# Multiple Access Channel Simulation 

## Citation for published version (APA):

Kurri, G. R., Ramachandran, V., Pillai, S. R. B., \& Prabhakaran, V. M. (2022). Multiple Access Channel Simulation. IEEE Transactions on Information Theory, 68(11), 7575-7603. Article 9801874.
https://doi.org/10.1109/TIT.2022.3184643

## Document license:

TAVERNE

## DOI:

10.1109/TIT.2022.3184643

## Document status and date:

Published: 01/11/2022

## Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.
Link to publication


## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25 fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

## Take down policy

If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.

# Multiple Access Channel Simulation 

Gowtham R. Kurri ${ }^{\ominus}$, Member, IEEE, Viswanathan Ramachandran ${ }^{( }$, Sibi Raj B. Pillai ${ }^{\ominus}$, Member, IEEE, and Vinod M. Prabhakaran ${ }^{\circledR}$, Member, IEEE


#### Abstract

We study the problem of simulating a two-user multiple-access channel (MAC) over a multiple access network of noiseless links. Two encoders observe independent and identically distributed (i.i.d.) copies of a source random variable each, while a decoder observes i.i.d. copies of a side-information random variable. There are rate-limited noiseless communication links between each encoder and the decoder, and there is independent pairwise shared randomness between all the three possible pairs of nodes. The decoder has to output approximately i.i.d. copies of another random variable jointly distributed with the two sources and the side information. We are interested in the rate tuples which permit this simulation. This setting can be thought of as a multi-terminal generalization of the point-topoint channel simulation problem studied by Bennett et al. (2002) and Cuff (2013). When the pairwise shared randomness between the encoders is absent, the setting reduces to a special case of MAC simulation using another MAC studied by Haddadpour et al. (2013). We establish that the presence of encoder shared randomness can strictly improve the communication rate requirements. We first show that the inner bound derived from Haddadpour et al. (2013) is tight when the sources at the encoders are conditionally independent given the side-information at the decoder. This result recovers the existing results on point-to-point channel simulation and function computation over such multiterminal networks. We then explicitly compute the communication rate regions for an example both with and without the encoder shared randomness and demonstrate that its presence strictly reduces the communication rates. Inner and outer bounds for the general case are also obtained.


Manuscript received 22 February 2021; revised 26 January 2022; accepted 6 June 2022. Date of publication 21 June 2022; date of current version 21 October 2022. This work was supported by the Department of Atomic Energy, Government of India, under Project RTI4001. The work of Viswanathan Ramachandran was supported in part by the Bharti Centre for Communication, IIT Bombay. The work of Sibi Raj B. Pillai was supported by the Department of Science and Technology (DST)-Imprint under Project 2018/001280. The work of Vinod M. Prabhakaran was supported by the Science and Engineering Research Board, India, under Grant MTR/2020/000308. An earlier version of this paper was presented in part at the 2021 IEEE International Symposium on Information Theory (ISIT) [DOI: 10.1109/ISIT45174.2021.9517982]. (Corresponding author: Gowtham R. Kurri.)

Gowtham R. Kurri was with the Tata Institute of Fundamental Research, Mumbai 400005, India. He is now with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287 USA (e-mail: gowthamkurri@gmail.com).

Viswanathan Ramachandran was with the Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India. He is now with the Department of Electrical Engineering, Eindhoven University of Technology, 5612 Eindhoven, The Netherlands (e-mail: v.ramachandran@tue.nl).

Sibi Raj B. Pillai is with the Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India (e-mail: bsraj@ee.iitb.ac.in).

Vinod M. Prabhakaran is with the School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai 400005, India (e-mail: vinodmp@tifr.res.in).

Communicated by M. Dalai, Associate Editor for Shannon Theory and Information Measures.
Digital Object Identifier 10.1109/TIT.2022.3184643

Index Terms-Channel simulation, strong coordination, pairwise shared randomness, multiple access channel, random binning.

## I. INTRODUCTION

WHAT is the minimum amount of communication required to create correlation remotely? The channel simulation problem seeks to answer this fundamental question. In the point-to-point formulation, an encoder observing an independent and identically distributed (i.i.d.) source $X^{n}$ with distribution $q_{X}$ sends a message through a noiseless link to a decoder. The decoder has to output $Y^{n}$ such that the total variation distance between the joint distribution on $\left(X^{n}, Y^{n}\right)$ and the i.i.d. joint distribution induced by passing the source $X^{n}$ through a discrete memoryless channel $q_{Y \mid X}$ vanishes asymptotically. This requirement that the synthesized joint distribution be close to the desired i.i.d. joint distribution in total variation distance has been termed as strong coordination [1], which is also the focus of this paper. A source of common randomness accessible to both the encoder and the decoder may assist them in the aforementioned task. This framework was first investigated by Bennett et al. [2] assuming unlimited common randomness, where they established a 'reverse Shannon theorem' to synthesize a noisy channel from a noiseless channel. ${ }^{1}$ It was shown that the minimum communication rate is nothing but the mutual information $I(X ; Y)$ of the joint distribution. Harsha et al. [3] studied the non-asymptotic version of this problem. Winter [4] studied the setting with limited common randomness, albeit only for a certain extremal operating point. Cuff [5] and Bennett et al. [6] independently determined the entire optimal trade-off between communication and shared randomness rates. Later, Wilde et al. [7] obtained a similar trade-off in the quantum information-theoretic setting, generalizing the above result. Yassaee et al. [8] established a similar trade-off for channel simulation in a point-to-point network with side-information at the decoder. Simulation of a channel using another channel (instead of the noiseless communication link) was studied by Haddadpour et al. [9] and Cervia et al. [10]. A weaker form of coordination, namely, empirical coordination, where only the empirical distribution of the sequence of samples is required to be close to the desired distribution, has also been studied in point-to-point networks [1], [8], [11]-[14].

Channel simulation problems may also be thought of as distributed computation of randomized functions.

[^0]

Fig. 1. Strong coordination over a multiple-access network. Encoder $j \in$ $\{1,2\}$ on observing the source $X_{j}^{n}$ and shared randomness $S_{0}, S_{j}$ sends a message $M_{j}$ over a noiseless link to the decoder, which has side-information $Z^{n}$. Here $\left(X_{1 i}, X_{2 i}, Z_{i}\right), i=1, \ldots, n$, are i.i.d. with $q_{X_{1} X_{2} Z}$. Also ( $X_{1}^{n}, X_{2}^{n}, Z^{n}$ ), $S_{0}, S_{1}$, and $S_{2}$ are mutually independent. The decoder has to output $Y^{n}$ such that $\left(X_{1 i}, X_{2 i}, Z_{i}, Y_{i}\right), i=1, \ldots, n$, are approximately i.i.d. with $q_{X_{1} X_{2} Z Y \text {. }}$

Distributed computation of (deterministic) functions has received much attention in the computer science literature (see, e.g., [15], [16] and references therein) and the information theory literature [17]-[32]. Two terminal interactive function computation was studied in [18], [19], [21], [22], and [26]. Distributed multi-terminal function computation in a multiple-access network of noiseless links was studied by Körner and Marton [17], Han and Kobayashi [20], Kuzuoka and Watanabe [29], Watanabe [32], and Sefidgaran et al. [30]. Function computation in more general graph networks where a single node seeks to compute a function of the inputs at the other nodes was studied by Appuswamy et al. [27], Kowshik and Kumar [28], and Sefidgaran and Tchamkerten [31]. A related line of work is function computation over multiple-access channels studied by Nazer and Gastpar [23], [24], and Gastpar [25], which, in contrast to the above works, exploit the computation performed by the channel itself. This is fundamentally different from computation over a network of noiseless links where the communication channel does not perform any computation.

There are relatively few conclusive results on channel simulation in multi-terminal networks where possibly randomized functions of the inputs need to be computed. A natural generalization of the point-to-point network to a multi-terminal setting is the cascade network [33], [34]. Satpathy and Cuff [33] considered a cascade network under a secrecy requirement and obtained the optimal trade-off between communication and common randomness rates. Vellambi et al. [34] obtained the optimal rates for specific settings, e.g., when the communication topology matches the coordination structure. Empirical coordination in multi-terminal networks has been studied in [1], [35], and [36].

In this paper, we study multiple access channel simulation (randomized function computation) over a multiple access network of noiseless links. In particular, there are two encoders who share noiseless communication links with a decoder see Figure 1. The two encoders and the decoder observe i.i.d.
copies of the sources $X_{1}$ and $X_{2}$, and the side-information $Z$, respectively, that are generated according to a given distribution $q_{X_{1} X_{2} Z}$. Each encoder has access to a resource of pairwise shared randomness with the decoder. There is another resource of pairwise shared randomness between the encoders. The three pairwise shared randomness resources are independent of each other and also jointly independent of the sources and side information. In addition, the encoders and the decoder may privately randomize. The encoders transmit messages through noiseless links to the decoder, whose output together with the input sources and decoder side-information should be approximately i.i.d. with $q_{X_{1} X_{2} Z Y}$.

The work which is closest to ours is by Haddadpour et al. [37, Sec. IV], who studied simulation of a multiple-access channel from another multiple-access channel, a more general resource than the multiple-access network of noiseless links considered here. The main difference is the presence of shared randomness between the encoders, i.e., $S_{0}$, in our model. Haddadpour et al. [37] obtained an inner bound to the rate-region for their setting. In this work, we investigate the role of this additional shared randomness resource and specifically ask the question: Can this additional pairwise shared randomness resource between the encoders (that is not available to the decoder) strictly improve the communication rate requirements for channel simulation? We answer this question in the affirmative. This complements the results of [5] and [37] which established that shared randomness between encoder(s) and decoder is a useful resource for channel simulation. The configuration of a shared randomness resource between the encoders is reminiscent of the setting of multiple access channels with partially cooperating encoders [38] - see also [39], [40]. However, in the present setting, the encoders are not allowed to cooperate after observing the sources.

The problem of finding the optimal communication rates for computing possibly randomized functions at the decoder in this multiple-access network of noiseless links remains largely open even for the case of independent sources. Sefidgaran and Tchamkerten [31] determined the optimal communication rates for computing deterministic functions ${ }^{2}$ when the sources at the encoders are conditionally independent given the sideinformation at the decoder, i.e., $I\left(X_{1} ; X_{2} \mid Z\right)=0$. Note that Sefidgaran and Tchamkerten [31] in fact studied a more general setting consisting of multiple terminals over a rooted multi-level directed tree, where the multiple-access network of noiseless links is a special case. Atif et al. [41] studied multiple access channel simulation in the presence of three-way common randomness instead of pairwise shared randomness as above and obtained an achievable inner bound. Atif et al. [42] obtained a similar inner bound in the quantum informationtheoretic setting. After the submission of this work, the authors learnt about a concurrent work by Atif et al. [43], which, like [37], considered pairwise shared randomness between each encoder and the decoder (i.e., no encoder shared randomness), and derived inner and outer bounds on the

[^1]rate region. Achievable schemes using algebraic-structured codes exploiting the specific structure of the function to be computed were further explored by Krithivasan and Pradhan [44] and Atif and Pradhan [45], where the latter considered the quantum setting, in the spirit of Körner and Marton [17].

Main Contributions: We analyze the role of pairwise shared randomness in multiple-access channel simulation and establish that the presence of encoder shared randomness can strictly improve the communication rates required.

- First, for the case when the encoder shared randomness is absent, we obtain the optimal trade-off between the communication rates and one of the two remaining shared randomness rates (under sufficiently large other shared randomness rate) when $I\left(X_{1} ; X_{2} \mid Z\right)=0$ (Theorem 3). This recovers the results on point-to-point channel simulation [5], and also shows that the inner bound of Haddadpour et al. [37, Th. 3] is tight for the special case under consideration. When specialized to deterministic function computation, our Theorem 3 also recovers the result of Sefidgaran and Tchamkerten [31] restricted to multiple-access network of noiseless links. However, the techniques used there do not readily generalize to randomized function computation (see the discussion below Theorem 3).
- Next, we explicitly compute the communication rate regions for an example both with and without the encoder shared randomness and show that its presence strictly reduces the communication rates required for coordination (see Example 1 and Section V-A).
- We also derive general inner and outer bounds on the rate region with all the three pairwise shared randomness resources present (Theorems 4, 5, and 6). Our inner bound proofs are mostly in the spirit of Haddadpour et al. [37, Th. 3].
The remainder of this paper is organized as follows. We present our system model in Section II, and give the results for the case when the encoder shared randomness is absent in Section III. In Section IV, we show through an example that the presence of encoder shared randomness can lead to a strict improvement of the communication rate region. General inner and outer bounds on the rate region are presented in Section V. The proofs of our main results are given in Section VI.


## II. System Model

We study the problem of strong coordination of signals in a three-node multiple-access network. There are two encoders with inputs $X_{1}^{n}$ and $X_{2}^{n}$, respectively, and a decoder with side-information $Z^{n}$, where $\left(X_{1 i}, X_{2 i}, Z_{i}\right), i=1, \ldots, n$, are independent and identically distributed (i.i.d.) with distribution $q_{X_{1} X_{2} Z}$, with $X_{1}, X_{2}$, and $Z$ taking values in finite alphabets $\mathcal{X}_{1}, \mathcal{X}_{2}$, and $\mathcal{Z}$, respectively. For $j=1,2$, encoder $j$ and the decoder have access to a pairwise shared randomness $S_{j}$ uniformly distributed on $\left[1: 2^{n R_{0 j}}\right]$. There is another pairwise shared randomness $S_{0}$ between the two encoders that is uniformly distributed on $\left[1: 2^{n R_{00}}\right]$. The random variables $S_{0}, S_{1}$ and $S_{2}$ are independent and also jointly
independent of $\left(X_{1}^{n}, X_{2}^{n}, Z^{n}\right)$. Encoder $j \in\{1,2\}$ observes $X_{j}^{n}$ and shared randomness $\left(S_{0}, S_{j}\right)$, and sends a message $M_{j} \in\left[1: 2^{n R_{j}}\right]$ over a noiseless communication link to the decoder. The decoder observes $M_{1}, M_{2}$, in addition to the shared randomness $S_{1}, S_{2}$, and side-information $Z^{n}$. The goal is to output $Y^{n}$ (where $Y_{i}, i=1, \ldots, n$, takes values in a finite alphabet $\mathcal{Y}$ ) which along with the input sources and decoder side-information is approximately distributed according to $q_{X_{1} X_{2} Z Y}^{(n)}\left(x_{1}^{n}, x_{2}^{n}, z^{n}, y^{n}\right):=\prod_{i=1}^{n} q_{X_{1} X_{2} Z Y}\left(x_{1 i}, x_{2 i}, z_{i}, y_{i}\right)$ (see Figure 1).

Definition 1: A $\left(2^{n R_{1}}, 2^{n R_{2}}, 2^{n R_{00}}, 2^{n R_{01}}, 2^{n R_{02}}, n\right)$ code consists of two randomized encoders $p^{E_{1}}\left(m_{1} \mid s_{0}, s_{1}, x_{1}^{n}\right)$ and $p^{E_{2}}\left(m_{2} \mid s_{0}, s_{2}, x_{2}^{n}\right)$ and a randomized decoder $p^{D}\left(y^{n} \mid s_{1}, s_{2}, m_{1}, m_{2}, z^{n}\right)$, where $s_{0} \in\left[\begin{array}{lll}1 & : & \left.2^{n R_{00}}\right] \text { and }\end{array}\right.$ $s_{j} \in\left[1: 2^{n R_{0 j}}\right], m_{j} \in\left[1: 2^{n R_{j}}\right], j=1,2$.

