous block $(b-1)$ (which are all resent in block b) using Wyner-Ziv compression [3] to account for the side-information at Terminal 2.

codeword not only encodes message $W_{k,(b)}$ but also a compression index $J^*_{k,(b-1)}$ that carries information about the block-$(b-1)$ state $S_{\bar{k},(b-1)}$. This compression index is then decoded at Terminal $\bar{k}$ after block $(b+1)$ simultaneously with message $W_{k,(b)}$. See Figure 6.3.

The analysis of the communication-part of our ISAC scheme is similar as in Han's scheme. Since the compression indices take parts of the place reserved for ordinary messages in Han's scheme, their rates $R_{\mathrm{WZ},1}$ and $R_{\mathrm{WZ},2}$ have to be subtracted from Han's communication rates. We thus have the following constraints for reliable communication and reliable decoding of the compression indices:

$$R_1 + R_{\text{WZ},1} \leq I(U_1; X_2, Z_2, \tilde{U}_2, \tilde{X}_2, \tilde{Z}_2) \tag{6.13a}$$

$$R_2 + R_{\text{WZ},2} \leq I(U_2; X_1, Z_1, \tilde{U}_1, \tilde{X}_1, \tilde{Z}_1). \tag{6.13b}$$

It remains to explain the compression and state estimation in more details. In our scheme, the index $J^*_{k,(b-1)}$ is obtained by means of a Wyner-Ziv compression [3] that lossily compresses the tuple $(x^N_{k,(b-1)}, z^N_{k,(b-1)}, u^N_{k,(b-2)}, x^N_{k,(b-2)}, z^N_{k,(b-2)})$ for a decoder that has side-information $(x^N_{\bar{k},(b-1)}, z^N_{\bar{k},(b-1)}, u^N_{\bar{k},(b-2)}, x^N_{\bar{k},(b-2)}, z^N_{\bar{k},(b-2)})$. In order for the decoder to be able to correctly reconstruct the compression codeword, the Wyner-Ziv codes need to be of rates at least [3]

$$R_{\text{WZ},k} > I(V_k; X_k, Z_k, \tilde{U}_k, \tilde{X}_k, \tilde{Z}_k | X_{\bar{k}}, Z_{\bar{k}}, \tilde{U}_{\bar{k}}, \tilde{X}_{\bar{k}}, \tilde{Z}_{\bar{k}}), \qquad k \in \{1, 2\}, \tag{6.14}$$

where the tuple $(U_1, U_2, X_1, X_2, Z_1, Z_2, \tilde{U}_1, \tilde{U}_2, \tilde{X}_1, \tilde{X}_2, \tilde{Z}_1, \tilde{Z}_2)$ refers to the auxiliary random variables chosen by Han's scheme of joint pmf as in (6.11) and V_1 and V_2 can be any random variables satisfying the Markov chains:

$$V_k - (X_k, Z_k, \tilde{U}_k, \tilde{X}_k, \tilde{Z}_k) \to (X_{\bar{k}}, Z_{\bar{k}}, \tilde{U}_{\bar{k}}, \tilde{X}_{\bar{k}}, \tilde{Z}_{\bar{k}}, S_k, S_{\bar{k}}). \tag{6.15}$$

In Wyner-Ziv coding, the encoder produces a codeword that is then reconstructed also at the Rx. We shall denote these codewords by $v^N_{k,(b-1)}(J^*_{k,(b-1)}, \ell_{k,(b-1)})$, for $k \in \{1, 2\}$, where $\ell_{k,(b-1)}$ denotes a binning-index that does not have to be conveyed to the Terminal $\bar{k}$ because this latter can recover it from its side-information. Thus, after block $(b+1)$ and after decoding index $J^*_{k,(b-1)}$, with high probability Terminal $\bar{k}$ can reconstruct the codeword $v^N_{k,(b-1)}(J^*_{k,(b-1)}, \ell_{k,(b-1)})$ chosen at Terminal k.

Terminal k can wait arbitrarily long to produce an estimate of the state-sequence S^N_k. We propose that it waits after the block-$(b+1)$ decoding to reconstruct the block-b state $S^N_{k,(b)}$ by applying an optimal symbol-by-symbol estimator to the related sequences of inputs, outputs, and channel codewords of blocks $b-1$ and b, as well as on the compression codeword $v^N_{\bar{k},(b)}$:

$$\hat{S}^N_{k,(b)} = \tilde{\phi}^{*\otimes N}_{2,k}\left(v^N_{\bar{k},(b)}, x^N_{k,(b)}, z^N_{k,(b)}, \hat{u}_{\bar{k},(b)}, u^N_{k,(b-1)}, x^N_{k,(b-1)}, z^N_{k,(b-1)}, \hat{u}_{\bar{k},(b-1)}\right), \tag{6.16}$$

where

$$\tilde{\phi}^*_{2,k}(v_{\bar{k}}, x_k, z_k, u_{\bar{k}}, \tilde{u}_k, \tilde{x}_k, \tilde{z}_k, \tilde{u}_{\bar{k}}) := \arg\min_{s'_k \in \hat{\mathcal{S}}_k} \sum_{s_k \in \mathcal{S}_k} P_{S_k|X_k Z_k U_{\bar{k}}}(s_k|x_k, z_k, u_{\bar{k}})\, d_k(s_k, s'_k). \tag{6.17}$$

By (6.13) and (6.14) and standard typicality arguments, one obtains the following theorem.

Theorem 20 (Inner Bound via Separate Source-Channel Coding)**.** *Any nonnegative rate-distortion quadruple* (R_1, R_2, D_1, D_2) *is achievable if it satisfies the following two rate-constraints*

$$R_1 \leq I(U_1; X_2, Z_2, \tilde{U}_2, \tilde{X}_2, \tilde{Z}_2) - I(V_1; X_1, Z_1, \tilde{U}_1, \tilde{X}_1, \tilde{Z}_1 | X_2, Z_2, \tilde{U}_2, \tilde{X}_2, \tilde{Z}_2) \tag{6.18a}$$

$$R_2 \leq I(U_2; X_1, Z_1, \tilde{U}_1, \tilde{X}_1, \tilde{Z}_1) - I(V_2; X_2, Z_2, \tilde{U}_2, \tilde{X}_2, \tilde{Z}_2 | X_1, Z_1, \tilde{U}_1, \tilde{X}_1, \tilde{Z}_1), \tag{6.18b}$$

and the two distortion constraints

$$\mathsf{E}\Big[d_1\big(S_1,\ \tilde{\phi}^*_{2,1}(V_2, X_1, Z_1, U_2, \tilde{U}_1, \tilde{X}_1, \tilde{Z}_1, \tilde{U}_2)\big)\Big] \leq D_1 \tag{6.18c}$$

$$\mathsf{E}\Big[d_2\big(S_2,\ \tilde{\phi}^*_{2,2}(V_1, X_2, Z_2, U_1, \tilde{U}_2, \tilde{X}_2, \tilde{Z}_2, \tilde{U}_2)\big)\Big] \leq D_2 \tag{6.18d}$$

for some choice of pmf $P_{U_1 U_2 X_1 X_2 Z_1 Z_2}$ *and functions* f_1 *and* f_2 *satisfying the stationarity condition* (6.9) *and* V_1, V_2 *satisfying the Markov chains* (6.15)*.*

Similarly to Shannon's inner bound, we can obtain the following corollary by setting $X_k = \tilde{U}_k$.

Corollary 9 (Inner Bound via Non-Adaptive Coding)**.** *Any nonnegative rate-distortion quadruple* (R_1, R_2, D_1, D_2) *is achievable if it satisfies the following two rate-constraints*

$$R_1 \leq I(X_1; X_2, Z_2) - I(V_1; X_1, Z_1 | X_2, Z_2) \tag{6.19a}$$

$$R_2 \leq I(X_2; X_1, Z_1) - I(V_2; X_2, Z_2 | X_1, Z_1), \tag{6.19b}$$

and the two distortion constraints

$$\mathsf{E}\Big[d_1\big(S_1,\ \tilde{\phi}^*_{2,1}(V_2, X_1, X_2, Z_1)\big)\Big] \leq D_1 \tag{6.19c}$$

$$\mathsf{E}\Big[d_2\big(S_2,\ \tilde{\phi}^*_{2,2}(V_1, X_1, X_2, Z_2)\big)\Big] \leq D_2 \tag{6.19d}$$

for some choice of pmfs P_{X_1}*,* P_{X_2}*,* $P_{V_1|X_1,Z_1}$*, and* $P_{V_2|X_2,Z_2}$*.*

As the following example shows, above corollary achieves the fundamental rate-distortion tradeoff for some channels.

Example 3. *Consider the following state-dependent two-way channel*

$$Z_1 = X_1 \oplus X_2 \oplus S_2 \qquad \text{and} \qquad Z_2 = X_1 \oplus X_2 \oplus S_1, \tag{6.20a}$$

where inputs, outputs, and states are binary and S_1 and S_2 are independent Bernoulli-p_1 and p_2 random variables, for $p_1, p_2 \in [0, 1/2]$. Notice that Terminal 1's outputs depend on the state desired at Terminal 2 and Terminal 2's outputs on the state desired at Terminal 1, which calls for collaborative sensing.

Whenever $D_{\bar{k}} < p_{\bar{k}}$, we choose

$$V_k = Z_k \oplus X_k \oplus B_k = X_{\bar{k}} \oplus S_{\bar{k}} \oplus B_k \tag{6.21}$$

where B_k is an independent Bernoulli-D_k random variable. If $D_k \geq p_k$, choose V_k a constant. Inputs X_1 and X_2 are chosen independent Bernoulli-$1/2$, i.e., capacity-achieving on channels with Bernoulli-noses. When $D_{\bar{k}} < p_{\bar{k}}$, the optimal symbol-by-symbol state-estimator is

$$\tilde{\phi}^*_{2,\bar{k}}(v_k, x_1, x_2, z_{\bar{k}}) = v_k \oplus x_{\bar{k}} \tag{6.22}$$

and otherwise it is the constant estimator $\tilde{\phi}^*_{2,\bar{k}}(v_k, x_1, x_2, z_{\bar{k}}) = 0$.

For the described choice of random variables, Corollary 9 evaluates to the set of rate-distortion tuples (R_1, R_2, D_1, D_2) satisfying

$$R_k \leq 1 - H_b(p_k) - \max\{0,\ H_b(p_{\bar{k}}) - H_b(D_{\bar{k}})\}, \qquad k \in \{1, 2\}, \tag{6.23}$$

and achieves the fundamental rate-distortion region. The region in (6.23) *is concave (because the rate-distortion function $\max\{0,\ H_b(p_{\bar{k}}) - H_b(D_{\bar{k}})\}$ is convex), and thus improves over classic time- and resource-sharing schemes. It also improves over a similar ISAC scheme without collaborative sensing where the compression codewords V_1 and V_2 are set to constants. In this latter case, only rate-distortion tuples are possible that satisfy $D_k \geq p_k$, for $k \in \{1, 2\}$.*

6.2.3 Collaborative Sensing and JSCC Scheme

In this scheme, we fully integrate the compression into the communication scheme, in a similar way that hybrid coding [25] uses a single codeword for compression and channel coding in source-channel coding applications.

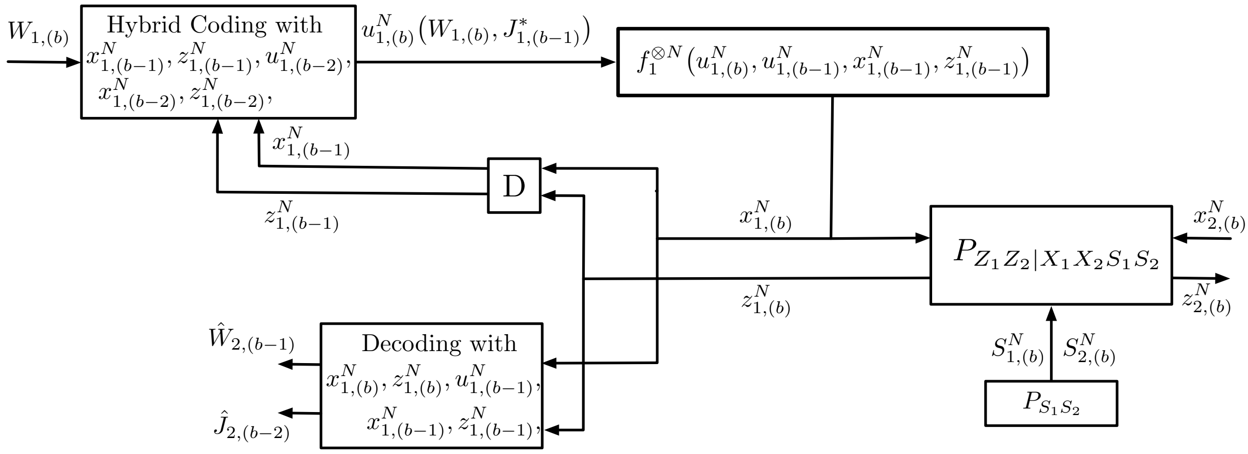


Figure 6.4: A ISAC scheme integrating collaborative sensing for D2D into Han's two-way coding scheme by means of hybrid coding. A single codeword is used both for compression and for channel coding.

Encoding and decoding in block b of the new scheme are depicted in Figure 6.4. The main difference compared to the scheme in the previous subsection is that here the block-b codeword $u^N_{1,(b)}$ is *correlated* with the inputs and outputs in the previous block $(b-1)$.[2] This correlation introduces additional dependence between blocks, which was previously missing because of the independence of the compression codewords and the codewords used for channel coding in the next block. To still obtain a stationary distribution on the codewords and channel inputs/outputs, which then allows for a single-letter characterization of the performance of the scheme, one has to choose a joint pmf $P_{U_1'U_2'Z_1Z_2X_1X_2U_1U_2}$, conditional pmfs $P_{U_1'|X_1Z_1\tilde{U}_1\tilde{X}_1\tilde{Z}_1}$ and $P_{U_2'|X_2Z_2\tilde{U}_2\tilde{X}_2\tilde{Z}_2}$ as well as functions f_1 and f_2 on appropriate domains satisfying the new stationarity condition

$$\begin{aligned}
&P_{U_1'U_2'Z_1Z_2X_1X_2U_1U_2}(u_1', u_2', z_1, z_2, x_1, x_2) \\
&\quad = \sum_{\tilde{u}_1,\tilde{u}_2,\tilde{x}_1,\tilde{x}_2,\tilde{z}_1,\tilde{z}_2} P_{U_1'|X_1Z_1\tilde{U}_1\tilde{X}_1\tilde{Z}_1}(u_1'|u_1, x_1, z_1, \tilde{u}_1, \tilde{x}_1, \tilde{z}_1) P_{U_2'|X_2Z_2\tilde{U}_2\tilde{X}_2\tilde{Z}_2}(u_2'|x_2, z_2, u_2, \tilde{u}_2, \tilde{x}_2, \tilde{z}_2) \\
&\qquad\qquad \cdot P_{Z_1Z_2|X_1X_2}(z_1, z_2|x_1, x_2)\mathbb{1}\{x_1 = f_1(u_1, \tilde{u}_1, \tilde{x}_1, \tilde{z}_1)\}\mathbb{1}\{x_2 = f_2(u_2, \tilde{u}_2, \tilde{x}_2, \tilde{z}_2)\} \\
&\qquad\qquad\qquad \cdot P_{U_1'U_2'Z_1Z_2X_1X_2U_1U_2}(u_1, u_2, \tilde{z}_1, \tilde{z}_2, \tilde{x}_1, \tilde{x}_2, \tilde{u}_1, \tilde{u}_2),
\end{aligned} \tag{6.24}$$

In the following, all mentioned conditional and marginal pmfs are with respect to the joint pmf

[2] In the previous scheme, the compression codeword $v^N_{1,(b)}$ was correlated with the block-$(b-1)$ signals but not the channel coding codeword $u_{1,(b)}$. Now the codeword $u^N_{1,(b)}$ acts both as a compression codeword and as a channel coding codeword.

$P_{U_1'U_2'Z_1Z_2X_1X_2U_1U_2\tilde{U}_1\tilde{U}_2\tilde{W}_1\tilde{W}_2\tilde{V}_1\tilde{V}_2}$ indicated by the summand in (6.24).

We next explain the code construction, encodings and decodings. For each $k \in \{1,2\}$, for each block $b \in \{1,\ldots,B+1\}$, and each message $m_k \in [2^{N\bar{R}_k}]$, choose a subcodebook $\{u_{k,(b)}^N(m_k,j) \colon j \in [2^{NR_k'}]\}$ by picking all entries i.i.d. $P_{U_k'}$. Terminal k then picks the codeword $u_{k,(b)}^N(W_{k,(b)},j)$ so that the following joint-typicality check is satisfied for some fixed $\epsilon > 0$:

$$\left(u_{k,(b)}^N(W_{k,(b)},j), x_{k,(b-1)}^N,\, z_{k,(b-1)}^N,\, u_{k,(b-2)}^n,\, x_{k,(b-2)}^N,\, z_{k,(b-2)}^N\right) \in \mathcal{T}_\epsilon^{(N)}\Big(P_{U_k'X_kZ_k\tilde{U}_k\tilde{X}_k\tilde{Z}_k}\Big), \tag{6.25}$$

and sets $J_{k,(b-1)}^* = j$. By standard arguments, such an index j exists with probability tending to 1 as $N \to \infty$ if

$$R_k' \geq I(U_k'; X_k, Z_k, \tilde{U}_k, \tilde{X}_k, \tilde{Z}_k), \quad k \in \{1,2\}. \tag{6.26}$$

Terminal k then sends the block-b input sequence

$$X_{k,(b)}^N = f_k^{\otimes N}\left(u_{k,(b)}^N\big(W_{k,(b)}J_{k,(b-1)}^*\big), u_{k,(b-1)}^N, x_{k,(b-1)}^n, z_{k,(b-1)}^N\right). \tag{6.27}$$

Decoding is again performed using a joint-typicality decoder. At the end of block b, Terminal k looks for indices $\hat{w}_{\bar{k}}$ and $\hat{j}_{\bar{k}}$ satisfying the two typicality checks

$$\left(u_{\bar{k},(b-1)}^N(\hat{w}_{\bar{k}},\hat{j}_{\bar{k}}), x_{k,(b)}^N,\, z_{k,(b)}^N,\, u_{k,(b-1)}^N\, x_{k,(b-1)}^N,\, z_{k,(b-1)}^N\right) \in \mathcal{T}_\epsilon^{(N)}\Big(P_{\tilde{U}_{\bar{k}}X_kZ_k\tilde{U}_k\tilde{X}_k\tilde{Z}_k}\Big) \tag{6.28}$$

and

$$\left(u_{\bar{k},(b-1)}^N(\hat{w}_{\bar{k}},\hat{j}_{\bar{k}}), x_{k,(b-2)}^N,\, z_{k,(b-2)}^N,\, u_{k,(b-3)}^n,\, x_{k,(b-3)}^N,\, z_{k,(b-3)}^N\right) \in \mathcal{T}_\epsilon^{(N)}\left(P_{U_{\bar{k}}X_kZ_k\tilde{U}_k\tilde{X}_k\tilde{Z}_k}\right). \tag{6.29}$$

If a unique pair of such element exists, set $\hat{W}_{\bar{k},(b-1)} = w_{\bar{k}}$ and $\hat{u}_{\bar{k},(b-1)}^N \triangleq u_{\bar{k},(b-1)}^N(\hat{w}_{\bar{k}},\hat{j}_{\bar{k}})$. Decoding is successful with probability tending to 0 as $N \to \infty$ if

$$\bar{R}_{\bar{k}} + R_{\bar{k}}' \leq I(\tilde{U}_{\bar{k}}; X_k, Z_k, \tilde{U}_k, \tilde{X}_k, \tilde{Z}_k) + I(U_{\bar{k}}; X_k, Z_k, \tilde{U}_k, \tilde{X}_k, \tilde{Z}_k), \quad k \in \{1,2\}. \tag{6.30}$$

State-estimation is similar to (6.16), but where Terminal k replaces the compression codeword $v_{k,(b)}^N$ by the joint source-channel codeword $u_{k,(b+1)}^N$ and similarly to hybrid coding also uses the inputs/outputs corresponding to the block where the codeword $u_{1,(b+1)}^N$ is sent, i.e., inputs and outputs in block $b+1$. Thus, Terminal k computes its estimate of the block-b state as:

$$\hat{s}^N_{k,(b)} = \phi^{*\otimes N}_{2,k}\big(\hat{u}^N_{\bar{k},(b+1)}, x^N_{k,(b+1)}, z^N_{k,(b+1)}, \hat{u}^N_{\bar{k},(b)}, x^N_{k,(b)}, z^N_{k,(b)}, u^N_{k,(b-1)}, x^N_{k,(b-1)}, z^N_{k,(b-1)}, \hat{u}^N_{\bar{k},(b-1)}\big), \tag{6.31}$$

where

$$\phi^*_{2,k}(u'_{\bar{k}}, x'_k, z'_k, u_{\bar{k}}, x_k, z_k, \tilde{u}_k, \tilde{x}_k, \tilde{z}_k, \tilde{u}_{\bar{k}}) := \arg\min_{s'_k \in \hat{\mathcal{S}}_k} \sum_{s_k \in \mathcal{S}_k} P_{S_k|X_k Z_k U_{\bar{k}}}(s_k|x_k, z_k, u_{\bar{k}})\ d_k(s_k, s'_k). \tag{6.32}$$

By standard arguments and because of the stationarity condition in (6.24) the probability of violating the distortion constraints tends to 0 as $N \to \infty$ if

$$\mathsf{E}\Big[d_k\big(S_k,\ \phi^*_{2,k}(U'_{\bar{k}}, X'_k, Z'_k, U_{\bar{k}}, X_k, Z_k, \tilde{U}_k, \tilde{X}_k, \tilde{Z}_k, \tilde{U}_{\bar{k}})\big)\Big] \leq D_k, \qquad k \in \{1,2\}, \tag{6.33a}$$

where $X'_1 = f_1(U'_1, U_1, X_1, Z_1)$ and $X'_2 = f_2(U'_2, U_2, X_2, Z_2)$ and the outputs Z'_1 and Z'_2 are obtained from X'_1 and X'_2 via the channel transition law $P_{Z_1 Z_2|X_1 X_2}$.

From above considerations and by eliminating the dummy rates R'_1 and R'_2, we obtain the following theorem.

Theorem 21 (Inner Bound via Joint Source-Channel Coding)**.** *Any nonnegative rate-distortion quadruple* (R_1, R_2, D_1, D_2) *is achievable if it satisfies the following two rate-constraints*

$$R_k \leq I(\tilde{U}_k; X_{\bar{k}}, Z_{\bar{k}}, \tilde{U}_{\bar{k}}, \tilde{X}_{\bar{k}}, \tilde{Z}_{\bar{k}}) - I(U_k; X_k, Z_k, \tilde{U}_k, \tilde{X}_k, \tilde{Z}_k | X_{\bar{k}}, Z_k, \tilde{U}_{\bar{k}}, \tilde{X}_{\bar{k}}, \tilde{Z}_{\bar{k}}), \quad k \in \{1,2\} \tag{6.34}$$

and the two distortion constraints in (6.33) *for some choice of pmf* $P_{U'_1 U'_2 Z_1 Z_2 X_1 X_2 U_1 U_2}$ *and functions* f_1 *and* f_2 *satisfying the stationarity condition* (6.9).

Remark 10. *We notice that the described compression technique does not use binning as in Wyner-Ziv coding [3]. Instead, decoder side-information is taken into account via the joint typicality check in* (6.29).

Remark 11. *For the choice* $U'_k = (U''_k, V_k)$ *with* $U''_k \sim P_{U_k}$ *independent of all other random variables and* V_1 *and* V_2 *satisfying the Markov chains in* (6.15)*, the inner bound in Theorem 21 achieved by our joint source-channel coding scheme specializes to the inner bound Theorem 20 achieved by separate source-channel coding. For above choice of auxiliary random variables, the reconstruction functions* g_1 *and* g_2 *can restrict their first arguments only to the* V_1*- and* V_2*-components without loss in performance.*

6.3 Conclusion

For D2D communication, we also proposed a joint source-channel coding (JSCC) scheme to integrate compression and coding into a single codeword as in hybrid coding. We extended Han's coding scheme to include also collaborative sensing. Each terminal compressed its block-b inputs and outputs to capture information about the other terminal's state and sensed this state-information in the next-following block. In this first collaborative sensing ISAC scheme that we presented, the sensing did not affect the communication. In the second scheme, we fully integrated the compression into the communication scheme, in a similar way that hybrid coding.

Chapter 7

Summary and Possible Research Directions

7.1 Summary

This thesis focused on the fundamental limits of information-theoretic ISAC over multi-Tx or multi-Rx networks. More precisely:

- In Chapter 2, we glanced at the sensing problem and communication separately. Then we reviewed the naive practical solutions to integrate the two tasks.

- In Chapter 3, we reviewed the first model of ISAC over state-dependent P2P channels. This work characterized capacity-distortion-cost tradeoff of SDMC by random coding construction and introduced the optimal estimator (a deterministic symbol-by-symbol estimator).

- Furthermore, in Chapter 3, we had a summary of related works. A line of the work similarly modeled ISAC where the Tx estimated the state in which the state was assumed slow-varying, which changed the estimation problem introduced in Section 3.1 to a hypothesis testing. Another line of works modeled ISAC where the Rx estimates the state in which the state was i.i.d. Thus, the tools and results had been dual to the model introduced in Section 3.1.

- Chapter 4 considered single-Tx two-Rx Bcs. we proposed the inner and outer bounds for the capacity-distortion tradeoff region of SDMBC. Specially, we characterized the capacity-distortion tradeoff region of physically degraded SDMBC. We evaluated the proposed schemes through various examples and verified that ISAC schemes improved the region over naive solutions built on resource-sharing. We also found the conditions that showed, in some cases, no tradeoff between communication and sensing arises.

- In Chapter 5, we introduced information-theoretic collaborative sensing schemes in which Txs could cooperate in communication and sensing where terminals convey data to not only the other terminals but also state-information that they extract from their previous observations. This state information can be exploited at the other terminals to improve their sensing performances. Furthermore, as we showed through examples, our schemes improve over the previous non-collaborative scheme regarding their achievable rate-distortion tradeoffs.
- In Chapter 6, we proposed two information-theoretic collaborative sensing ISAC schemes, one where compression of state information was separated from channel coding and one where it was integrated via a hybrid coding approach.

7.2 Possible Research Directions

Various interesting future research directions arise as follows:

- Our results are built on the assumption of i.i.d. states. In practice, it seems that the states at different time slots can be correlated. Similar to [26, 61], we can extend our results of ISAC over BCs and MAC with a slow-varying state. Also, it is worth looking at the correlated state during the communication time slots rather than the fixed or i.i.d state.
- As far as our knowledge, there has not been any study on ISAC when the Rx estimates the slow-varying or correlated states. So, there is an exciting line of research to look at the model introduced in [22, 30, 33, 42] with slow-varying/correlated states instead of i.i.d state.
- Another outgrowth of our work is the finite-blocklength derivation of achievability and converse bounds. In practice, assessing the backoff from capacity required to sustain the desired error probability at a given fixed finite block length is vital. Readers are encouraged to refer to [62]
- The JSCC scheme proposed for ISAC D2D communication could be integrated into our ISAC MAC scheme.
- Another exciting research direction for the MAC scheme is to include state estimation at the Rx. In this respect, it would be interesting to include an additional superposition compression layer to generate compression information that is only decoded by the Rx but not the other Tx.
- For D2D communication, an interesting extension would be to consider specific channel models and replace Han's result with two-way communication schemes tailored to these particular channels.

- There is still lots to know in the secure ISAC whether the estimation happens at Rx, Tx, or both.

Appendix A

Proofs

A.1 Proofs of Chapter 3

A.1.1 Proof of Lemma 2

Recall that $\hat{S}^n = h(X^n, Z^n)$, and write for each $i = 1, \cdots, n$:

$$\begin{aligned}
\mathbb{E}\left[d(S_i, \hat{S}_i)\right] &= \mathbb{E}_{X^n,Z^n}\left[\mathbb{E}[d(S_i, \hat{S}_i)|X^n, Z^n]\right] \\
&\overset{\text{(a)}}{=} \sum_{x^n,z^n} P_{X^nZ^n}(x^n, z^n) \sum_{\hat{s}\in\mathcal{S}} P_{\hat{S}_i|X^nZ^n}(\hat{s}|x^n, z^n) \sum_{s} P_{S_i|X_iZ_i}(s|x_i, z_i) d(s, \hat{s}) \\
&\geq \sum_{x^n,z^n} P_{X^nZ^n}(x^n, z^n) \min_{\hat{s}\in\mathcal{S}} \sum_{s} P_{S_i|X_iZ_i}(s|x_i, z_i) d(s, \hat{s}) \\
&= \mathbb{E}[d(S_i, \hat{s}^*(X_i, Z_i))], \qquad \text{(A.1)}
\end{aligned}$$

where (a) holds by the Markov chain

$$\left(X^{i-1}, X_{i+1}^n, Z^{i-1}, Z_{i+1}^n, \hat{S}_i\right) - \circ - (X_i, Z_i) - \circ - S_i. \qquad \text{(A.2)}$$

Summing over all $i = 1, \ldots, n$, we thus obtain:

$$\begin{aligned}
\Delta^{(n)} &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[d(S_i, \hat{S}_i)\right] \qquad \text{(A.3)} \\
&\geq \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[d(S_i, \hat{s}^*(X_i, Z_i))], \qquad \text{(A.4)}
\end{aligned}$$

which yields the desired conclusion.

A.1.2 Proof of Theorem 1

Converse

Fix a sequence (in n) of $(2^{n\mathsf{R}}, n)$ codes such that Limits (3.5) hold. By Fano's inequality there exists a sequence $\epsilon_n \to 0$ as $n \to \infty$ so that:

$$\begin{aligned} n\mathsf{R} &\leq I(W; Y^n, S^n) + n\epsilon_n \\ &= I(W; Y^n \mid S^n) + n\epsilon_n \\ &= \sum_{i=1}^{n} H(Y_i \mid Y^{i-1}, S^n) - H(Y_i \mid W, Y^{i-1}, S^n) + n\epsilon_n \\ &\overset{(a)}{\leq} \sum_{i=1}^{n} H(Y_i \mid S_i) - H(Y_i \mid X_i, Y^{i-1}, W, S^n) + n\epsilon_n \\ &\overset{(b)}{=} \sum_{i=1}^{n} H(Y_i \mid S_i) - H(Y_i \mid X_i, S_i) + n\epsilon_n \\ &= \sum_{i=1}^{n} I(X_i; Y_i \mid S_i) + n\epsilon_n \end{aligned} \tag{A.5}$$

where (a) holds because conditioning can only reduce entropy; and (b) holds because $(W, Y^{i-1}, S^{i-1}, S_{i+1}^n) - (S_i, X_i) - Y_i$ form a Markov chain. We continue as:

$$\begin{aligned} \mathsf{R} &\leq \frac{1}{n}\sum_{i=1}^{n} I(X_i; Y_i \mid S_i) + \epsilon_n \\ &\overset{(c)}{\leq} \frac{1}{n}\sum_{i=1}^{n} \mathcal{C}_{\text{inf}}\left(\sum_x P_{X_i}(x)c(x),\ \sum_x P_{X_i}(x)b(x)\right) + \epsilon_n \\ &\overset{(d)}{\leq} \mathcal{C}_{\text{inf}}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_x P_{X_i}(x)c(x),\ \frac{1}{n}\sum_{i=1}^{n}\sum_x P_{X_i}(x)b(x)\right) + \epsilon_n \\ &\overset{(e)}{\leq} \mathcal{C}_{\text{inf}}(\mathsf{D}, \mathsf{B}) \end{aligned} \tag{A.6}$$

where (c) holds by the definition of $\mathcal{C}_{\text{inf}}(\mathsf{D}, \mathsf{B})$, and (d) and (e) hold by Lemma 1.

Achievability

Fix $P_X(\cdot)$ and functions $\hat{h}(x, z)$ that achieve $C(\mathsf{D}/(1+\epsilon), \mathsf{B})$, where D is the desired distortion and B is the target cost, for a small positive number $\epsilon > 0$. We define the joint pmf $P_{SXY} := P_S P_X P_{Y|SX}$.

Codebook generation Generate $2^{n\mathsf{R}}$ sequences $\{x^n(w)\}_{w=1}^{2^{n\mathsf{R}}}$ by randomly and independently drawing each entry according to P_X. This defines the codebook $\mathcal{C} = \{x^n(w)\}_{w=1}^{2^{n\mathsf{R}}}$, which is revealed to the encoder and the decoder.

Encoding To send a message $w \in \mathcal{W}$, the encoder transmits $x^n(w)$.

Decoding Upon observing outputs $Y^n = y^n$ and state sequence $S^n = s^n$, the decoder looks for an index $\hat{w}$ such that

$$(s^n, x^n(\hat{w}), y^n) \in \mathcal{T}_\epsilon^{(n)}(P_{SXY}). \tag{A.7}$$

If exactly one such index exists, it declares $\hat{W} = \hat{w}$. Otherwise, it declares an error.

Estimation Assuming that it sent the input sequence $X^n = x^n$ and observed the feedback signal $Z^n = z^n$, the encoder computes the reconstruction sequence as:

$$\hat{S}^n = (\hat{s}^*(x_1, z_1), \hat{s}^*(x_2, z_2), \ldots, \hat{s}^*(x_n, z_n)). \tag{A.8}$$

Analysis We start by analyzing the probability of error and the distortion averaged over the random code construction. Given the symmetry of the code construction, we can condition on the event $W = 1$.

We then notice that the decoder makes an error, i.e., declares nothing or $\hat{W} \neq 1$ if, and only if, one or both of the following events occur:

$$\mathcal{E}_1 = \left\{(S^n, X^n(1), Y^n) \notin \mathcal{T}_\epsilon^{(n)}(P_{SXY})\right\} \tag{A.9}$$

$$\mathcal{E}_2 = \left\{(S^n, X^n(w'), Y^n) \in \mathcal{T}_\epsilon^{(n)}(P_{SXY}) \text{ for some } w' \neq 1\right\}. \tag{A.10}$$

where we defined $P_{SXY} := P_S P_X P_{Y|SX}$. Thus, by the union bound:

$$P_e^{(n)} = P(\mathcal{E}_1 \cup \mathcal{E}_2) \leq P(\mathcal{E}_1) + P(\mathcal{E}_2). \tag{A.11}$$

The first term goes to zero as $n \to \infty$ by the weak law of large numbers. The second term also tends to zero as $n \to \infty$ if $\mathsf{R} < I(X;Y|S)$ by the independence of the codewords and the packing lemma [20, Lemma 3.1]. Therefore, $P_e^{(n)}$ tends to zero as $n \to \infty$ whenever $R < I(X;Y|S)$.

The expected distortion (averaged over the random codebook, state and channel noise) can be upper bounded as

$$\Delta^{(n)} = \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}\Big[d(S_i,\hat{S}_i)\Big] \tag{A.12}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}\Big[d(S_i,\hat{S}_i)\big|\hat{W}\neq 1\Big]\,\Pr(\hat{W}\neq 1)$$

$$+ \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}\Big[d(S_i,\hat{S}_i)\big|\hat{W}= 1\Big]\,\Pr(\hat{W}= 1) \tag{A.13}$$

$$\leq \mathsf{D}_{\max}P_e + \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}\Big[d(S_i,\hat{S}_i)\big|\hat{W}= 1\Big]\cdot(1-P_e). \tag{A.14}$$

In the event of correct decoding, i.e., $\hat{W}=1$,

$$(S^n, X^n(1), Y^n) \in \mathcal{T}_\epsilon^{(n)}(P_S P_X P_{Y|SX}), \tag{A.15}$$

and since $\hat{S}_i = \hat{s}^*(X_i, Z_i)$, also

$$(S^n, X^n(1), \hat{S}^n) \in \mathcal{T}_\epsilon^{(n)}\left(P_{SX\hat{S}}\right), \tag{A.16}$$

where $P_{SX\hat{S}}$ denotes the joint marginal pmf of $P_{SXZ\hat{S}}(s,x,z,\hat{s}) := P_S(s)P_X(x)P_{Z|SX}(z|s,x)\mathbb{1}\{\hat{s} = \hat{s}^*(x,z)\}$. Then,

$$\varlimsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}\Big[d(S_i,\hat{S}_i)|\hat{W}=1\Big] \leq (1+\epsilon)\mathsf{E}\Big[d(S,\hat{S})\Big], \tag{A.17}$$

for $(S,\hat{S})$ following the marginal of the pmf $P_{SXZ\hat{S}}$ defined above. Assuming that $\mathsf{R} < I(X;Y|S)$, and thus $P_e \to 0$ as $n\to\infty$, we obtain from (A.14) and (A.17):

$$\varlimsup_{n\to\infty} \Delta^{(n)} = (1+\epsilon)\mathsf{E}\Big[d(S,\hat{S})\Big]. \tag{A.18}$$

Taking finally $\epsilon \downarrow 0$, we can conclude that the error probability and distortion constraint (3.5a), (3.5b) hold (averaged over the random code constructions, the random states, and the noise in the channel) whenever

$$\mathsf{R} < I(X;Y\mid S), \tag{A.19}$$

$$\mathsf{E}\Big[d(S,\hat{S})\Big] < \mathsf{D}. \tag{A.20}$$

Notice that the cost constraint (3.5c) is fulfilled by construction. By standard arguments it can then be shown that there must exist at least one sequence of deterministic code books $\mathcal{C}_n$ so that constraints (3.5) hold.

A.1.3 Proof of Remark 2

Converse

Fix a sequence (in n) of $(2^{n\mathsf{R}}, n)$ codes such that Limits (3.5) hold. By Fano's inequality there exists a sequence $\epsilon_n \to 0$ as $n \to \infty$ so that:

$$\begin{aligned}
n\mathsf{R} &\leq I(W; Y^n, S_R^n) + n\epsilon_n \\
&= I(W; Y^n \mid S_R^n) + n\epsilon_n \\
&= \sum_{i=1}^{n} H(Y_i \mid Y^{i-1}, S_R^n) - H(Y_i \mid W, Y^{i-1}, S_R^n) + n\epsilon_n \\
&\overset{(a)}{\leq} \sum_{i=1}^{n} H(Y_i \mid S_{R,i}) - H(Y_i \mid X_i, Y^{i-1}, W, S_R^n) + n\epsilon_n \\
&\overset{(b)}{=} \sum_{i=1}^{n} H(Y_i \mid S_{R,i}) - H(Y_i \mid X_i, S_{R,i}) + n\epsilon_n \\
&= \sum_{i=1}^{n} I(X_i; Y_i \mid S_{R,i}) + n\epsilon_n
\end{aligned} \tag{A.21}$$

where (a) holds because conditioning can only reduce entropy; and (b) holds because $(W, Y^{i-1}, S_R^{i-1}, S_{R,i+1}^n) - (S_{R,i}, X_i) - Y_i$ form a Markov chain.

Define

$$\mathcal{C}_{\text{inf}}^{\text{imp}}(\mathsf{D}, \mathsf{B}) := \max_{P_X \in \mathcal{P}_{\mathsf{D}} \cap \mathcal{P}_{\mathsf{B}}} I(X; Y \mid S_R). \tag{A.22}$$

Then, we have

$$\begin{aligned}
\mathsf{R} &\leq \frac{1}{n} \sum_{i=1}^{n} I(X_i; Y_i \mid S_{R,i}) + \epsilon_n \\
&\overset{(c)}{\leq} \frac{1}{n} \sum_{i=1}^{n} \mathcal{C}_{\text{inf}}^{\text{Imp}} \left(\sum_x P_{X_i}(x) c(x), \ \sum_x P_{X_i}(x) b(x) \right) + \epsilon_n \\
&\overset{(d)}{\leq} \mathcal{C}_{\text{inf}}^{\text{Imp}} \left(\frac{1}{n} \sum_{i=1}^{n} \sum_x P_{X_i}(x) c(x), \ \frac{1}{n} \sum_{i=1}^{n} \sum_x P_{X_i}(x) b(x) \right) + \epsilon_n \\
&\overset{(e)}{\leq} \mathcal{C}_{\text{inf}}^{\text{Imp}}(\mathsf{D}, \mathsf{B})
\end{aligned} \tag{A.23}$$

where (c) holds by the definition of $\mathcal{C}_{\text{inf}}^{\text{Imp}}(\mathsf{D}, \mathsf{B})$ in (A.22), and (d) and (e) hold by similar monotonicity and concavity properties as stated in Lemma 1.

Achievability

Fix $P_X(\cdot)$ and a function $\hat{h}(x,z)$ that achieve $C(\mathsf{D}/(1+\epsilon),\mathsf{B})$, where D is the desired distortion and B is the target cost, for a small positive number $\epsilon > 0$. We define the joint pmf $P_{SS_RXY} := P_{SS_R}P_X P_{Y|SS_RX}$. Codebook generation, encoding, and estimation are as described in the proof of Theorem 1; the only difference is in the decoding at the Rx, where the state S^n has to be replaced by S_R. In more details:

Decoding Upon observing outputs $Y^n = y^n$ and state sequence $S_R^n = s_R^n$, the decoder looks for an index $\hat{w}$ such that

$$(s_R^n, x^n(\hat{w}), y^n) \in \mathcal{T}_\epsilon^{(n)}(P_{S_RXY}) \tag{A.24}$$

where $P_{S_RXY} = \sum_{\mathcal{S}} P_{SS_RXY}$. If exactly one such index exists, it declares $\hat{W} = \hat{w}$. Otherwise, it declares an error.

Analysis We start by analyzing the probability of error and the distortion averaged over the random code construction. Given the symmetry of the code construction, we can condition on the event $W = 1$. We then notice that the decoder makes an error, i.e., declares nothing or $\hat{W} \neq 1$ if, and only if, one or both of the following events occur:

$$\mathcal{E}_1 = \big\{(S_R^n, X^n(1), Y^n) \notin \mathcal{T}_\epsilon^{(n)}(P_{XS_RY})\big\} \tag{A.25}$$

or

$$\mathcal{E}_2 = \big\{(S_R^n, X^n(w'), Y^n) \in \mathcal{T}_\epsilon^{(n)}(P_{XS_RY}) \text{ for some } w' \neq 1\big\}. \tag{A.26}$$

Thus, by the union bound:

$$P_e^{(n)} = P(\mathcal{E}_1 \cup \mathcal{E}_2) \leq P(\mathcal{E}_1) + P(\mathcal{E}_2), \tag{A.27}$$

where we consider the average probability of error not only over the random channel noise and states but also over the random codeconstruction. The first term goes to zero as $n \to \infty$ by the weak law of large numbers. By the independence of the codewords and the packing lemma [20, Lemma 3.1], the second term also tends to zero as $n \to \infty$

$$R < I(X;Y|S_R). \tag{A.28}$$

Following similar steps as in the analysis in Appendix A.1.2, and using the fact that by the weak law of large numbers with probability tending to 1 as $n \to \infty$:

$$(S^n, S_R^n, X^n(1), Y^n) \in \mathcal{T}_\epsilon^{(n)}(P_X P_S P_{S_R} P_{Y|SS_RX}), \tag{A.29}$$

it can be shown that

$$\overline{\lim_{n\to\infty}}\,\Delta^{(n)} = (1+\epsilon)\mathsf{E}[d(S,\hat{s}^*(X,Z))]\,. \tag{A.30}$$

Thus when $\epsilon \downarrow 0$, the distortion constraint (3.5b) holds (averaged over the random code constructions, the random states, and the noise in the channel) whenever

$$\mathsf{E}[d(S,\hat{s}^*(X,Z))] < \mathsf{D}. \tag{A.31}$$

Notice that the cost constraint (3.5c) is fulfilled by construction.

By standard arguments it can then be shown that there must exist at least one sequence of deterministic code books $\mathcal{C}_n$ so that constraints (3.5) are satisfied under conditions (A.28) and (A.31).

A.2 Proof of Chapter 4

A.2.1 Converse Proof of Theorem 11

Fix a sequence (in n) of $(2^{n\mathsf{R}_0}, 2^{n\mathsf{R}_1}, 2^{n\mathsf{R}_2}, n)$ codes satisfying (6.3). Fix a blocklength n and start with Fano's inequality:

$$\begin{aligned}
\mathsf{R}_0+\mathsf{R}_2 &= \frac{1}{n}H(W_0,W_2)\\
&\le \frac{1}{n}\sum_{i=1}^{n} I(W_0,W_2;Y_{2i}\mid Y_2^{i-1},)+\epsilon_n\\
&\le \frac{1}{n}\sum_{i=1}^{n} I(W_0,W_2,Y_2^{i-1};Y_{2,i})+\epsilon_n\\
&= I(W_0,W_2,Y_2^{T-1};Y_{2,T}\mid T)+\epsilon_n\\
&\le I(W_0,W_2,Y_2^{T-1},T;Y_{2,T})+\epsilon_n\\
&\overset{(a)}{=} I(U;Y_2)+\epsilon_n,
\end{aligned} \tag{A.32}$$

where T is chosen uniformly over $\{1,\cdots,n\}$ and independent of $X^n, Y_1^n, Y_2^n, W_0, W_1, W_2, S_1^n, S_2^n$; ϵ_n is a sequence that tends to 0 as $n\to\infty$; and $U := (W_0, W_2, Y_2^{T-1}, S_2^{T-1}, T)$, $Y_2 := Y_{2,T}$ and $S_2 := S_{2,T}$. Here (a) holds because $S_2 \sim P_{S_2}$ independent of (U,X), where we define $X := X_T$.

Following similar steps, we obtain:

$$R_1 = \frac{1}{n}H(W_1\mid W_0,W_2)$$

$$
\begin{aligned}
&\leq \frac{1}{n} I(W_1; Y_1^n \mid W_0, W_2) + \epsilon_n \\
&\leq \frac{1}{n} I(W_1; Y_1^n, Y_2^n \mid W_0, W_2) + \epsilon_n \\
&= \frac{1}{n} \sum_{i=1}^{n} I(W_1; Y_{1,i}, Y_{2,i} \mid Y_1^{i-1}, Y_2^{i-1}, W_0, W_2) + \epsilon_n \\
&\leq \frac{1}{n} \sum_{i=1}^{n} I(X_i, W_1, Y_1^{i-1}, ; Y_{1,i}, Y_{2,i} \mid Y_2^{i-1}, W_0, W_2) + \epsilon_n \\
&\overset{(b)}{=} \frac{1}{n} \sum_{i=1}^{n} I(X_i; Y_{1,i} \mid Y_2^{i-1}, W_0, W_2) + \epsilon_n &\text{(A.33)} \\
&= I(X_T; Y_{1T} \mid Y_2^{T-1}, W_0, W_2, T) + \epsilon_n \\
&\overset{(c)}{=} I(X; Y_1 \mid U) + \epsilon_n, &\text{(A.34)}
\end{aligned}
$$

where we defined $Y_1 := Y_{1,T}$; and where (b) holds by the physically degradedness of the SDMBC which implies the Markov chain $(W_0, W_2, W_1, Y_1^{i-1}, Y_2^{i-1}) \to X_i \to (Y_{1,i}) \to (Y_{2,i})$, and (c) holds.

Recall that we assume the optimal estimators (4.5) in Lemma 3. Using the definitions of T, X, S_k above and defining $Z := Z_T$, we can write the average expected distortions as:

$$
\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[d_k(S_{k,i}, \hat{s}_k^*(X_i, Z_i)] = \mathbb{E}[d_k(S_k, \hat{s}_k^*(X, Z)]. \tag{A.35}
$$

Combining (A.32), (A.34), and (A.35) and letting $n \to \infty$, we obtain that there exists a limiting pmf P_{UX} such that the tuple $(U, X, S_1, S_2, Y_1, Y_2, Z) \sim P_{UX} P_{S_1 S_2} P_{Y_1 Y_2 Z | S_1 S_2 X}$ satisfies the rate-constraints

$$
\mathsf{R}_0 + \mathsf{R}_2 \leq I(U; Y_2) \tag{A.36}
$$

$$
\mathsf{R}_1 \leq I(X; Y_1 \mid U) \tag{A.37}
$$

and the distortion constraints

$$
\mathbb{E}[d_k(S_k, \hat{s}_k^*(X, Z)] \leq \mathsf{D}_k, \quad k = 1, 2, \tag{A.38}
$$

This completes the proof.

A.2.2 Proof of Theorem 12

Fix a sequence (in n) of $(2^{n\mathsf{R}_0}, 2^{n\mathsf{R}_1}, 2^{n\mathsf{R}_2}, n)$ codes satisfying (6.3). Fix then a blocklength n and consider an enhanced SDMBC where Rx 1 observes the pair of states $\tilde{S}_1 = (S_1, S_2)$ and the pair of outputs $\tilde{Y}_1 = (Y_1, Y_2)$. The enhanced SDMBC is clearly physically degraded because for any input pmf P_X the Markov chain

$$X -\!\!\circ\!\!- (\tilde{S}_1, \tilde{Y}_1) -\!\!\circ\!\!- (S_2, Y_2) \tag{A.39}$$

holds.

Following the steps in the previous Appendix A.2.1, we can conclude that

$$\mathsf{R}_0 + \mathsf{R}_2 \leq I(U_2; Y_2 \mid S_2) + \epsilon_n \tag{A.40}$$

$$\mathsf{R}_0 + \mathsf{R}_1 + \mathsf{R}_2 \leq I(X; Y_1, Y_2 \mid S_1, S_2) + \epsilon_n \tag{A.41}$$

and for $k = 1, 2$

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[d_k(S_{k,i}, \hat{s}_k^*(X_i, Z_i)] = \mathbb{E}[d_k(S_k, \hat{s}_k^*(X, Z)]. \tag{A.42}$$

Consider next a reversely enhanced SDMBC where Rx 1 observes only (Y_1, S_1) but Rx 2 observes both state sequences $\tilde{S}_2 := (S_1, S_2)$ and both outputs $\tilde{Y}_2 := (Y_1, Y_2)$. Following again the steps in the previous Appendix A.2.1, but now with exchanged indices 1 and 2, we obtain:

$$\mathsf{R}_0 + \mathsf{R}_1 \leq I(U_1; Y_1 \mid S_1) + \epsilon_n \tag{A.43}$$

$$\mathsf{R}_0 + \mathsf{R}_1 + \mathsf{R}_2 \leq I(X; Y_1, Y_2 \mid S_1, S_2) + \epsilon_n. \tag{A.44}$$

Combining all these inequalities and letting first $n \to \infty$ and then $\epsilon_n \downarrow 0$, establishes the desired converse result.

A.2.3 Proofs for Dueck's State-Dependent BC

Optimal Estimator of Lemma 3

We first derive the optimal estimator $\hat{s}_k^*(x_1, x_2, y_1', y_2')$ of Lemma 3 for this example.

Case $y_1' = y_2' = 1$: In this case, $S_1 = S_2 = 1$ deterministically, and thus

$$\hat{s}_k^*(x_1, x_2, 1, 1) = 1, \quad \forall (x_1, x_2), \qquad k = 1, 2. \tag{A.45}$$

Case $y_1 = 1'$ and $y_2' = 0$: In this case, $S_1 = 1$ deterministically and

$$\hat{s}_1^*(x_1, x_2, 1, 0) = 1, \quad \forall (x_1, x_2). \tag{A.46}$$

To derive the optimal estimator for state S_2, we notice that $y_1' = 1$ implies $x_1 \oplus N = 1$, i.e., $N = x_1 \oplus 1$. As a consequence,

$$y_2' = (x_2 \oplus x_1 \oplus 1) S_2. \tag{A.47}$$

So, for $x_2 = x_1$ we have $y_2' = S_2 = 0$ and the optimal estimator sets

$$\hat{s}_2^*(x_1, x_2, 1, 0) = 0, \quad x_1 = x_2. \tag{A.48}$$

Instead for $x_2 \neq x_1$, the feedback output $y_2' = 0$, irrespective of the state S_2. The optimal estimator then is the constant estimator

$$\hat{s}_2^*(x_1, x_2, 1, 0) = \operatorname*{argmax}_{\hat{s} \in \{0,1\}} P_S(\hat{s}), \quad x_1 \neq x_2. \tag{A.49}$$

Case $y_1' = 1, y_2' = 0$: Symmetric to the previous case $y_1' = 0, y_2' = 1$. The optimal estimators are as in (A.48) and (A.49), but with exchanged indices 1 and 2.

Case $y_1' = y_2' = 0$: To find the optimal estimators, we calculate the conditional probabilities $P_{S_k|X_1X_2Y_1'Y_2'}(\cdot|x_1, x_2, y_1', y_2')$ for $y_1' = y_2' = 0$.