The joint distribution of $\left(S_{0}, S_{1}, S_{2}, X_{1}^{n}, X_{2}^{n}, Z^{n}, M_{1}\right.$, $M_{2}, Y^{n}$ ) and the resulting induced joint distribution on $\left(X_{1}^{n}, X_{2}^{n}, Z^{n}, Y^{n}\right)$ are respectively given by

$$
\begin{aligned}
& p\left(s_{0}, s_{1}, s_{2}, x_{1}^{n}, x_{2}^{n}, z^{n}, m_{1}, m_{2}, y^{n}\right) \\
& \quad=\frac{1}{2^{n\left(R_{00}+R_{01}+R_{02}\right)}} p\left(x_{1}^{n}, x_{2}^{n}, z^{n}\right) \prod_{j=1}^{2} p^{\mathrm{E}_{j}}\left(m_{j} \mid s_{0}, s_{j}, x_{j}^{n}\right) \\
& \quad \times p^{\mathrm{D}}\left(y^{n} \mid s_{1}, s_{2}, m_{1}, m_{2}, z^{n}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
& p^{\mathrm{ind}}\left(x_{1}^{n}, x_{2}^{n}, z^{n}, y^{n}\right) \\
& \quad=\sum_{s_{0}, s_{1}, s_{2}, m_{1}, m_{2}} p\left(s_{0}, s_{1}, s_{2}, x_{1}^{n}, x_{2}^{n}, z^{n}, m_{1}, m_{2}, y^{n}\right)
\end{aligned}
$$

Recall that the total variation between two p.m.f.'s $p_{X}$ and $q_{X}$ on the same alphabet $\mathcal{X}$ is defined as

$$
\left\|p_{X}-q_{X}\right\|_{1} \triangleq \frac{1}{2} \sum_{x \in \mathcal{X}}\left|p_{X}(x)-q_{X}(x)\right|
$$

Definition 2: A rate tuple $\left(R_{1}, R_{2}, R_{00}, R_{01}, R_{02}\right)$ is said to be achievable for a distribution $q_{X_{1} X_{2} Z Y}$ if there exists a sequence of $\left(2^{n R_{1}}, 2^{n R_{2}}, 2^{R_{00}}, 2^{n R_{01}}, 2^{n R_{02}}, n\right)$ codes such that

$$
\begin{equation*}
\lim _{n \rightarrow \infty}\left\|p_{X_{1}^{n}, X_{2}^{n}, Z^{n}, Y^{n}}^{\mathrm{ind}}-q_{X_{1} X_{2} Z Y}^{(n)}\right\|_{1}=0 \tag{1}
\end{equation*}
$$

where $q_{X_{1} X_{2} Z Y}^{(n)}$ is the product distribution given by

$$
q_{X_{1} X_{2} Z Y}^{(n)}\left(x_{1}^{n}, x_{2}^{n}, z^{n}, y^{n}\right):=\prod_{i=1}^{n} q_{X_{1} X_{2} Z Y}\left(x_{1 i}, x_{2 i}, z_{i}, y_{i}\right)
$$

Remark 1: If $q_{X_{1} X_{2} Z Y}$ is such that $Y$ is a deterministic function of ( $X_{1}, X_{2}, Z$ ), then the pairwise shared randomness and the private randomness at the encoders and the decoder do not have any effect on the communication rates. In fact, more generally, common randomness available to both the encoders and the decoder does not help to improve communication rates in this case.

Definition 3: The rate region $\mathcal{R}_{\mathrm{MAC} \text {-coord }}$ is the closure of the set of all achievable rate tuples $\left(R_{1}, R_{2}, R_{00}, R_{01}, R_{02}\right)$.


Fig. 2. Strong coordination over a multiple-access network with no encoder shared randomness. Encoder $j \in\{1,2\}$ on observing the source $X_{j}^{n}$ and shared randomness $S_{j}$ sends a message $M_{j}$ over a noiseless link to the decoder, which has side-information $Z^{n}$. Here $\left(X_{1 i}, X_{2 i}, Z_{i}\right), i=1, \ldots, n$, are i.i.d. with $q_{X_{1} X_{2} Z}$. Also $\left(X_{1}^{n}, X_{2}^{n}, Z^{n}\right), S_{1}$, and $S_{2}$ are mutually independent. The decoder has to output $Y^{n}$ such that $\left(X_{1 i}, X_{2 i}, Z_{i}, Y_{i}\right)$, $i=1, \ldots, n$, are approximately i.i.d. with $q_{X_{1} X_{2} Z Y}$.

Let $\mathcal{R}_{\text {MAC-coord, UL- }\left(S_{0}, S_{1}, S_{2}\right)}$ be the rate region when all the three pairwise shared randomness are unlimited, i.e.,

$$
\begin{align*}
& \mathcal{R}_{\mathrm{MAC}-\text { coord, UL- }\left(S_{0}, S_{1}, S_{2}\right)}=\left\{\left(R_{1}, R_{2}\right): \exists R_{00}, R_{01} \text { and } R_{02}\right. \\
& \text { s.t. } \left.\left(R_{1}, R_{2}, R_{00}, R_{01}, R_{02}\right) \in \mathcal{R}_{\mathrm{MAC} \text {-coord }}\right\} . \tag{2}
\end{align*}
$$

For purposes of comparison, we separately consider a special case when the encoder shared randomness $S_{0}$ is absent. A code, an achievable rate tuple, and the rate region can be defined analogously. In particular, the code and an achievable rate tuple can be defined similar to Definitions 1 and 2 by removing the respective coordinates containing $R_{00}$ and by treating $S_{0}=\emptyset$. The rate region $\mathcal{R}_{\text {MAC-coord }}^{\mathrm{NO}-S_{0}}$ is the closure of the set of all achievable rate tuples $\left(R_{1}, R_{2}, R_{01}, R_{02}\right)$ when $S_{0}$ is absent. Let $\mathcal{R}_{\text {MAC-coord, UL- } S_{2}}^{\mathrm{NO}-S_{0}}$ be the rate region when the shared randomness $S_{2}$ is unlimited, i.e.,

$$
\begin{align*}
& \mathcal{R}_{\text {MAC-coord, UL-S }}^{\text {NO- } S_{0}}=\left\{\left(R_{1}, R_{2}, R_{01}\right): \exists R_{02}\right. \text { s.t. } \\
& \left.\left(R_{1}, R_{2}, R_{01}, R_{02}\right) \in \mathcal{R}_{\text {MAC-coord }}^{\mathrm{NO}-S_{0}}\right\} . \tag{3}
\end{align*}
$$

The communication rate region $\mathcal{R}_{\mathrm{MAC} \text {-coord, } \mathrm{UL}-\left(S_{1}, S_{2}\right)}^{\mathrm{NO}-\mathrm{S}_{0}}$ is given by $\left\{\left(R_{1}, R_{2}\right): \exists R_{01}\right.$ and $R_{02}$ s.t. $\left(R_{1}, R_{2}, R_{01}, R_{02}\right) \in$ $\left.\mathcal{R}_{\text {MAC-coord }}^{\mathrm{NO}-S_{0}}\right\}$. The communication rate region can be thought of as the trade-off between the communication rates $R_{1}$ and $R_{2}$ under sufficiently large pairwise shared randomness rates $R_{01}$ and $R_{02}$.

## III. Rate Region With No Encoder Shared Randomness

In this section, we present our results for the special case when the encoder shared randomness $S_{0}$ is absent (see Figure 2). Our main result here is a complete characterization of the rate region $\mathcal{R}_{\text {MAC-coord, UL-S }}^{\mathrm{NO}-\mathrm{S}_{2}}$ when the sources are conditionally independent given the side-information. We first present an inner bound to the rate region $\mathcal{R}_{\text {MAC-coord }}^{\mathrm{NO}-S_{0}}$ that follows from Haddadpour et al. [37, Th. 3] which studies multiple-access channel simulation using another multipleaccess channel as a resource (instead of a multiple-access network of noiseless links as in this work).

Theorem 1 (Inner Bound With No Encoder Shared Randomness): Given a p.m.f. $q_{X_{1} X_{2} Z Y}$, the rate tuple ( $R_{1}, R_{2}, R_{01}, R_{02}$ ) is in $\mathcal{R}_{\text {MAC-coord }}^{N O-S_{0}}$ if

$$
\begin{aligned}
& R_{1} \geq I\left(U_{1} ; X_{1} \mid U_{2}, Z, T\right) \\
& R_{2} \geq I\left(U_{2} ; X_{2} \mid U_{1}, Z, T\right) \\
& R_{1}+R_{2} \geq I\left(U_{1}, U_{2} ; X_{1}, X_{2} \mid Z, T\right) \\
& R_{1}+R_{01} \geq I\left(U_{1} ; X_{1}, X_{2}, Y \mid Z, T\right) \\
&-I\left(U_{1} ; U_{2} \mid Z, T\right) \\
& R_{2}+R_{02} \geq I\left(U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right) \\
&-I\left(U_{1} ; U_{2} \mid Z, T\right) \\
& R_{1}+R_{2}+R_{01} \geq I\left(U_{1} ; X_{1}, X_{2}, Y \mid Z, T\right) \\
&+I\left(U_{2} ; X_{2} \mid U_{1}, Z, T\right) \\
& R_{1}+R_{2}+R_{02} \geq I\left(U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right) \\
&+I\left(U_{1} ; X_{1} \mid U_{2}, Z, T\right) \\
& R_{1}+R_{2}+R_{01}+R_{02} \geq I\left(U_{1}, U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right)
\end{aligned}
$$

for some p.m.f.

$$
\begin{align*}
& p\left(x_{1}, x_{2}, z, t, u_{1}, u_{2}, y\right)= \\
& \quad p\left(x_{1}, x_{2}, z\right) p(t) p\left(u_{1} \mid x_{1}, t\right) p\left(u_{2} \mid x_{2}, t\right) p\left(y \mid u_{1}, u_{2}, z, t\right) \tag{4}
\end{align*}
$$

such that

$$
\sum_{u_{1}, u_{2}} p\left(x_{1}, x_{2}, z, u_{1}, u_{2}, y \mid t\right)=q\left(x_{1}, x_{2}, z, y\right), \text { for all } t
$$

Remark 2: The inner bound in Theorem 1 without sideinformation $Z$ follows as a corollary of Haddadpour et al. [37, Th. 3]. In particular, let the resource mutiple access channel in [37, Th. 3] consists of two independent channels which can be converted to two noiseless links by operating at the rates of respective channel capacities. The details are analogous to how the inner bound of point-to-point channel simulation using a noiseless link [5] can be recovered from that of point-to-point channel simulation using another channel [37, Th. 1] (see [37, Remark 2]). However, for completeness, we present a proof of Theorem 1 incorporating the sideinformation $Z$ with minor differences to that of [37, Th. 3] in Appendix A.

The intuition behind the auxiliary random variables $U_{1}$ and $U_{2}$ is analogous to the auxiliary random variable in the point-to-point channel simulation setting [5], [6]. In particular, $U_{1}$ and $U_{2}$ may be thought of as quantized versions of the observations $X_{1}$ and $X_{2}$ respectively (with respect to the corresponding shared random variables). The Markov conditions on $U_{1}$ and $U_{2}$ in (4) arises naturally due to the information structure of the problem. We note that the inner bound in Theorem 1 with $Z=\emptyset$ also appears in the concurrent work by Atif et al. [43, Th. 1]. We now provide an outer bound to the region $\mathcal{R}_{\text {MAC-coord }}^{\mathrm{NO}-S_{0}}$.

Theorem 2 (Outer Bound With No Encoder Shared Randomness): Given a p.m.f. $q_{X_{1} X_{2} Z Y}$, any rate tuple $\left(R_{1}, R_{2}, R_{01}, R_{02}\right)$ in $\mathcal{R}_{\text {MAC-coord }}^{N O-S_{0}}$ satisfies, for every $\epsilon \in\left(0, \frac{1}{4}\right]$,

$$
\begin{gathered}
R_{1} \geq \max \left\{I\left(U_{1} ; X_{1} \mid Z, T\right)\right. \\
\left.I\left(U_{1} ; X_{1} \mid U_{2}, X_{2}, Z, T\right)\right\} \\
R_{2} \geq \max \left\{I\left(U_{2} ; X_{2} \mid Z, T\right)\right. \\
\left.\quad I\left(U_{2} ; X_{2} \mid U_{1}, X_{1}, Z, T\right)\right\}
\end{gathered}
$$

$$
\begin{aligned}
R_{1}+R_{2} & \geq I\left(U_{1}, U_{2} ; X_{1}, X_{2} \mid Z, T\right) \\
R_{1}+R_{01} & \geq I\left(U_{1} ; X_{1}, X_{2}, Y \mid Z, T\right)-g(\epsilon) \\
R_{2}+R_{02} & \geq I\left(U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right)-g(\epsilon) \\
R_{1}+R_{2}+R_{01}+R_{02} & \geq I\left(U_{1}, U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right)-g(\epsilon),
\end{aligned}
$$

with $g(\epsilon)=2 \sqrt{\epsilon}\left(H_{q}\left(X_{1}, X_{2}, Y, Z\right)+\log \frac{\left(\left|\mathcal{X}_{1}\left\|\mathcal{X}_{2}\right\| \mathcal{Y} \| \mathcal{Z}\right|\right)}{\epsilon}\right)$ (which tends to 0 as $\epsilon \rightarrow 0$ ), for some p.m.f.

$$
\begin{aligned}
& p\left(x_{1}, x_{2}, z, t, u_{1}, u_{2}, y\right)= \\
& \quad p\left(x_{1}, x_{2}, z\right) p(t) p\left(u_{1}, u_{2} \mid x_{1}, x_{2}, t\right) p\left(y \mid u_{1}, u_{2}, z, t\right)
\end{aligned}
$$

such that

$$
\begin{array}{r}
p\left(u_{1} \mid x_{1}, x_{2}, z, t\right)=p\left(u_{1} \mid x_{1}, t\right) \\
p\left(u_{2} \mid x_{1}, x_{2}, z, t\right)=p\left(u_{2} \mid x_{2}, t\right) \\
\left\|p\left(x_{1}, x_{2}, y, z \mid t\right)-q\left(x_{1}, x_{2}, y, z\right)\right\|_{1} \leq \epsilon \text { for all } t
\end{array}
$$

When the random variables $X_{1}$ and $X_{2}$ are conditionally independent given $Z$, and the shared randomness rate $R_{02}$ is unlimited, we can show that the inner bound of Theorem 1 is tight by obtaining an outer bound which is strictly stronger than that in Theorem 2 along with cardinality bounds on the auxiliary random variables that allows us to prove the continuity of the outer bound at $\epsilon=0$, thereby completely characterizing the rate region $\mathcal{R}_{\text {MAC-coord, UL-S }}^{\mathrm{NO}-S_{0}}$.