We again distinguish the two cases $x_1 = x_2$ and $x_1 \neq x_2$ and start by considering $x_1 = x_2$. In this case, $x_1 \oplus N = x_2 \oplus N$, and so if $S_k = 1$ then $y_1' = y_2' = 0$ only if $x_1 \oplus N = x_2 \oplus N = 0$, which happens with probability $1/2$ because N is Bernoulli-$1/2$. By the independence of the states and the inputs for $x_1 = x_2$ and $k = 1, 2$:

$$\begin{aligned} P_{S_k|X_1X_2Y_1'Y_2'}(1|x_1, x_2, 0, 0) &= \frac{P_{S_k}(1) P_{Y_1'Y_2'|X_1X_2S_k}(0, 0|x_1, x_2, 1)}{P_{Y_1'Y_2'|X_1X_2}(0, 0|x_1, x_2)} \\ &= \frac{P_S(1) 1/2}{P_{Y_1'Y_2'|X_1X_2}(0, 0|x_1, x_2)}. \end{aligned} \tag{A.50a}$$

Let $\bar{k} := 3 - k$ for $k = 1, 2$. If $S_k = 0$, then $y_1' = y_2' = 0$ happens when either $x_1 \oplus N = x_2 \oplus N = 0$ or when $S_{\bar{k}} = 0$ and $x_1 \oplus N = x_2 \oplus N = 1$. Since these are exclusive events and have total probability of $1/2 + P_S(0)1/2$, we obtain for $x_1 = x_2$ and $k \in \{1, 2\}$:

$$
\begin{aligned}
P_{S_k|X_1X_2Y_1'Y_2'}(0|x_1, x_2, 0, 0) &= \frac{P_{S_k}(0)P_{Y_1'Y_2'|X_1X_2S_k}(0, 0|x_1, x_2, 0)}{P_{Y_1'Y_2'|X_1X_2}(0, 0|x_1, x_2)} \\
&= \frac{P_S(0)(1/2 + P_S(0)1/2)}{P_{Y_1'Y_2'|X_1X_2}(0, 0|x_1, x_2)}.
\end{aligned} \tag{A.50b}
$$

We conclude from (A.50) that for $y_1' = y_2' = 0$ and $x = x_1 = x_2$, the optimal estimators are

$$
\hat{s}_k^*(x, x, 0, 0) = \mathbb{1}\left\{P_S(0)(1 + P_S(0)) < P_S(1)\right\}, \ k = 1, 2. \tag{A.51}
$$

We turn to the case $x_1 \neq x_2$, where $x_1 \oplus N = 1 \oplus (x_2 \oplus N)$. As before, if $S_k = 1$, then $Y_k' = 0$ only if $x_1 \oplus N = 0$, which happens with probability $1/2$. Now this implies $x_2 \oplus N = 1$, and thus $Y_{\bar{k}}' = 0$ only if $S_{\bar{k}} = 0$, which happens with probability $P_S(0)$. We thus obtain for $x_1 \neq x_2$ and $k = 1, 2$:

$$
\begin{aligned}
P_{S_k|X_1X_2Y_1'Y_2'}(1|x_1, x_2, 0, 0) &= \frac{P_{S_k}(1)P_{Y_1'Y_2'|X_1X_2S_k}(0, 0|x_1, x_2, 1)}{P_{Y_1'Y_2'|X_1X_2}(0, 0|x_1, x_2)} \\
&= \frac{P_S(1)P_S(0)1/2}{P_{Y_1'Y_2'|X_1X_2}(0, 0|x_1, x_2)}.
\end{aligned} \tag{A.52a}
$$

If $S_k = 0$, then $Y_1' = Y_2' = 0$ happens when $x_{\bar{k}} \oplus N = 0$ or when $x_{\bar{k}} \oplus N = 1$ and $S_{\bar{k}} = 0$. Since these are exclusive events with total probability $1/2 + P_S(0)1/2$, we obtain for $x_1 \neq x_2$ and $k = 1, 2$:

$$
\begin{aligned}
P_{S_k|X_1X_2Y_1'Y_2'}(0|x_1, x_2, 0, 0) &= \frac{P_{S_k}(0)P_{Y_1'Y_2'|X_1X_2S_k}(0, 0|x_1, x_2, 0)}{P_{Y_1'Y_2'|X_1X_2}(0, 0|x_1, x_2)} \\
&= \frac{P_S(0)(1/2 + P_S(0)1/2)}{P_{Y_1'Y_2'|X_1X_2}(0, 0|x_1, x_2)}.
\end{aligned} \tag{A.52b}
$$

Since $P_S(1) < 1 + P_S(0)$, we conclude that for $y_1' = 0, y_2' = 0$ and $x_1 \neq x_2$, the optimal estimator is

$$
\hat{s}_k^*(x_1, x_2, 0, 0) = 0, \qquad x_1 \neq x_2, \ k = 1, 2. \tag{A.53}
$$

Minimum distortion

We evaluate the expected distortion of the optimal estimators in (4.47), for a given input pmf $P_{X_0X_1X_2}$. Let $t := \Pr[X_1 \neq X_2]$. We first consider the distortion on state S_2:

$$
\mathbb{E}[d(S_2, \hat{s}_2^*(X_1, X_2, Y_1', Y_2'))]
$$

$$= \sum_{(x_1,x_2,y_1',y_2')\in\{0,1\}^4} P_{X_1X_2Y_1'Y_2'}(x_1,x_2,y_1',y_2') \cdot \Pr\big[S_2 \neq \hat{s}_2^*(x_1,x_2,y_1',y_2') \mid X_1 = x_1, X_2 = X_2, Y_1' = y_1, Y_2' = y_2\big] \tag{A.54}$$

$$\stackrel{(a)}{=} \sum_{(x_1,x_2,y_1',y_2')\in\{0,1\}^4} P_{X_1X_2Y_1'Y_2'}(x_1,x_2,y_1',y_2') \cdot \min_{\hat{s}\in\{0,1\}} P_{S_2|X_1X_2Y_1'Y_2'}(\hat{s}|x_1,x_2,y_1',y_2'), \tag{A.55}$$

where (a) follows by the definition of the function $\hat{s}_2^*$.

In the previous Subsection A.2.3, we argued that for $y_2' = 1$ or for $(y_2' = 0, y_1' = 1, x_1 = x_2)$, the state S_2 is deterministic ($S_2 = 1$ in the former case and $S_2 = 0$ in the latter) and thus $\min_{\hat{s}\in\{0,1\}} P_{S_2|X_1X_2Y_1'Y_2'}(\hat{s}|x_1,x_2,y_1',y_2') = 0$. We further argued that for $(y_1' = 1, y_2' = 0, x_1 \neq x_2)$ the transmitter learns nothing about state S_2, which is thus still distributed according to P_S. Based on these observations, we continue from (A.55) as:

$$\begin{aligned}
&\mathbb{E}[d(S_2, \hat{s}_2^*(X_1,X_2,Y_1',Y_2'))] \\
&= \Pr\big[X_1 \neq X_2, Y_1' = 1, Y_2' = 0\big] \min\{P_S(0), P_S(1)\} \\
&\quad + \sum_{(x_1,x_2)\in\{0,1\}^2} P_{X_1X_2Y_1'Y_2'}(x_1,x_2,0,0) \\
&\qquad\qquad \min\{P_{S_1|X_1X_2Y_1'Y_2'}(0|x_1,x_2,0,0), P_{S_1|X_1X_2Y_1'Y_2'}(1|x_1,x_2,0,0)\} \\
&\stackrel{(b)}{=} \Pr[X_1 \neq X_2, N = X_1 \oplus 1, S_1 = 1] \min\{P_S(0), P_S(1)\} \\
&\quad + \Pr[X_1 = X_2]\, \frac{1}{2} \min\{P_S(1), P_S(0)(1 + P_S(0))\} + \Pr[X_1 \neq X_2]\, \frac{1}{2} P_S(0)P_S(1) && \text{(A.56)} \\
&= \frac{1}{2} tq\Big(\min\{q,(1-q)\} + (1-q)\Big) + \frac{1}{2}(1-t)q \min\{q,(1-q)(2-q)\}. && \text{(A.57)}
\end{aligned}$$

where in (b) we used (A.50)–(A.53) and the fact that when $X_1 \neq X_2$, then event $\{Y_1' = 1, Y_2' = 0\}$ is equivalent to event $\{N = X_1 \oplus 1, S_1 = 1\}$.

Proof of the Outer Bound

The outer bound is based on Theorem 12, as detailed out in the following. The single-rate constraints (4.30a) specialize to

$$\mathsf{R}_k \leq I(U_k; Y_k', X_0 \mid S_1, S_2) \tag{A.58}$$

$$\stackrel{(c)}{=} I(U_k; X_0) \tag{A.59}$$

$$\leq 1, \tag{A.60}$$

where the equality holds by the chain rule, because (U_1, X_0) and (S_1, S_2) are independent, and because $I(U_1; Y_1' \mid X_0, S_1, S_2) = 0$ due to the Bernoulli-$1/2$ noise N.

Defining $t := \Pr[X_1 \neq X_2]$, Bound (4.30b) specializes to:

$$\begin{aligned}
\mathsf{R}_1 + \mathsf{R}_2 &\leq I(X_0, X_1, X_2; Y_1', Y_2', X_0 \mid S_1, S_2) && \text{(A.61)}\\
&\overset{(d)}{=} H(X_0) + I(X_1, X_2; Y_2' \mid S_1, S_2, Y_1', X_0) && \text{(A.62)}\\
&\overset{(e)}{=} H(X_0) + I(X_1, X_2; Y_2' \mid S_1 = 1, S_2 = 1, Y_1', X_0) && \text{(A.63)}\\
&= H(X_0) + I(X_1, X_2; Y_2' \oplus Y_1' \mid S_1 = 1, S_2 = 1, X_0) && \text{(A.64)}\\
&\overset{(f)}{\leq} H(X_0) + P_{S_1 S_2}(1,1) H(X_1 \oplus X_2) && \text{(A.65)}\\
&\leq 1 + q^2 H_b(t). && \text{(A.66)}
\end{aligned}$$

where (d) holds by the chain rule and because $I(X_1, X_2; Y_1' \mid X_0, S_1, S_2) = 0$ due to the Bernoulli-$1/2$ noise N; (e) holds because for $(s_1, s_2) \neq (1,1)$ the mutual information term $I(X_1, X_2; Y_2' \mid S_1 = s_1, S_2 = s_2, Y_1', X_0) = 0$ due to the Bernoulli-$1/2$ noise N; and (f) holds because for $S_1 = S_2 = 1$ we have $Y_2' \oplus Y_1' = (X_2 \oplus N) \oplus (X_1 \oplus N) = X_2 \oplus X_1$ and because conditioning can only reduce entropy.

The sum-rate constraint (A.66) is maximized for $t = 1/2$, which combined with (A.60) establishes the converse to the capacity region in (4.52).

Proof of Achievability Results

We evaluate Proposition 2 for different choices of the involved random variables. Since we ignore the common rate R_0, bound (4.32d) is not active and can be ignored.

- **First choice**
  - X_0, X_1, X_2 Bernoulli-$1/2$ with X_0 independent of (X_1, X_2) and $X_1 = X_2 = x$ with probability $\frac{1-t}{2}$ for all $x \in \{0, 1\}$;
  - $U_k = X_k$, for $k = 0, 1, 2$;
  - $V_1 = (X_0, X_1)$, $V_2 = (X_0, X_2)$, $V_0 = X_1 \oplus Y_1'$.

We plug this choice into Proposition 2. Constraint (4.32a) evaluates to:

$$\begin{aligned}
\mathsf{R}_1 &\le I(U_0,U_1;Y_1,V_1 \mid S_1) - I(U_0,U_1,U_2,Z;V_0,V_1 \mid S_1,Y_1) && \text{(A.67)}\\
&= I(X_0,X_1;X_0,Y_1',S_1,S_2,X_1 \mid S_1)\\
&\quad -I(X_0,X_1,X_2,Y_1',Y_2';X_1\oplus Y_1',X_0,X_1 \mid S_1,S_2,X_0,Y_1') && \text{(A.68)}\\
&\overset{(e)}{=} H(X_0)+H(X_1)-H(X_1 \mid Y_1') && \text{(A.69)}\\
&= H(X_0) = 1 && \text{(A.70)}
\end{aligned}$$

where (e) holds because Y_1' is independent of X_1 due to the Bernoulli-$1/2$ noise N.

Constraint (4.32b) evaluates to:

$$\begin{aligned}
\mathsf{R}_2 &\le I(U_0,U_2;Y_2,V_2 \mid S_2) - I(U_0,U_1,U_2,Z;V_0,V_2|S_2,Y_2) && \text{(A.71)}\\
&= I(X_0,X_2;X_0,Y_2',S_1,S_2,X_2 \mid S_1)\\
&\quad -I(X_0,X_1,X_2,Y_1',Y_2';X_1\oplus Y_1',X_0,X_2 \mid S_1,S_2,X_0,Y_2') && \text{(A.72)}\\
&\overset{(f)}{=} H(X_0)+H(X_2)-H(X_2)-H(X_1\oplus Y_1' \mid S_1,S_2,X_0,Y_2',X_2) && \text{(A.73)}\\
&\overset{(g)}{=} 1-(1-q)(H_{\mathrm{b}}(t)+q), && \text{(A.74)}
\end{aligned}$$

where (f) holds because of the chain rule and the independence of X_2 and Y_2'; and (g) holds because for $S_1=0$ the XOR $X_1\oplus Y_1' = X_1$ and thus $H(X_1\oplus Y_1' \mid S_1,S_2,X_0,Y_2',X_2) = H(X_1 \mid X_2)$, for $S_1=S_2=1$ the XOR $X_1\oplus Y_1' = X_2\oplus Y_2'$, and finally for $S_1=1$ and $S_2=0$ the XOR $X_1\oplus Y_1' = N$ independent of $(Y_2'=0,X_2)$.

Constraint (4.32b) evaluates to:

$$\begin{aligned}
&\mathsf{R}_1+\mathsf{R}_2\\
&\le I(U_1;Y_1,V_1 \mid U_0,S_1)+I(U_2;Y_2,V_2 \mid U_0,S_2)+\min_{k\in\{1,2\}} I(U_0;Y_k,V_k \mid S_k)-I(U_1;U_2|U_0)\\
&\quad -I(U_0,U_1,U_2,Z;V_1 \mid V_0,S_1,Y_1)-I(U_0,U_1,U_2,Z;V_2 \mid V_0,S_2,Y_2)\\
&\quad -\max_{k\in\{1,2\}} I(U_0,U_1,U_2,Z;V_0 \mid S_k,Y_k) && \text{(A.75)}\\
&= \underbrace{I(X_1;X_0,Y_1',S_1,S_2,X_1 \mid X_0,S_1)}_{=H(X_1)}+\underbrace{I(X_2;X_0,Y_2',S_1,S_2,X_2 \mid X_0,S_2)}_{=H(X_2)}
\end{aligned}$$

$$
\begin{aligned}
&+ \min_{k\in\{1,2\}} \underbrace{I(X_0; X_0, Y_k', S_1, S_2, X_k \mid S_k)}_{=H(X_0)} \\
&- \quad \underbrace{I(X_1; X_2)}_{=H(X_1)-H(X_1|X_2)} \quad - \underbrace{I(\underline{X}, Y_1', Y_2'; X_0, X_1 \mid X_1 \oplus Y_1', X_0, \underline{S}, Y_1')}_{=0} \\
&- \underbrace{I(\underline{X}, Y_1', Y_2'; X_0, X_2 \mid X_1 \oplus Y_1', X_0, \underline{S}, Y_2')}_{=H(X_2|X_1\oplus Y_1', S_1, S_2, Y_2')} - \max_{k\in\{1,2\}} \underbrace{I(\underline{X}, Y_1', Y_2'; X_1 \oplus Y_1' \mid \underline{S}, X_0, Y_k')}_{=H(X_1\oplus Y_1'|S_1,S_2,Y_k')} && \text{(A.76)} \\
&\overset{(h)}{=} 2 + H_\mathrm{b}(t) - H(X_2 \mid X_1 \oplus Y_1', S_1, S_2, Y_2') - H(X_1 \oplus Y_1') && \text{(A.77)} \\
&\overset{(i)}{=} 1 + H_\mathrm{b}(t) - (1-q)H_\mathrm{b}(t) - q(1-q), && \text{(A.78)}
\end{aligned}
$$

where we used the abbreviations $\underline{X} = (X_0, X_1, X_2)$ and $\underline{S} = (S_1, S_2)$ and (h) holds because $X_1 \oplus Y_1$ is independent of (S_1, S_2, Y_k'), for $k = 1, 2$; and (i) holds because for $S_1 = S_2 = 1$ we have $X_2 = Y_2' \oplus Y_1' \oplus X_1$ and thus $H(X_2 \mid X_1 \oplus Y_1', S_1 = 1, S_2 = 1, Y_2') = 0$, for $S_1 = 0$ the XOR $X_1 \oplus Y_1' = X_1$ and thus $H(X_2 \mid X_1 \oplus Y_1', S_1 = 1, S_2, Y_2') = H(X_2|X_1) = H_\mathrm{b}(t)$, and finally for $S_1 = 1$ and $S_2 = 0$, we have $X_1 \oplus Y_1' = N$ and $Y_2' = 0$ and thus $H(X_2 \mid X_1 \oplus Y_1', S_1 = 1, S_2 = 0, Y_2') = H(X_2) = 1$.

The presented choice of parameters can thus achieve all rate-distortion tuples $(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ satisfying the distortion constraints in (4.49) (which only depends on the probability $t :=$ Pr$[X_1 \neq X_2]$) and

$$
\begin{aligned}
\mathsf{R}_1 &\leq 1 && \text{(A.79a)} \\
\mathsf{R}_2 &\leq 1 - (1-q)(H_\mathrm{b}(t) + q) && \text{(A.79b)} \\
\mathsf{R}_1 + \mathsf{R}_2 &\leq 1 + qH_\mathrm{b}(t) - q(1-q). && \text{(A.79c)}
\end{aligned}
$$

- **Second choice**

Same as the first choice except that $V_0 = X_2 \oplus Y_2'$. Following symmetric arguments as above, we conclude that for this choice the constraints in (4.32) evaluate to:

$$
\begin{aligned}
\mathsf{R}_1 &\leq 1 - (1-q)(H_\mathrm{b}(t) + q) && \text{(A.80a)} \\
\mathsf{R}_2 &\leq 1 && \text{(A.80b)} \\
\mathsf{R}_1 + \mathsf{R}_2 &\leq 1 + qH_\mathrm{b}(t) - q(1-q). && \text{(A.80c)}
\end{aligned}
$$

- **Combining the Choices and Time-Sharing**

From the two previous subsections, we conclude that for any $t \in [0, 1]$ the set of rate-distortion tuples

$(\mathsf{R}_0, \mathsf{R}_1, \mathsf{R}_2, \mathsf{D}_1, \mathsf{D}_2)$ is achievable if it satisfies (4.49) and

$$\mathsf{R}_0 + \mathsf{R}_1 \leq 1 \tag{A.81a}$$

$$\mathsf{R}_0 + \mathsf{R}_2 \leq 1 \tag{A.81b}$$

$$\mathsf{R}_0 + \mathsf{R}_1 + \mathsf{R}_2 \leq 1 + qH_{\mathsf{b}}(t) - q(1-q). \tag{A.81c}$$

As previously discussed, for $q \leq 1/2$, the distortion constraints (4.49) do not depend on t, and thus without loss in optimality in (A.81) one can set $t = 1/2$, which results in a sum-rate constraint

$$\mathsf{R}_0 + \mathsf{R}_1 + \mathsf{R}_2 \leq 1 + q^2. \tag{A.82}$$

Combined with (A.81a) and (A.81b), this sum-rate bound establishes the achievability of the capacity region in (4.52).

For $q > 1/2$ the distortion constraints (4.49) are either increasing or decreasing in t. The set of achievable rate-distortion tuples is then obtained by varying t either over $[0, 1/2]$ or over $[1/2, 1]$. Numerical results indicate that the so obtained set is not convex and the convex hull is obtained by considering convex combinations between different values of $t > 0$ and $t = 0$ for $q \in [2/3, 1]$ and $t = 1$ for $q \in [1/2, 2/3]$.

A.2.4 Proof of Lemma 3

Recall that $\hat{S}_k^n$ is a function of X^n, Z^n and write for each $i = 1, \cdots, n$:

$$\begin{aligned}
\mathbb{E}\left[d_k(S_{k,i}, \hat{S}_{k,i})\right] &= \mathbb{E}_{X^n,Z^n}\left[\mathbb{E}[d_k(S_{k,i}, \hat{S}_{k,i})|X^n, Z^n]\right] && \text{(A.83)}\\
&\overset{(a)}{=} \sum_{x^n,z^n} P_{X^nZ^n}(x^n, z^n) \sum_{\hat{s}_k \in \mathcal{S}_k} P_{\hat{S}_{k,i}|X^nZ^n}(\hat{s}_k|x^n, z^n) \\
&\qquad \sum_{s_k} P_{S_{k,i}|X_iZ_i}(s_k|x_i, z_i) d(s_k, \hat{s}_k) && \text{(A.84)}\\
&\geq \sum_{x^n,z^n} P_{X^nZ^n}(x^n, z^n) \cdot \min_{\hat{s}_k \in \mathcal{S}_k} \sum_{s_k} P_{S_{k,i}|X_iZ_i}(s_k|x_i, z_i) d(s_k, \hat{s}_k) \\
&= \mathbb{E}[d(S_{k,i}, \hat{s}^*_{k,i}(X_i, Z_i))], && \text{(A.85)}
\end{aligned}$$

where (a) holds by the Markov chain

$$\left(X^{i-1}, X_{i+1}^n, Z^{i-1}, Z_{i+1}^n, \hat{S}_{k,i}\right) -\!\!\circ\!\!- (X_i, Z_i) -\!\!\circ\!\!- S_{k,i}.$$

Summing over all $i = 1, \dots, n$, we thus obtain:

$$\Delta_k^{(n)} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[d(S_{k,i}, \hat{S}_{k,i})\right] \geq \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[d(S_{k,i}, \hat{s}_k^*(X_i, Z_i))], \tag{A.86}$$

which yields the desired conclusion.

A.2.5 Proof of Proposition 3

It suffices to show that under the described conditions, the distortion constraint (4.31) does not depend on P_X. To this end, we define $T_k = \psi_k(X, Z)$ for $k = 1, 2$ and rewrite the expected distortion as:

$$\begin{aligned}
\mathbb{E}[d(S_k, \hat{S}_k)] &= \sum_{(x,z)\in\mathcal{X}\times\mathcal{Z}} P_{XZ}(x,z) \sum_{s\in\mathcal{S}_k} P_{S_k|XZ}(s \mid x,z) \cdot d(s, \hat{s}_k^*(x,z)) && \text{(A.87)}\\
&\stackrel{(a)}{=} \sum_{(x,z)\in\mathcal{X}\times\mathcal{Z}} P_{XZ}(x,z) \cdot \min_{s'\in\hat{\mathcal{S}}_k} \sum_{(s,t)\in\mathcal{S}_k\times\mathcal{T}_k} P_{S_kT_k|XZ}(s,t \mid x,z) d(s, s'_k) && \text{(A.88)}\\
&\stackrel{(b)}{=} \sum_{(x,z,t)\in\mathcal{X}\times\mathcal{Z}\times\mathcal{T}_k} P_{XZ}(x,z)\mathbb{1}\{t = \psi(x,z)\} \cdot \min_{s'\in\hat{\mathcal{S}}_k} \sum_{s\in\mathcal{S}_k} P_{S_k|T_k}(s \mid t) d(s,s') && \text{(A.89)}\\
&= \sum_{t\in\mathcal{T}_k} P_{T_k}(t) \min_{s'\in\hat{\mathcal{S}}_k} \sum_{s\in\mathcal{S}_k} P_{S_k|T_k}(s \mid t) d(s,s') && \text{(A.90)}
\end{aligned}$$

where (a) holds by the definition of $\hat{s}_k^*(x,z)$ and the law of total probability; and (b) by the Markov chain $S_k \multimap T_k \multimap (X,Z)$, see (4.35), and because T_k is a function of X, Z. The independence of the pair (T_k, S_k) with X from (4.36), together with the above expression implies that the expected distortion does not depend on the choice of the input distribution P_X. Hence, we can conclude that for any given $\mathsf{B} \geq 0$, the rate-distortion tradeoff function $\mathcal{C}(\mathsf{D}, \mathsf{B})$ is constant over all $\mathsf{D} \geq \mathsf{D}_{\min}$ and coincides with the capacity of the SDMC $\mathcal{C}_{\text{NoEst}}(\mathsf{B})$.

A.3 Proof of Chapter 5

A.3.1 Proof of Theorem 17

To derive an upper bound on the average error probability (averaged over the random code construction and the state and channel realizations), we enlarge the error event to the event that for some $k = 1, 2$ and $b = 1, \dots, B$:

$$\hat{W}_{k,c,(b)} \neq W_{k,c,(b)} \quad \text{or} \quad \hat{W}_{k,p,(b)} \neq W_{k,p,(b)} \quad \text{or} \quad \hat{W}_{k,c,(b)}^{(\bar{k})} \neq W_{k,c,(b)} \tag{A.91}$$

or

$$J^*_{k,(b)} = -1 \quad \text{or} \quad \hat{J}_{k,(b)} \neq J^*_{k,(b)} \quad \text{or} \quad \hat{J}^{(\bar{k})}_{k,(b)} \neq J^*_{k,(b)}. \tag{A.92}$$

For ease of notation, we define the block-b Tx-error events for $k = 1, 2$ and $b = 1, \ldots, B$:

$$\mathcal{E}_{\text{Tx},k,(b)} := \left\{ \hat{W}^{(k)}_{\bar{k},c,(b)} \neq W_{\bar{k},c,(b)} \text{ or } \hat{J}^{(k)}_{\bar{k},(b-1)} \neq J^*_{\bar{k},(b-1)} \text{ or } J^*_{k,b} = -1 \right\}, \tag{A.93}$$

and

$$\mathcal{E}_{\text{Tx},k,(B+1)} := \left\{ \hat{J}^{(k)}_{\bar{k},(B)} \neq J^*_{\bar{k},(B)} \right\}, \qquad k \in \{1, 2\}. \tag{A.94}$$

Define also the Rx-error events for $k = 1, 2$ and block $b = 1, \ldots, B+1$:

$$\mathcal{E}_{\text{Rx},(b)} := \left\{ \hat{W}_{k,c,(b-1)} \neq W_{k,c,(b-1)} \text{ or } \hat{W}_{k,p,(b)} \neq W_{k,p,(b)} \text{ or } \hat{J}_{k,(b-1)} \neq J^*_{k,(b-1)} \colon \ k = 1, 2 \right\}. \tag{A.95}$$

By the union bound and basic probability, we find:

$$\begin{aligned} \Pr\left(\hat{W}_1 \neq W_1 \text{ or } \hat{W}_2 \neq W_2\right) \leq & \sum_{b=1}^{B+1} \Pr\left(\mathcal{E}_{\text{Rx},(b)} \middle| \bigcup_{b'=1}^{B+1} \{\bar{\mathcal{E}}_{\text{Tx},1,(b')},\ \bar{\mathcal{E}}_{\text{Tx},2,(b')}\}\right) \\ & + \sum_{b=1}^{B+1} \Pr\left(\mathcal{E}_{\text{Tx},1,(b)} \middle| \bigcup_{b'=1}^{b-1} \{\bar{\mathcal{E}}_{\text{Tx},1,(b')},\ \bar{\mathcal{E}}_{\text{Tx},2,(b')}\}\right) \\ & + \sum_{b=1}^{B+1} \Pr\left(\mathcal{E}_{\text{Tx},2,(b)} \middle| \bigcup_{b'=1}^{b-1} \{\bar{\mathcal{E}}_{\text{Tx},1,(b')},\ \bar{\mathcal{E}}_{\text{Tx},2,(b')}\}\right). \end{aligned} \tag{A.96}$$

We analyze the three sums separately. The first sum is related to Tx 1's error event, the second sum to Tx 2's error event, and the third sum to the Rx's error event.

Analysis of Tx 1's error event

To simplify notations, we define for each block $b \in \{2, \ldots, B+1\}$ and each triple of indices $(j_1^*, \hat{w}_2, \hat{j}_2)$ the event $\mathcal{F}_{\text{Tx}1,(b)}(j_1^*, \hat{w}_2, \hat{j}_2)$ that the following two conditions (A.97) and (A.98) (only Condition (A.97) for $b = 1$) hold:

$$\Bigg(u^N_{0,(b)}\left(W_{1,c,(b-1)}, \hat{W}^{(1)}_{2,c,(b-1)}\right),\ u^N_{1,(b)}\left(W_{1,c,(b)}, J^*_{1,(b-1)} \,\middle|\, W_{1,c,(b-1)}, W_{2,c,(b-1)}\right)$$
$$u^N_{2,(b)}\left(\hat{w}_2, \hat{j}_2 \,\middle|\, W_{1,c,(b-1)}, W_{2,c,(b-1)}\right),\ x^N_{1,(b)}\left(W_{1,p,(b)} \,\middle|\, W_{1,c,(b)}, J^*_{1,(b-1)}, W_{1,c,(b-1)}, W_{2,c,(b-1)}\right),$$

$$v^N_{1,(b)}\Big(j_1^* \;\Big|\; J^*_{1,(b-1)}, W_{1,c,(b)}, \hat{w}_2, \hat{j}_2, W_{1,c,(b-1)}, W_{2,c,(b-1)}\Big),\; Z^N_{1,(b)}\Big) \in \mathcal{T}^N_\epsilon(P_{U_0U_1U_2X_1V_1Z_1}) \qquad \text{(A.97)}$$

and if $b > 1$

$$\Big(u^N_{0,(b-1)}\Big(W_{1,c,(b-2)}, W_{2,c,(b-2)}\Big), u^N_{1,(b-1)}\Big(W_{1,c,(b-1)}, J^*_{1,(b-1)} \;\Big|\; W_{1,c,(b-2)}, W_{2,c,(b-2)}\Big) u^N_{2,(b-1)}\Big(W_{2,c,(b-1)}, J_{2,(b-2)} \Big| W_{1,c,(b-2)}, W_{2,c,(b-2)}\Big), x^N_{1,(b-1)}\Big(W_{1,p,(b-1)} \;\Big|\; W_{1,c,(b-1)}, J^*_{1,(b-2)}, W_{1,c,(b-2)}, W_{2,c,(b-2)}\Big), v^N_{2,(b-1)}\Big(\hat{j}_2 \;\Big|\; W_{1,c,(b-1)}, J^*_{1,(b-2)}, W_{2,c,(b-1)}, J^*_{2,(b-2)}, W_{1,c,(b-2)}, W_{2,c,(b-2)}\Big), Z^N_{1,(b-1)}\Big) \in \mathcal{T}^N_\epsilon(P_{U_0U_1U_2X_1V_2Z_1}). \qquad \text{(A.98)}$$

Notice that compared to (5.19) and (5.20), here we replaced the triple $(\hat{W}^{(1)}_{2,c,(b-2)}, \hat{W}^{(1)}_{2,c,(b-1)}, \hat{J}^{(1)}_{2,(b-2)})$ by their correct values $W_{2,c,(b-2)}, W_{2,c,(b-1)}, J^*_{2,(b-2)})$. Similarly, define the event $\mathcal{F}_{\text{Tx1},(B+1)}(\hat{j}_2)$ as the event that the following two conditions are satisfied:

$$\Big(u^N_{0,(B+1)}\Big(W_{1,c,(B)}, W_{2,c,(B)}\Big), u^N_{1,(B+1)}\Big(1, J^*_{1,(B)} \;\Big|\; W_{1,c,(B)}, W_{2,c,(B)}\Big),\; u^N_{2,(B+1)}\Big(1, \hat{j}_2 \;\Big|\; W_{1,c,(B)}, W_{2,c,(B)}\Big), x^N_{1,(B+1)}\Big(1 \;\Big|\; W_{1,c,(B+1)}, J^*_{1,(B)}, W_{1,c,(B)}, W_{2,c,(B)}\Big),\; Z^N_{1,(B+1)}\Big) \in \mathcal{T}^N_\epsilon(P_{U_0U_1U_2X_1Z_1}) \qquad \text{(A.99)}$$

and

$$\Big(u^N_{0,(B)}\Big(W_{1,c,(B-1)}, W_{2,c,(B-1)}\Big), u^N_{1,(B)}\Big(W_{1,c,(B)}, J^*_{1,(B)} \;\Big|\; W_{1,c,(B-1)}, W_{2,c,(B-1)}\Big), u^N_{2,(B)}\Big(W^{(1)}_{2,c,(B)}, J^{(1)}_{2,(B-1)} \;\Big|\; W_{1,c,(B-1)}, \hat{W}^{(1)}_{2,c,(B-1)}\Big), x^N_{1,(B)}\Big(W_{1,p,(B)} \;\Big|\; W_{1,c,(B)}, J^*_{1,(B-1)}, W_{1,c,(B-1)}, \hat{W}^{(1)}_{2,c,(B-1)}\Big), v^N_{2,(B)}\Big(\hat{j}_2 \;\Big|\; W_{1,c,(B)}, J^*_{1,(B-1)}, \hat{W}^{(1)}_{2,c,(B)}, J^{(1)}_{2,(B-1)}, W_{1,c,(B-1)}, W_{2,c,(B-1)}\Big), Z^N_{1,(B)}\Big) \in \mathcal{T}^N_\epsilon(P_{U_0U_1U_2X_1V_2Z_1}) \qquad \text{(A.100)}$$

We continue by noticing that event $\bigcup_{b'=1}^{b-1}\left\{\bar{\mathcal{E}}_{\mathrm{Tx},1,(b')},\ \bar{\mathcal{E}}_{\mathrm{Tx},2,(b')}\right\}$ implies that for all $b'=1,\ldots,b-1$, $k=1,2$:

$$\hat{W}^{(k)}_{\bar{k},c,(b')} = W_{\bar{k},c,(b')} \tag{A.101}$$

$$J^*_{k,(b')} \neq -1 \tag{A.102}$$

$$\hat{J}^{(\bar{k})}_{k,(b'-1)} = J^*_{k,(b'-1)}. \tag{A.103}$$

Moreover, for any block $b = 1,\ldots,B+1$, event $\bar{\mathcal{E}}_{\mathrm{Tx},1,(b)}$ is implied by the event that $\mathcal{F}_{\mathrm{Tx}1,(b)}(j_1^*, \hat{w}_2, \hat{j}_2)$ *is not* satisfied for any tuple $(j_1^*, \hat{w}_2, \hat{j}_2)$ with $(\hat{w}_2, \hat{j}_2) = (W_{2,c,(b)}, J^*_{2,(b-1)})$ or it *is* satisfied for some triple $(j_1^*, \hat{w}_2, \hat{j}_2)$ with $(\hat{w}_2, \hat{j}_2) \neq (W_{2,c,(b)}, J^*_{2,(b-1)})$. Thus, the sequence of inequalities on top of the next page holds, where the inequalities hold by the union bound. By the Covering Lemma [63], the way we construct the codebooks and the weak law of large numbers, and because we condition on event $\bar{\mathcal{E}}_{\mathrm{Tx},2,(b-1)}$ implying $J^*_{2,b-1} \neq -1$, the first summand in (A.104c) tends to 0 as $N \to \infty$ if

$$R_{1,v} > I(V_1; X_1 Z_1 \mid U_0 U_1 U_2). \tag{A.105}$$

By the way we constructed the codebooks, and standard information-theoretic arguments [63], the sum in the second line of (A.104c) tends to 0 as $N \to \infty$, if

$$R_{1,v}+R_{2,v} + R_{2,c} < I(U_2 V_1; Z_1 X_1 \mid U_0 U_1) + I(V_2; Z_1 X_1 \mid U_0 U_1 U_2), \tag{A.106}$$

the sum in the third line of (A.104c) tends to 0 as $N \to \infty$ if

$$R_{1,v}+R_{2,v} < I(U_2 V_1; Z_1 X_1 \mid U_0 U_1) + I(V_2; Z_1 X_1 \mid U_0 U_1 U_2), \tag{A.107}$$

and the sum in the fourth line of (A.104c) tends to 0 as $N \to \infty$ if

$$R_{1,v}+R_{2,c} < I(Z_1 X_1; U_2 V_1 \mid U_0 U_1). \tag{A.108}$$

Since Condition (A.107) is obsolete in view of (A.106), we conclude that for any finite B the sum of the probability of errors $\sum_{b=1}^{B+1} \Pr\left(\mathcal{E}_{\mathrm{Tx},1,(b)} \middle| \bigcup_{b'=1}^{b-1}\left\{\bar{\mathcal{E}}_{\mathrm{Tx},1,(b')},\ \bar{\mathcal{E}}_{\mathrm{Tx},2,(b')}\right\}\right)$ tends to 0 as $N \to \infty$ if Conditions (A.105), (A.106), and (A.108) are satisfied.

$$\begin{aligned}
&\Pr\left(\mathcal{E}_{\mathrm{Tx},1,(b)} \,\middle|\, \bigcup_{b'=1}^{b-1}\left\{\bar{\mathcal{E}}_{\mathrm{Tx},1,(b')},\ \bar{\mathcal{E}}_{\mathrm{Tx},2,(b')}\right\}\right) \\
&= \Pr\left(\left(\bigcap_{j_1^*\in[2^{nR_{v,1}}]} \bar{\mathcal{F}}_{\mathrm{Tx1},(b)}(j_1^*, W_{2,c,(b)}, J^*_{2,(b-1)})\right)\right. \\
&\qquad \left.\cup\left(\bigcup_{\substack{(j_1^*,\hat{w}_2,\hat{j}_2):\\(\hat{w}_2,\hat{j}_2)\neq(W_{2,c,(b)},J^*_{2,(b-1)})}} \mathcal{F}_{\mathrm{Tx1},(b)}(j_1^*, \hat{w}_2, \hat{j}_2)\right) \,\middle|\, \bigcup_{b'=1}^{b-1}\left\{\bar{\mathcal{E}}_{\mathrm{Tx},1,(b')},\ \bar{\mathcal{E}}_{\mathrm{Tx},2,(b')}\right\}\right)
\end{aligned} \tag{A.104a}$$

$$\begin{aligned}
&\le \Pr\left(\bigcap_{j_1^*\in[2^{nR_{v,1}}]} \bar{\mathcal{F}}_{\mathrm{Tx1},(b)}(j_1^*, W_{2,c,(b)}, J^*_{2,(b-1)}) \,\middle|\, \bigcup_{b'=1}^{b-1}\left\{\bar{\mathcal{E}}_{\mathrm{Tx},1,(b')},\ \bar{\mathcal{E}}_{\mathrm{Tx},2,(b')}\right\}\right) \\
&\quad + \Pr\left(\bigcup_{\substack{(j_1^*,\hat{w}_2,\hat{j}_2):\\(\hat{w}_2,\hat{j}_2)\neq(W_{2,c,(b)},J^*_{2,(b-1)})}} \mathcal{F}_{\mathrm{Tx1},(b)}(j_1^*, \hat{w}_2, \hat{j}_2) \,\middle|\, \bigcup_{b'=1}^{b-1}\left\{\bar{\mathcal{E}}_{\mathrm{Tx},1,(b')},\ \bar{\mathcal{E}}_{\mathrm{Tx},2,(b')}\right\}\right)
\end{aligned} \tag{A.104b}$$

$$\begin{aligned}
&\le \Pr\left(\bigcap_{j_1^*\in[2^{nR_{v,1}}]} \bar{\mathcal{F}}_{\mathrm{Tx1},(b)}(j_1^*, W_{2,c,(b)}, J^*_{2,(b-1)}) \,\middle|\, \bigcup_{b'=1}^{b-1}\left\{\bar{\mathcal{E}}_{\mathrm{Tx},1,(b')},\ \bar{\mathcal{E}}_{\mathrm{Tx},2,(b')}\right\}\right) \\
&\quad + \sum_{\substack{(j_1^*,\hat{w}_2,\hat{j}_2):\\ \hat{w}_2\neq W_{2,c,(b)},\\ \hat{j}_2\neq J^*_{2,(b-1)}}} \Pr\left(\mathcal{F}_{\mathrm{Tx1},(b)}(j_1^*, \hat{w}_2, \hat{j}_2) \,\middle|\, \bigcup_{b'=1}^{b-1}\left\{\bar{\mathcal{E}}_{\mathrm{Tx},1,(b')},\ \bar{\mathcal{E}}_{\mathrm{Tx},2,(b')}\right\}\right) \\
&\quad + \sum_{\substack{(j_1^*,\hat{j}_2):\\ \hat{j}_2\neq J^*_{2,(b-1)}}} \Pr\left(\mathcal{F}_{\mathrm{Tx1},(b)}(j_1^*, W_{2,c,(b)}, \hat{j}_2) \,\middle|\, \bigcup_{b'=1}^{b-1}\left\{\bar{\mathcal{E}}_{\mathrm{Tx},1,(b')},\ \bar{\mathcal{E}}_{\mathrm{Tx},2,(b')}\right\}\right) \\
&\quad + \sum_{\substack{(j_1^*,\hat{w}_2):\\ \hat{w}_2\neq W_{2,c,(b)}}} \Pr\left(\mathcal{F}_{\mathrm{Tx1},(b)}(j_1^*, \hat{w}_2, J^*_{2,(b-1)}) \,\middle|\, \bigcup_{b'=1}^{b-1}\left\{\bar{\mathcal{E}}_{\mathrm{Tx},1,(b')},\ \bar{\mathcal{E}}_{\mathrm{Tx},2,(b')}\right\}\right),
\end{aligned} \tag{A.104c}$$

Analysis of Tx 2's error event

By similar arguments, one can also prove that for finite B the sum of the probability of errors $\sum_{b=1}^{B+1} \Pr\left(\mathcal{E}_{\text{Tx},2,(b)} \middle| \bigcup_{b'=1}^{b-1} \{\bar{\mathcal{E}}_{\text{Tx},1,(b')},\ \bar{\mathcal{E}}_{\text{Tx},2,(b')}\}\right)$ tends to 0 as $N \to \infty$ if Conditions (5.25a), (5.25b), and (5.25c), are satisfied for $k = 2$.

Analysis of Rx's error event

For each block $b = 2, \ldots, B$ and each tuple $(w_{1,c}, w_{2,c}, w_{1,p}, w_{2,p}, j_1, j_2)$ define $\mathcal{F}_{\text{Rx},(b)}(w_{1,c}, w_{2,c}, w_{1,p}, w_{2,p}, j_1, j_2)$ as the event

$$\begin{aligned}\Bigg(& u_{0,(b)}^N(w_{1,c}, w_{2,c}),\ u_{1,(b)}^N\left(W_{1,c,(b)}, j_1 \,\middle|\, w_{1,c}, w_{2,c}\right),\ u_{2,(b)}^N\left(W_{2,c,(b)}, j_2 \,\middle|\, w_{1,c}, w_{2,c}\right), \\ & x_{1,(b)}^N\left(w_{1,p} \,\middle|\, W_{1,c,(b)}, j_1, w_{1,c}, w_{2,c}\right),\ x_{2,(b)}^N\left(w_{2,p} \,\middle|\, W_{2,c,(b)}, j_2, w_{1,c}, w_{2,c}\right) \\ & v_{1,(b)}^N\left(J_{1,(b)} \,\middle|\, W_{1,c,(b)}, W_{2,c,(b)}, w_{1,c}, j_1, w_{2,c}, j_2\right),\ v_{2,(b)}^N(J_{2,(b)} \mid W_{1,c,(b)}, W_{2,c,(b)}, w_{1,c}, j_1, w_{2,c}, j_2), \\ & Y_{(b)}^N\Bigg) \in \mathcal{T}_{2\epsilon}(P_{U_0U_1U_2X_1X_2Y}).\end{aligned} \tag{A.109}$$

We continue by noticing that for $b = 2, \ldots, B$ event $\bar{\mathcal{E}}_{\text{Rx},(b)}$ is equivalent to the event that $\mathcal{F}_{\text{Rx},(b)}(w_{1,c}, w_{2,c}, w_{1,p}, w_{2,p}, j_1, j_2)$ *is not* satisfied for the tuple $(w_{1,c}, w_{2,c}, w_{1,p}, w_{2,p}, j_1, j_2) = (W_{1,c,(b-1)}, W_{2,c,(b-1)},\ W_{1,p,(b)}, W_{2,p,(b)}, J_{1,(b-1)}^*, J_{2,(b-1)}^*)$ or it *is* satisfied for some tuple $(w_{1,c}, w_{2,c}, w_{1,p}, w_{2,p}, j_1, j_2) \neq (W_{1,c,(b-1)}, W_{2,c,(b-1)},\ W_{1,p,(b)}, W_{2,p,(b)}, J_{1,(b-1)}^*, J_{2,(b-1)}^*)$. Similarly for events $\bar{\mathcal{E}}_{\text{Rx},(1)}$ and $\bar{\mathcal{E}}_{\text{Rx},(B+1)}$. Thus, for $b \in \{2, \ldots, B\}$, the sequence of (in)equalities (A.110) holds,

$$\begin{aligned}& \Pr\left(\mathcal{E}_{\text{Rx},(b)} \,\middle|\, \bigcup_{b'=1}^{B+1} \{\bar{\mathcal{E}}_{\text{Tx},1,(b')},\ \bar{\mathcal{E}}_{\text{Tx},2,(b')}\}\right) \\ & = \Pr\Bigg(\Bigg(\bigcup_{\substack{(w_{1,c},w_{2,c},w_{1,p},w_{2,p},j_1,j_2)\neq \\ (W_{1,c,(b-1)},W_{2,c,(b-1)},W_{1,p,(b)},W_{2,p,(b)},J_{1,b-1}^*,J_{2,(b-1)}^*)}} \mathcal{F}_{\text{Rx},(b)}(w_{1,c}, w_{2,c}, w_{1,p}, w_{2,p}, j_1, j_2)\Bigg) \\ & \qquad \cup\ \mathcal{F}_{\text{Rx},(b)}\left(W_{1,c,(b-1)}, W_{2,c,(b-1)}, W_{1,p,(b)}, W_{2,p,(b)}, J_{1,b-1}^*, J_{2,(b-1)}^*\right) \\ & \qquad\qquad \Bigg| \bigcup_{b'=1}^{B+1} \{\bar{\mathcal{E}}_{\text{Tx},1,(b')},\ \bar{\mathcal{E}}_{\text{Tx},2,(b')}\}\Bigg) \qquad \text{(A.110a)} \\ & \leq \sum_{\substack{(w_{1,c},w_{2,c},w_{1,p},w_{2,p},j_1,j_2)\neq \\ (W_{1,c,(b-1)},W_{2,c,(b-1)},W_{1,p,(b)}, \\ W_{2,p,(b)},J_{1,b-1}^*,J_{2,(b-1)}^*)}} \Pr\left(\mathcal{F}_{\text{Rx},(b)}(w_{1,c}, w_{2,c}, w_{1,p}, w_{2,p}, j_1, j_2)\right) \Bigg| \bigcup_{b'=1}^{B+1} \{\bar{\mathcal{E}}_{\text{Tx},1,(b')},\ \bar{\mathcal{E}}_{\text{Tx},2,(b')}\}\Bigg)\end{aligned}$$

$$+ \quad \Pr\left(\mathcal{F}_{\mathrm{Rx},(b)}\left(W_{1,c,(b-1)}, W_{2,c,(b-1)}, W_{1,p,(b)}, W_{2,p,(b)}, J^*_{1,b-1}, J^*_{2,(b-1)}\right) \middle| \bigcup_{b'=1}^{B+1} \left\{\bar{\mathcal{E}}_{\mathrm{Tx},1,(b')},\ \bar{\mathcal{E}}_{\mathrm{Tx},2,(b')}\right\}\right) \tag{A.110b}$$

where the inequalities hold by the union bound.

By the event in the conditioning and the way we construct the codebooks, and by the weak law of large numbers and the Covering Lemma, both summands tend to 0 as $N \to \infty$ if (5.25) hold.

The scheme satisfies the distortion constraints (6.3b) because of (5.25j) and by the weak law of large numbers.

A.3.2 Fourier-Motzkin Elimination

We apply the Fourier-Motzkin Elimination Algorithm to show that Constraints (5.25) are equivalent to the constraints in Theorem 17. For ease of notation, define

$$I_0 := I(V_1; X_1X_2Y \mid \underline{U}) + I(V_2; X_1X_2YV_1 \mid \underline{U}) \tag{A.111a}$$

$$I_1 := I(V_1; X_1Z_1 \mid \underline{U}) \tag{A.111b}$$

$$I_2 := I(V_2; X_2Z_2 \mid \underline{U}) \tag{A.111c}$$

$$I_3 := I(U_1; X_2Z_2 \mid U_0U_2) \tag{A.111d}$$

$$I_4 := I(U_2; X_1Z_1 \mid U_0U_1) \tag{A.111e}$$

$$I_5 := I(V_1; X_2Z_2 \mid \underline{U}) \tag{A.111f}$$

$$I_6 := I(V_2; X_1Z_1 \mid \underline{U}) \tag{A.111g}$$

$$I_7 := I(X_1X_2; YV_1V_2 \mid \underline{U}) \tag{A.111h}$$

$$I_8 := I(X_1; YV_1V_2 \mid \underline{U}X_2) \tag{A.111i}$$

$$I_9 := I(X_2; YV_1V_2 \mid \underline{U}X_1) \tag{A.111j}$$

$$I_{10} := I(X_1; Y \mid U_0X_2) \tag{A.111k}$$

$$I_{11} := I(X_2; Y \mid U_0X_1) \tag{A.111l}$$

$$I_{12} := I(X_1X_2; Y \mid U_0U_2) \tag{A.111m}$$

$$I_{13} := I(X_1X_2; Y \mid U_0U_1) \tag{A.111n}$$

$$I_{14} := I(X_1X_2; Y \mid U_0) \tag{A.111o}$$

$$I_{15} := I(X_1X_2; Y). \tag{A.111p}$$

Setting $R_{k,c} = R_k - R_{k,p}$, with above definitions we can rewrite Constraints (5.25) as:

$$R_{1,v} > I_1 \tag{A.112a}$$

$$R_{2,v} > I_2 \tag{A.112b}$$

$$R_{2,v}+R_1 - R_{1,p} < I_2 + I_3 \tag{A.112c}$$

$$R_{1,v}+R_2 - R_{2,p} < I_1 + I_4 \tag{A.112d}$$

$$R_{1,v} + R_{2,v}+R_1 - R_{1,p} < I_2 + I_3 + I_5 \tag{A.112e}$$

$$R_{1,v}+R_{2,v} + R_2 - R_{2,p} < I_1 + I_4 + I_6 \tag{A.112f}$$

$$R_{1,p} + R_{2,p} < I_7 \tag{A.112g}$$

$$R_{1,p} < I_8 \tag{A.112h}$$

$$R_{2,p} < I_9 \tag{A.112i}$$

$$R_{1,v} + R_{1,p} < I_{10} + I_0 \tag{A.112j}$$

$$R_{2,v} + R_{2,p} < I_{11} + I_0 \tag{A.112k}$$

$$R_{1,v} + R_{1,p} + R_{2,p} < I_{12} + I_0 \tag{A.112l}$$

$$R_{2,v} + R_{1,p} + R_{2,p} < I_{13} + I_0 \tag{A.112m}$$

$$R_{1,v} + R_{1,p} + R_{2,v} + R_{2,p} < I_{14} + I_0 \tag{A.112n}$$

$$R_{1,v} + R_1 + R_{2,v} + R_2 < I_{15} + I_0. \tag{A.112o}$$

In a next step we eliminate the variables $R_{1,v}$ and $R_{2,v}$ to obtain:

$$R_1 - R_{1,p} < I_3 \tag{A.113a}$$

$$R_2 - R_{2,p} < I_4 \tag{A.113b}$$

$$R_1 - R_{1,p} < I_3 + I_5 - I_1 \tag{A.113c}$$

$$R_2 - R_{2,p} < I_4 + I_6 - I_2 \tag{A.113d}$$

$$R_{1,p} < \min\{I_8, I_{10} + I_0 - I_1\} \tag{A.113e}$$

$$R_{2,p} < \min\{I_9, I_{11} + I_0 - I_2\} \tag{A.113f}$$

$$\begin{aligned} R_{1,p} + R_{2,p} < \min\{&I_7, I_{12} + I_0 - I_1, \\ &I_{13} + I_0 - I_2, I_{14} + I_0 - I_1 - I_2\} \end{aligned} \tag{A.113g}$$

$$R_1 + R_2 < I_{15} + I_0 - I_1 - I_2 \tag{A.113h}$$

Notice that $I_1 \geq I_5$ and $I_2 \geq I_6$ because $V_1 - (Z_1 X_1 \underline{U}) - (X_2 Z_2)$ form a Markov chain, and thus Constraints (A.113a) and (A.113b) are inactive in view of Constraints (A.113c) and (A.113d). We thus neglect (A.113a) and (A.113b) in the following. Eliminating next variable $R_{1,p}$, where we take into account the nonnegativity of $R_{1,p}$ and $R_1 - R_{1,p}$, we obtain:

$$R_1 < I_3 + I_5 - I_1 + \min\{I_8, I_{10} + I_0 - I_1\} \tag{A.114a}$$

$$\begin{aligned} R_1 + R_{2,p} < I_3 + I_5 - I_1 + \min\{&I_7, I_{12} + I_0 - I_1, \\ &I_{13} + I_0 - I_2, I_{14} + I_0 - I_1 - I_2\} \end{aligned} \tag{A.114b}$$

$$R_2 - R_{2,p} < I_4 + I_6 - I_2 \tag{A.114c}$$

$$R_{2,p} < \min\{I_9, I_{11} + I_0 - I_2\} \tag{A.114d}$$

$$\begin{aligned} R_{2,p} < \min\{&I_7, I_{12} + I_0 - I_1, \\ &I_{13} + I_0 - I_2, I_{14} + I_0 - I_1 - I_2\} \end{aligned} \tag{A.114e}$$

$$R_1 + R_2 < I_{15} + I_0 - I_1 - I_2 \tag{A.114f}$$

and

$$I_3 + I_5 > I_1 \tag{A.114g}$$

$$I_{10} + I_0 > I_1. \tag{A.114h}$$

Notice that $I_7 > I_9$ and $I_{13} > I_{11}$ and therefore the two Constraints (A.114d) and (A.114e) combine to

$$\begin{aligned} R_{2,p} < \min\{&I_9, I_{11} + I_0 - I_2, \\ &I_{12} + I_0 - I_1, I_{14} + I_0 - I_1 - I_2\}. \end{aligned} \tag{A.115}$$

Eliminating finally $R_{2,p}$ (while taking into account the nonnegativity of $R_{2,p}$ and $R_2 - R_{2,p}$) results in:

$$R_1 < I_3 + I_5 - I_1 + \min\{I_8,\ I_{10} + I_0 - I_1\} \tag{A.116a}$$

$$\begin{aligned} R_1 < I_3 + I_5 - I_1 + \min\{&I_7, I_{12} + I_0 - I_1, \\ &I_{13} + I_0 - I_2, I_{14} + I_0 - I_1 - I_2\} \end{aligned} \tag{A.116b}$$

$$R_2 < I_4 + I_6 - I_2 + \min\{I_9,\ I_{11} + I_0 - I_2$$
$$I_{12} + I_0 - I_1, I_{14} + I_0 - I_1 - I_2\} \tag{A.116c}$$
$$R_1 + R_2 < I_4 + I_6 - I_2 + I_3 + I_5 - I_1$$
$$+ \min\{I_7,\ I_{12} + I_0 - I_1,$$
$$I_{13} + I_0 - I_2, I_{14} + I_0 - I_1 - I_2\} \tag{A.116d}$$
$$R_1 + R_2 < I_{15} + I_0 - I_1 - I_2 \tag{A.116e}$$

and

$$I_3 + I_5 > I_1 \tag{A.116f}$$
$$I_4 + I_6 > I_2 \tag{A.116g}$$
$$I_{14} + I_0 > I_1 + I_2 \tag{A.116h}$$
$$I_{10} + I_0 > I_1 \tag{A.116i}$$
$$I_{11} + I_0 > I_2 \tag{A.116j}$$
$$I_{12} + I_0 > I_1. \tag{A.116k}$$

Notice that $I_{12} > I_{10}$ and thus (4.42) is obsolete in view of (A.116i). Moreover, since also $I_7 > I_8$, Constraints (A.116a) and (A.116b) combine to

$$R_1 < I_3 + I_5 - I_1 + \min\{I_8,\ I_{10} + I_0 - I_1,$$
$$I_{13} + I_0 - I_2, I_{14} + I_0 - I_1 - I_2\}. \tag{A.117}$$

The final expression is thus given by constraints:

$$R_1 < I_3 + I_5 - I_1 + \min\{I_8,\ I_{10} + I_0 - I_1$$
$$I_{13} + I_0 - I_2, I_{14} + I_0 - I_1 - I_2\} \tag{A.118a}$$
$$R_2 < I_4 + I_6 - I_2 + \min\{I_9,\ I_{11} + I_0 - I_2$$
$$I_{12} + I_0 - I_1, I_{14} + I_0 - I_1 - I_2\} \tag{A.118b}$$
$$R_1 + R_2 < I_4 + I_6 - I_2 + I_3 + I_5 - I_1$$
$$+ \min\{I_7,\ I_{12} + I_0 - I_1,$$

$$I_{13} + I_0 - I_2, I_{14} + I_0 - I_1 - I_2\} \tag{A.118c}$$

$$R_1 + R_2 < I_{15} + I_0 - I_1 - I_2 \tag{A.118d}$$

and

$$I_3 + I_5 > I_1 \tag{A.118e}$$

$$I_4 + I_6 > I_2 \tag{A.118f}$$

$$I_{14} + I_0 > I_1 + I_2 \tag{A.118g}$$

$$I_{10} + I_0 > I_1 \tag{A.118h}$$

$$I_{11} + I_0 > I_2. \tag{A.118i}$$

A.4 Proof for Example 2

In this part, we explicitly evaluate Theorem 17 and Corollary 16 for Example 2. To omit the unnecessary complexity, we assume Rx observes the output and states, so we replace Y by $Y = (Y', S_1, S_2)$.