Theorem 3 (Rate Region - Conditionally Independent Sources Given the Side Information): Consider a p.m.f. $q_{X_{1} X_{2} Z Y}$ such that the random variables $X_{1}$ and $X_{2}$ are conditionally independent given $Z$, i.e., $I\left(X_{1} ; X_{2} \mid Z\right)=0$. Then the rate region $\mathcal{R}_{\mathrm{MAC} \text {-coord, UL-S }}^{\mathrm{NO} \mathrm{S}_{2}}$ is given by the set of all rate tuples $\left(R_{1}, R_{2}, R_{01}\right)$ such that

$$
\begin{aligned}
R_{1} & \geq I\left(U_{1} ; X_{1} \mid Z, T\right) \\
R_{2} & \geq I\left(U_{2} ; X_{2} \mid Z, T\right) \\
R_{1}+R_{01} & \geq I\left(U_{1} ; X_{1}, Y \mid X_{2}, Z, T\right)
\end{aligned}
$$

for some p.m.f.

$$
\begin{align*}
& p\left(x_{1}, x_{2}, z, t, u_{1}, u_{2}, y\right)= \\
& p(z) p\left(x_{1} \mid z\right) p\left(x_{2} \mid z\right) p(t) p\left(u_{1} \mid x_{1}, t\right) p\left(u_{2} \mid x_{2}, t\right) p\left(y \mid u_{1}, u_{2}, z, t\right) \tag{5}
\end{align*}
$$

such that $\sum_{u_{1}, u_{2}} p\left(x_{1}, x_{2}, z, u_{1}, u_{2}, y \mid t\right)=q\left(x_{1}, x_{2}, z, y\right)$, for all $t$, with $\left|\mathcal{U}_{1}\right| \leq\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\right||\mathcal{Y}||\mathcal{Z}|,\left|\mathcal{U}_{2}\right| \leq\left|\mathcal{U}_{1}\right|\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\right||\mathcal{Y}||\mathcal{Z}|$, and $|\mathcal{T}| \leq 3$.

In Remark 5 (on page 7580), we show that the outer bound implicit in Theorem 3 is strictly stronger than that of Theorem 2 (even after assuming its continuity at $\epsilon=$ 0 ). The non-trivial part in the converse of Theorem 3 is that we single-letterize the distributed protocol in order to obtain a p.m.f. structure matching that of the inner bound in Theorem 1, particularly leveraging the conditional independence of the sources given the side-information. In general, obtaining single-letter forms matching the inner bound is known to be notoriously difficult for distributed source coding problems [46], [47]. It is interesting to note that for the case of deterministic function computation when the sources are
conditionally independent given the side-information, the inner and outer bounds of Sefidgaran and Tchamkerten [31] specialized to the two-user multiple-access network of noiseless links match, analogous to a result of Gastpar [49]. However, for randomized function computation, the inner and outer bounds in Theorems 1 and 2 (again after assuming the continuity of the outer bound at $\epsilon=0$ for the sake of comparison) do not match for sources conditionally independent given sideinformation and we need a strictly stronger outer bound to show Theorem 3. A detailed proof of Theorem 3 is given in Section VI-A. 1 (achievability) and Section VI-B. 3 (converse).

Remark 3: As mentioned earlier (see Remark 1), when $Y$ is a deterministic function of $\left(X_{1}, X_{2}, Z\right)$, pairwise shared randomness (or common randomness shared by all users) does not have any effect on the communication rates. Indeed, the rate constraints in Theorem 3 involving shared randomness rates become redundant as we show in Appendix C. We also show in Appendix C that Theorem 3 reduces to the main result of Sefidgaran and Tchamkerten [31, Th. 3] specialized to the two-user multiple-access network of noiseless links, also reported in [50, Th. 3]. In appendix B, we show that for deterministic function computation, Theorem 1 reduces to Theorem 2 of Sefidgaran and Tchamkerten [31] specialized to the two-user multiple-access network of noiseless links, also reported in [50, Proposition 1].

Remark 4: Theorem 3 recovers the point-to-point channel simulation results [5, Th. II.1], [6, Th. 1], [8, Th. 1] (specialized to a single round of interaction) when $X_{2}=\emptyset$.

Sefidgaran and Tchamkerten [50] already observed that their inner bound (and hence, our Theorem 1, which recovers their inner bound as shown in Appendix B) is not tight, in general. The deterministic function computation problem of Körner and Marton [17] illustrates this (see [51, Example 2]). For computing the mod- 2 sum of binary $X_{1}$ and $X_{2}$ with symmetric input distribution, Körner and Marton [17] showed that structured codes can strictly outperform standard random coding schemes. Achievable schemes using algebraic-structured codes exploiting the specific structure of the function to be computed were further explored by Krithivasan and Pradhan [44] and Atif and Pradhan [45], where the latter considered the quantum setting.

## IV. Encoder Shared Randomness Can Strictly Reduce the Communication Rates

In this section, we show that if the encoders share additional independent randomness (see Figure 1), the communication rates in some cases can be strictly improved, even if the additional randomness is not available to the decoder. This is done via an example for which we first explicitly compute the communication rate region of Theorem 3 when there is no shared randomness between the encoders, assuming sufficiently large pairwise shared randomness rates. Then we show that a rate pair outside this region is achievable in the presence of shared randomness between the encoders.

Example 1: Let $X_{1}=\left(X_{11}, X_{12}\right)$ be a vector of two independent and uniformly distributed binary random variables. Similarly, let $X_{2}=\left(X_{21}, X_{22}\right)$ be another vector of

Fig. 3. The communication rate regions for $X_{k}=\left(X_{k 1}, X_{k 2}\right), k \in\{1,2\}$ ( $X_{11}, X_{12}, X_{21}, X_{22}$ are mutually independent, uniform binary random variables) and $Y=\left(X_{1 J}, X_{2 J}\right)$, where $J$ is uniform on $\{1,2\}$ and independent of $\left(X_{1}, X_{2}\right)$. The region is defined by the constraints $R_{1} \geq 1$, $R_{2} \geq 1$, and $R_{1}+R_{2} \geq 3$ (shown via solid line) when the shared randomness between each encoder and the decoder are of sufficiently large rates. When an additional shared randomness is available between the encoders, the region is defined by the constraints $R_{1} \geq 1$ and $R_{2} \geq 1$ (shown via dotted line) provided the shared randomness between all the three pairs are of sufficiently large rates (see Section V-A).
two independent and uniformly distributed binary random variables independent of $X_{1}$. Consider simulating a channel $q_{Y \mid X_{1} X_{2}}$ with $Y=\left(X_{1 J}, X_{2 J}\right)$, where $J$ is a random variable uniformly distributed on $\{1,2\}$ and independent of $\left(X_{1}, X_{2}\right)$. For simplicity, we let $Z=\emptyset$, i.e. there is no side information at the decoder. Let us assume unlimited rates $R_{01}$ and $R_{02}$.

When there is no additional shared randomness between the encoders, from Theorem 3, the communication rate region $\mathcal{R}_{\text {MAC-coord, UL- }\left(S_{1}, S_{2}\right)}^{\mathrm{NO}-S_{0}}$ is given by the set of all rate pairs ( $R_{1}, R_{2}$ ) such that

$$
\begin{align*}
& R_{1} \geq I\left(U_{1} ; X_{1} \mid T\right),  \tag{6}\\
& R_{2} \geq I\left(U_{2} ; X_{2} \mid T\right), \tag{7}
\end{align*}
$$

for some p.m.f.

$$
\begin{align*}
& p\left(x_{1}, x_{2}, t, u_{1}, u_{2}, y\right)= \\
& \quad p\left(x_{1}\right) p\left(x_{2}\right) p(t) p\left(u_{1} \mid x_{1}, t\right) p\left(u_{2} \mid x_{2}, t\right) p\left(y \mid u_{1}, u_{2}, t\right) \tag{8}
\end{align*}
$$

satisfying

$$
\begin{equation*}
\sum_{u_{1}, u_{2}} p\left(x_{1}, x_{2}, u_{1}, u_{2}, y \mid t\right)=q\left(x_{1}, x_{2}, y\right) \tag{9}
\end{equation*}
$$

for all $t$. The following proposition (proved at the end of this section) explicitly characterizes the communication rate region for this $q_{X_{1} X_{2} Y}$.

Proposition 1: For the joint distribution $q_{X_{1} X_{2} Y}$ in Example 1, the communication rate region of Theorem 3 under sufficiently large shared randomness rates is equal to the region defined by the constraints $R_{1} \geq 1, R_{2} \geq 1$, and $R_{1}+R_{2} \geq 3$.

Now we show that if there exists an additional source of shared randomness between the encoders, then the rate pair $\left(R_{1}, R_{2}\right)=(1,1)$ is achievable (see Figure 3). In particular, we prove that if this additional shared randomness is of rate at least 1 , then a rate pair $\left(R_{1}, R_{2}\right)=(1,1)$ is achievable under sufficiently large shared randomness rates $R_{01}$ and $R_{02}$. To see this, notice that both the encoders, using a shared randomness of rate 1 , can sample a sequence $W_{1}, W_{2}, \cdots, W_{n}$ i.i.d. distributed on $\{1,2\}$ such that $p_{W}(1)=p_{W}(2)=$ 0.5 and independent of $\left(X_{1}^{n}, X_{2}^{n}\right)$. Now we invoke Theorem 1 with $\left(X_{i}, W\right)$ as the input source to the encoder- $i, i=1,2$.

This implies that a rate pair $\left(R_{1}, R_{2}\right)$ is achievable under sufficiently large shared randomness rates $R_{01}$ and $R_{02}$ if

$$
\begin{aligned}
R_{1} & \geq I\left(U_{1} ; X_{1}, W \mid U_{2}\right) \\
R_{2} & \geq I\left(U_{2} ; X_{2}, W \mid U_{1}\right) \\
R_{1}+R_{2} & \geq I\left(U_{1}, U_{2} ; X_{1}, X_{2}, W\right),
\end{aligned}
$$

for some p.m.f.

$$
\begin{align*}
& p\left(x_{1}, x_{2}, w, u_{1}, u_{2}, y\right)= \\
& \quad q\left(x_{1}, x_{2}, w\right) p\left(u_{1} \mid x_{1}, w\right) p\left(u_{2} \mid x_{2}, w\right) p\left(y \mid u_{1}, u_{2}\right) \tag{10}
\end{align*}
$$

It is easy to see that $U_{1}=X_{1 W}$ and $U_{2}=X_{2 W}$ satisfy the conditions on the structure of the probability distribution $p\left(x_{1}, x_{2}, w, u_{1}, u_{2}, y\right)$. This gives

$$
\begin{aligned}
I\left(X_{1 W} ; X_{1}, W \mid X_{2 W}\right) & =1 \\
I\left(X_{2 W} ; X_{2}, W \mid X_{1 W}\right) & =1 \\
I\left(X_{1 W}, X_{2 W} ; X_{1}, X_{2}, W\right) & =2
\end{aligned}
$$

which imply that a rate pair $(1,1)$ is achievable, thereby strictly improving over the rate region without encoder shared randomness, defined by the constraints $R_{1} \geq 1, R_{2} \geq 1$, and $R_{1}+R_{2} \geq 3$.

Remark 5: We remark that the outer bound implicit in Theorem 3 is strictly stronger than that of Theorem 2 (even after assuming its continuity at $\epsilon=0$ ) for the p.m.f. $q_{X_{1} X_{2} Z Y}$ in Example 1. We first observe that that the communication rate pair $(1,1)$ is contained in the outer bound given by Theorem 2 (under unlimited shared randomness rates and the assumption of continuity at $\epsilon=0$ ). To see this, first notice that the choice of $U_{1}=\left(X_{1 J}, J\right), U_{2}=\left(X_{2 J}, J\right)$, where $J$ is a random variable uniformly distributed on $\{1,2\}$ and independent of $\left(X_{1}, X_{2}\right)$, satisfies the conditions on the structure of the p.m.f. in the outer bound of Theorem 2. Now we evaluate the bounds on communication rates in Theorem 2 for this choice of auxiliary random variables.

$$
\begin{aligned}
\max \left\{I\left(X_{1 J}, J ; X_{1}\right), I\left(X_{1 J} ; X_{1} \mid X_{2 J}, J, X_{2}\right)\right\} & =1 \\
\max \left\{I\left(X_{2 J}, J ; X_{2}\right), I\left(X_{2 J} ; X_{2} \mid X_{1 J}, J, X_{1}\right)\right\} & =1 \\
I\left(X_{1 J}, X_{2 J}, J ; X_{1}, X_{2}\right) & =2
\end{aligned}
$$

This implies that a rate pair $\left(R_{1}, R_{2}\right)=(1,1)$ is contained in the outer bound on the communication rate region implied by Theorem 2 (under sufficiently large shared randomness rates and the assumption of continuity at $\epsilon=0$ ). However, by Proposition 1, this rate pair lies outside the communication rate region implied by Theorem 3.

We conclude the section with a proof of Proposition 1.
Proof of Proposition 1: For the achievability, it suffices to show that the corner points $(1,2)$ and $(2,1)$ are in the region defined by (6) - (9). By symmetry, it is enough to show that there exists a p.m.f.

$$
p\left(x_{1}, x_{2}, u_{1}, u_{2}, y\right)=p\left(x_{1}, x_{2}\right) p\left(u_{1} \mid x_{1}\right) p\left(u_{2} \mid x_{2}\right) p\left(y \mid u_{1}, u_{2}\right)
$$

such that $\sum_{u_{1}, u_{2}} p\left(x_{1}, x_{2}, u_{1}, u_{2}, y\right)=q\left(x_{1}, x_{2}, y\right)$, $I\left(U_{1} ; X_{1}\right)=2$, and $I\left(U_{2} ; X_{2}\right)=1$. It is easy to see that $U_{1}=X_{1}$ and $U_{2}=\left(J, X_{2 J}\right)$, where $J$ is a random variable uniformly distributed on $\{1,2\}$ and independent of
( $X_{1}, X_{2}$ ), satisfy the conditions on the structure of the joint probability distribution $p\left(x_{1}, x_{2}, u_{1}, u_{2}, y\right)$. Now $I\left(U_{1} ; X_{1}\right)=$ $H\left(X_{1}\right)=2$ and $I\left(U_{2} ; X_{2}\right)=I\left(J, X_{2 J} ; X_{2}\right)=$ $I\left(X_{2 J} ; X_{2} \mid J\right)=1$.

For the converse, it suffices to show that for any p.m.f. $p\left(x_{1}, x_{2}, u_{1}, u_{2}, y\right)=p\left(x_{1}, x_{2}\right) p\left(u_{1} \mid x_{1}\right) p\left(u_{2} \mid x_{2}\right) p\left(y \mid u_{1}, u_{2}\right)$ such that $\sum_{u_{1}, u_{2}} p\left(x_{1}, x_{2}, u_{1}, u_{2}, y\right)=q\left(x_{1}, x_{2}, y\right)$, we have $I\left(U_{1} ; X_{1}\right) \geq 1, I\left(U_{2} ; X_{2}\right) \geq 1$, and $I\left(U_{1} ; X_{1}\right)+I\left(U_{2} ;\right.$ $\left.X_{2}\right) \geq 3$. Consider the p.m.f. in (8) for a fixed value of $t$, i.e.,
$p\left(x_{1}, x_{2}, u_{1}, u_{2}, y\right)=p\left(x_{1}\right) p\left(x_{2}\right) p\left(u_{1} \mid x_{1}\right) p\left(u_{2} \mid x_{2}\right) p\left(y \mid u_{1}, u_{2}\right)$
such that

$$
\begin{equation*}
\sum_{u_{1}, u_{2}} p\left(x_{1}, x_{2}, u_{1}, u_{2}, y\right)=q\left(x_{1}, x_{2}, y\right) \tag{11b}
\end{equation*}
$$

Note that the independence of $X_{1}$ and $X_{2}$ along with the long Markov chain $U_{1} \rightarrow X_{1} \rightarrow X_{2} \rightarrow U_{2}$ implies the independence of $\left(U_{1}, X_{1}\right)$ and $\left(U_{2}, X_{2}\right)$. The following three cases now arise based on $H\left(X_{1} \mid U_{1}\right)$ and $H\left(X_{2} \mid U_{2}\right)$.