A.4.1 Distortion 2 in Example 2

We compute the distortion directly by choosing the state which makes minimum distortion at each case mentioned in the proof of Example 2. The detailed steps of the distortion computation is as follows:

$$\begin{aligned}
D_2^{(5.32)} &= \sum_{\mathcal{S}_2,\mathcal{U}_1,\mathcal{X}_2,\mathcal{Z}_2,\mathcal{V}_1} P_{S_2,U_1X_2Z_2V_1}(s_2,u_1,x_2,z_2,v_1)d(\hat{s}_2,s_2)\\
&= \Pr[X_2=0]\min\{p_s,\bar{p}_s\}\\
&+ \min\Big\{\Pr[S_2=0,U_1=0,X_2=1,\underbrace{Z_2=1,V_1=1}_{Y'=1,B=0}],\Pr[S_2=1,U_1=0,X_2=1,Z_2=1,V_1=1]\Big\}\\
&+ \min\Big\{\Pr[S_2=0,U_1=1,X_2=1,Z_2=1,V_1=1],\Pr[S_2=1,U_1=1,X_2=1,Z_2=1,V_1=1]\Big\}\\
&+ \min\Big\{\Pr[S_2=0,U_1=0,X_2=1,Z_2=2,V_1=1],\Pr[S_2=1,U_1=0,X_2=1,Z_2=2,V_1=1]\Big\}\\
&+ \min\Big\{\Pr[S_2=0,U_1=1,X_2=1,Z_2=2,V_1=1],\Pr[S_2=1,U_1=1,X_2=1,Z_2=2,V_1=1]\Big\}\\
&= \Pr[X_2=0]\min\{p_s,\bar{p}_s\}
\end{aligned}$$

$$
\begin{aligned}
&+ \min\Big\{P_{S_2}(0)P_{U_1,X_2}(0,1)\underbrace{P_{Y'|S_2U_1X_2}(1\mid 0,0,1)}_{P_{S_1}(1)P_{X_1|U_1}(1|0)}\underbrace{P_{Z_2|Y'}(1\mid 1)}_{P_B(0)}, \\
&\qquad\qquad P_{S_2}(1)P_{U_1,X_2}(0,1)\underbrace{P_{Y'|S_2U_1X_2}(1\mid 1,0,1)}_{1-P_{S_1}(1)P_{X_1|U_1}(1|0)}\underbrace{P_{Z_2|Y'}(1\mid 1)}_{P_B(0)}\Big\} \\
&+ \min\Big\{P_{S_2}(0)P_{U_1,X_2}(1,1)\underbrace{P_{Y'|S_2U_1X_2}(1\mid 0,1,1)}_{P_{S_1}(1)P_{X_1|U_1}(1|1)}\underbrace{P_{Z_2|Y'}(1\mid 1)}_{P_B(0)} \\
&\qquad\qquad P_{S_2}(1)P_{U_1,X_2}(1,1)\underbrace{P_{Y'|S_2U_1X_2}(1\mid 1,1,1)}_{1-P_{S_1}(1)P_{X_1|U_1}(1|1)}\underbrace{P_{Z_2|Y'}(1\mid 1)}_{P_B(0)}\Big\} \\
&+ \min\Big\{P_{S_2}(0)P_{U_1,X_2}(0,1)\underbrace{P_{Y'|S_2U_1X_2}(1\mid 0,0,1)}_{P_{S_1}(1)P_{X_1|U_1}(1|0)}\underbrace{P_{Z_2|Y'}(2\mid 1)}_{P_B(1)}, \\
&\qquad\qquad P_{S_2}(1)P_{U_1,X_2}(0,1)\underbrace{P_{Y'|S_2U_1X_2}(1\mid 1,0,1)}_{1-P_{S_1}(1)P_{X_1|U_1}(1|0)}\underbrace{P_{Z_2|Y'}(2\mid 1)}_{P_B(1)}\Big\} \\
&+ \min\Big\{P_{S_2}(0)P_{U_1,X_2}(1,1)\underbrace{P_{Y'|S_2U_1X_2}(1\mid 0,1,1)}_{P_{S_1}(1)P_{X_1|U_1}(1|1)}\underbrace{P_{Z_2|Y'}(2\mid 1)}_{P_B(1)}, \\
&\qquad\qquad P_{S_2}(1)P_{U_1,X_2}(1,1)\underbrace{P_{Y'|S_2U_1X_2}(1\mid 1,1,1)}_{1-P_{S_1}(1)P_{X_1|U_1}(1|1)}\underbrace{P_{Z_2|Y'}(2\mid 1)}_{P_B(1)}\Big\}
\end{aligned} \tag{A.119}
$$

which simplifies to (5.34)

$$
\begin{aligned}
D_2^{(5.32)} &= P_{X_2}(0)\min\{p_s,\bar{p}_s\} \\
&\quad + P_{U_1X_2}(0,1)\cdot\min\Big\{\bar{p}_s\cdot p_s r_1\bar{t}_1\bar{t}_2 + (1-p_s r_1)t_1t_2,\quad p_s\cdot(1-p_s r_1)\bar{t}_1\bar{t}_2\Big\}
\end{aligned} \tag{A.120}
$$

$$
\quad + P_{U_1X_2}(1,1)\cdot\min\Big\{\bar{p}_s\cdot p_s\bar{r}_1\bar{t}_1\bar{t}_2 + (1-p_s\bar{r}_1)t_1t_2,\quad p_s\cdot(1-p_s\bar{r}_1)\bar{t}_1\bar{t}_2\Big\} \tag{A.121}
$$

We compute the distortion directly according to the scheme introduced in Corollary (16) without using r.v V_1:

$$
\begin{aligned}
D_2^{\text{Corollary16}} &= \sum_{\mathcal{S}_2,\mathcal{U}_1,\mathcal{X}_2,\mathcal{Z}_2} P_{S_2,U_1X_2Z_2}(s_2,u_1,x_2,z_2)d(\hat{s}_2,s_2) \\
&= P_{X_2}(0)\min\{P_{S_2}(1),P_{S_2}(0)\} \\
&\quad + \min\{\Pr[S_2=0,X_2=1,U_1=0,Z_2=1],\Pr[S_2=1,X_2=1,U_1=0,Z_2=1]\} \\
&\quad + \min\{\Pr[S_2=0,X_2=1,U_1=1,Z_2=1],\Pr[S_2=1,X_2=1,U_1=1,Z_2=1]\} \\
&\quad + \min\{\Pr[S_2=0,X_2=1,U_1=0,Z_2=2],\Pr[S_2=1,X_2=1,U_1=0,Z_2=2]\}
\end{aligned}
$$

$$
\begin{aligned}
&\quad + \min\{\Pr[S_2 = 0, X_2 = 1, U_1 = 1, Z_2 = 2], \Pr[S_2 = 1, X_2 = 1, U_1 = 1, Z_2 = 2]\}\\
&= P_{U_1X_2}(0,1)\min\{\bar{p}_s(p_s r_1\bar{t} + t(1-p_s r_1)), \bar{t}(1-p_s r_1)p_s\}\\
&+ P_{U_1X_2}(1,1)\min\{\bar{p}_s(p_s \bar{r}_1\bar{t} + t(1-p_s \bar{r}_1)), \bar{t}(1-p_s \bar{r}_1)p_s\}\\
&+ P_{U_1X_2}(0,1)\min\{p_s(p_s r_1\bar{t} + t(1-p_s r_1))\}\\
&+ P_{U_1X_2}(1,1)\min\{p_s(p_s \bar{r}_1\bar{t} + t(1-p_s \bar{r}_1))\}\\
&= P_{U_1X_2}(0,1)\Big\{ \min\{\bar{p}_s(p_s r_1\bar{t} + t(1-p_s r_1)), \bar{t}(1-p_s r_1)p_s\} \qquad (A.122)\\
&\qquad\qquad + \min\{p_s(p_s r_1\bar{t} + t(1-p_s r_1))\}\Big\}\\
&+ P_{U_1X_2}(1,1)\Big\{ \min\{\bar{p}_s(p_s \bar{r}_1\bar{t} + t(1-p_s \bar{r}_1)), \bar{t}(1-p_s \bar{r}_1)p_s\} \qquad (A.123)\\
&\qquad\qquad + \min\{p_s(p_s \bar{r}_1\bar{t} + t(1-p_s \bar{r}_1))\}\Big\}
\end{aligned}
$$

A.4.2 Computing Rate Constraints for Example 2

As we mentioned after (5.32), we need to introduce E_k with probability p_{e_k} and thus, replace V_k by E_k, V_k in terms (A.111a)-(A.111j). By using E_k, we regulate how much we sacrifice the rate of Tx_k to reduce $\mathrm{Tx}_{\bar{k}}$'s distortion, otherwise we just achieve the minimum distortion.

To compute I_0

The evaluation of the mutual information term (A.111a) is detailed out as follows:

$$
\begin{aligned}
I_0 &:= I(V_1E_1; X_1X_2Y \mid \underline{U}) + I(V_2E_2; X_1X_2YV_1E_1 \mid \underline{U})\\
&= \bar{p}_{e_1}\Big[H(V_1 \mid U_1U_2) - \underbrace{H(V_1 \mid X_1X_2Y'S_1S_2)}_{H(B_1|B_0)}\Big] + \bar{p}_{e_2}\Big[H(V_2 \mid U_1U_2) - \underbrace{H(V_2 \mid X_1X_2Y'S_1S_2)}_{H(B_2|B_0)}\Big]\\
&= \bar{p}_{e_1}\Big[P_{U_1U_2}(0,0)f_0(r_1,r_2,t_1) + P_{U_1U_2}(1,0)f_0(\bar{r}_1,r_2,t_1) + P_{U_1U_2}(0,1)f_0(r_1,\bar{r}_2,t_1)\\
&\qquad + P_{U_1U_2}(1,1)f_0(\bar{r}_1,\bar{r}_2,t_1) - H(t_1,\bar{t}_1)\Big]\\
&\quad \bar{p}_{e_2}\Big[P_{U_1U_2}(0,0)f_0(r_2,r_1,t_2) + P_{U_1U_2}(1,0)f_0(\bar{r}_2,r_1,t_2) + P_{U_1U_2}(0,1)f_0(r_2,\bar{r}_1,t_2)\\
&\qquad + P_{U_1U_2}(1,1)f_0(\bar{r}_2,\bar{r}_1,t_2) - H(t_2,\bar{t}_2)\Big] \qquad (A.124)
\end{aligned}
$$

where

$$
\begin{aligned}
f_0(r_1, r_2, t) \triangleq H\Bigg(& (1-p_s r_1)(1-p_s r_2)\bar{t} + p_s r_1 p_s r_2 t, \\
& p_s r_1(1-p_s r_2)\bar{t} + (1-p_s r_1)p_s r_2 \bar{t} + (1-p_s r_1)(1-p_s r_2)t, \\
& p_s r_1 p_s r_2 \bar{t} + p_s r_1(1-p_s r_2)t + (1-p_s r_1)p_s r_2 t\Bigg),
\end{aligned} \tag{A.125}
$$

in which for each pair of (U_1, U_2), we find the conditional probability of $P_{V_1|U_1U_2}(v_1 \mid u_1, u_2)$:

$$
\begin{aligned}
P_{V_1|U_1U_2}(0 \mid 0,0) &= \underbrace{P_{Z_1|U_1U_2}(0 \mid 0,0)}_{\Pr[S_1X_1=0 \,\&\, S_2X_2=0 \,\&\, B_1=0|U_1=U_2=0]} + \underbrace{P_{Z_1|U_1U_2}(3 \mid 0,0)}_{\Pr[S_1X_1=1 \,\&\, S_2X_2=1 \,\&\, B_1=1|U_1=U_2=0]} \\
&= \big(1-\Pr[S_1X_1=1 \mid U_1=0]\big)\big(1-\Pr[S_2X_2=1 \mid U_2=0]\big)P_{B_1}(0) \\
&+ \Pr[S_1X_1=1 \mid U_1=0]\Pr[S_2X_2=1 \mid U_2=0]P_{B_1}(1) \\
&= (1-p_s r_1)(1-p_s r_2)\bar{t}_1 + p_s r_1 p_s r_2 t_1, && \text{(A.126a)} \\
P_{V_1|U_1U_2}(1 \mid 0,0) &= \Pr\Big[S_1X_1 + S_2X_2 + B_1 = 1 \mid U_1 = U_2 = 0\Big] \\
&= \Pr[S_1X_1=1 \mid U_1=0]\big(1-\Pr[S_2X_2=1 \mid U_2=0]\big)P_{B_1}(0) \\
&+ \big(1-\Pr[S_1X_1=1 \mid U_2=0]\big)\Pr[S_2X_2=1 \mid U_1=0]P_{B_1}(0) \\
&+ \big(1-\Pr[S_1X_1=1 \mid U_1=0]\big)\big(1-\Pr[S_2X_2=1 \mid U_2=0]\big)P_{B_1}(1) \\
&= p_s r_1(1-p_s r_2)\bar{t}_1 + (1-p_s r_1)p_s r_2 \bar{t}_1 + (1-p_s r_1)(1-p_s r_2)t_1, && \text{(A.126b)} \\
P_{V_1|U_1U_2}(2 \mid 0,0) &= \Pr\Big[S_1X_1 + S_2X_2 + B_1 = 2 \mid U_1 = U_2 = 0\Big] \\
&= \Pr[S_1X_1=1 \mid U_1=0]\Pr[S_2X_2=1 \mid U_2=0]P_{B_1}(0) \\
&+ \Pr[S_1X_1=1 \mid U_1=0](1-\Pr[S_2X_2=1 \mid U_2=0])P_{B_1}(1) \\
&+ (1-\Pr[S_1X_1=1 \mid U_1=0])\Pr[S_2X_2=1 \mid U_2=0]P_{B_1}(1) \\
&= p_s r_1 p_s r_2 \bar{t}_1 + p_s r_1(1-p_s r_2)t_1 + (1-p_s r_1)p_s r_2 t_1. && \text{(A.126c)}
\end{aligned}
$$

To compute I_1

The mutual information term in (A.111b) is evaluated in following steps:

$$I_1 := I(E_1V_1; X_1Z_1 \mid \underline{U})$$

$$= \bar{p}_{e_1}\left[H(V_1 \mid U_2X_1) - \underbrace{H(V_1 \mid U_2X_1Z_1)}_{0}\right]$$

$$= \bar{p}_{e_1}\Bigg[P_{U_2X_1}(0,0)f_1(r_2,t_1) + P_{U_2X_1}(1,0)f_1(\bar{r}_2,t_1)$$

$$+P_{U_2X_1}(0,1)G_1(r_2,t_1) + P_{U_2X_1}(1,1)G_1(\bar{r}_2,t_1)\Bigg] \quad \text{(A.127)}$$

where

$$f_1(r_2,t_1) \triangleq H\Bigg((1-p_sr_2)\bar{t}_1, \quad p_sr_2\bar{t}_1 + (1-p_sr_2)t_1, \quad p_sr_2t_1\Bigg) \quad \text{(A.128)}$$

and

$$G_1(r_2,t_1) \triangleq H\Bigg(\bar{p}_s(1-p_sr_2)\bar{t}_1 + p_sp_sr_2t_1, \quad p_s(1-p_sr_2)\bar{t}_1 + \bar{p}_sp_sr_2\bar{t}_1 + \bar{p}_s(1-p_sr_2)t_1,$$

$$p_s^2r_2\bar{t}_1 + p_s(1-p_sr_2)t_1 + \bar{p}_sp_sr_2t_1\Bigg) \quad \text{(A.129)}$$

The function $f_1(r_2,t_1) \triangleq H(V_1 \mid u_2, x_1 = 0)$ is obtainable by calculating the conditional probability $P_{V_1|U_2X_1}$ as follows:

$$P_{V_1|U_2X_1}(0 \mid 0,0) = \underbrace{\Pr[S_2X_2 + B_1 = 0 \mid U_2 = 0, X_1 = 0]}_{\Pr[S_2X_2=0 \ \&\ B_1=0|U_2=0,X_1=0]} + \underbrace{\Pr[S_2X_2 + B_1 = 3 \mid U_2 = 0, X_1 = 0]}_{0}$$

$$= (1-p_sr_2)\bar{t}_1 \quad \text{(A.130a)}$$

$$P_{V_1|U_2X_1}(1 \mid 0,0) = \Pr[S_2X_2 + B_1 = 1 \mid U_2 = 0, X_1 = 0] = p_sr_2\bar{t}_1 + (1-p_sr_2)t_1 \quad \text{(A.130b)}$$

$$P_{V_1|U_2X_1}(2 \mid 0,0) = \Pr[S_2X_2 + B_1 = 2 \mid U_2 = 0, X_1 = 0] = p_sr_2t_1 \quad \text{(A.130c)}$$

and $G_1(r_2,t_1) \triangleq H(V_1 \mid U_2 = 0, X_1 = 1)$ is the conditional entropy function computable by following conditional probabilities:

$$P_{V_1|U_2X_1}(0 \mid 0,1) = \underbrace{\Pr[S_1 + S_2X_2 + B_1 = 0 \mid U_2 = 0, X_1 = 1]}_{\Pr[S_1=0 \ \&\ S_2X_2=0 \ \&\ B_1=0|U_2=0,X_1=1]} + \underbrace{\Pr[S_1 + S_2X_2 + B_1 = 3 \mid U_2 = 0, X_1 = 1]}_{\Pr[S_1=1 \ \&\ S_2X_2=1 \ \&\ B_1=1|U_2=0,X_1=1]}$$

$$= \bar{p}_s(1-p_sr_2)\bar{t}_1 + p_sp_sr_2t_1 \quad \text{(A.131a)}$$

$$
\begin{aligned}
P_{V_1|U_2X_1}(1 \mid 0,1) &= \Pr[S_1 + S_2X_2 + B_1 = 1 \mid U_2 = 0, X_1 = 1] \\
&= \Pr[S_1 = 1, S_2X_2 = 0, B_1 = 0 \mid U_2 = 0, X_1 = 1] \\
&\quad + \Pr[S_1 = 0, S_2X_2 = 1, B_1 = 0 \mid U_2 = 0, X_1 = 1] \\
&+ \Pr[S_1 = 0, S_2X_2 = 0, B_1 = 1 \mid U_2 = 0, X_1 = 1] \\
&= p_s(1 - p_sr_2)\bar{t}_1 + \bar{p}_sp_sr_2\bar{t}_1 + \bar{p}_s(1 - p_sr_2)t_1 \qquad \text{(A.131b)}
\end{aligned}
$$

$$
\begin{aligned}
P_{V_1|U_2X_1}(2 \mid 0,1) &= \Pr[S_1 + S_2X_2 + B_1 = 2 \mid U_2 = 0, X_1 = 1] \\
&= \Pr[S_1 = 1, S_2X_2 = 1, B_1 = 0 \mid U_2 = 0, X_1 = 1] \\
&\quad + \Pr[S_1 = 1, S_2X_2 = 0, B_1 = 1 \mid U_2 = 0, X_1 = 1] \\
&\quad + \Pr[S_1 = 0, S_2X_2 = 1, B_1 = 1 \mid U_2 = 0, X_1 = 1] \\
&= p_s^2r_2\bar{t}_1 + p_s(1 - p_sr_2)t_1 + \bar{p}_sp_sr_2t_1 \qquad \text{(A.131c)}
\end{aligned}
$$

To compute I_2

The mutual information term (A.111c) specializes to:

$$
\begin{aligned}
I_2 = \bar{p}_{e_2}\Bigg[&P_{U_1X_2}(0,0)f_1(r_1,t_2) + P_{U_1X_2}(1,0)f_1(\bar{r}_1,t_2) \\
&+P_{U_1X_2}(0,1)G_1(r_1,t_2) + P_{U_1X_2}(1,1)G_1(\bar{r}_2,t_2)\Bigg] \qquad \text{(A.132)}
\end{aligned}
$$

To Compute I_3

The rate constraint (A.111d) specialize to as detailed out in the following:

$$
\begin{aligned}
I_3 = &P_{U_0X_2}(0,0)f_3(k_1) + P_{U_0X_2}(1,0)f_3(\bar{k}_1) + P_{U_0X_2}(0,1)G_3(k_1) + P_{U_0X_2}(1,1)G_3(\bar{k}_1) \\
&-P_{U_1X_2}(0,0)f_3(\bar{r}_1) - P_{U_1X_2}(1,0)f_3(r_1) - P_{U_1X_2}(0,1)G_3(\bar{r}_1) - P_{U_1X_2}(1,1)G_3(r_1) \quad \text{(A.133a)}
\end{aligned}
$$

where

$$
f_3(k_1) \triangleq H\bigg((1 - p_s\bar{k}_1)\bar{t}_2, \quad (1 - p_s\bar{k}_1)t_2 + p_s\bar{k}_1\bar{t}_2, \quad p_s\bar{k}_1t_2\bigg) \qquad \text{(A.133b)}
$$

$$G_3(k) \triangleq H\bigg((1-p_s\bar{k})p_s\bar{t}_2, \quad p_s\bar{k}\bar{p}_s\bar{t}_2 + (1-p_s\bar{k})[p_s\bar{t}_2 + \bar{p}_s\bar{t}_2], \\ (1-p_s\bar{k})p_st_2 + p_s\bar{k}\bar{p}_st_2 + p_s\bar{k}p_s\bar{t}_2, \quad p_s\bar{k}p_st_2\bigg) \tag{A.133c}$$

in which we have computed the following conditional entropies:

- To compute $H(Z_2 \mid U_0X_2)$, we need the related conditional probabilities
  - First, we start by fixing $(U_0 = 0, X_2 = 0)$

$$\begin{aligned}\Pr\Big[Z_2 = 0 \mid U_0 = 0, X_2 = 0\Big] &= \Pr\Big[S_1X_1 + B_2 = 0 \mid U_0 = 0, X_2 = 0\Big] \\ &= \Big(1 - p_s(1)P_{X_1|U_0}(1 \mid 0)\Big)P_{B_2}(0) \\ &= (1-p_s\bar{k}_1)\bar{t}_2 \end{aligned} \tag{A.134a}$$

$$\Pr\Big[Z_2 = 1 \mid U_0 = 0, X_2 = 0\Big] = (1-p_s\bar{k}_1)t_2 + p_s\bar{k}_1\bar{t}_2 \tag{A.134b}$$

$$\Pr\Big[Z_2 = 2 \mid U_0 = 0, X_2 = 0\Big] = p_s\bar{k}_1t_2 \tag{A.134c}$$

  - Second, we fix $X_2 = 1$ and calculate the conditional probabilities similarly:

$$\begin{aligned}\Pr\Big[Z_2 = 0 \mid U_0 = 0, X_2 = 1\Big] &= \Pr\Big[S_1X_1 + S_2 + B_2 = 0 \mid U_0 = 0\Big] \\ &= \Big(1 - P_{S_1}(1)P_{X_1|U_0}(1 \mid 0)\Big)P_{S_2}(0)P_{B_2}(0) \\ &= (1-p_s\bar{k}_1)p_s\bar{t}_2 \end{aligned} \tag{A.135}$$

$$\begin{aligned}\Pr\Big[S_1X_1 + S_2 + B_2 = 1 \mid U_0 = 0, X_1 = 1\Big] &= \big(P_{S_1}(1)P_{X_1|U_0}(1 \mid 0)\big)P_{S_2}(0)P_{B_2}(0) \\ &\quad + \Big(1 - P_{S_1}(1)P_{X_1|U_0}(1 \mid 0)\Big)P_{S_2}(1)P_{B_2}(0) \\ &\quad + \Big(1 - P_{S_1}(1)P_{X_1|U_0}(1 \mid 0)\Big)P_{S_2}(0)P_{B_2}(1) \\ &= p_s\bar{k}_1\bar{p}_s\bar{t}_2 + (1-p_s\bar{k}_1)[p_s\bar{t}_2 + \bar{p}_s\bar{t}_2] \end{aligned} \tag{A.136}$$

$$\begin{aligned}\Pr\Big[S_1X_1 + S_2 + B = 2 \mid U_0 = 0, X_1 = 1\Big] &= \Big(1 - P_{S_1}(1)P_{X_1|U_0}(1 \mid 0)\Big)P_{S_2}(1)P_{B_2}(1) \\ &\quad + \big(P_{S_1}(1)P_{X_1|U_0}(1 \mid 0)\big)P_{S_2}(0)P_{B_2}(1) \\ &\quad + \big(P_{S_1}(1)P_{X_1|U_0}(1 \mid 0)\big)P_{S_2}(1)P_{B_2}(0) \\ &= (1-p_s\bar{k}_1)p_st_2 + p_s\bar{k}_1\bar{p}_st_2 + p_s\bar{k}_1p_s\bar{t}_2 \end{aligned} \tag{A.137}$$

$$\begin{aligned}\Pr\Big[S_1X_1 + S_2 + B = 3 \mid U_0 = 0, X_2 = 1\Big] &= P_{S_1}(1)P_{X_1|U_0}(1 \mid 0)P_{S_2}(1)P_{B_2}(1) \\ &= p_s\bar{k}_1p_st_2 \end{aligned} \tag{A.138}$$