Case 1: $H\left(X_{1} \mid U_{1}\right)=0$,
Case 2: $H\left(X_{2} \mid U_{2}\right)=0$,
Case 3: $H\left(X_{1} \mid U_{1}\right)>0$ and $H\left(X_{2} \mid U_{2}\right)>0$.
Case $1\left(H\left(X_{1} \mid U_{1}\right)=0\right)$ :
We have $I\left(X_{1} ; U_{1}\right)=H\left(X_{1}\right)=2$. Now we prove that $I\left(U_{2} ; X_{2}\right) \geq 1$. We show this by contradiction. Suppose that $I\left(U_{2} ; X_{2}\right)<1$. Then $H\left(X_{2} \mid U_{2}\right)=H\left(X_{2}\right)-I\left(U_{2} ; X_{2}\right)=$ $2-I\left(U_{2} ; X_{2}\right)>1$ and hence there exists a $u_{2}$ with $P\left(U_{2}=\right.$ $\left.u_{2}\right)>0$ such that $p_{X_{2} \mid U_{2}=u_{2}}$ has a support whose size is larger than 2. Notice that the Markov chain $Y \rightarrow\left(X_{1}, U_{2}\right) \rightarrow$ $X_{2}$ holds because

$$
\begin{aligned}
I\left(Y ; X_{2} \mid X_{1}, U_{2}\right) & \leq I\left(Y, U_{1} ; X_{2} \mid X_{1}, U_{2}\right) \\
& =I\left(U_{1} ; X_{2} \mid X_{1}, U_{2}\right)+I\left(Y ; X_{2} \mid U_{1}, U_{2}, X_{1}\right) \\
& \leq I\left(U_{1}, X_{1} ; U_{2}, X_{2}\right)+I\left(Y ; X_{1}, X_{2} \mid U_{1}, U_{2}\right) \\
& =0
\end{aligned}
$$

where the last equality follows because $\left(U_{1}, X_{1}\right)$ is independent of $\left(U_{2}, X_{2}\right)$ and the Markov chain $Y \rightarrow\left(U_{1}, U_{2}\right) \rightarrow$ ( $X_{1}, X_{2}$ ) holds.

Suppose $p_{X_{2} \mid U_{2}=u_{2}}$ has the support which is a superset of $\{(0,0),(1,0),(0,1)\}$. Consider the induced distribution $p_{Y \mid X_{1}=(0,1), U_{2}=u_{2}}$. This is well-defined because $P\left(X_{1}=\right.$ $\left.(0,1), U_{2}=u_{2}\right)>0$ as $X_{1}$ is independent of $U_{2}$ and $P\left(X_{1}=\right.$ $(0,1)), P\left(U_{2}=u_{2}\right)>0$. Since $P\left(X_{2}=(0,1) \mid X_{1}=\right.$ $\left.(0,1), U_{2}=u_{2}\right)=P\left(X_{2}=(0,1) \mid U_{2}=u_{2}\right)>0$ and $Y \rightarrow\left(X_{1}, U_{2}\right) \rightarrow X_{2}$, we have $P\left(Y=(1,0) \mid X_{1}=\right.$ $\left.(0,1), U_{2}=u_{2}\right)=P\left(Y=(0,1) \mid X_{1}=(0,1), U_{2}=\right.$ $\left.u_{2}\right)=0$. Since $P\left(X_{2}=(1,0) \mid X_{1}=(0,1), U_{2}=u_{2}\right)=$ $P\left(X_{2}=(1,0) \mid U_{2}=u_{2}\right)>0$ and $Y \rightarrow\left(X_{1}, U_{2}\right) \rightarrow X_{2}$, we have $P\left(Y=(1,1) \mid X_{1}=(0,1), U_{2}=u_{2}\right)=P(Y=$ $\left.(0,0) \mid X_{1}=(0,1), U_{2}=u_{2}\right)=0$. This is a contradiction since $p_{Y \mid X_{1}=(0,1), U_{2}=u_{2}}$ has to be a probability distribution.

Suppose $p_{X_{2} \mid U_{2}=u_{2}}$ has the support which is a superset of $\{(0,0),(1,1),(0,1)\}$. Since $P\left(X_{2}=(0,0) \mid X_{1}=\right.$ $\left.(0,0), U_{2}=u_{2}\right)=P\left(X_{2}=(0,0) \mid U_{2}=u_{2}\right)>0$ and
$Y \rightarrow\left(X_{1}, U_{2}\right) \rightarrow X_{2}$, we have $P\left(Y=(1,0) \mid X_{1}=\right.$ $\left.(0,0), U_{2}=u_{2}\right)=P\left(Y=(0,1) \mid X_{1}=(0,0), U_{2}=u_{2}\right)=$ $P\left(Y=(1,1) \mid X_{1}=(0,0), U_{2}=u_{2}\right)=0$. Since $P\left(X_{2}=\right.$ $\left.(1,1) \mid X_{1}=(0,0), U_{2}=u_{2}\right)=P\left(X_{2}=(1,1) \mid U_{2}=u_{2}\right)>0$ and $Y \rightarrow\left(X_{1}, U_{2}\right) \rightarrow X_{2}$, we have $P(Y=(0,0) \mid$ $\left.X_{1}=(0,0), U_{2}=u_{2}\right)=0$. This is a contradiction since $p_{Y \mid X_{1}=(0,0), U_{2}=u_{2}}$ is a probability distribution.

The other supports $\{(1,1),(0,1),(1,0)\}$ and $\{(0,0),(1,1)$, $(1,0)\}$ can be analysed in a similar manner to arrive at a contradiction. Hence, $I\left(U_{2} ; X_{2}\right) \geq 1$.

Case $2\left(H\left(X_{2} \mid U_{2}\right)=0\right)$ :
By symmetry, the analysis for this case is similar to that of Case 1.

Case $3\left(H\left(X_{1} \mid U_{1}\right)>0\right.$ and $\left.H\left(X_{2} \mid U_{2}\right)>0\right)$ :
We prove the following claim in Appendix D.
Claim 1: When $H\left(X_{1} \mid U_{1}\right)>0$ and $H\left(X_{2} \mid U_{2}\right)>0$, there exists a $k \in\{1,2\}$ such that for all $u_{1}$ and $u_{2}$ with $P\left(U_{1}=\right.$ $\left.u_{1}\right)>0, P\left(U_{2}=u_{2}\right)>0$, we have $H\left(X_{1 k} \mid U_{1}=u_{1}\right)=0$ and $H\left(X_{2 k} \mid U_{2}=u_{2}\right)=0$.

Claim 1 roughly states that, under Case $3, U_{1}$ and $U_{2}$ always reveal the $k^{\text {th }}$ components of $X_{1}$ and $X_{2}$, respectively, for a fixed $k \in\{1,2\}$. Let us assume $k=1$ without loss of generality, i.e., $U_{i}$ conveys atleast $X_{i 1}$ losslessly, for $i=1,2$. Let $A_{i}$ be the event " $U_{i}$ conveys both $X_{i 1}$ and $X_{i 2}$ losslessly", i.e., for $i=1,2$,

$$
\begin{equation*}
A_{i}=\bigcup_{u_{i}: H\left(X_{i 1} \mid U_{i}=u_{i}\right)=H\left(X_{i 2} \mid U_{i}=u_{i}\right)=0}\left(U_{i}=u_{i}\right) \tag{12}
\end{equation*}
$$

Let, for $i=1,2$,

$$
p_{i}:=P\left(A_{i}\right)=\sum_{\substack{u_{i}:}} P\left(U_{i}=u_{i}\right) .
$$

For $i=1,2$, it follows that $A_{i}^{c}$ is the event " $U_{i}$ conveys only $X_{i 1}$ losslessly" from Claim 1 along with our assumption that $k=1$. Now, in view of the independence of $\left(U_{1}, X_{1}\right)$ and $\left(U_{2}, X_{2}\right)$, it follows that $1-p_{1} p_{2}=1-P\left(A_{1} \cap A_{2}\right)=$ $P\left(A_{1}^{c} \cup A_{2}^{c}\right)$. Notice that this event $A_{1}^{c} \cup A_{2}^{c}$ must be a subset of the event $(J=1)$, otherwise the correctness condition (11b) is violated. Thus we have

$$
\begin{equation*}
1-p_{1} p_{2} \leq P(J=1)=0.5 \tag{13}
\end{equation*}
$$

Also, we have

$$
\begin{align*}
& I\left(U_{1} ; X_{1}\right) \\
& =H\left(X_{1}\right)-H\left(X_{1} \mid U_{1}\right) \\
& =2-\sum_{u_{1}} P\left(U_{1}=u_{1}\right) H\left(X_{1} \mid U_{1}=u_{1}\right) \\
& \stackrel{(a)}{=} 2-\sum_{\substack{u_{1}: \\
H\left(X_{11} \mid U_{1}=u_{1}\right)=0, H\left(X_{12} \mid U_{1}=u_{1}\right)>0}} P\left(U_{1}=u_{1}\right) H\left(X_{1} \mid U_{1}=u_{1}\right) \\
& -\sum_{\substack{u_{1}: \\
H\left(X_{11} \mid \\
H\left(X_{12} \mid U_{1}=u_{1}\right)=0, u_{1}\right)=0}} P\left(U_{1}=u_{1}\right) H\left(X_{1} \mid U_{1}=u_{1}\right) \\
& \geq 2-\sum_{\substack{u_{1}: \\
H\left(X_{11} \mid U_{1}=u_{1}\right)=0, H\left(X_{12} \mid U_{1}=u_{1}\right)>0}} P\left(U_{1}=u_{1}\right) H\left(X_{12} \mid U_{1}=u_{1}\right) \\
& \geq 2-\left(1-p_{1}\right) \\
& =1+p_{1} \text {, } \tag{14}
\end{align*}
$$

where (a) follows from Claim 1 along with our assumption that $k=1$, since $U_{i}$ reveals atleast $X_{i 1}$ losslessly for $i=1,2$. Similarly, we have $I\left(U_{2} ; X_{2}\right) \geq 1+p_{2}$. From (13), we have $p_{1} p_{2} \geq 0.5$. Now since

$$
\begin{equation*}
\min _{\substack{p_{1}, p_{2} \\ 0 \leq p_{1}, p_{2} \leq 1 \text { and } \\ p_{1} p_{2} \geq 0.5}} p_{1}+p_{2}>1 \tag{15}
\end{equation*}
$$

we have $I\left(U_{1} ; X_{1}\right)+I\left(U_{2} ; X_{2}\right) \geq 2+p_{1}+p_{2} \geq 3$. This proves that the communication rate region is equal to the region defined by the constraints $R_{1} \geq 1, R_{2} \geq 1$, and $R_{1}+R_{2} \geq 3$.

## V. Inner and Outer Bounds on the Rate Region

In this section, we present our results for the general setup when a pairwise shared randomness of limited rate $R_{00}$ is present between the encoders (see Figure 1). In addition, for Example 1, we will show that the (achievable) rate region defined by $R_{1} \geq 1, R_{2} \geq 1$ is indeed optimal with sufficiently large pairwise shared randomness rates (all three pairs). We will also exploit common components [52] between the two sources, i.e. random variables $X_{0}$ such that there exist deterministic functions $f_{1}$ and $f_{2}$ with

$$
\begin{equation*}
X_{0}=f_{1}\left(X_{1}\right)=f_{2}\left(X_{2}\right) \text { a.s. } \tag{16}
\end{equation*}
$$

The following theorem provides an inner bound to the region $\mathcal{R}_{\text {MAC-coord. }}$.

Theorem 4 (Inner Bound With Encoder Shared Randomness): Given a p.m.f. $q_{X_{1} X_{2} Z Y}$, the rate tuple ( $\left.R_{1}, R_{2}, R_{00}, R_{01}, R_{02}\right)$ is in $\mathcal{R}_{\text {MAC-coord }}$ if

$$
\begin{align*}
& R_{00} \geq H\left(U_{0} \mid X_{0}, T\right)  \tag{17}\\
& R_{1} \geq I\left(U_{1} ; X_{1}, U_{0} \mid U_{2}, Z, T\right)  \tag{18}\\
& R_{2} \geq I\left(U_{2} ; X_{2}, U_{0} \mid U_{1}, Z, T\right)  \tag{19}\\
& R_{1}+R_{2} \geq I\left(U_{1}, U_{2} ; X_{1}, X_{2}, U_{0} \mid Z, T\right)  \tag{20}\\
& R_{1}+R_{01} \geq I\left(U_{1} ; X_{1}, X_{2}, U_{0}, Y \mid Z, T\right) \\
&-I\left(U_{1} ; U_{2} \mid Z, T\right)  \tag{21}\\
& R_{2}+R_{02} \geq I\left(U_{2} ; X_{1}, X_{2}, U_{0}, Y \mid Z, T\right) \\
& \quad-I\left(U_{1} ; U_{2} \mid Z, T\right)  \tag{22}\\
& R_{1}+R_{2}+R_{01} \geq I\left(U_{1} ; X_{1}, X_{2}, U_{0}, Y \mid Z, T\right) \\
&+I\left(U_{2} ; X_{2}, U_{0} \mid U_{1}, Z, T\right)  \tag{23}\\
& R_{1}+R_{2}+R_{02} \geq I\left(U_{2} ; X_{1}, X_{2}, U_{0}, Y \mid Z, T\right) \\
& \quad+I\left(U_{1} ; X_{1}, U_{0} \mid U_{2}, Z, T\right)  \tag{24}\\
& R_{1}+R_{2}+R_{01}+R_{02} \geq I\left(U_{1}, U_{2} ; X_{1}, X_{2}, U_{0}, Y \mid Z, T\right) \tag{25}
\end{align*}
$$

for some p.m.f.

$$
\begin{aligned}
& p\left(x_{1}, x_{2}, z, t, u_{0}, u_{1}, u_{2}, y\right)= \\
& p\left(x_{1}, x_{2}, z\right) p(t) p\left(u_{0} \mid x_{0}, t\right) p\left(u_{1} \mid x_{1}, u_{0}, t\right) \\
& \quad \times p\left(u_{2} \mid x_{2}, u_{0}, t\right) p\left(y \mid u_{1}, u_{2}, z, t\right)
\end{aligned}
$$

such that

$$
\sum_{u_{0}, u_{1}, u_{2}} p\left(x_{1}, x_{2}, u_{0}, u_{1}, u_{2}, y, z \mid t\right)=q\left(x_{1}, x_{2}, z, y\right), \text { for all } t
$$

The main idea behind the proof is to make use of the shared randomness between the encoders in order to simulate a common description of $X_{0}^{n}$, viz. $U_{0}^{n}$ at both the encoders approximately distributed according to $q_{U_{0}^{n} \mid X_{0}^{n}}^{(n)}\left(u_{0}^{n} \mid x_{0}^{n}\right):=$ $\prod_{i=1}^{n} p_{U_{0} \mid X_{0}}\left(u_{0 i} \mid x_{0 i}\right)$. Then we invoke Theorem 1 with $X_{j}$ replaced by $\left(X_{j}, U_{0}\right)$ for $j \in\{1,2\}$. A detailed proof is given in Section VI-A.2.