– In an analogous way, we have following conditional probabilities for $H(Z_1 \mid U_1 X_2)$:

$$\begin{aligned}\Pr\Big[Z_2 = 0 \mid U_1 = 0, X_2 = 0\Big] &= \Pr\Big[S_1 X_1 + B_2 = 0 \mid U_1 = 0, X_2 = 0\Big] \\ &= \Big(1 - P_{S_1}(1) P_{X_1|U_1}(1 \mid u_1)\Big) P_{B_2}(0) \\ &= (1 - p_s r_1)\bar{t}_2 \end{aligned} \tag{A.139}$$

$$\tag{A.140}$$

$$\begin{aligned}\Pr\Big[Z_2 = 1 \mid U_1 = 0, X_2 = 0\Big] &= \Pr\Big[S_1 X_1 + B_2 = 1 \mid U_1 = 0\Big] \\ &= (1 - p_s r_1) t_2 + p_s r_1 \bar{t}_2 \end{aligned} \tag{A.141}$$

$$\Pr\Big[Z_2 = 2 \mid U_1 = 0, X_2 = 0\Big] = \Pr\Big[S_1 X_1 + B_2 = 2 \mid U_1 = 0, X_2 = 0\Big] = p_s r_1 t_1 \tag{A.142}$$

To compute I_4

In an analogous way to (A.111d), the mutual information term (A.111e) specializes to:

$$\begin{aligned} I_4 = &P_{U_0 X_1}(0,0) f_4(k_1) + P_{U_0 X_1}(1,0) f_4(\bar{k}_1) + P_{U_0 X_1}(0,1) G_4(k_1) + P_{U_0 X_1}(1,1) G_4(\bar{k}_1) \\ &- P_{U_2 X_1}(0,0) f_4(\bar{r}_1) - P_{U_2 X_1}(1,0) f_4(r_1) - P_{U_2 X_1}(0,1) G_4(\bar{r}_1) - P_{U_2 X_1}(1,1) G_4(r_1), \end{aligned} \tag{A.143}$$

where

$$f_4(k) \triangleq H\Big((1 - p_s \bar{k})\bar{t}_2, \quad (1 - p_s \bar{k}) t_2 + p_s \bar{k} \bar{t}_2, \quad p_s \bar{k} t_2\Big) \tag{A.144}$$

$$\begin{aligned} G_4(r) \triangleq H\Big(&(1 - p_s \bar{r}) p_s \bar{t}_2, \quad p_s \bar{r} \bar{p}_s \bar{t}_2 + (1 - p_s \bar{r})[p_s \bar{t}_2 + \bar{p}_s \bar{t}_2], \\ &(1 - p_s \bar{r}) p_s t_2 + p_s \bar{r} \bar{p}_s t_2 + p_s \bar{r} p_s \bar{t}_2, \quad p_s \bar{r} p_s t_2\Big) \end{aligned} \tag{A.145}$$

To compute I_5

The final expression of the mutual information term (A.111f) specializes to:

$$\begin{aligned} I_5 := &H(V_1 \mid U_1 U_2) - P_{X_2 Z_2}(0,0) H(t_1, \bar{t}_1) - P_{X_2 Z_2}(1,3) H(t_1, \bar{t}_1) - P_{X_2 Z_2}(0,2) H(t_1, \bar{t}_1) \\ &- P_{U_1 X_2 Z_2}(0,0,1) G_5(r_1, t_1, t_2) - P_{U_1 X_2 Z_2}(1,0,1) G_5(\bar{r}_1, t_1, t_2) \\ &- P_{U_1 X_2 Z_2}(0,1,1) Q_5(r_1) - P_{U_1 X_2 Z_2}(1,1,1) Q_5(\bar{r}_1) \\ &- P_{U_1 X_2 Z_2}(0,1,2) f_5(r_1) - P_{U_1 X_2 Z_2}(1,1,2) f_5(\bar{r}_1) \end{aligned} \tag{A.146}$$

where

$$G_5(r_1,t_1,t_2) \triangleq H(V_1 \mid u_1=0,x_2=0,z_2=1)$$
$$= \Big(\frac{(1-p_sr_1)\bar{t}_1t_2}{(p_sr_1)\bar{t}_2+(1-p_sr_1)t_2},,\ \frac{p_sr_1\bar{t}_1\bar{t}_2+(1-p_sr_1)t_1t_2}{(p_sr_1)\bar{t}_2+(1-p_sr_1)t_2},\ \frac{p_sr_1t_1\bar{t}_2}{(p_sr_1)\bar{t}_2+(1-p_sr_1)t_2}\Big) \tag{A.147}$$

and we define three following functions:

$$Q_5(r_1,t_1,t_2) \triangleq H(V_1 \mid u_1=0,x_2=1,z_2=1) = \left(\frac{(1-p_sr_1)\bar{p}_s\bar{t}_1t_2}{p_sr_1\bar{p}_s\bar{t}_2+(1-p_sr_1)p_s\bar{t}_2+(1-p_sr_1)\bar{p}_st_2},\ \frac{(1-p_sr_1)\bar{p}_st_1t_2+(p_sr_1\bar{p}_s+(1-p_Sr_1)p_s)\bar{t}_1\bar{t}_2}{p_sr_1\bar{p}_s\bar{t}_2+(1-p_sr_1)p_s\bar{t}_2+(1-p_sr_1)\bar{p}_st_2},\ \frac{\big(p_sr_1\bar{p}_s+(1-p_sr_1)p_s\big)t_1\bar{t}_2}{p_sr_1\bar{p}_s\bar{t}_2+(1-p_sr_1)p_s\bar{t}_2+(1-p_sr_1)\bar{p}_st_2}\right) \tag{A.148}$$

and

$$f_5(r_1,t_1,t_2) \triangleq H(V_1 \mid u_1=0,x_2=1,z_2=2) = \left(\frac{p_sr_1t_1\bar{t}_2}{(p_sr_1\bar{p}_s+(1-p_sr_1)p_s)t_2+p_s^2r_1bart_2}\ \frac{\Big(p_sr_1\bar{p}_s+(1-p_sr_1)p_s\Big)\bar{t}_1t_2}{(p_sr_1\bar{p}_s+(1-p_sr_1)p_s)t_2+p_s^2r_1\bar{t}_2},\ \frac{p_s^2r_1\bar{t}_1\bar{t}_2+\Big(p_sr_1\bar{p}_s+(1-p_sr_1)p_s\Big)t_1t_2}{\Big(p_sr_1\bar{p}_s+(1-p_sr_1)p_s\Big)t_2+p_s^2r_1\bar{t}_2}\right) \tag{A.149}$$

Th detailed steps are as follows:

$$\begin{aligned}
I_5 := I(V_1;X_2Z_2 \mid \underline{U}) &= H(V_1 \mid U_1U_2) - H(V_1 \mid U_1X_2Z_2) = H(V_1 \mid U_1U_2)\\
&-P_{U_1X_2Z_2}(0,0,0)H(V_1 \mid u_1=0,x_2=0,z_2=0) - P_{U_1X_2Z_2}(1,0,0)H(V_1 \mid u_1=1,x_2=0,z_2=0)\\
&-\underbrace{P_{U_1X_2Z_2}(0,0,3)H(V_1 \mid u_1=0,x_2=0,z_2=3) - P_{U_1X_2Z_2}(1,0,3)H(V_1 \mid u_1=1,x_2=0,z_2=3)}_{0}\\
&-P_{U_1X_2Z_2}(0,1,3)H(V_1 \mid u_1=0,x_2=1,z_2=3) - P_{U_1X_2Z_2}(1,1,3)H(V_1 \mid u_1=1,x_2=1,z_2=3)\\
&-P_{U_1X_2Z_2}(0,0,1)H(V_1 \mid u_1=0,x_2=0,z_2=1) - P_{U_1X_2Z_2}(1,0,1)H(V_1 \mid u_1=1,x_2=0,z_2=1)\\
&-P_{U_1X_2Z_2}(0,1,1)H(V_1 \mid u_1=0,x_2=1,z_2=1) - P_{U_1X_2Z_2}(1,1,1)H(V_1 \mid u_1=1,x_2=1,z_2=1)\\
&-P_{U_1X_2Z_2}(0,1,2)H(V_1 \mid u_1=0,x_2=1,z_2=2) - P_{U_1X_2Z_2}(1,1,2)H(V_1 \mid u_1=1,x_2=1,z_2=2)
\end{aligned} \tag{A.150}$$

We have already calculated $H(V_1 \mid U_1U_2)$in (A.126a). To compute functions $G_5(r_1), Q_5(r_1), f_5(r_1)$ based on each observation of (U_1, X_2, Z_2), we have the steps as follows:

- For case $(U_1, X_2, Z_2) = (0, 0, 0)$, the conditional entropy is given by

$$H(V_1 \mid u_1 = 0, x_2 = 0, z_2 = 0) = H\Big(\bar{t}_1,\ t_1\Big) \tag{A.151}$$

  where one can compute it as follows:

  - $\Pr[V_1 = 0 \mid U_1 = 0, X_2 = 0, Z_2 = 0] = \Pr[Z_1 = 0 \mid Z_2 = 0] = \Pr[B_1 = 0 \mid B_2 = 0] = \bar{t}_1,$
  - $\Pr[V_1 = 1 \mid U_1 = 0, X_2 = 0, Z_2 = 0] = \Pr[Z_1 = 1 \mid Z_2 = 0] = \Pr[B_1 = 1 \mid B_2 = 0] = t_1,$
  - $\Pr[V_1 = 2 \mid U_1 = 0, X_2 = 0, Z_2 = 0] = 0,$

  The last step is concluded by assuming that the feedback noises are independent.

- For case $(U_1, X_2, Z_2) = (1, 0, 0)$, the steps are similar to previous case

$$H(V_1 \mid u_1 = 1, x_2 = 0, z_2 = 0) = H\Big(\bar{t}_1,\ t_1\Big) \tag{A.152}$$

- For case $(U_1, X_2, Z_2) = (0, 1, 3)$, we have

  - $\Pr[V_1 = 0 \mid U_1 = 0, X_2 = 1, Z_2 = 3] = \underbrace{\Pr[Z_1 = 0 \mid U_1 = 0, X_2 = 1, Z_2 = 3]}_{0} + \Pr[Z_1 = 3 \mid U_1 = 0, X_2 = 1, Z_2 = 3] = t_1$
  - $\Pr[V_1 = 1 \mid U_1 = 0, X_2 = 1, Z_2 = 3] = 0$, because $Z_2 = 3$ means that $S_1X_1 + S_2X_2 = 2$ which means $Z_1 \geq 2$
  - $\Pr[V_1 = 2 \mid U_1 = 0, X_2 = 1, Z_2 = 3] = \bar{t}_1$

- For case $(U_1, X_2, Z_2) = (1, 1, 3)$, we go through similar step as if we have $U_1 = 0$.

- For case $(U_1, X_2, Z_2) = (0, 0, 1)$, we have the following steps:

$$\begin{aligned}&\Pr\Big[V_1 = 0 \mid U_1 = 0, X_2 = 0, Z_2 = 1\Big] = \\ &\Pr\Big[Z_1 = 0 \mid U_1 = 0, X_2 = 0, Z_2 = 1\Big] + \underbrace{\Pr\Big[Z_1 = 3 \mid U_1 = 0, X_2 = 0, Z_2 = 1\Big]}_{0, \text{ because } Z_2 = 1,\ Z_2 \text{ can not be } 3}\end{aligned}$$

$$\begin{aligned}
&= \Pr[S_1X_1 + B_1 = 0 \mid U_1 = 0, X_2, Z_2 = S_1X_1 + B_2 = 1] \\
&= \frac{\Pr[S_1X_1 + B_1 = 0 \quad \& \quad S_1X_1 + B_2 = 1 \mid U_1 = 0]}{\Pr[S_1X_1 + B_2 = 1 \mid U_1 = 0]} \\
&= \frac{(1 - p_s r_1) P_{B_1B_2}(0,1)}{(p_s r_1) P_{B_2}(0) + (1 - p_s r_1) P_{B_2}(1)} \\
&= \frac{(1 - p_s r_1)\bar{t}_1 t_2}{(p_s r_1)\bar{t}_2 + (1 - p_s r_1) t_2},
\end{aligned} \tag{A.153}$$

$$\begin{aligned}
&\Pr\Big[V_1 = 1 \mid U_1 = 0, X_2 = 0, Z_2 = 1\Big] \\
&= \Pr\Big[S_1X_1 + B_1 = 1 \mid U_1 = 0, X_2 = 0, S_1X_1 + B_2 = 1\Big] \\
&= \frac{\Pr[S_1X_1 + B_1 = 1 \quad \& \quad S_1X_1 + B_2 = 1 \mid U_1 = 0]}{\Pr[S_1X_1 + B_1 = 1 \mid U_1 = 0]} \\
&= \frac{p_s P_{X_1|U_1}(1 \mid 0) P_{B_1B_2}(0,0) + (1 - p_s P_{X_1|U_1}(1 \mid 0)) P_{B_1B_2}(1,1)}{(p_s r_1) P_{B_2}(0) + (1 - p_s r_1) P_{B_2}(1)} \\
&= \frac{p_s r_1 \bar{t}_1 \bar{t}_2 + (1 - p_s r_1) t_1 t_2}{(p_s r_1)\bar{t}_2 + (1 - p_s r_1) t_2}
\end{aligned} \tag{A.154}$$

and

$$\begin{aligned}
&\Pr\Big[V_1 = 2 \mid U_1 = 0, X_2 = 0, Z_2 = 1\Big] \\
&= \Pr\Big[S_1X_1 + B_1 = 2 \mid U_1 = 0, X_2 = 0, S_1X_1 + B_2 = 1\Big] \\
&= \frac{\Pr\Big[S_1X_1 + B_1 = 2 \quad \& \quad S_1X_1 + B_2 = 1 \mid U_1 = 0\Big]}{Pr\Big[S_1X_1 + B_1 = 2 \mid U_1 = 0\Big]} \\
&= \frac{p_s P_{X_1|U_1}(1 \mid 0) P_{B_1B_2}(1,0)}{(p_s P_{X_1|U_1}(1 \mid 0)) P_{B_2}(0) + (1 - p_s P_{X_1|U_1}(1 \mid 0)) P_{B_2}(1)} \\
&= \frac{p_s r_1 t_1 \bar{t}_2}{(p_s r_1)\bar{t}_2 + (1 - p_s r_1) t_2}
\end{aligned} \tag{A.155}$$

- The computation for $(U_1, X_2, Z_2) = (1, 0, 1)$ is similar to the case in which $U_1 = 0$ and it is obtainable by replacing r_1 into $\bar{r}_1$ and vice versa.$(= G_5(\bar{r}_1))$
- For case $(U_1, X_2, Z_2) = (0, 1, 1)$, we have

$$H\Bigg(\frac{(1 - p_s r_1)\bar{p}_s \bar{t}_1 t_2}{p_s r_1 \bar{p}_s \bar{t}_2 + (1 - p_s r_1) p_s \bar{t}_2 + (1 - p_s r_1)\bar{p}_s t_2},$$

$$\frac{(1-p_sr_1)\bar{p}_sP_{B_1B_2}(1,1)+(p_sr_1\bar{p}_s+(1-p_Sr_1)p_s)P_{B_1B_2}(0,0)}{p_sr_1\bar{p}_s\bar{t}_2+(1-p_sr_1)p_s\bar{t}_2+(1-p_sr_1)\bar{p}_st_2}, \\ \left.\frac{\big(p_sr_1\bar{p}_s+(1-p_sr_1)p_s\big)t_1\bar{t}_2}{p_sr_1\bar{p}_s\bar{t}_2+(1-p_sr_1)p_s\bar{t}_2+(1-p_sr_1)\bar{p}_st_2}\right) \tag{A.156}$$

where we compute the following conditional probabilities:

$$\begin{aligned}
&\Pr[V_1=0\mid U_1=0,X_2=1,Z_2=1]\\
&=\Pr[Z_1=0\mid U_1=0,X_2=1,Z_2=1]+=\underbrace{\Pr[Z_1=3\mid U_1=0,X_2=1,Z_2=1]}_{0}\\
&=\Pr[S_1X_1+S_2+B_1=0\mid U_1=0,X_2=1,S_1X_1+S_2+B_2=1]\\
&=\frac{\Pr[S_1X_1+S_2+B_1=0\quad\&\quad S_1X_1+S_2+B_2=1\mid U_1=0]}{\Pr[S_1X_1+S_2+B_2=1\mid U_1=0]}\\
&=\frac{(1-p_sr_1)\bar{p}_sP_{B_1B_2}(0,1)}{p_sr_1\bar{p}_sP_{B_2}(0)+(1-p_sr_1)p_sP_{B_2}(0)+(1-p_sr_1)\bar{p}_sP_{B_2}(1)}\\
=&\frac{(1-p_sr_1)\bar{p}_s\bar{t}_1t_2}{p_sr_1\bar{p}_s\bar{t}_2+(1-p_sr_1)p_s\bar{t}_2+(1-p_sr_1)\bar{p}_st_2}
\end{aligned} \tag{A.157}$$

$$\begin{aligned}
&\Pr[V_1=1\mid U_1=0,X_2=1,Z_2=1]\\
&=\Pr[S_1X_1+S_2+B_1=1\mid U_1=0,X_2=1,S_1X_1+S_2+B_2=1]\\
&=\frac{(1-p_sr_1)\bar{p}_sP_{B_1B_2}(1,1)+(p_sr_1\bar{p}_s+(1-p_Sr_1)p_s)P_{B_1B_2}(0,0)}{p_sr_1\bar{p}_s\bar{t}_2+(1-p_sr_1)p_s\bar{t}_2+(1-p_sr_1)\bar{p}_st_2}
\end{aligned} \tag{A.158}$$

$$\begin{aligned}
&\Pr[V_1=2\mid U_1=0,X_2=1,Z_2=1]\\
&=\frac{\big(p_sr_1\bar{p}_s+(1-p_sr_1)p_s\big)P_{B_1B_2(1,0)}}{p_sr_1\bar{p}_sP_{B_2}(0)+(1-p_sr_1)p_sP_{B_2}(0)+(1-p_sr_1)\bar{p}_sP_{B_2}(1)}\\
&=\frac{\big(p_sr_1\bar{p}_s+(1-p_sr_1)p_s\big)t_1\bar{t}_2}{p_sr_1\bar{p}_s\bar{t}_2+(1-p_sr_1)p_s\bar{t}_2+(1-p_sr_1)\bar{p}_st_2}
\end{aligned} \tag{A.159}$$

- The case $(U_1,X_2,Z_2)=(1,1,1)$ is analogous to the previous case when $U_1=0$. Thus we replace $r_1,\bar{r}_1$ by $\bar{r}_1,r_1$ respectively.

- For case $(U_1,X_2,Z_2)=(0,0,2)$, we simply have $H(\bar{t}_1,t_1)$ on account of the following computations:

$$\Pr[V_1=0\mid U_1=0,X_2=1,Z_2=1]=P_{B_1|B_2}(0\mid 1)=0 \tag{A.160}$$

$$\Pr[V_1 = 1 \mid U_1 = 0, X_2 = 1, Z_2 = 1] = P_{B_1|B_2}(0 \mid 1) = \bar{t}_1 \tag{A.161}$$

$$\Pr[V_1 = 2 \mid U_1 = 0, X_2 = 1, Z_2 = 1] = P_{B_1|B_2}(1 \mid 1) = t_1 \tag{A.162}$$

- For $(U_1, X_2, Z_2) = (1, 0, 2)$, the calculation and final result is similar as previous case.

- For case $(U_1, X_2, Z_2) = (0, 1, 2)$, we obtain the following entropy:

$$H\Bigg(\frac{p_s^2 r_1 t_1 \bar{t}_2}{(p_s r_1 \bar{p}_s + (1-p_s r_1)p_s)t_2 + p_s^2 r_1 \bar{t}_2}, \frac{\Big(p_s r_1 \bar{p}_s + (1-p_s r_1)p_s\Big)\bar{t}_1 t_2}{(p_s r_1 \bar{p}_s + (1-p_s r_1)p_s)t_2 + p_s^2 r_1 \bar{t}_2}, \frac{p_s^2 r_1 \bar{t}_1 \bar{t}_2 + \Big(p_s r_1 \bar{p}_s + (1-p_s r_1)p_s\Big)t_1 t_2}{\Big(p_s r_1 \bar{p}_s + (1-p_s r_1)p_s\Big)t_2 + p_s^2 r_1 \bar{t}_2}\Bigg) \tag{A.163}$$

where each conditional probability is comuted as follows:

$$\begin{aligned}
&\Pr[V_1 = 0 \mid U_1 = 0, X_2 = 1, Z_2 = 2] \\
&= \underbrace{\Pr[Z_1 = 0 \mid U_1 = 0, X_2 = 1, Z_2 = 2]}_{=0} + \Pr[Z_1 = 3 \mid U_1 = 0, X_2 = 1, Z_2 = 2] && \text{(A.164)} \\
&= \frac{\Pr[Z_1 = 3, U_1 = 0, X_2 = 1, Z_2 = 2]}{\Pr[U_1 = 0, X_2 = 1, Z_2 = 2]} \\
&= \frac{\Pr[S_1X_1 + S_2X_2 + B_1 = 3, S_1X_1 + S_2X_2 + B_2 = 2 \mid U_1 = 0, X_2 = 1]}{\Pr[Z_2 = 2 \mid U_1 = 0, X_2 = 1]} \\
&= \frac{\Pr[S_1X_1 + S_2 + B_1 = 3, S_1X_1 + S_2 + B_2 = 2 \mid U_1 = 0]}{\Pr[S_1X_1 + S_2 + B_2 = 2 \mid U_1 = 0]} \\
&= \frac{\Pr[S_1X_1 + S_2 = 2 \mid U_1 = 0]P_{B_1B_2}(1,0)}{(p_s r_1 \bar{p}_s + (1-p_s r_1)p_s)P_{B_2}(1) + p_s^2 r_1 P_{B_2}(0)} \\
&= \frac{p_s^2 r_1 t_1 \bar{t}_2}{(p_s r_1 \bar{p}_s + (1-p_s r_1)p_s)t_2 + p_s^2 r_1 \bar{t}_2} && \text{(A.165)}
\end{aligned}$$

$$\begin{aligned}
\Pr[V_1 = 1 \mid U_1 = 0, X_2 = 1, Z_2 = 2] &= \frac{\Big(p_s r_1 \bar{p}_s + (1-p_s r_1)p_s\Big)P_{B_1B_2}(0,1)}{(p_s r_1 \bar{p}_s + (1-p_s r_1)p_s)P_{B_2}(1) + p_s^2 r_1 P_{B_2}(0)} \\
&= \frac{\Big(p_s r_1 \bar{p}_s + (1-p_s r_1)p_s\Big)\bar{t}_1 t_2}{(p_s r_1 \bar{p}_s + (1-p_s r_1)p_s)t_2 + p_s^2 r_1 \bar{t}_2} && \text{(A.166)}
\end{aligned}$$

$$\begin{aligned}\Pr[V_1 = 2 \mid U_1 = 0, X_2 = 1, Z_2 = 2] &= \frac{p_s^2 r_1 P_{B_1B_2}(0,0) + \Big(p_s r_1 \bar{p}_s + (1-p_s r_1)p_s\Big) P_{B_1B_2}(1,1)}{\Big(p_s r_1 \bar{p}_s + (1-p_s r_1)p_s\Big) P_{B_2}(1) + p_s^2 r_1 P_{B_2}(0)} \\ &= \frac{p_s^2 r_1 \bar{t}_1 \bar{t}_2 + \Big(p_s r_1 \bar{p}_s + (1-p_s r_1)p_s\Big) t_1 t_2}{\Big(p_s r_1 \bar{p}_s + (1-p_s r_1)p_s\Big) t_2 + p_s^2 r_1 \bar{t}_2}, \end{aligned} \tag{A.167}$$

To compute I_6

We replace (r_1, t_1, t_2) by (r_2, t_2, t_1) in functions $f_5(\cdot), Q_5(\cdot), G_5(\cdot)$, respectively in A.149, A.147, A.148 to compute mutual information term (A.111g):

$$\begin{aligned} I_6 = H(V_2 \mid U_1U_2) &- P_{X_1Z_1}(0,0)H(t_2,\bar{t}_2) - P_{X_1Z_1}(1,3)H(t_2,\bar{t}_2) - P_{X_1Z_1}(0,2)H(t_2,\bar{t}_2) \\ &- P_{U_2X_1Z_1}(0,0,1)G_5(r_2,t_2,t_1) - P_{U_2X_1Z_1}(1,0,1)G_6(\bar{r}_2,t_2,t_1) \\ &- P_{U_2X_1Z_1}(0,1,1)Q_6(r_2,t_2,t_1) - P_{U_2X_1Z_1}(1,1,1)Q_6(\bar{r}_2,t_2,t_1) \\ &- P_{U_2X_1Z_1}(0,1,2)f_6(r_2,t_2,t_1) - P_{U_2X_1Z_1}(1,1,2)f_6(\bar{r}_2,t_2,t_1) \end{aligned} \tag{A.168}$$

To compute I_7

The mutual information term (A.111h) specializes to:

$$I_7 = p_{e_1}p_{e_2}f_{7,1} + \bar{p}_{e_1}p_{e_2}f_{7,2} + p_{e_1}\bar{p}_{e_2}f_{7,3} + \bar{p}_{e_1}\bar{p}_{e_2}f_{7,4} \tag{A.169}$$