We now provide an outer bound to the region $\mathcal{R}_{\text {MAC-coord }}$.
Theorem 5 (Outer Bound With Encoder Shared Randomness): Given a p.m.f. $q_{X_{1} X_{2} Z Y}$, any rate tuple $\left(R_{1}, R_{2}\right.$, $\left.R_{00}, \quad R_{01}, \quad R_{02}\right) \quad$ in $\quad \mathcal{R}_{\text {MAC-coord }}$ satisfies, for every $\epsilon \in\left(0, \frac{1}{4}\right]$,

$$
\begin{aligned}
& R_{1} \geq \max \left\{I\left(U_{1} ; X_{1} \mid U_{0}, Z, T\right)\right. \\
& \left.\quad I\left(U_{1} ; X_{1} \mid U_{0}, U_{2}, X_{2}, Z, T\right)\right\}
\end{aligned}
$$

$$
\begin{equation*}
R_{2} \geq \max \left\{I\left(U_{2} ; X_{2} \mid U_{0}, Z, T\right)\right. \tag{26}
\end{equation*}
$$

$$
\begin{equation*}
\left.I\left(U_{2} ; X_{2} \mid U_{0}, U_{1}, X_{1}, Z, T\right)\right\} \tag{27}
\end{equation*}
$$

$$
\begin{equation*}
R_{1}+R_{2} \geq I\left(U_{1}, U_{2} ; X_{1}, X_{2} \mid U_{0}, Z, T\right) \tag{28}
\end{equation*}
$$

$$
\begin{equation*}
R_{1}+R_{01} \geq I\left(U_{1} ; X_{1}, X_{2}, Y \mid Z, T\right)-g(\epsilon) \tag{29}
\end{equation*}
$$

$$
\begin{equation*}
R_{2}+R_{02} \geq I\left(U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right)-g(\epsilon) \tag{30}
\end{equation*}
$$

$$
R_{00}+R_{1}+R_{01} \geq I\left(U_{0}, U_{1} ; X_{1}, X_{2}, Y \mid Z, T\right)-g(\epsilon)
$$

$$
R_{00}+R_{2}+R_{02} \geq I\left(U_{0}, U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right)-g(\epsilon)
$$

$$
\begin{equation*}
R_{1}+R_{2}+R_{01}+R_{02} \geq I\left(U_{1}, U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right)-g(\epsilon) \tag{32}
\end{equation*}
$$

$$
\begin{align*}
R_{00}+R_{1}+R_{2}+R_{01}+ & R_{02}+g(\epsilon)  \tag{33}\\
& \geq I\left(U_{0}, U_{1}, U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right) \tag{34}
\end{align*}
$$

with $g(\epsilon)=2 \sqrt{\epsilon}\left(H_{q}\left(X_{1}, X_{2}, Y, Z\right)+\log \frac{\left(\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\|\mathcal{Y}\| \mathcal{Z}\right|\right)}{\epsilon}\right)$ (which tends to 0 as $\epsilon \rightarrow 0$ ), for some p.m.f.

$$
\begin{align*}
& p\left(x_{1}, x_{2}, z, t, u_{0}, u_{1}, u_{2}, y\right)= \\
& p\left(x_{1}, x_{2}, z\right) p(t) p\left(u_{0} \mid t\right) p\left(u_{1}, u_{2} \mid x_{1}, x_{2}, u_{0}, t\right) p\left(y \mid u_{1}, u_{2}, z, t\right) \tag{35}
\end{align*}
$$

such that

$$
\begin{align*}
p\left(u_{1} \mid x_{1}, x_{2}, u_{0}, z, t\right) & =p\left(u_{1} \mid x_{1}, u_{0}, t\right)  \tag{36}\\
p\left(u_{2} \mid x_{1}, x_{2}, u_{0}, z, t\right) & =p\left(u_{2} \mid x_{2}, u_{0}, t\right)  \tag{37}\\
\left\|p\left(x_{1}, x_{2}, y, z \mid t\right)-q\left(x_{1}, x_{2}, y, z\right)\right\|_{1} & \leq \epsilon \text { for all } t \tag{38}
\end{align*}
$$

A detailed proof is given in Section VI-B.1. Once again, the outer bound in Theorem 5 is only an epsilon rate region, whose continuity at $\epsilon=0$ is unknown.

When the random variables $X_{1}$ and $X_{2}$ are conditionally independent given $Z$, and the shared randomness rates
( $R_{00}, R_{01}, R_{02}$ ) are unlimited, we obtain a potentially stronger outer bound which is also continuous at $\epsilon=0$.

Theorem 6 (Outer Bound - Conditionally Independent Sources Given the Side Information): Consider a p.m.f. $q_{X_{1} X_{2} Z Y}$ such that the random variables $X_{1}$ and $X_{2}$ are conditionally independent given $Z$, i.e., $I\left(X_{1} ; X_{2} \mid Z\right)=0$. Then any rate tuple in $\mathcal{R}_{\text {MAC-coord, UL- }\left(S_{0}, S_{1}, S_{2}\right)}$ satisfies

$$
\begin{aligned}
& R_{1} \geq I\left(U_{0}, U_{1} ; X_{1} \mid Z, T\right) \\
& R_{2} \geq I\left(U_{0}, U_{2} ; X_{2} \mid Z, T\right) \\
& R_{1}+R_{2} \geq I\left(U_{0}, U_{1}, U_{2} ; X_{1}, X_{2} \mid Z, T\right)
\end{aligned}
$$

for some p.m.f.

$$
\begin{align*}
& p\left(x_{1}, x_{2}, z, t, u_{0}, u_{1}, u_{2}, y\right)= \\
& p(z) p\left(x_{1} \mid z\right) p\left(x_{2} \mid z\right) p(t) p\left(u_{0} \mid t\right) p\left(u_{1} \mid x_{1}, u_{0}, t\right) p\left(u_{2} \mid x_{2}, u_{0}, t\right) \\
& \quad \times p\left(y \mid u_{1}, u_{2}, z, t\right) \tag{39}
\end{align*}
$$

such that $\sum_{u_{0}, u_{1}, u_{2}} p\left(x_{1}, x_{2}, z, u_{0}, u_{1}, u_{2}, y \mid t\right)=q\left(x_{1}, x_{2}, z, y\right)$, for all $t$, with $\left|\mathcal{U}_{0}\right| \leq\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\right||\mathcal{Y} \| \mathcal{Z}|, \quad\left|\mathcal{U}_{1}\right| \leq$ $\left|\mathcal{U}_{0}\right|\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\right||\mathcal{Y}||\mathcal{Z}|,\left|\mathcal{U}_{2}\right| \leq\left|\mathcal{U}_{0}\right|\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\right||\mathcal{Y}||\mathcal{Z}|$, and $|\mathcal{T}| \leq 3$.

Notice that the improvement is in the structure of the p.m.f. compared to that of Theorem 5. A detailed proof can be found in Section VI-B.4.

## A. Optimal Region for Example 1 With Unlimited Shared Randomness Between All Three Pairs

Here, we show that in the setting of Example 1, the region $R_{1} \geq 1$ and $R_{2} \geq 1$ is indeed the optimal rate region (not just achievable as shown in Section IV) with unlimited pairwise shared randomness (all three pairs). The achievability can also be inferred from Theorem 4 with the choice of $U_{0}=J$, $U_{1}=X_{1 J}$ and $U_{2}=X_{2 J}$, where $J$ is a random variable uniformly distributed on $\{1,2\}$ and independent of $\left(X_{1}, X_{2}\right)$. To prove the converse, first note that Theorem 6 (with $Z=\emptyset$ ) implies that any achievable rate pair $\left(R_{1}, R_{2}\right)$ must satisfy

$$
\begin{aligned}
R_{1} & \geq I\left(U_{0}, U_{1} ; X_{1} \mid T\right) \\
R_{2} & \geq I\left(U_{0}, U_{2} ; X_{2} \mid T\right) \\
R_{1}+R_{2} & \geq I\left(U_{0}, U_{1}, U_{2} ; X_{1}, X_{2} \mid T\right),
\end{aligned}
$$

for some p.m.f.

$$
\begin{aligned}
p\left(x_{1}, x_{2},\right. & \left.t, u_{0}, u_{1}, u_{2}, y\right)=p\left(x_{1}\right) p\left(x_{2}\right) p(t) p\left(u_{0} \mid t\right) \\
& \times p\left(u_{1} \mid x_{1}, u_{0}, t\right) p\left(u_{2} \mid x_{2}, u_{0}, t\right) p\left(y \mid u_{1}, u_{2}, t\right)
\end{aligned}
$$

such that $\sum_{u_{0}, u_{1}, u_{2}} p\left(x_{1}, x_{2}, u_{0}, u_{1}, u_{2}, y \mid t\right)=q\left(x_{1}, x_{2}, y\right)$, for all $t$.

For the converse, it suffices to show that for any p.m.f.

$$
\begin{aligned}
& p\left(x_{1}, x_{2}, u_{0}, u_{1}, u_{2}, y\right)= \\
& \quad p\left(x_{1}\right) p\left(x_{2}\right) p\left(u_{0}\right) p\left(u_{1} \mid x_{1}, u_{0}\right) p\left(u_{2} \mid x_{2}, u_{0}\right) p\left(y \mid u_{1}, u_{2}\right)
\end{aligned}
$$

with $\sum_{u_{0}, u_{1}, u_{2}} p\left(x_{1}, x_{2}, u_{0}, u_{1}, u_{2}, y\right)=q\left(x_{1}, x_{2}, y\right)$, we have $I\left(U_{0}, U_{1} ; X_{1}\right) \geq 1$ and $I\left(U_{0}, U_{2} ; X_{2}\right) \geq 1$. Equivalently, by
marginalizing away $U_{2}$ and letting $\left(U_{0}, U_{1}\right) \triangleq U$, it suffices to show that for any p.m.f.

$$
p\left(x_{1}, x_{2}, u, y\right)=p\left(x_{1}\right) p\left(x_{2}\right) p\left(u \mid x_{1}\right) p\left(y \mid u, x_{2}\right)
$$

with $\sum_{u} p\left(x_{1}, x_{2}, u, y\right)=q\left(x_{1}, x_{2}, y\right)$, we have $I\left(U ; X_{1}\right) \geq 1$ (The condition $I\left(U_{0}, U_{2} ; X_{2}\right) \geq 1$ can be shown analogously.). This can be established by proving that for each $U=u$ such that $H\left(X_{1} \mid U=u\right)>0$, there exists $k(u) \in\{1,2\}$ such that $H\left(X_{1 k} \mid U=u\right)=0$. Indeed this yields

$$
\begin{align*}
I\left(U ; X_{1}\right) & =H\left(X_{1}\right)-H\left(X_{1} \mid U\right) \\
& =2-\sum_{u} P(U=u) H\left(X_{1} \mid U=u\right) \\
& =2-\sum_{u} P(U=u)\left(H\left(X_{1 k} \mid U=u\right)\right. \\
& \left.\quad+H\left(X_{1 k^{\prime}} \mid U=u, X_{1 k}\right)\right) \\
& \geq 2-\sum_{u} P(U=u)(0+1) \\
& =1 \tag{40}
\end{align*}
$$

We have the following claim.
Claim 2: When $H\left(X_{1} \mid U\right)>0$, for all $u$ with $H\left(X_{1} \mid U=\right.$ $u)>0$ there exists a $k$ in $\{1,2\}$ such that $H\left(X_{1 k} \mid U=u\right)=0$.

Proof of Claim 2: We prove this by contradiction. Suppose $H\left(X_{1 i} \mid U=u\right)>0$, for $i=1,2$. Then the support of $p_{X_{1} \mid U=u}$ has to be a superset of either $\{(0,1),(1,0)\}$ or $\{(0,0),(1,1)\}$. However, it turns out that, the support cannot be a superset of $\{(0,0),(1,1)\}$. To see this, first notice that $P\left(X_{1}=(0,0) \mid U=u, X_{2}=(0,0)\right)=P\left(X_{1}=(0,0) \mid U=\right.$ $u)>0$, where the equality follows from the independence of $\left(U, X_{1}\right)$ and $X_{2}$. Similarly, $P\left(X_{1}=(1,1) \mid U=u, X_{2}=\right.$ $(0,0))>0$. Now since $X_{1 J}-\left(U, X_{2}\right)-X_{1}$, we have $P\left(X_{1 J}=0 \mid U=u, X_{2}=(0,0)\right)=P\left(X_{1 J}=0 \mid U=\right.$ $\left.u, X_{2}=(0,0), X_{1}=(1,1)\right)=0$, where the last equality follows from the correctness of the output $Y=\left(X_{1 J}, X_{2 J}\right)$. Similarly, $P\left(X_{1 J}=1 \mid U=u, X_{2}=(0,0)\right)=P\left(X_{1 J}=\right.$ $\left.1 \mid U=u, X_{2}=(0,0), X_{1}=(0,0)\right)=0$. This is a contradiction since $p_{Y_{1} \mid U=u, X_{2}=(0,0)}$ has to be a probability distribution. The only other possibility is that $p_{X_{1} \mid U=u}$ has a support that is a superset of $\{(0,1),(1,0)\}$. Consider the induced distribution $p_{Y \mid X_{2}=(0,1), U=u}$. This is well-defined because $P\left(X_{2}=(0,1), U=u\right)>0$ as $X_{2}$ is independent of $U$ and $P\left(X_{2}=(0,1)\right), P(U=u)>0$. Since $P\left(X_{1}=\right.$ $\left.(0,1) \mid X_{2}=(0,1), U=u\right)=P\left(X_{1}=(0,1) \mid U=u\right)>0$ and $Y \rightarrow\left(X_{2}, U\right) \rightarrow X_{1}$, we have $P\left(Y=(1,0) \mid X_{2}=\right.$ $(0,1), U=u)=P\left(Y=(0,1) \mid X_{2}=(0,1), U=u\right)=0$. Since $P\left(X_{1}=(1,0) \mid X_{2}=(0,1), U=u\right)=P\left(X_{1}=\right.$ $(1,0) \mid U=u)>0$ and $Y \rightarrow\left(X_{2}, U\right) \rightarrow X_{1}$, we have $P(Y=$ $\left.(1,1) \mid X_{2}=(0,1), U=u\right)=P\left(Y=(0,0) \mid X_{2}=(0,1)\right.$, $U=u)=0$. This is a contradiction since $p_{Y \mid X_{2}=(0,1), U=u}$ has to be a valid probability distribution.

This proves that the optimal communication rate region with unlimited pairwise shared randomness (all three pairs) $\mathcal{R}_{\text {MAC-coord, UL- }\left(S_{0}, S_{1}, S_{2}\right)}$ is indeed defined by the constraints $R_{1} \geq 1$ and $R_{2} \geq 1$.

## VI. Proofs

## A. Achievability Proofs

The proof of Theorem 1 is in Appendix A.

1) Achievability Proof of Theorem 3: Proof: We argue that achievability follows from Theorem 1 by enforcing the constraint $p\left(x_{1}, x_{2}, z\right)=p(z) p\left(x_{1} \mid z\right) p\left(x_{2} \mid z\right)$ along with unlimited shared randomness rate $R_{02}$. In this case, the joint distribution on $\left(X_{1}, X_{2}, Z, T, U_{1}\right.$, $\left.U_{2}, Y\right)$ decomposes as $p\left(x_{1}, x_{2}, z, t, u_{1}, u_{2}, y\right)=$ $p(z) p\left(x_{1} \mid z\right) p\left(x_{2} \mid z\right) p(t) p\left(u_{1} \mid x_{1}, t\right) p\left(u_{2} \mid x_{2}, t\right) p\left(y \mid u_{1}, u_{2}, z, t\right)$. We first write down the inner bound from Theorem 1 for this case.

$$
\begin{align*}
& R_{1} \geq I\left(U_{1} ; X_{1} \mid U_{2}, Z, T\right) \\
&=I\left(U_{1} ; X_{1} \mid Z, T\right) \\
& R_{2} \geq I\left(U_{2} ; X_{2} \mid U_{1}, Z, T\right) \\
&=I\left(U_{2} ; X_{2} \mid Z, T\right) \\
& R_{1}+R_{2} \geq I\left(U_{1}, U_{2} ; X_{1}, X_{2} \mid Z, T\right) \\
&=I\left(U_{1} ; X_{1} \mid Z, T\right)+I\left(U_{2} ; X_{2} \mid Z, T\right) \\
& R_{1}+R_{01} \geq I\left(U_{1} ; X_{1}, X_{2}, Y \mid Z, T\right) \\
& \quad-I\left(U_{1} ; U_{2} \mid Z, T\right) \\
&=I\left(U_{1} ; X_{1}, X_{2}, Y \mid Z, T\right) \\
&=I\left(U_{1} ; X_{1}, Y \mid X_{2}, Z, T\right) \\
& R_{2}+R_{02} \geq I\left(U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right) \\
& \quad-I\left(U_{1} ; U_{2} \mid Z, T\right) \\
&=I\left(U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right) \\
&=I\left(U_{2} ; X_{2}, Y \mid X_{1}, Z, T\right) \\
& R_{1}+R_{2}+R_{01} \geq I\left(U_{1} ; X_{1}, X_{2}, Y \mid Z, T\right) \\
& \quad+I\left(U_{2} ; X_{2} \mid U_{1}, Z, T\right) \\
&=I\left(U_{1} ; X_{1}, X_{2}, Y \mid Z\right) \\
& \quad+I\left(U_{2} ; X_{2} \mid Z, T\right) \\
& R_{1}+R_{2}+R_{01}+R_{02} \geq I\left(U_{1}, U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right) . \\
& \quad \quad+I(41)
\end{align*}
$$

Notice that the constraints on $R_{1}+R_{2}, R_{1}+R_{2}+R_{01}$ and $R_{1}+R_{2}+R_{02}$ are redundant. This region is an inner bound to $\mathcal{R}_{\mathrm{MAC}}^{\mathrm{NO}-\mathrm{S}_{0}} \mathrm{R}^{2}$. Considering only the constraints that exclude $R_{02}$ completes the achievability of Theorem 3.
2) Proof of Theorem 4: Proof: We prove the achievability for $|\mathcal{T}|=1$, and the rest of the proof follows by using time sharing argument similar to that in Theorem 1 (in particular, see the paragraph after Lemma 8). Firstly, we show that the shared randomness between the encoders can be harnessed to simulate $U_{0}^{n}$ approximately distributed according to $q_{U_{0}^{n} \mid X_{0}^{n}}^{(n)}\left(u_{0}^{n} \mid x_{0}^{n}\right)=\prod_{i=1}^{n} p_{U_{0} \mid X_{0}}\left(u_{0 i} \mid x_{0 i}\right)$. The rate of shared randomness needed here will turn out to be $R_{00} \geq H\left(U_{0} \mid X_{0}\right)$, i.e. the constraint (17) on $R_{00}$ in Theorem 4.


Fig. 4. Implications of soft covering [5].

We make use of the following setup from Cuff [5, Corollary VII.5]. Let $X_{0}^{n}$ be an i.i.d. sequence with distribution $p_{X_{0}}$ and $p_{U_{0} \mid X_{0}} p_{V \mid U_{0}, X_{0}}$ be a memoryless channel. A deterministic encoder $f(\cdot)$ receives both $X_{0}^{n}$ and a uniformly distributed random variable $S_{0} \in\left[1: 2^{n R_{00}}\right]$. Cuff [5] gave the following sufficient condition on the rate $R_{00}$ so that the induced distribution on the channel output in Figure 4 is i.i.d. in the limit of large $n$.

Lemma 1: [5, Corollary VII.5] Let $p_{V^{n}}$ be the induced distribution on the channel output in Figure 4 and $q_{V^{n}}$ be the i.i.d. output distribution specified by $\sum_{u_{0}, x_{0}} p_{X_{0}} p_{U_{0} \mid X_{0}} p_{V \mid U_{0}, X_{0}}$. Then if

$$
\begin{equation*}
R \geq I\left(X_{0}, U_{0} ; V\right)-H\left(X_{0}\right) \tag{42}
\end{equation*}
$$

we have

$$
\begin{equation*}
\lim _{n \rightarrow \infty}\left\|p_{V^{n}}-q_{V^{n}}\right\|_{1}=0 \tag{43}
\end{equation*}
$$

For our purposes, we take $p_{V \mid U_{0}, X_{0}}$ to be an identity channel i.e. $V=\left(U_{0}, X_{0}\right)$ in Lemma 1. Hence

$$
\begin{equation*}
\lim _{n \rightarrow \infty}\left\|p_{U_{0}^{n}, X_{0}^{n}}-q_{U_{0}^{n}, X_{0}^{n}}\right\|_{1}=0 \tag{44}
\end{equation*}
$$

provided that $R_{00}$ satisfies

$$
\begin{align*}
R_{00} & \geq I\left(X_{0}, U_{0} ; X_{0}, U_{0}\right)-H\left(X_{0}\right) \\
& =H\left(U_{0} \mid X_{0}\right) . \tag{45}
\end{align*}
$$

Suppose if $\left(X_{1}^{n}, X_{2}^{n}, U_{0}^{n}\right)$ are generated exactly i.i.d., then we can invoke Theorem 1 with $X_{j}$ replaced by $\left(X_{j}, U_{0}\right)$ for $j \in\{1,2\}$ and it can be verified that this exactly yields the eight rate constraints (18)-(25) involving ( $R_{1}, R_{2}, R_{01}, R_{02}$ ) in Theorem 4. Here, we show that the same set of rate constraints suffices even if $\left(U_{0}^{n}, X_{0}^{n}\right)$ are approximately i.i.d. in the sense of (44). All that remains to be shown is that under the rate constraints in Theorem 4, there exists a sequence of codes with an induced p.m.f. on $\left(X_{1}^{n}, X_{2}^{n}, U_{0}^{n}, Z^{n}, Y^{n}\right)$ such that the total variation distance between this p.m.f. and the desired i.i.d. p.m.f. $\prod q_{X_{1}, X_{2}, U_{0}, Z, Y}$ vanishes in the limit of large blocklength $n$. This can be argued out as follows.

If $\left(X_{1}^{n}, X_{2}^{n}, U_{0}^{n}\right)$ were exactly i.i.d., by invoking Theorem 1 there exists

$$
\begin{aligned}
& p\left(m_{1}, m_{2}, y^{n} \mid x_{1}^{n}, x_{2}^{n}, u_{0}^{n}, z^{n}, s_{1}, s_{2}\right) \\
& :=p\left(m_{1} \mid s_{1}, x_{1}^{n}, u_{0}^{n}\right) p\left(m_{2} \mid s_{2}, x_{2}^{n}, u_{0}^{n}\right) p\left(y^{n} \mid m_{1}, m_{2}, s_{1}, s_{2}, z^{n}\right)
\end{aligned}
$$

such that

$$
\begin{equation*}
\left\|p_{X_{1}^{n}, X_{2}^{n}, U_{0}^{n}, Z^{n}, Y^{n}}-\prod q_{X_{1} X_{2} U_{0} Z Y}\right\|_{1} \leq \epsilon \tag{46}
\end{equation*}
$$

On the other hand, let $\tilde{p}\left(x_{1}^{n}, x_{2}^{n}, u_{0}^{n}, z^{n}, s_{1}, s_{2}\right)$ be the distribution in which $\left(U_{0}^{n}, X_{0}^{n}\right)$ are approximately i.i.d. in the sense of (44). Let us define

$$
\begin{align*}
& \tilde{p}\left(x_{1}^{n}, x_{2}^{n}, u_{0}^{n}, z^{n}, s_{1}, s_{2}, m_{1}, m_{2}, y^{n}\right) \\
& \quad:=\tilde{p}\left(x_{1}^{n}, x_{2}^{n}, u_{0}^{n}, z^{n}, s_{1}, s_{2}\right) \\
& \quad \times p\left(m_{1}, m_{2}, y^{n} \mid x_{1}^{n}, x_{2}^{n}, u_{0}^{n}, z^{n}, s_{1}, s_{2}\right) . \tag{47}
\end{align*}
$$

Now, the total variation distance of interest can be bounded using triangle inequality as

$$
\begin{align*}
& \left\|\tilde{p}_{X_{1}^{n}, X_{2}^{n}, U_{0}^{n}, Z^{n}, Y^{n}}-\prod q_{X_{1} X_{2} U_{0} Z Y}\right\|_{1} \\
& \leq\left\|\tilde{p}_{X_{1}^{n}, X_{2}^{n}, U_{0}^{n}, Z^{n}, Y^{n}}-p_{X_{1}^{n}, X_{2}^{n}, U_{0}^{n}, Z^{n}, Y^{n}}\right\|_{1} \\
& \quad \quad+\left\|p_{X_{1}^{n}, X_{2}^{n}, U_{0}^{n}, Z^{n}, Y^{n}}-\prod q_{X_{1} X_{2} U_{0} Y}\right\|_{1} \\
& \quad \leq\left\|\tilde{p}_{X_{1}^{n}, X_{2}^{n}, U_{0}^{n}, Z^{n}, Y^{n}}-p_{X_{1}^{n}, X_{2}^{n}, U_{0}^{n}, Z^{n}, Y^{n}}\right\|_{1}+\epsilon, \tag{48}
\end{align*}
$$

where (48) follows from (46). Next consider the first term on the RHS of (48).

$$
\begin{align*}
& \left\|\tilde{p}_{X_{1}^{n}, X_{2}^{n}, U_{0}^{n}, Z^{n}, Y^{n}}-p_{X_{1}^{n}, X_{2}^{n}, U_{0}^{n}, Z^{n}, Y^{n}}\right\|_{1} \\
& =\sum_{\substack{x_{1}^{n}, x_{2}^{n}, u_{n}^{n}, z^{n}, y^{n}}} \mid \sum_{\substack{m_{1}, m_{2}, s_{1}, s_{2}}} \tilde{p}\left(x_{1}^{n}, x_{2}^{n}, u_{0}^{n}, z^{n}, y^{n}, m_{1}, m_{2}, s_{1}, s_{2}\right) \\
& -p\left(x_{1}^{n}, x_{2}^{n}, u_{0}^{n}, z^{n}, y^{n}, m_{1}, m_{2}, s_{1}, s_{2}\right) \mid \\
& \stackrel{(a)}{\leq} \sum_{\substack{x_{1}^{n}, x_{2}^{n}, u_{0}^{n}, z^{n}, y^{n}, m_{1}, m_{2}, s_{1}, s_{2}}} \frac{p\left(y^{n}, m_{1}, m_{2} \mid x_{1}^{n}, x_{2}^{n}, u_{0}^{n}, z^{n}, s_{1}, s_{2}\right)}{2^{n\left(R_{01}+R_{02}\right)}} \\
& \times\left|\tilde{p}\left(x_{1}^{n}, x_{2}^{n}, u_{0}^{n}, z^{n}\right)-p\left(x_{1}^{n}, x_{2}^{n}, u_{0}^{n}, z^{n}\right)\right| \\
& =\sum_{x_{1}^{n}, x_{2}^{n}, u_{0}^{n}, z^{n}}\left|\tilde{p}\left(x_{1}^{n}, x_{2}^{n}, u_{0}^{n}, z^{n}\right)-p\left(x_{1}^{n}, x_{2}^{n}, u_{0}^{n}, z^{n}\right)\right| \\
& \stackrel{(b)}{\leq} \sum_{x_{1}^{n}, x_{2}^{n}, x_{0}^{n}, u_{0}^{n}, z^{n}} \mid \tilde{p}\left(x_{1}^{n}, x_{2}^{n}, x_{0}^{n}, u_{0}^{n}, z^{n}\right)- \\
& p\left(x_{1}^{n}, x_{2}^{n}, x_{0}^{n}, u_{0}^{n}, z^{n}\right) \mid \\
& =\sum_{x_{1}^{n}, x_{2}^{n}, x_{0}^{n}, u_{0}^{n}, z^{n}} p\left(x_{1}^{n}, x_{2}^{n}, x_{0}^{n}, z^{n}\right)\left|\tilde{p}\left(u_{0}^{n} \mid x_{0}^{n}\right)-p\left(u_{0}^{n} \mid x_{0}^{n}\right)\right| \\
& =\sum_{x_{0}^{n}, u_{0}^{n}}\left|\tilde{p}\left(x_{0}^{n}, u_{0}^{n}\right)-p\left(x_{0}^{n}, u_{0}^{n}\right)\right| \\
& \leq \epsilon \text {, } \tag{49}
\end{align*}
$$

where (a) follows from the triangle inequality and the fact that $S_{1}$ and $S_{2}$ are uniformly distributed on their respective ranges,
(b) follows since $X_{0}$ is a common function of $X_{1}$ and $X_{2}$, and
(49) follows from the statement below (46). From expressions
(48) and (49), we conclude that

$$
\left\|\tilde{p}_{X_{1}^{n}, X_{2}^{n}, U_{0}^{n}, Z^{n}, Y^{n}}-\prod q_{X_{1}, X_{2}, U_{0}, Z, Y}\right\|_{1} \leq 2 \epsilon
$$

## B. Converse Proofs

We will show that Theorem 2 is a direct consequence of Theorem 5, which we prove first.

1) Proof of Theorem 5: Proof: Consider a code that induces a joint distribution on $\left(X_{1}^{n}, X_{2}^{n}, Z^{n}, Y^{n}\right)$ such that

$$
\begin{equation*}
\left\|p_{X_{1}^{n}, X_{2}^{n}, Z^{n}, Y^{n}}-q_{X_{1} X_{2} Z Y}^{(n)}\right\|_{1}<\epsilon, \tag{50}
\end{equation*}
$$

for $\epsilon \in\left(0, \frac{1}{4}\right]$. For ease of notation, for a vector $A^{n}$, we write $A_{\sim i} \triangleq\left(A^{i-1}, A_{i+1}^{n}\right)$. Also let $M_{[0: 2]} \triangleq\left(M_{0}, M_{1}, M_{2}\right)$ and $S_{[0: 2]} \triangleq\left(S_{0}, S_{1}, S_{2}\right)$. We quote the following lemmas that will prove useful in the outer bound.