In I_7, we have to be careful because it is the only term that contains (Y', V_1, V_2, E_1, E_2) in which (Y', V_1, V_2) are not independent because Y', Z_1 and Z_2 are not necessarily independent. We go through the following steps:

$$\begin{aligned} I_7 &:= I(X_1X_2; Y'S_1S_2V_1V_2E_1E_2 \mid \underline{U}) \\ &= H(Y'S_1S_2V_1V_2 \mid E_1E_2\underline{U}) - H(Y'S_1S_2V_1V_2 \mid E_1E_2X_1X_2) \\ &= P_{E_1E_2}(1,1)\Big[H(Y' \mid \underline{U}S_1S_2) - \underbrace{H(Y' \mid X_1X_2S_1S_2)}_{H(t_0,\bar{t}_0)}\Big] \\ &+ P_{E_1E_2}(0,1)\Big[H(Y'V_1 \mid \underline{U}S_1S_2) - \underbrace{H(Y'V_1 \mid X_1X_2S_1S_2)}_{H(t_0,\bar{t}_0)+H(t_1,\bar{t}_1)}\Big] \\ &+ P_{E_1E_2}(1,0)\Big[H(Y'V_2 \mid \underline{U}S_1S_2) - \underbrace{H(Y'V_2 \mid X_1X_2S_1S_2)}_{H(t_0,\bar{t}_0)+H(t_2,\bar{t}_2)}\Big] \end{aligned}$$

$$+\ P_{E_1E_2}(0,0)\Big[H(Y'V_1V_2 \mid \underline{U}S_1S_2) - \underbrace{H(Y'V_1V_2 \mid X_1X_2S_1S_2)}_{H(t_0,\bar{t}_0)+H(t_1,\bar{t}_1)+H(t_2,\bar{t}_2)}\Big] \tag{A.170}$$

$$= p_{e_1}p_{e_2}f_{7,1} + \bar{p}_{e_1}p_{e_2}f_{7,2} + p_{e_1}\bar{p}_{e_2}f_{7,3} + \bar{p}_{e_1}\bar{p}_{e_2}f_{7,4} \tag{A.171}$$

Here, we compute $f_{7,1}, f_{7,2}, f_{7,3}, f_{7,4}$ as some individual terms as detailed out in the following steps:

$$f_{7,1} \triangleq \bar{p}_s^2 H(t_0,\bar{t}_0) - H(t_0,\bar{t}_0)$$

$$+p_s\bar{p}_s \underbrace{P_{U_1}(0)}_{p*\bar{q}_1} H(X_1+B_0 \mid U_1=0) + p_s\bar{p}_s \underbrace{P_{U_1}(0)}_{p*q_1} H(X_1+B_0 \mid U_1=1)$$

$$+\bar{p}_s p_s \underbrace{P_{U_2}(0)}_{p*\bar{q}_2} H(X_2+B_0 \mid U_2=0) + +\bar{p}_s p_s \underbrace{P_{U_2}(1)}_{p*q_2} H(X_2+B_0 \mid U_2=1)$$

$$+\ p_s^2 \underbrace{P_{U_1U_2}(0,0)}_{pq_1q_2+\bar{p}\bar{q}_1\bar{q}_2} H(X_1+X_2+B_0 \mid U_1=1, U_2=0) \tag{A.172}$$

$$+p_s^2 \underbrace{P_{U_1U_2}(0,1)}_{pq_1\bar{q}_2+\bar{p}\bar{q}_1q_2} H(X_1+X_2+B_0 \mid U_1=0, U_2=1) \tag{A.173}$$

$$+p_s^2 \underbrace{P_{U_1U_2}(1,0)}_{p\bar{q}_1q_2+\bar{p}q_1\bar{q}_2} H(X_1+X_2+B_0 \mid U_1=U_2=0) \tag{A.174}$$

$$+p_s^2 \underbrace{P_{U_1U_2}(1,1)}_{p\bar{q}_1\bar{q}_2+\bar{p}q_1q_2} H(X_1+X_2+B_0 \mid U_1=U_2=1) \tag{A.175}$$

$$= \big(\bar{p}_s^2-1\big)H(t_0,\bar{t}_0) + p_s\bar{p}_s\Big((p*\bar{q}_1)H(\bar{r}_1\bar{t}_0, r_1*t_0, r_1t_0) + (p*q_1)H(r_1\bar{t}_0, \bar{r}_1*t_0, \bar{r}_1t_0)\Big)$$

$$+\bar{p}_s p_s\Big[(p*\bar{q}_2)H(\bar{r}_2\bar{t}_0, r_2*t_0, r_2t_0) + (p*q_2)H(r_2\bar{t}_0, \bar{r}_2*t_0, \bar{r}_2t_0)\Big]$$

$$+p_s^2\Big[(pq_1q_2+\bar{p}\bar{q}_1\bar{q}_2)H\Big(\bar{r}_1\bar{r}_2\bar{t}_0, (r_1*r_2)\bar{t}_0+\bar{r}_1\bar{r}_2t_0, r_1r_2\bar{t}_0+(r_1*r_2)t_0, r_1r_2t_0\Big)$$

$$+(pq_1\bar{q}_2+\bar{p}\bar{q}_1q_2)H\Big(\bar{r}_1r_2\bar{t}_0, (r_1*\bar{r}_2)\bar{t}_0+\bar{r}_1r_2t_0, r_1\bar{r}_2\bar{t}_0+(r_1*\bar{r}_2)t_0, r_1\bar{r}_2t_0\Big)$$

$$+(p\bar{q}_1q_2+\bar{p}q_1\bar{q}_2)H\Big(r_1\bar{r}_2\bar{t}_0, (\bar{r}_1*r_2)\bar{t}_0+r_1\bar{r}_2t_0, \bar{r}_1r_2\bar{t}_0+(\bar{r}_1*r_2)t_0, \bar{r}_1r_2t_0\Big)$$

$$+(p\bar{q}_1\bar{q}_2+\bar{p}q_1q_2)H\Big(r_1r_2\bar{t}_0, (\bar{r}_1*\bar{r}_2)\bar{t}_0+r_1r_2t_0, \bar{r}_1\bar{r}_2\bar{t}_0+(\bar{r}_1*\bar{r}_2)t_0, \bar{r}_1\bar{r}_2t_0\Big)\Big] \tag{A.176}$$

$$f_{7,2} \triangleq \Big[\bar{p}_s^2\big(H(t_0,\bar{t}_0)+H(t_1,\bar{t}_1)\big) - H(t_0,\bar{t}_0) - H(t_1,\bar{t}_1)$$

$$+p_s\bar{p}_s H(X_1+B_0, X_1+B_1 \mid U_1) + \bar{p}_s p_s H(X_2+B_0, X_2+B_1 \mid U_2)$$

$$+p_s^2 H(X_1+X_2+B_0, X_1+X_2+B_1 \mid U_1U_2)\Big] \tag{A.177}$$

$$= (\bar{p}_s^2-1)\Big(H(t_0,\bar{t}_0)+H(t_1,\bar{t}_1)\Big)$$

$$+p_s\bar{p}_s\left[P_{U_1}(0)\underbrace{H\Big(\bar{r}_1\bar{t}_0\bar{t}_1,\ \bar{r}_1\bar{t}_0t_1,\ \bar{r}_1t_0\bar{t}_1,\ r_1\bar{t}_0\bar{t}_1+\bar{r}_1t_0t_1,\ r_1\bar{t}_0t_1,\ r_1t_0\bar{t}_1,\ r_1t_0t_1\Big)}_{\triangleq G_{7,2}(r_1)}+P_{U_1}(1)G_{7,2}(\bar{r}_1)\right]$$
$$+\bar{p}_sp_s\Big[P_{U_2}(0)G_{7,2}(r_2)+P_{U_2}(1)G_{7,2}(\bar{r}_2)\Big]$$
$$+p_s^2\Big[P_{U_1U_2}(0,0)H_{7,2}(r_1,r_2)+P_{U_1U_2}(0,1)H_{7,2}(r_1,\bar{r}_2)$$
$$+P_{U_1U_2}(1,0)H_{7,2}(\bar{r}_1,r_2)+P_{U_1U_2}(1,1)H_{7,2}(\bar{r}_1,\bar{r}_2)\Big] \tag{A.178}$$

where

$$H_{7,2}\triangleq H\Big(\bar{r}_1\bar{r}_2\bar{t}_0\bar{t}_1,\ \bar{r}_1\bar{r}_2\bar{t}_0t_1,\ \bar{r}_1\bar{r}_2t_0\bar{t}_1,(r_1*r_2)\bar{t}_0\bar{t}_1+\bar{r}_1\bar{r}_2t_0t_1,\ (r_1*r_2)\bar{t}_0t_1,\ (r_1*r_2)t_0\bar{t}_1,$$
$$r_1r_2\bar{t}_0\bar{t}_1+(r_1*r_2)t_0t_1,\ r_1r_2\bar{t}_0t_1,\ r_1r_2t_0\bar{t}_1,\ r_1r_2t_0t_1\Big) \tag{A.179}$$

The calculation of $f_{7,3}$ is similar to $f_{7,2}$ by replacing $r_1\to r_2$ and $t_1\to t_2$.

$$f_{7,3}\triangleq\Big[\bar{p}_s^2\big(H(t_0,\bar{t}_0)+H(t_2,\bar{t}_2)\big)-H(t_0,\bar{t}_0)-H(t_2,\bar{t}_2)$$
$$+p_s\bar{p}_sH(X_1+B_0,X_1+B_2\mid U_1)+\bar{p}_sp_sH(X_2+B_0,X_2+B_2\mid U_2)$$
$$+p_s^2H\big(X_1+X_2+B_0,\quad X_1+X_2+B_2\mid U_1U_2\big)\Big] \tag{A.180}$$

$$f_{7,4}\triangleq\Big[\bar{p}_s^2\Big(H(t_0,\bar{t}_0)+H(t_1,\bar{t}_1)+H(t_2,\bar{t}_2)\Big)-H(t_0,\bar{t}_0)-H(t_1,\bar{t}_1)-H(t_2,\bar{t}_2)$$
$$+p_s\bar{p}_sH(X_1+B_0,X_1+B_1,X_1+B_2\mid U_1)+\bar{p}_sp_sH(X_2+B_0,X_2+B_1,X_2+B_2\mid U_2)$$
$$+p_s^2H\big(X_1+X_2+B_0,\quad X_1+X_2+B_1\quad X_1+X_2+B_2\mid U_1U_2\big)\Big] \tag{A.181}$$
$$=\big(\bar{p}_s^2-1\big)\Big(H(t_0,\bar{t}_0)+H(t_1,\bar{t}_1)+H(t_2,\bar{t}_2)\Big)$$
$$+\ p_s\bar{p}_sG_{7,4}(r_1)+\bar{p}_sp_sG_{7,4}(r_2)$$
$$+p_s^2\Big[P_{U_1U_2}(0,0)G_{7,4}(r_1,r_2)+P_{U_1U_2}(0,1)G_{7,4}(r_1,\bar{r}_2)$$
$$+P_{U_1U_2}(1,0)G_{7,4}(\bar{r}_1,r_2)+P_{U_1U_2}(1,1)G_{7,4}(\bar{r}_1,\bar{r}_2)\Big] \tag{A.182}$$

where

$$G_{7,4}(r) \triangleq H\Big(\bar{r}\bar{t}_0\bar{t}_1\bar{t}_2,\ \bar{r}\bar{t}_0\bar{t}_1t_2,\ \bar{r}\bar{t}_0t_1\bar{t}_2,\ \bar{r}\bar{t}_0t_1t_2,\ \bar{r}t_0\bar{t}_1\bar{t}_2,\ \bar{r}t_0\bar{t}_1t_2,\ \bar{r}t_0t_1\bar{t}_2,$$
$$r\bar{t}_0\bar{t}_1\bar{t}_2 + \bar{r}t_0t_1t_2,\ r\bar{t}_0\bar{t}_1t_2,\ r\bar{t}_0t_1\bar{t}_2,\ r\bar{t}_0t_1t_2,\ rt_0\bar{t}_1\bar{t}_2,\ rt_0\bar{t}_1t_2,\ rt_0t_1\bar{t}_2,\ rt_0t_1t_2\Big) \tag{A.183}$$

and

$$G_{7,5}(a,b) \triangleq H\Bigg(\bar{a}\bar{b}\bar{t}_0\bar{t}_1\bar{t}_2,\ \bar{a}\bar{b}\bar{t}_0\bar{t}_1t_2,\ ,\bar{a}\bar{b}\bar{t}_0t_1\bar{t}_2,\ ,\bar{a}\bar{b}\bar{t}_0t_1t_2,\ ,\bar{a}\bar{b}t_0\bar{t}_1\bar{t}_2,\ \bar{a}\bar{b}t_0\bar{t}_1t_2,\ ,\bar{a}\bar{b}t_0t_1\bar{t}_2,$$
$$(a*b)\bar{t}_0\bar{t}_1\bar{t}_2 + \bar{a}\bar{b}t_0t_1t_2,\ ,(a*b)\bar{t}_0\bar{t}_1t_2,\ (a*b)\bar{t}_0t_1\bar{t}_2,\ ,(a*b)\bar{t}_0t_1t_2,\ ,(a*b)t_0\bar{t}_1\bar{t}_2,$$
$$,(a*b)t_0\bar{t}_1t_2,\ (a*b)t_0t_1\bar{t}_2\ ,abt_0t_1t_2 + (a*b)t_0t_1t_2\ ,ab\bar{t}_0\bar{t}_1t_2,$$
$$ab\bar{t}_0t_1\bar{t}_2,\ ab\bar{t}_0t_1t_2,\ ,abt_0\bar{t}_1\bar{t}_2,\ abt_0\bar{t}_1t_2,\ abt_0t_1\bar{t}_2,\ abt_0t_1t_2\Bigg). \tag{A.184}$$

To compute I_8, I_9

The final expression of mutual information term (A.111i) is given by:

$$\begin{aligned}
I_8 = {} & p_{e_1}p_{e_2}\Big[p_s\big[(p*\bar{q}_1)H(\bar{r}_1\bar{t}_0, r_1*t_0, r_1t_0) + (p*q_1)H(r_1\bar{t}_0, \bar{r}_1*t_0, \bar{r}_1t_0)\big]\Big] \\
& + \bar{p}_{e_1}p_{e_2}\Big[\bar{p}_s\big[H(t_0,\bar{t}_0) + H(t_1,\bar{t}_1)\big] + p_sH\Big(\bar{r}_1(\bar{t}_0\bar{t}_1 + t_0*t_1), r_1\bar{t}_0\bar{t}_1 + \bar{r}_1t_0t_1, r_1(t_0*t_1), r_1t_0t_1\Big) \\
& \qquad -H(t_0,\bar{t}_0) - H(t_1,\bar{t}_1)\Big] \\
& + p_{e_1}\bar{p}_{e_2}\Big[\bar{p}_s\big(H(t_0,\bar{t}_0) + H(t_2,\bar{t}_2)]\big) - H(t_0,\bar{t}_0) - H(t_2,\bar{t}_2) \\
& \qquad +p_sH\Big(\bar{r}_1(\bar{t}_0\bar{t}_2 + t_0*t_2), r_1\bar{t}_0\bar{t}_2 + \bar{r}_1t_0t_2, r_1(t_0*t_2), r_1t_0t_2\Big)\Big] \\
& + \bar{p}_{e_1}\bar{p}_{e_2}\Big[\bar{p}_s\big(H(t_0,\bar{t}_0) + H(t_1,\bar{t}_1) + H(t_2,\bar{t}_2)\big) - H(t_0,\bar{t}_0) - H(t_1,\bar{t}_1) - H(t_2,\bar{t}_2) \\
& + p_s\big((p*\bar{q}_1)f_8(r_1) + (p*q_1f_8(\bar{r}_1))\Big]
\end{aligned} \tag{A.185}$$

One can see the detailed steps as follows:

$$\begin{aligned}
I_8 &:= I(X_1; YV_1V_2E_1E_2 \mid \underline{U}X_2) \\
&= P_{E_1E_2}(1,1)\Big[I(X_1 : Y'V_1 = V_2 = \emptyset \mid U_1X_2S_1S_2)\Big]
\end{aligned}$$

$$
\begin{aligned}
&+\ P_{E_1E_2}(0,1)\Big[I(X_1:Y' \underbrace{V_1}_{=1\{Z_1=1\}+2\cdot 1\{Z_1=2\}} \mid U_1X_2S_1S_2)\Big] \\
&+\ P_{E_1E_2}(1,0)\Big[I(X_1:Y' \underbrace{V_2}_{=1\{Z_2=1\}+2\cdot 1\{Z_2=2\}} \mid U_1X_2S_1S_2)\Big] \\
&+\ P_{E_1E_2}(0,0)\Big[I(X_1:Y' \underbrace{V_1}_{=1\{Z_1=1\}+2\cdot 1\{Z_1=2\}}, \underbrace{V_2}_{=1\{Z_2=1\}+2\cdot 1\{Z_2=2\}} \mid U_1X_2S_1S_2)\Big] && \text{(A.186)} \\
&=\ P_{E_1E_2}(1,1)\Big[\underbrace{H(Y' \mid U_1X_2S_1S_2)}_{P_{S_1}(0)H(B_0)+P_{S_1}(1)H(X_1+B_0|U_1)} - H(t_0,\bar{t}_0)\Big] \\
&+\ P_{E_1E_2}(0,1)\Big[\underbrace{H(Y',V_1 \mid U_1S_1S_2X_2)}_{P_{S_1}(0)(H(B_0,B_1))+P_{S_1}(1)H(X_1+B_0,X_1+B_1|U_1)} - \underbrace{H(B_0,B_1)}_{H(t_0,\bar{t}_0)+H(t_1,\bar{t}_1)} \\
&+\ P_{E_1E_2}(1,0)\Big[\underbrace{H(Y',V_2 \mid U_1S_1S_2X_2)}_{P_{S_1}(0)(H(B_0,B_2))+P_{S_1}(1)H(X_1+B_0,X_1+B_1|U_1)} - \underbrace{H(B_0,B_2)}_{H(t_0,\bar{t}_0)+H(t_2,\bar{t}_2)}\Big] \\
&+\ P_{E_1E_2}(0,0)\Big[\underbrace{H(Y',V_1,V_2 \mid U_1S_1S_2X_2)}_{P_{S_1}(0)H(B_0,B_1,B_2)+P_{S_1}(1)H(X_1+B_0,X_1+B_1,X_1+B_2|U_1)} - H(B_0,B_1,B_2)\Big] && \text{(A.187)} \\
&=\ P_{E_1E_2}(1,1)\Big[p_s\big[(p*\bar{q}_1)H(\bar{r}_1\bar{t}_0, r_1*t_0, r_1t_0) + (p*q_1)H(r_1\bar{t}_0, \bar{r}_1*t_0, \bar{r}_1t_0)\big]\Big] \\
&+\ P_{E_1E_2}(1,0)\Big[\bar{p}_s\big[H(t_0,\bar{t}_0)+H(t_1,\bar{t}_1)\big] + p_sH\Big(\bar{r}_1(\bar{t}_0\bar{t}_1+t_0*t_1), r_1\bar{t}_0\bar{t}_1+\bar{r}_1t_0t_1, r_1(t_0*t_1), r_1t_0t_1\Big) \\
&\qquad\qquad - H(t_0,\bar{t}_0) - H(t_1,\bar{t}_1)\Big] \\
&+\ P_{E_1E_2}(1,0)\Big[\bar{p}_s\big(H(t_0,\bar{t}_0)+H(t_2,\bar{t}_2)]\big) - H(t_0,\bar{t}_0) - H(t_2,\bar{t}_2) \\
&\qquad\qquad + p_sH\Big(\bar{r}_1(\bar{t}_0\bar{t}_2+t_0*t_2), r_1\bar{t}_0\bar{t}_2+\bar{r}_1t_0t_2, r_1(t_0*t_2), r_1t_0t_2\Big)\Big] \\
&+\ P_{E_1E_2}(0,0)\Big[\bar{p}_s\big(H(t_0,\bar{t}_0)+H(t_1,\bar{t}_1)+H(t_2,\bar{t}_2)\big) - H(t_0,\bar{t}_0) - H(t_1,\bar{t}_1) - H(t_2,\bar{t}_2) \\
&+\ p_s\big((p*\bar{q}_1)f_8(r_1) + (p*q_1f_8(\bar{r}_1)\big)\Big] && \text{(A.188)}
\end{aligned}
$$

where

$$
\begin{aligned}
f_8(r_1) \triangleq H\Big(&\bar{r}_1\bar{t}_0\bar{t}_1\bar{t}_2, \quad \bar{r}_1\bar{t}_0\bar{t}_1t_2+\bar{r}_1\bar{t}_0t_1\bar{t}_2+\bar{r}_1t_0\bar{t}_1\bar{t}_2+r_1\bar{t}_0\bar{t}_1\bar{t}_2, \\
&\bar{r}_1\bar{t}_0t_1t_2+\bar{r}_1t_0\bar{t}_1t_2+\bar{r}_1t_0t_1\bar{t}_2+, \quad r_1\bar{t}_0\bar{t}_1t_2+r_1\bar{t}_0t_1\bar{t}_2+r_1t_0\bar{t}_1\bar{t}_2, \\
&r_1\bar{t}_0t_1t_2+r_1t_0\bar{t}_1t_2+r_1t_0t_1\bar{t}_2, \quad \bar{r}_1t_0t_1t_2+r_1\bar{t}_0\bar{t}_1\bar{t}_2, \quad r_1t_0t_1t_2\Big)
\end{aligned}
\qquad \text{(A.189)}
$$

Due to symmetry of the Txs, we have similar steps for I_9 with parameters q_2, r_2, t_2 corresponding to Z_2, X_2:

$$\begin{aligned}
I_9 = p_{e_1}p_{e_2}&\Bigg[p_s\big[(p*\bar{q}_2)H(\bar{r}_2\bar{t}_0, r_2*t_0, r_2t_0) + (p*q_2)H(r_2\bar{t}_0, \bar{r}_2*t_0, \bar{r}_2t_0)\big]\Bigg] \\
+ p_{e_1}\bar{p}_{e_2}&\Bigg[\bar{p}_s\big[H(t_0,\bar{t}_0) + H(t_2,\bar{t}_2)\big] + p_sH\Bigg(\bar{r}_2(\bar{t}_0\bar{t}_2 + t_0*t_2), r_2\bar{t}_0\bar{t}_2 + \bar{r}_2t_0t_2, r_2(t_0*t_2), r_2t_0t_2\Bigg) \\
&\qquad -H(t_0,\bar{t}_0) - H(t_1,\bar{t}_1)\Bigg] \\
+ \bar{p}_{e_1}p_{e_2}&\Bigg[\bar{p}_s\big(H(t_0,\bar{t}_0) + H(t_1,\bar{t}_1)]\big) - H(t_0,\bar{t}_0) - H(t_1,\bar{t}_1) \\
&\qquad +p_sH\Bigg(\bar{r}_2(\bar{t}_0\bar{t}_1 + t_0*t_1), r_2\bar{t}_0\bar{t}_1 + \bar{r}_2t_0t_1, r_2(t_0*t_1), r_2t_0t_1\Bigg)\Bigg] \\
+ \bar{p}_{e_1}\bar{p}_{e_2}&\Bigg[\bar{p}_s\big(H(t_0,\bar{t}_0) + H(t_1,\bar{t}_1) + H(t_2,\bar{t}_2)\big) - H(t_0,\bar{t}_0) - H(t_1,\bar{t}_1) - H(t_2,\bar{t}_2) \\
+ p_s\big((p*\bar{q}_2)&f_8(r_2) + (p*q_2f_8(\bar{r}_2)\big)\Bigg]
\end{aligned} \tag{A.190}$$

To compute I_{10}

The mutual information term (A.111k) is given by:

$$\begin{aligned}
I_{10} &:= I(X_1;Y \mid U_0X_2) = H(Y' \mid S_1S_2U_0X_2) - H(Y' \mid S_1S_2U_0X_2X_1) \\
&= P_{S_1}(0)H(t_0,\bar{t}_0) + P_{S_1}(1)H(X_1 + B_0 \mid U_0, X_2) - H(t_0,\bar{t}_0) \\
&= p_s\Bigg[\bar{p}H\Bigg((q*\bar{r}_1)\bar{t}_0, \quad (q*\bar{r}_1)t_0 + (q*r_1)\bar{t}_0, \quad (q*r_1)t_0\Bigg) \\
&\quad +pH\Bigg((q*r_1)\bar{t}_0, \quad (q*r_1)t_0 + (q*\bar{r}_1)\bar{t}_0, \quad (q*\bar{r}_1)t_0\Bigg)\Bigg] - p_sH(t_0,\bar{t}_0)
\end{aligned} \tag{A.191}$$

To compute I_{11}

The mutual information term (A.111l) is calculated as follows:

$$\begin{aligned}
I_{11} &:= I(X_2;Y \mid U_0X_1) = H(Y' \mid U_0X_1S_1S_2) - \underbrace{H(Y' \mid U_0X_1X_2S_1S_2)}_{H(t_0,\bar{t}_0)} \\
&= P_{S_2}(0)H(t_0,\bar{t}_0) + P_{S_2}(1)H(X_2 + B_0 \mid U_0, X_1) - H(t_0,\bar{t}_0) \\
&= p_s\Bigg[\bar{p}H\Bigg((q*\bar{r}_2)\bar{t}_0, \quad (q*\bar{r}_2)t_0 + (q*r_2)\bar{t}_0, \quad (q*r_2)t_0\Bigg)
\end{aligned}$$

$$+pH\Bigg((q*r_2)\bar{t}_0, \quad (q*r_2)t_0+(q*\bar{r}_2)\bar{t}_0, \quad (q*\bar{r}_2)t_0\Bigg)\Bigg] - p_s H(t_0,\bar{t}_0) \tag{A.192}$$

To compute I_{13}, I_{12}

The rate constraint (A.111m) specialize as detailed out in the following where we use $f_{13}(\cdot)$ defined later by different inputs to reduce the complexity of the calculations :

$$\begin{aligned}
I_{12} &:= I(X_1X_2;Y \mid U_0U_2) = H(Y' \mid U_0U_2S_1S_2) - \underbrace{H(Y' \mid U_0U_2S_1S_2X_1X_2)}_{H(t_0,\bar{t}_0)} \\
&= p_s^2\Big[P_{U_0U_2}(0,0)f_{13}(r_2,k_1,t_0) + P_{U_0U_2}(0,1)f_{13}(\bar{r}_2,k_1,t_0) \\
&\qquad +P_{U_0U_2}(1,0)f_{13}(r_2,\bar{k}_1,t_0) + P_{U_0U_2}(1,1)f_{13}(\bar{r}_2,\bar{k}_1,t_0)\Big] \\
&+ p_s\bar{p}_s\Big[P_{U_2}(0)H(\bar{r}_2\bar{t}_0,\bar{r}_2t_0+r_2\bar{t}_0,r_2t_0) + P_{U_2}(1)H(r_2\bar{t}_0,r_2t_0+\bar{r}_2t_0,\bar{r}_2t_0)\Big] \\
&+ \bar{p}_sp_s\Big[P_{U_2}(0)H(k_1\bar{t}_0,k_1t_0+\bar{k}_1t_0,\bar{k}_1t_0) + P_{U_2}(1)H(\bar{k}_1\bar{t}_0,\bar{k}_1t_0+k_1\bar{t}_0,k_1t_0)\Big] \\
&+ (\bar{p}_s^2-1)H(t_0,\bar{t}_0)
\end{aligned} \tag{A.193}$$