Lemma 2: [10, Lemma 6] Let $p_{X^{n}}$ be such that $\| p_{X^{n}}-$ $q_{X}^{(n)} \|_{1} \leq \epsilon$, where $q_{X}^{(n)}\left(x^{n}\right)=\prod_{i=1}^{n} q_{X}\left(x_{i}\right)$, then

$$
\begin{equation*}
\sum_{i=1}^{n} I_{p}\left(X_{i} ; X_{\sim i}\right) \leq n g_{1}(\epsilon) \tag{51}
\end{equation*}
$$

where $g_{1}(\epsilon)=2 \sqrt{\epsilon}\left(H(X)+\log |\mathcal{X}|+\log \frac{1}{\sqrt{\epsilon}}\right) \rightarrow 0$ as $\epsilon \rightarrow 0$.
Lemma 3: [5, Lemma VI.3] Let $p_{X^{n}}$ be such that $\| p_{X^{n}}-$ $q_{X}^{(n)} \|_{1} \leq \epsilon$, where $q_{X}^{(n)}\left(x^{n}\right)=\prod_{i=1}^{n} q_{X}\left(x_{i}\right)$, then for any RV $T \in[1: n]$ independent of $X^{n}$,

$$
\begin{equation*}
I_{p}\left(T ; X_{T}\right) \leq g_{2}(\epsilon) \tag{52}
\end{equation*}
$$

where $g_{2}(\epsilon)=4 \epsilon\left(\log |\mathcal{X}|+\log \frac{1}{\epsilon}\right) \rightarrow 0$ as $\epsilon \rightarrow 0$.
Notice that for $\epsilon \in\left(0, \frac{1}{4}\right]$, we have $\max \left\{g_{1}(\epsilon), g_{2}(\epsilon)\right\} \leq$ $g(\epsilon):=2 \sqrt{\epsilon}\left(H(X)+\log |\mathcal{X}|+2 \log \frac{1}{\sqrt{\epsilon}}\right)$. So, we can replace $g_{1}(\epsilon)$ and $g_{2}(\epsilon)$ in Lemmas 2 and 3 by $g(\epsilon)$, which also tends to 0 as $\epsilon \rightarrow 0$. Let us consider the first lower bound on $R_{j}$ for $j \in\{1,2\}$ in (26)-(27).

$$
\begin{align*}
n R_{j} & \geq H\left(M_{j}\right) \\
& \geq H\left(M_{j} \mid S_{0}, S_{j}, Z^{n}\right) \\
& \geq I\left(M_{j} ; X_{j}^{n} \mid S_{0}, S_{j}, Z^{n}\right) \\
& \stackrel{(a)}{=} I\left(M_{j}, S_{j} ; X_{j}^{n} \mid S_{0}, Z^{n}\right) \\
& =\sum_{i=1}^{n} I\left(M_{j}, S_{j} ; X_{j i} \mid X_{j, i+1}^{n}, S_{0}, Z_{i}, Z_{\sim i}\right) \\
& \stackrel{(b)}{=} \sum_{i=1}^{n} I\left(M_{j}, S_{j}, X_{j, i+1}^{n}, Z_{\sim i} ; X_{j i} \mid S_{0}, Z_{i}\right) \\
& \geq \sum_{i=1}^{n} I\left(M_{j}, S_{j}, X_{j, i+1}^{n}, Z^{i-1} ; X_{j i} \mid S_{0}, Z_{i}\right) \\
& \stackrel{(c)}{\geq} \sum_{i=1}^{n} I\left(U_{j i} ; X_{j i} \mid U_{0 i}, Z_{i}\right) \\
& \stackrel{(d)}{=} n I\left(U_{j T} ; X_{j T} \mid U_{0 T}, Z_{T}, T\right) \\
& \stackrel{(e)}{=} n I\left(U_{j} ; X_{j} \mid U_{0}, Z, T\right) \tag{53}
\end{align*}
$$

where (a) follows since $S_{j}$ is independent of $\left(S_{0}, X_{j}^{n}, Z^{n}\right)$, (b) follows since $\left(X_{j i}, Z_{i}\right), i=1, \ldots, n$, are jointly i.i.d. and $S_{0}$ is independent of $\left(X_{j}^{n}, Z^{n}\right)$, (c) follows by defining $U_{0 i}=S_{0}, U_{1 i}=\left(M_{1}, S_{1}, X_{1, i+1}^{n}, Z^{i-1}\right)$ and $U_{2 i}=$ ( $M_{2}, S_{2}, X_{2, i+1}^{n}$ ), (d) follows by introducing a uniform timesharing random variable $T \in[1: n]$ that is independent of everything else, while (e) follows by defining $U_{0}:=U_{0 T}$, $U_{1}:=U_{1 T}, U_{2}:=U_{2 T}, X_{1}:=X_{1 T}, X_{2}:=X_{2 T}, Y:=Y_{T}$
and $Z:=Z_{T}$. The second lower bound on $R_{1}$ in (26) is obtained as follows.

$$
\begin{align*}
n R_{1} & \geq H\left(M_{1}\right) \\
& \geq H\left(M_{1} \mid X_{2}^{n}, S_{[0: 2]}, Z^{n}\right) \\
& \geq I\left(M_{1} ; X_{1}^{n} \mid X_{2}^{n}, S_{[0: 2]}, Z^{n}\right) \\
& \stackrel{(a)}{=} I\left(M_{[1: 2]} ; X_{1}^{n} \mid X_{2}^{n}, S_{[0: 2]}, Z^{n}\right) \\
& \stackrel{(b)}{=} I\left(M_{[1: 2]}, S_{[1: 2]} ; X_{1}^{n} \mid S_{0}, X_{2}^{n}, Z^{n}\right) \\
& =\sum_{i=1}^{n} I\left(M_{[1: 2]}, S_{[1: 2]} ; X_{1 i} \mid X_{1, i+1}^{n}, S_{0}, X_{2}^{n}, Z^{n}\right) \\
& \stackrel{(c)}{=} \sum_{i=1}^{n} I\left(M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, \sim i}, Z_{\sim i} ; X_{1 i} \mid S_{0}, X_{2 i}, Z_{i}\right) \\
& \geq \sum_{i=1}^{n} I\left(M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{i-1} ; X_{1 i} \mid S_{0}, X_{2 i}, Z_{i}\right) \\
& \stackrel{(d)}{=} \sum_{i=1}^{n} I\left(U_{1 i}, U_{2 i} ; X_{1 i} \mid U_{0 i}, X_{2 i}, Z_{i}\right) \\
& \stackrel{(e)}{=} \sum_{i=1}^{n} I\left(U_{1 i} ; X_{1 i} \mid U_{0 i}, U_{2 i}, X_{2 i}, Z_{i}\right) \\
& =n I\left(U_{1 T} ; X_{1 T} \mid U_{0 T}, U_{2 T}, X_{2 T}, Z_{T}, T\right) \\
& =n I\left(U_{1} ; X_{1} \mid U_{0}, U_{2}, X_{2}, Z, T\right) \tag{54}
\end{align*}
$$

where (a) follows from the Markov chain $M_{2} \rightarrow$ $\left(X_{2}^{n}, S_{0}, S_{2}\right) \rightarrow\left(M_{1}, S_{1}, X_{1}^{n}, Z^{n}\right)$, (b) follows since $S_{[1: 2]}$ is independent of $\left(S_{0}, X_{1}^{n}, X_{2}^{n}, Z^{n}\right)$, (c) follows since $S_{0}$ is independent of $\left(X_{1}^{n}, X_{2}^{n}, Z^{n}\right)$ and the fact that $\left(X_{1 i}, X_{2 i}, Z_{i}\right)$, $i=1, \cdots, n$ are i.i.d., (d) follows from the identifications $U_{0 i}=S_{0}, U_{1 i}=\left(M_{1}, S_{1}, X_{1, i+1}^{n}, Z^{i-1}\right)$ and $U_{2 i}=$ ( $M_{2}, S_{2}, X_{2, i+1}^{n}$ ), while (e) follows from the Markov chain $U_{2 i} \rightarrow\left(U_{0 i}, X_{2 i}\right) \rightarrow\left(X_{1 i}, Z_{i}\right)$. Similarly, we obtain

$$
\begin{equation*}
n R_{2} \geq n I\left(U_{2} ; X_{2} \mid U_{0}, U_{1}, X_{1}, Z, T\right) \tag{55}
\end{equation*}
$$

We next derive the lower bound on $\left(R_{1}+R_{2}\right)$ in (28).

$$
\begin{align*}
& n\left(R_{1}+R_{2}\right) \\
& \geq H\left(M_{[1: 2]}\right) \\
& \geq H\left(M_{[1: 2]} \mid S_{[0: 2]}, Z^{n}\right) \\
& \geq I\left(M_{[1: 2]} ; X_{1}^{n}, X_{2}^{n} \mid S_{[0: 2]}, Z^{n}\right) \\
& \stackrel{(a)}{=} I\left(M_{[1: 2]}, S_{[1: 2]} ; X_{1}^{n}, X_{2}^{n} \mid S_{0}, Z^{n}\right) \\
& =\sum_{i=1}^{n} I\left(M_{[1: 2]}, S_{[1: 2]} ; X_{1 i}, X_{2 i} \mid X_{1, i+1}^{n}, X_{2, i+1}^{n}, S_{0}, Z^{n}\right) \\
& \stackrel{(b)}{=} \sum_{i=1}^{n} I\left(M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z_{\sim i} ; X_{1 i}, X_{2 i} \mid S_{0}, Z_{i}\right) \\
& \geq \sum_{i=1}^{n} I\left(M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{i-1} ; X_{1 i}, X_{2 i} \mid S_{0}, Z_{i}\right) \\
& =\sum_{i=1}^{n} I\left(U_{1 i}, U_{2 i} ; X_{1 i}, X_{2 i} \mid U_{0 i}, Z_{i}\right) \\
& =n I\left(U_{1 T}, U_{2 T} ; X_{1 T}, X_{2 T} \mid U_{0 T}, Z_{T}, T\right) \\
& =n I\left(U_{1}, U_{2} ; X_{1}, X_{2} \mid U_{0}, Z, T\right), \tag{56}
\end{align*}
$$

where (a) follows since $S_{[1: 2]}$ is independent of $\left(S_{0}, X_{1}^{n}, X_{2}^{n}, Z^{n}\right)$ while (b) follows since $\left(X_{1 i}, X_{2 i}, Z_{i}\right)$, $i=1, \ldots, n$, are jointly i.i.d. and $S_{0}$ is independent of $\left(X_{1}^{n}, X_{2}^{n}, Z^{n}\right)$. We next derive the lower bound on ( $R_{j}+R_{0 j}$ ) for $j \in\{1,2\}$ in (29)-(30).

$$
\begin{align*}
& n\left(R_{j}+R_{0 j}\right) \\
& \geq H\left(M_{j}, S_{j}\right) \\
& \geq H\left(M_{j}, S_{j} \mid Z^{n}\right) \\
& \geq I\left(M_{j}, S_{j} ; X_{1}^{n}, X_{2}^{n}, Y^{n} \mid Z^{n}\right) \\
& =\sum_{i=1}^{n} I\left(M_{j}, S_{j} ; X_{1 i}, X_{2 i}, Y_{i} \mid X_{1, i+1}^{n}, X_{2, i+1}^{n}, Y_{i+1}^{n}, Z^{n}\right) \\
& =\sum_{i=1}^{n} I\left(M_{j}, S_{j}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Y_{i+1}^{n}, Z_{\sim i} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right) \\
& -\sum_{i=1}^{n} I\left(X_{1, i+1}^{n}, X_{2, i+1}^{n}, Y_{i+1}^{n}, Z_{\sim i} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right) \\
& \stackrel{(a)}{\geq} \sum_{i=1}^{n} I\left(M_{j}, S_{j}, X_{j, i+1}^{n}, Z^{i-1} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right)-n g(\epsilon) \\
& \stackrel{(b)}{\geq} \sum_{i=1}^{n} I\left(U_{j i} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right)-n g(\epsilon) \\
& =n I\left(U_{j T} ; X_{1 T}, X_{2 T}, Y_{T} \mid Z_{T}, T\right)-n g(\epsilon) \\
& =n I\left(U_{j} ; X_{1}, X_{2}, Y \mid Z, T\right)-n g(\epsilon) \text {, } \tag{57}
\end{align*}
$$

where (a) follows since

$$
\begin{align*}
& \sum_{i=1}^{n} I\left(X_{1, i+1}^{n}, X_{2, i+1}^{n}, Y_{i+1}^{n}, Z_{\sim i} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right) \\
& \quad \leq \sum_{i=1}^{n} I\left(X_{1 \sim i}, X_{2 \sim i}, Y_{\sim i}, Z_{\sim i} ; X_{1 i}, X_{2 i}, Y_{i}, Z_{i}\right) \\
& \quad \leq n g(\epsilon) \tag{58}
\end{align*}
$$

by (50) and Lemma 2, while (b) follows from the identifications $U_{1 i}=\left(M_{1}, S_{1}, X_{1, i+1}^{n}, Z^{i-1}\right)$ and $U_{2 i}=$ $\left(M_{2}, S_{2}, X_{2, i+1}^{n}\right)$.

We next derive the lower bound on $\left(R_{00}+R_{j}+R_{0 j}\right)$ for $j \in\{1,2\}$ in (31)-(32).

$$
\begin{aligned}
& n\left(R_{00}+R_{j}+R_{0 j}\right) \\
& \geq H\left(M_{j}, S_{0}, S_{j}\right) \\
& \geq H\left(M_{j}, S_{0}, S_{j} \mid Z^{n}\right) \\
& \geq I\left(M_{j}, S_{0}, S_{j} ; X_{1}^{n}, X_{2}^{n}, Y^{n} \mid Z^{n}\right) \\
& =\sum_{i=1}^{n} I\left(M_{j}, S_{0}, S_{j} ; X_{1 i}, X_{2 i}, Y_{i} \mid X_{1, i+1}^{n}, X_{2, i+1}^{n}, Y_{i+1}^{n}, Z^{n}\right) \\
& =\sum_{i=1}^{n} I\left(M_{j}, S_{0}, S_{j}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Y_{i+1}^{n}, Z_{\sim i} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right) \\
& \quad-\sum_{i=1}^{n} I\left(X_{1, i+1}^{n}, X_{2, i+1}^{n}, Y_{i+1}^{n}, Z_{\sim i} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right) \\
& \left(\stackrel{(a)}{\geq} \sum_{i=1}^{n} I\left(M_{j}, S_{0}, S_{j}, X_{j, i+1}^{n}, Z^{i-1} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right)-n g(\epsilon)\right.
\end{aligned}
$$

$$
\begin{align*}
& \stackrel{(b)}{\geq} \sum_{i=1}^{n} I\left(U_{0 i}, U_{j i} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right)-n g(\epsilon) \\
& =n I\left(U_{0 T}, U_{j T} ; X_{1 T}, X_{2 T}, Y_{T} \mid Z_{T}, T\right)-n g(\epsilon) \\
& =n I\left(U_{0}, U_{j} ; X_{1}, X_{2}, Y \mid Z, T\right)-n g(\epsilon), \tag{59}
\end{align*}
$$

where (a) follows from (58), while (b) follows from the identifications $U_{0 i}=S_{0}, U_{1 i}=\left(M_{1}, S_{1}, X_{1, i+1}^{n}, Z^{i-1}\right)$ and $U_{2 i}=\left(M_{2}, S_{2}, X_{2, i+1}^{n}\right)$. We next derive the lower bound on ( $R_{1}+R_{2}+R_{01}+R_{02}$ ) in (33).