The final expression of (A.111n) is as follows:

$$\begin{aligned}
I_{13} &= p_s^2\Big[P_{U_0U_1}(0,0)f_{13}(r_1,k_2,t_0) + P_{U_0U_1}(0,1)f_{13}(\bar{r}_1,k_2,t_0) \\
&+ P_{U_0U_1}(1,0)f_{13}(r_1,\bar{k}_2,t_0) + P_{U_0U_1}(1,1)f_{13}(\bar{r}_1,\bar{k}_2,t_0)\Big] \\
&+ p_s\bar{p}_s\Big[P_{U_1}(0)G_{13}(r_1) + P_{U_1}(1)G_{13}(\bar{r}_1)\Big] + \bar{p}_sp_s\Big[P_{U_1}(0)G_{13}(\bar{k}_2) + P_{U_1}(1)G_{13}(k_2)\Big] \\
&+ (\bar{p}_s^2-1)H(t_0,\bar{t}_0)
\end{aligned} \tag{A.194}$$

The term (A.111n) specialize s as detailed out in the following:

$$\begin{aligned}
I_{13} &:= I(X_1X_2;Y \mid U_0U_1) = H(Y' \mid U_0U_1S_1S_2) - \underbrace{H(Y' \mid U_0U_1S_1S_2X_1X_2)}_{H(t_0,\bar{t}_0)} \\
&= P_{S_1S_2}(1,1)\Big[P_{U_0U_1}(0,0)f_{13}(r_1,k_2,t_0) + P_{U_0U_1}(0,1)f_{13}(\bar{r}_1,k_2,t_0) \\
&+ P_{U_0U_1}(1,0)f_{13}(r_1,\bar{k}_2,t_0) + P_{U_0U_1}(1,1)f_{13}(\bar{r}_1,\bar{k}_2,t_0)\Big] \\
&+ P_{S_1S_2}(1,0)\Big[P_{U_1}(0)G_{13}(r_1) + P_{U_1}(1)G_{13}(\bar{r}_1)\Big] + P_{S_1S_2}(0,1)\Big[P_{U_1}(0)G_{13}(\bar{k}_2) + P_{U_1}(1)G_{13}(k_2)\Big] \\
&+ (P_{S_1S_2}(0,0)-1)H(t_0,\bar{t}_0)
\end{aligned}$$

(A.195)

$$
\begin{aligned}
&= p_s^2\Big[P_{U_0U_1}(0,0)f_{13}(r_1,k_2,t_0) + P_{U_0U_1}(0,1)f_{13}(\bar{r}_1,k_2,t_0)\\
&+ P_{U_0U_1}(1,0)f_{13}(r_1,\bar{k}_2,t_0) + P_{U_0U_1}(1,1)f_{13}(\bar{r}_1,\bar{k}_2,t_0)\Big]\\
&+ p_s\bar{p}_s\Big[P_{U_1}(0)G_{13}(r_1) + P_{U_1}(1)G_{13}(\bar{r}_1)\Big] + \bar{p}_s p_s\Big[P_{U_1}(0)G_{13}(\bar{k}_2) + P_{U_1}(1)G_{13}(k_2)\Big]\\
&+ (\bar{p}_s^2 - 1)H(t_0,\bar{t}_0)
\end{aligned} \tag{A.196}
$$

where

$$
f_{13}(r,k,t) \triangleq H\Big(\bar{r}k\bar{t},\quad rk\bar{t} + \bar{r}\bar{k}\bar{t} + \bar{r}kt,\quad r\bar{k}\bar{t} + rkt + \bar{r}\bar{k}t,\quad r\bar{k}t_0\Big), \tag{A.197}
$$

and

$$
G_{13}(r_1) = H(\bar{r}_1\bar{t}_0, \bar{r}_1 t_0 + r_1\bar{t}_0, r_1 t_0). \tag{A.198}
$$

To compute I_{14}

The mutual information term (A.111o) specializes as detailed out in the following:

$$
\begin{aligned}
I_{14} = I(X_1X_2;Y'S_1S_2 \mid U_0) &= H(Y' \mid S_1S_2U_0) - \underbrace{H(Y' \mid X_1X_2S_1S_2U_0)}_{H(t_0,\bar{t}_0)}\\
&= p_s^2\Big[P_{U_0}(0)f_{14}(k_1,k_2) + P_{U_0}(1)f_{14}(\bar{k}_1,\bar{k}_2)\Big] + (\bar{p}_s^2 - 1)H(t_0,\bar{t}_0)\\
&+ \bar{p}_s p_s\Big[P_{U_0}(0)H\big(k_2\bar{t}_0,\ k_2t_0 + \bar{k}_2\bar{t}_0,\ \bar{k}_2t_0\big) + P_{U_0}(1)H\big(\bar{k}_2\bar{t}_0,\ \bar{k}_2t_0 + k_2\bar{t}_0,\ k_2t_0\big)\Big]\\
&+ p_s\bar{p}_s\Big[P_{U_0}(0)H\big(k_1\bar{t}_0,\ k_1t_0 + \bar{k}_1\bar{t}_0,\ \bar{k}_1t_0\big) + P_{U_0}(1)H\big(\bar{k}_1\bar{t}_0,\ \bar{k}_1t_0 + k_1\bar{t}_0,\ k_1t_0\big)\Big]
\end{aligned} \tag{A.199}
$$

where

$$
f_{14}(k_1,k_2) \triangleq H\Big(k_1k_2\bar{t}_0,\quad k_1k_2t_0 + (k_1\bar{k}_2 + \bar{k}_1k_2)\bar{t}_0,\quad \bar{k}_1\bar{k}_2\bar{t}_0 + (k_1\bar{k}_2 + \bar{k}_1k_2)t_0,\quad \bar{k}_1\bar{k}_2t_0\Big). \tag{A.200}
$$

To compute I_{15}

The mutual information term (A.111p) is calculated as follows:

$$I_{15} = p_s^2 H\Bigg((pk_1k_2 + \bar{p}\bar{k}_1\bar{k}_2)\bar{t}_0,\ \bar{p}\Big[(k_1\bar{k_2} + \bar{k}_1k_2)\bar{t}_0 + \bar{k}_1\bar{k}_2t_0\Big] + p\Big[(k_1\bar{k_2} + \bar{k}_1k_2)\bar{t}_0 + k_1k_2t_0\Big],$$
$$\bar{p}\Big[(k_1\bar{k_2} + \bar{k}_1k_2)t_0 + k_1k_2\bar{t}_0\Big] + p\Big[(k_1\bar{k_2} + \bar{k}_1k_2)t_0 + \bar{k}_1\bar{k}_2\bar{t}_0\Big],\ (\bar{p}k_1k_2 + p\bar{k}_1\bar{k}_2)t_0\Bigg)$$
$$+ \bar{p}_sp_sH\Bigg(\Big[(p*\bar{q}_1)\bar{r}_1 + (p*q_1)r_1\Big]\bar{t}_0,\ \Big[(p*\bar{q}_1)r_1 + (p*q_1)\bar{r}_1\Big]t_0$$
$$,\ \Big[(p*\bar{q}_1)\bar{r}_1 + (p*q_1)r_1\Big]t_0 + \Big[(p*\bar{q}_1)r_1 + (p*q_1)\bar{r}_1\Big]\bar{t}_0\Bigg)$$
$$+ p_s\bar{p}_sH\Bigg([(p*\bar{q}_2)\bar{r}_2 + (p*q_2)r_2\Big]\bar{t}_0,,\ \Big[(p*\bar{q}_2)r_2 + (p*q_2)\bar{r}_2\Big]t_0$$
$$,\ \Big[(p*\bar{q}_2)\bar{r}_2 + (p*q_2)r_2\Big]t_0 + \Big[(p*\bar{q}_2)r_2 + (p*q_1)\bar{r}_2\Big]\bar{t}_0\Bigg)$$
$$+(\bar{p}_s^2 - 1)H(t_0, \bar{t}_0) \tag{A.201}$$

The mutual information term (A.111p) specializes to

$$\begin{aligned}
I_{15} &:= I(X_1X_2; Y'S_1S_2) = H(Y' \mid S_1S_2) - \underbrace{H(Y' \mid S_1S_2X_1X_2)}_{H(t_0,\bar{t}_0)} \\
&= H(S_1X_1 + S_2X_2 + B_0 \mid S_1S_2) - H(t_0, \bar{t}_0) \\
&= P_{S_1S_2}(1,1)H\Bigg(\underbrace{\Pr[X_1 + X_2 + B_0 = 0]}_{P_{X_1X_2}(0,0)\bar{t}_0},\ \underbrace{\Pr[X_1 + X_2 + B_0 = 1],}_{P_{X_1X_2}(0,0)t_0 + P_{X_1X_2}(X_1\neq X_2)\bar{t}_0} \\
&\qquad \underbrace{\Pr[X_1 + X_2 + B_0 = 2],}_{P_{X_1X_2}(X_1\neq X_2)t_0 + P_{X_1X_2}(1,1)\bar{t}_0}\quad \underbrace{\Pr[X_1 + X_2 + B_0 = 3]}_{P_{X_1X_2}(1,1)t_0}\Bigg) \\
&+\ P_{S_1S_2}(0,1)H\Big[P_{X_2}(0)\bar{t}_0,\ P_{X_2}(0)t_0 + P_{X_2}(1)\bar{t}_0,\ P_{X_2}(1)t_0\Big] \\
&+\ P_{S_1S_2}(1,0)H\Big[P_{X_1}(0)\bar{t}_0,\ P_{X_1}(0)t_0 + P_{X_1}(1)\bar{t}_0,\ P_{X_1}(1)t_0\Big] \\
&+\ P_{S_1S_2}(0,0)H(t_0, \bar{t}_0) - H(t_0, \bar{t}_0) \\
&= p_s^2 \cdot H\Bigg((pk_1k_2 + \bar{p}\bar{k}_1\bar{k}_2)\bar{t}_0, \qquad\qquad (\text{A.202}) \\
&\qquad \Big(\bar{p}\Big[(k_1\bar{k_2} + \bar{k}_1k_2)\bar{t}_0 + \bar{k}_1\bar{k}_2t_0\Big] + p\Big[(k_1\bar{k_2} + \bar{k}_1k_2)\bar{t}_0 + k_1k_2t_0\Big]\Big), \\
&\qquad \bar{p}\Big[(k_1\bar{k_2} + \bar{k}_1k_2)t_0 + k_1k_2\bar{t}_0\Big] + p\Big[(k_1\bar{k_2} + \bar{k}_1k_2)t_0 + \bar{k}_1\bar{k}_2\bar{t}_0\Big],\ (\bar{p}k_1k_2 + p\bar{k}_1\bar{k}_2)t_0\Bigg)
\end{aligned}$$

$$
\begin{aligned}
&+\bar{p}_s p_s H\Bigg(\Big[(p*\bar{q}_1)\bar{r}_1+(p*q_1)r_1\Big]\bar{t}_0,\ \Big[(p*\bar{q}_1)r_1+(p*q_1)\bar{r}_1\Big]t_0,\\
&\qquad\Big[(p*\bar{q}_1)\bar{r}_1+(p*q_1)r_1\Big]t_0+\Big[(p*\bar{q}_1)r_1+(p*q_1)\bar{r}_1\Big]\bar{t}_0\Bigg)\\
&+p_s\bar{p}_s\cdot H\Bigg(\big[(p*\bar{q}_2)\bar{r}_2+(p*q_2)r_2\Big]\bar{t}_0,\ \Big[(p*\bar{q}_2)r_2+(p*q_2)\bar{r}_2\Big]t_0,\\
&\qquad\Big[(p*\bar{q}_2)\bar{r}_2+(p*q_2)r_2\Big]t_0+\Big[(p*\bar{q}_2)r_2+(p*q_1)\bar{r}_2\Big]\bar{t}_0\Bigg)\\
&+(\bar{p}_s^2-1)H(t_0,\bar{t}_0).
\end{aligned}
\tag{A.203}
$$

Now, we have all terms to apply rate region inequalities in (A.116a)-(A.116e) with constraints (A.116f)-(A.116k). It suffices to do optimization over all parameter values.

Bibliography

[1] M. Kobayashi, G. Caire, and G. Kramer, “Joint state sensing and communication: Optimal tradeoff for a memoryless case,” in *Proc. IEEE Int. Symp. Info. Theory (ISIT)*, 2018, pp. 111–115.

[2] M. Kobayashi, H. Hamad, G. Kramer, and G. Caire, “Joint state sensing and communication over memoryless multiple access channels,” in *Proc. IEEE Int. Symp. Info. Theory (ISIT)*, 2019, pp. 270–274.

[3] A. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[4] C. de Lima, D. Belot, R. Berkvens, A. Bourdoux, A. Dardari, M. Guillaud, E.-S. Isomursu, M.and Lohan, Y. Miao, A. N. Barreto, M. R. K. Aziz, J. Saloranta, T. Sanguanpuak, G. Sarieddeen, H.and Seco-Granados, J. Suutala, M. Svensson, T.and Valkama, H. Wymeersch, and B. E. van Liempd, “6g white paper on localization and sensing [white paper],” *University of Oulu*, no. 12, 2020.

[5] L. Zheng, M. Lops, Y. C. Eldar, and X. Wang, “Radar and communication co-existence: An overview: A review of recent methods,” vol. 36, no. 5, pp. 85–99, Sep. 2019.

[6] F. Liu, C. Masouros, A. P. Petropulu, H. Griffiths, and L. Hanzo, “Joint radar and communication design: Applications, state-of-the-art, and the road ahead,” *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3834–3862, 2020.

[7] C. Sturm and W. Wiesbeck, “Waveform design and signal processing aspects for fusion of wireless communications and radar sensing,” vol. 99, no. 7, pp. 1236–1259, July 2011.

[8] L. Gaudio, M. Kobayashi, G. Caire, and G. Colavolpe, “On the effectiveness of OTFS for joint radar parameter estimation and communication,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 5951–5965, 2020.

[9] D. W. Bliss, “Cooperative radar and communications signaling: The estimation and information theory odd couple,” in *Radar Conf., 2014 IEEE*. IEEE, 2014, pp. 0050–0055.

[10] A. R. Chiriyath, B. Paul, G. M. Jacyna, and D. W. Bliss, “Inner bounds on performance of radar and communications co-existence.” *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 464–474, 2016.

[11] B. Paul, A. R. Chiriyath, and D. W. Bliss, “Survey of RF communications and sensing convergence research,” *IEEE Access*, vol. 5, pp. 252–270, 2017.

[12] P. Kumari, J. Choi, N. González-Prelcic, and R. W. Heath, “IEEE 802.11ad-based radar: an approach to joint vehicular communication-radar system,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 4, pp. 3012–3027, 2018.

[13] P. Kumari, D. H. Nguyen, and R. W. Heath, “Performance trade-off in an adaptive IEEE 802.11 ad waveform design for a joint automotive radar and communication system,” in *IEEE Int. Conf. Acoustics, Speech and Signal Proc. (ICASSP)*. IEEE, 2017, pp. 4281–4285.

[14] F. Liu, L. Zhou, C. Masouros, A. Li, W. Luo, and A. Petropulu, “Toward dual-functional radar-communication systems: Optimal waveform design,” *IEEE Transactions on Signal Processing*, vol. 66, no. 16, pp. 4264–4279, 2018.

[15] D. Guo, S. Shamai, and S. Verdu, “Mutual information and minimum mean-square error in gaussian channels,” *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261–1282, 2005.

[16] M. Ahmadipour, M. Kobayashi, M. Wigger, and G. Caire, “An information-theoretic approach to joint sensing and communication,” *IEEE Transactions on Information Theory*, pp. 1–1, 2022.

[17] O. Shayevitz and M. Wigger, “On the capacity of the discrete memoryless broadcast channel with feedback,” *IEEE Trans. Info. Theory*, vol. 59, no. 3, pp. 1329–1345, 2013.

[18] R. Venkataramanan and S. S. Pradhan, “An achievable rate region for the broadcast channel with feedback,” *IEEE Trans. Info. Theory*, vol. 59, no. 10, pp. 6175–6191, 2013.

[19] M. Gastpar, A. Lapidoth, Y. Steinberg, and M. Wigger, “Coding schemes and asymptotic capacity for the gaussian broadcast and interference channels with feedback,” *IEEE Trans. Info. Theory*, vol. 60, no. 1, pp. 54–71, 2014.

[20] A.El Gamal and Y.H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[21] K. Marton, “A coding theorem for the discrete memoryless broadcast channel,” *IEEE Transactions on Information Theory*, vol. 25, no. 3, pp. 306–311, 1979.

[22] C. Choudhuri, Y.-H. Kim, and U. Mitra, "Causal state communication," *IEEE Trans. Info. Theory*, vol. 59, no. 6, pp. 3709–3719, 2013.

[23] A. Lapidoth and Y. Steinberg, "The multiple-access channel with causal side information: Double state," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1379–1393, 2013.

[24] ——, "The multiple-access channel with causal side information: Common state," *IEEE Transactions on Information Theory*, vol. 59, no. 1, pp. 32–50, 2013.

[25] P. Minero, S. H. Lim, and Y.-H. Kim, "A unified approach to hybrid coding," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1509–1523, 2015.

[26] H. Joudeh and F. M. J. Willems, "Joint communication and binary state detection," *IEEE Journal on Selected Areas in Information Theory*, vol. 3, no. 1, pp. 113–124, 2022.

[27] H. Wu and H. Joudeh, "On joint communication and channel discrimination," in *2022 IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 3321–3326.

[28] M.-C. Chang, , Erdogan, S.-Y. Wang, and M. R. Bloch, "Rate and detection error-exponent tradeoffs of joint communication and sensing," in *2022 2nd IEEE International Symposium on Joint Communications and Sensing (JCS)*, 2022, pp. 1–6.

[29] O. Gunlu, M. Bloch, R. F. Schaefer, and A. Yener, "Secure joint communication and sensing," in *2022 IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 844–849.

[30] Y.-H. Kim, A. Sutivong, and T. M. Cover, "State amplification," *IEEE Trans. Info. Theory*, vol. 54, no. 5, pp. 1850–1859, 2008.

[31] S. I. Bross and A. Lapidoth, "The Gaussian source-and-data-streams problem," *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5618–5628, 2019.

[32] V. Ramachandran, S. R. B. Pillai, and V. M. Prabhakaran, "Joint state estimation and communication over a state-dependent gaussian multiple access channel," *IEEE Transactions on Communications*, vol. 67, no. 10, pp. 6743–6752, 2019.

[33] W. Zhang, S. Vedantam, and U. Mitra, "Joint transmission and state estimation: A constrained channel coding approach," *IEEE Trans. Info. Theory*, vol. 57, no. 10, pp. 7084–7095, 2011.

[34] H. V. Poor, in *An Introduction to Signal Detection and Estimation*, 1994.

[35] R. M. Mealey, "A method for calculating error probabilities in a radar communication system," *IEEE Transactions on Space Electronics and Telemetry*, vol. 9, no. 2, pp. 37–42, 1963.

[36] M. R.Winkler, "Chirp signals for communications," *IEEE WESCON Cony. Rec*, 1962.

[37] A. I. Bemi and W. Grcgg, "On the utility of chirp modulation for digital signaling," *IEEE Trans. Cnmmun*, vol. COM-21, pp. 748–751, I973.

[38] Q. Zhang, H. Sun, X. Gao, X. Wang, and Z. Feng, "Time-division isac enabled connected automated vehicles cooperation algorithm design and performance evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 7, pp. 2206–2218, 2022.

[39] C. Shi, F. Wang, M. Sellathurai, J. Zhou, and S. Salous, "Power minimization-based robust ofdm radar waveform design for radar and communication systems in coexistence," *IEEE Transactions on Signal Processing*, vol. 66, no. 5, pp. 1316–1330, 2018.

[40] S. Sodagari, A. Khawar, T. C. Clancy, and R. McGwier, "A projection based approach for radar and telecommunication systems coexistence," in *2012 IEEE Global Communications Conference (GLOBECOM)*, 2012, pp. 5010–5014.

[41] G. Kramer, "Topics in multi-user information theory," *Foundations and Trends® in Communications and Information Theory*, vol. 4, no. 4—5, pp. 265–444, 2008.

[42] W. Zhang, S. Vedantam, and U. Mitra, "A constrained channel coding approach to joint communication and channel estimation," in *2008 IEEE International Symposium on Information Theory*, 2008, pp. 930–934.

[43] C. Choudhuri and U. M. Ming, "On non-causal side information at the encoder," in *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2012, pp. 648–655.

[44] A. Sutivong, M. Chiang, T. M. Cover, and Y.-H. Kim, "Channel capacity and state estimation for state-dependent Gaussian channels," *IEEE Trans. Info. Theory*, vol. 51, no. 4, pp. 1486–1495, 2005.

[45] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *The Bell System Technical Journal*, vol. 53, no. 9, pp. 1681–1721, 1974.

[46] C. C. Wang, "On the capacity of 1-to-k broadcast packet erasure channels with channel output feedback," *IEEE Trans. Info. Theory*, vol. 58, no. 2, pp. 931–956, 2012.

[47] M. Gatzianas, L. Georgiadis, and L. Tassiulas, “Multiuser broadcast erasure channel with feedback: capacity and algorithms,” *IEEE Trans. Info. Theory*, vol. 59, no. 9, pp. 5779–5804, 2013.

[48] G. Dueck, “Partial feedback for two-way and broadcast channels,” *Information and Control*, vol. 46, no. 1, pp. 1–15, 1980.

[49] M. H. Yassaee, M. R. Aref, and A. Gohari, “Achievability proof via output statistics of random binning,” *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6760–6786, 2014.

[50] A. Salimi, W. Zhang, S. Vedantam, and U. Mitra, “The capacity-distortion function for multihop channels with state,” in *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 2228–2232.

[51] R. Ahlswede, “Multi-waycommunicationchannels,” in *in Proc. Second Int. Symp. Inform. Transmission, Tsahkadsor, Armenia, USSR, Hungarian Press,*, 1971.

[52] L. Ozarow, “The capacity of the white gaussian multiple access channel with feedback,” *IEEE Transactions on Information Theory*, vol. 30, no. 4, pp. 623–629, 1984.

[53] F. Willems, “The feedback capacity region of a class of discrete memoryless multiple access channels (corresp.),” *IEEE Transactions on Information Theory*, vol. 28, no. 1, pp. 93–95, 1982.

[54] G. Kramer, “Capacity results for the discrete memoryless network,” *IEEE Transactions on Information Theory*, vol. 49, no. 1, pp. 4–21, 2003.

[55] F. Willems, E. van der Meulen, and J. Schalkwijk, “Achievable rate region for the multiple access channel with generalized feedback,” in *Proc. Annual Allerton Conf. on Communication, Control and Computing*, 1983, pp. 284–292.

[56] E. Tuncel, “Slepian-wolf coding over broadcast channels,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1469–1–482, 2006.

[57] M. Costa, “Writing on dirty paper (corresp.),” *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, 1983.

[58] T. Han, “A general coding scheme for the two-way channel,” *IEEE Transactions on Information Theory*, vol. 30, no. 1, pp. 35–44, 1984.

[59] G. Kramer, “Directed information for channels with feedback,” *Ph.D. dissertation, Swiss Federal Institute of Technology Zurich*, 1998.

[60] C. E. Shannon, *TwoWay Communication Channels*, 1993, pp. 351–384.

[61] M.-C. Chang, , Erdogan, S.-Y. Wang, and M. R. Bloch, "Rate and detection error-exponent tradeoffs of joint communication and sensing," in *2022 2nd IEEE International Symposium on Joint Communications and Sensing (JCS)*, 2022, pp. 1–6.

[62] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[63] A. E. Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.

ECOLE DOCTORALE

Titre: Une approche basée sur la théorie de l'information pour l'estimation et la communication intégrées

Mots clés: ISAC, JSC, Communication, Détection, Théorie de l'information, réseau multi-terminal

Résumé: Les réseaux sans fil de la prochaine génération devraient prendre en charge les techniques de détection. Des exemples importants sont les systèmes de transport intelligents, où les véhicules détectent en permanence les changements environnementaux et échangent des informations avec les véhicules ou les serveurs centraux. Il existe des solutions naïves pour réaliser ces deux tâches, qui proposent de partager les ressources entre les deux. Cependant, les coûts élevés de spectre et de matériel de ces approches encouragent l'intégration des tâches de détection et de communication (ISAC) via une seule forme d'onde et une seule plateforme matérielle. Cette thèse se concentre sur l'ISAC théorique de l'information. Nous examinons le premier modèle informationnel théorique pour ISAC dans [1] où un canal sans mémoire dépendant de l'état (SDMC) avec des signaux de rétroaction généralisés est observé au niveau de l'émetteur (Tx). Notre première contribution est de caractériser le compromis fondamental entre les taux de communication et la distorsion de détection des canaux de diffusion (BC) dépendants de l'état, mono-Tx et bi-Rx, qui sont physiquement dégradés. Nous fournissons également des limites intérieures et extérieures sur les compromis taux-distorsion réalisables pour les canaux de diffusion généraux. La stratégie optimale de détection des Tx uniques est un simple estimateur symbole par symbole et l'optimalité de cet estimateur découle du fait que les canaux de rétroaction généralisés et la séquence d'état sont tous deux sans mémoire. Ce n'est pas nécessairement le cas dans les configurations avec plus d'une Tx. Plus précisément, pour le MAC, nous proposons une détection collaborative où chaque Tx compresse d'abord les sorties et les entrées obtenues pour extraire les informations d'état, puis transmet l'indice de compression à l'aide d'un code de canal pur aux autres Tx. Nous décrivons également deux schémas ISAC collaboratifs pour D2D, basés sur la séparation source-canal/le schéma de canal bidirectionnel de Han et basés sur le codage conjoint source-canal (JSCC). Dans le scénario MAC et D2D, nos schémas ISAC sont strictement concaves dans les paires taux-distorsion et améliorent donc également les stratégies classiques de partage du temps ou des ressources.

Title: An Information Theory-Based Approach to Integrated Sensing and Communication

Keywords: ISAC, JSC, Communication, Sensing, Information theory, multi-terminal network

Abstract: Next-generation wireless networks are expected to support sensing techniques. Important examples are intelligent transport systems, where vehicles continuously sense environmental changes and exchange information with vehicles or central servers. There are some naive solutions to do both tasks which propose to share the resources between the two. But, the high spectrum and hardware costs of these approaches encourage to integrate the sensing and communication (ISAC) tasks via a single waveform and a single hardware platform. This thesis focuses on information-theoretic ISAC. We review the first information-theoretic model for ISAC in [1] where a state-dependent memoryless channel (SDMC) with generalized feedback signals observed at the transmitter (Tx). Our first contribution is to characterize the fundamental tradeoff between communication rates and sensing distortion of state-dependent single-Tx two-Rx broadcast channels (BC) that are physically degraded. We also provide inner and outer bounds on the achievable rate-distortion tradeoffs for general BCs. The single-Txs' optimal sensing strategy is a simple symbol-by-symbol estimator and the optimality of this estimator stems from the fact that the generalized feedback channels and the state-sequence both are memoryless. This is not necessarily the case in setups with more than one Tx. Specifically, for the MAC, we propose collaborative sensing where each Tx first compresses the obtained outputs and inputs to extract state information, then transmits the compression index using a pure channel code to the other Tx. Also, we describe two collaborative ISAC schemes for D2D, based on source-channel separation/Han's two-way channel scheme and based on joint source-channel coding (JSCC). In both the MAC and the D2D scenario, our ISAC schemes are strictly concave in the rate-distortion pairs and thus also improve over classical time- or resource-sharing strategies.

Institut Polytechnique de Paris
91120 Palaiseau, France