$$
\begin{align*}
& n\left(R_{1}+R_{2}+R_{01}+R_{02}\right) \\
& \geq H\left(M_{[1: 2]}, S_{[1: 2]}\right) \\
& \geq H\left(M_{[1: 2]}, S_{[1: 2]} \mid Z^{n}\right) \\
& \geq I\left(M_{[1: 2]}, S_{[1: 2]} ; X_{1}^{n}, X_{2}^{n}, Y^{n} \mid Z^{n}\right) \\
& =\sum_{i=1}^{n} I\left(M_{[1: 2]}, S_{[1: 2]} ; X_{1 i}, X_{2 i}, Y_{i} \mid X_{1, i+1}^{n}, X_{2, i+1}^{n}, Y_{i+1}^{n}, Z^{n}\right) \\
& =\sum_{i=1}^{n} I\left(M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Y_{i+1}^{n}, Z_{\sim i} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right) \\
& -\sum_{i=1}^{n} I\left(X_{1, i+1}^{n}, X_{2, i+1}^{n}, Y_{i+1}^{n}, Z_{\sim i} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right) \\
& \stackrel{(a)}{\geq} \sum_{i=1}^{n} I\left(M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{i-1} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right) \\
& -n g(\epsilon) \\
& =\sum_{i=1}^{n} I\left(U_{1 i}, U_{2 i} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right)-n g(\epsilon) \\
& =n I\left(U_{1 T}, U_{2 T} ; X_{1 T}, X_{2 T}, Y_{T} \mid Z_{T}, T\right)-n g(\epsilon) \\
& =n I\left(U_{1}, U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right)-n g(\epsilon) \text {, } \tag{60}
\end{align*}
$$

where (a) follows from (58). We finally derive the lower bound on $\left(R_{00}+R_{1}+R_{2}+R_{01}+R_{02}\right)$ in (34).

$$
\begin{align*}
& n\left(R_{00}+R_{1}+R_{2}+R_{01}+R_{02}\right) \\
& \geq H\left(M_{[1: 2]}, S_{[0: 2]}\right) \\
& \geq H\left(M_{[1: 2]}, S_{[0: 2]} \mid Z^{n}\right) \\
& \geq I\left(M_{[1: 2]}, S_{[0: 2]} ; X_{1}^{n}, X_{2}^{n}, Y^{n} \mid Z^{n}\right) \\
& =\sum_{i=1}^{n} I\left(M_{[1: 2]}, S_{[0: 2]} ; X_{1 i}, X_{2 i}, Y_{i} \mid X_{1, i+1}^{n}, X_{2, i+1}^{n}, Y_{i+1}^{n}, Z^{n}\right) \\
& =\sum_{i=1}^{n} I\left(M_{[1: 2]}, S_{[0: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Y_{i+1}^{n}, Z_{\sim i} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right) \\
& \quad-\sum_{i=1}^{n} I\left(X_{1, i+1}^{n}, X_{2, i+1}^{n}, Y_{i+1}^{n}, Z_{\sim i} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right) \\
& \begin{array}{l}
\geq \\
\begin{array}{l}
(a) \\
\geq
\end{array} \sum_{i=1}^{n} I\left(M_{[1: 2]}, S_{[0: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{i-1} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right) \\
\quad-n g(\epsilon) \\
=
\end{array} \sum_{i=1}^{n} I\left(U_{0 i}, U_{1 i}, U_{2 i} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right)-n g(\epsilon) \\
& =n I\left(U_{0 T}, U_{1 T}, U_{2 T} ; X_{1 T}, X_{2 T}, Y_{T} \mid Z_{T}, T\right)-n g(\epsilon) \\
& =n I\left(U_{0}, U_{1}, U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right)-n g(\epsilon),
\end{align*}
$$

where (a) follows from (58).

We now prove the Markov chains $U_{1 i} \rightarrow\left(X_{1 i}, U_{0 i}\right) \rightarrow$ $\left(X_{2 i}, Z_{i}\right), U_{2 i} \rightarrow\left(X_{2 i}, U_{0 i}\right) \rightarrow\left(X_{1 i}, Z_{i}\right),\left(U_{1 i}, U_{2 i}\right) \rightarrow$ $\left(X_{1 i}, X_{2 i}, U_{0 i}\right) \rightarrow Z_{i}$ and $Y_{i} \rightarrow\left(U_{1 i}, U_{2 i}, Z_{i}\right) \rightarrow\left(X_{1 i}\right.$, $\left.X_{2 i}, U_{0 i}\right)$. Note that this implies that the joint p.m.f. satisfies (35)-(37). Recall the auxiliary random variable identifications

$$
\begin{aligned}
& U_{0 i}=S_{0} \\
& U_{1 i}=\left(M_{1}, S_{1}, X_{1, i+1}^{n}, Z^{i-1}\right), \\
& U_{2 i}=\left(M_{2}, S_{2}, X_{2, i+1}^{n}\right)
\end{aligned}
$$

Let us first show that $\left(U_{1 i}, U_{2 i}\right) \rightarrow\left(X_{1 i}, X_{2 i}, U_{0 i}\right) \rightarrow Z_{i}$.

$$
\begin{align*}
& I\left(U_{1 i}, U_{2 i} ; Z_{i} \mid X_{1 i}, X_{2 i}, U_{0 i}\right) \\
& \quad=I\left(M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{i-1} ; Z_{i} \mid X_{1 i}, X_{2 i}, S_{0}\right) \\
& \quad \leq I\left(M_{[1: 2]}, S_{[1: 2]}, X_{1 \sim i}, X_{2 \sim i}, Z^{i-1} ; Z_{i} \mid X_{1 i}, X_{2 i}, S_{0}\right) \\
& \quad=I\left(S_{[1: 2]}, X_{1 \sim i}, X_{2 \sim i}, Z^{i-1} ; Z_{i} \mid X_{1 i}, X_{2 i}, S_{0}\right) \\
& \quad+I\left(M_{[1: 2]} ; Z_{i} \mid X_{1}^{n}, X_{2}^{n}, S_{[0: 2]}, Z^{i-1}\right) \\
& \quad \stackrel{(a)}{=} 0+0=0, \tag{62}
\end{align*}
$$

where in (a), the first term is zero since $S_{[0: 2]}$ is independent of $\left(X_{1}^{n}, X_{2}^{n}, Z^{n}\right)$ and $\left(X_{1 i}, X_{2 i}, Z_{i}\right), i=1, \ldots, n$, are jointly i.i.d., while the second term is zero because of the Markov chain $M_{[1: 2]} \rightarrow\left(X_{1}^{n}, X_{2}^{n}, S_{[0: 2]}\right) \rightarrow Z^{n}$.
We now show that $U_{1 i} \rightarrow\left(X_{1 i}, U_{0 i}\right) \rightarrow\left(X_{2 i}, Z_{i}\right)$ holds.

$$
\begin{align*}
& I\left(U_{1 i} ; X_{2 i}, Z_{i} \mid X_{1 i}, U_{0 i}\right) \\
& \quad=I\left(M_{1}, S_{1}, X_{1, i+1}^{n}, Z^{i-1} ; X_{2 i}, Z_{i} \mid X_{1 i}, S_{0}\right) \\
& \quad \leq I\left(M_{1}, S_{1}, X_{1 \sim i}, Z^{i-1} ; X_{2 i}, Z_{i} \mid X_{1 i}, S_{0}\right) \\
& \quad=I\left(S_{1}, X_{1 \sim i}, Z^{i-1} ; X_{2 i}, Z_{i} \mid X_{1 i}, S_{0}\right) \\
& \quad+I\left(M_{1} ; X_{2 i}, Z_{i} \mid X_{1}^{n}, S_{1}, Z^{i-1}, S_{0}\right) \\
& \quad \stackrel{(a)}{=} 0+0=0 \tag{63}
\end{align*}
$$

where in (a), the first term is zero since $\left(S_{0}, S_{1}\right)$ is independent of $\left(X_{1}^{n}, X_{2}^{n}, Z^{n}\right)$ and $\left(X_{1 i}, X_{2 i}, Z_{i}\right), i=1, \ldots, n$, are jointly i.i.d., while the second term is zero because of the Markov chain $M_{1} \rightarrow\left(X_{1}^{n}, S_{0}, S_{1}\right) \rightarrow\left(X_{2}^{n}, Z^{n}\right)$. In a similar fashion, we can show that $U_{2 i} \rightarrow\left(X_{2 i}, U_{0 i}\right) \rightarrow\left(X_{1 i}, Z_{i}\right)$ holds as well. Finally, let us show that $Y_{i} \rightarrow\left(U_{1 i}, U_{2 i}, Z_{i}\right) \rightarrow$ $\left(X_{1 i}, X_{2 i}, U_{0 i}\right)$ is a Markov chain as well.

$$
\begin{aligned}
& I\left(Y_{i} ; X_{1 i}, X_{2 i}, U_{0 i} \mid U_{1 i}, U_{2 i}, Z_{i}\right) \\
& =I\left(Y_{i} ; X_{1 i}, X_{2 i}, S_{0} \mid M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{i-1}, Z_{i}\right) \\
& \leq I\left(Y_{i}, Z_{i+1}^{n} ; X_{1 i}, X_{2 i}, S_{0} \mid M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{i}\right) \\
& =I\left(Z_{i+1}^{n} ; X_{1 i}, X_{2 i}, S_{0} \mid M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{i}\right) \\
& +I\left(Y_{i} ; X_{1 i}, X_{2 i}, S_{0} \mid M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{n}\right) \\
& \stackrel{(a)}{=} I\left(Z_{i+1}^{n} ; X_{1 i}, X_{2 i}, S_{0} \mid M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{i}\right) \\
& \leq I\left(Z_{i+1}^{n} ; X_{1}^{i}, X_{2}^{i}, S_{0} \mid M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{i}\right) \\
& \leq I\left(Z_{i+1}^{n} ; X_{1}^{i}, X_{2}^{i}, M_{[1: 2]}, S_{0} \mid S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{i}\right) \\
& =I\left(Z_{i+1}^{n} ; X_{1}^{i}, X_{2}^{i}, S_{0} \mid S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{i}\right) \\
& +I\left(Z_{i+1}^{n} ; M_{[1: 2]} \mid S_{[0: 2]}, X_{1}^{n}, X_{2}^{n}, Z^{i}\right) \\
& \stackrel{(b)}{=} 0+0=0 \text {, }
\end{aligned}
$$

where (a) follows from the Markov chain $Y^{n} \rightarrow$ $\left(M_{[1: 2]}, S_{[1: 2]}, Z^{n}\right) \rightarrow\left(S_{0}, X_{1}^{n}, X_{2}^{n}\right)$, and in (b), the first
term is zero since $S_{[0: 2]}$ is independent of $\left(X_{1}^{n}, X_{2}^{n}, Z^{n}\right)$ and $\left(X_{1 i}, X_{2 i}, Z_{i}\right), i=1, \ldots, n$, are jointly i.i.d., while the second term is zero because of the Markov chain $M_{[1: 2]} \rightarrow$ $\left(X_{1}^{n}, X_{2}^{n}, S_{[0: 2]}\right) \rightarrow Z^{n}$. Note that, for all $t \in[1: n]$,

$$
\left\|p_{X_{1 t}, X_{2 t}, Z_{t}, Y_{t} \mid T=t}-q_{X_{1}, X_{2}, Z, Y \mid T=t}\right\|_{1}
$$

$$
\begin{aligned}
& \stackrel{(a)}{\leq}\left\|p_{X_{1}^{n}, X_{2}^{n}, Z^{n}, Y^{n}}-q_{X_{1}, X_{2}, Z, Y}^{(n)}\right\|_{1} \\
& \stackrel{(b)}{<} \epsilon,
\end{aligned}
$$

where (a) follows from [5, Lemma V.1] and the fact that $T$ is independent of all other random variables, while (b) follows from (71). Therefore,

$$
\begin{aligned}
& \left\|p_{X_{1}, X_{2}, Z, Y \mid T=t}-q_{X_{1}, X_{2}, Z, Y \mid T=t}\right\|_{1} \\
& \quad=\left\|p_{X_{1 T}, X_{2 T}, Z_{T}, Y_{T} \mid T=t}-q_{X_{1}, X_{2}, Z, Y \mid T=t}\right\|_{1} \\
& \quad=\left\|p_{X_{1 t}, X_{2 t}, Z_{t}, Y_{t} \mid T=t}-q_{X_{1}, X_{2}, Z, Y \mid T=t}\right\|_{1} \\
& \quad \leq \epsilon .
\end{aligned}
$$

2) Proof of Theorem 2: Proof: Note that by Definitions 2, if $\left(R_{1}, R_{2}, R_{01}, R_{02}\right) \in \mathcal{R}_{\text {MAC-coord }}^{\mathrm{NO}-S_{0}}$, then we have $\left(R_{1}, R_{2}, 0, R_{01}, R_{02}\right) \in \mathcal{R}_{\mathrm{MAC}-\text { coord }}$. In other words, an outer bound for the case without shared randomness between the encoders can be obtained by invoking Theorem 5 with $R_{00}=0$. The proof of the implication would be complete by proving that the resulting outer bound is exactly identical to Theorem 2.

With $R_{00}=0$, Theorem 5 reduces to the set of ( $R_{1}, R_{2}, R_{01}, R_{02}$ ) such that

$$
\begin{array}{r}
R_{1} \geq \max \left\{I\left(U_{1} ; X_{1} \mid U_{0}, Z, T\right),\right. \\
\left.I\left(U_{1} ; X_{1} \mid U_{0}, U_{2}, X_{2}, Z, T\right)\right\} \\
R_{2} \geq \max \left\{I\left(U_{2} ; X_{2} \mid U_{0}, Z, T\right),\right. \\
\left.I\left(U_{2} ; X_{2} \mid U_{0}, U_{1}, X_{1}, Z, T\right)\right\} \\
R_{1}+R_{2} \geq I\left(U_{1}, U_{2} ; X_{1}, X_{2} \mid U_{0}, Z, T\right) \\
R_{1}+R_{01} \geq I\left(U_{0}, U_{1} ; X_{1}, X_{2}, Y \mid Z, T\right) \\
R_{2}+R_{02} \geq I\left(U_{0}, U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right) \\
R_{1}+R_{2}+R_{01}+R_{02} \geq I\left(U_{0}, U_{1}, U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right),
\end{array}
$$

for some p.m.f.

$$
\begin{aligned}
& p\left(x_{1}, x_{2}, z, t, u_{0}, u_{1}, u_{2}, y\right)= \\
& p\left(x_{1}, x_{2}, z\right) p(t) p\left(u_{0} \mid t\right) p\left(u_{1}, u_{2} \mid x_{1}, x_{2}, u_{0}, t\right) p\left(y \mid u_{1}, u_{2}, z, t\right)
\end{aligned}
$$

such that

$$
\begin{aligned}
& p\left(u_{1} \mid x_{1}, x_{2}, u_{0}, z, t\right)=p\left(u_{1} \mid x_{1}, u_{0}, t\right) \\
& p\left(u_{2} \mid x_{1}, x_{2}, u_{0}, z, t\right)=p\left(u_{2} \mid x_{2}, u_{0}, t\right) \\
&\left\|p\left(x_{1}, x_{2}, y, z \mid t\right)-q\left(x_{1}, x_{2}, y, z\right)\right\|_{1} \leq \epsilon \text { for all } t
\end{aligned}
$$

Let us define $U_{j}^{\prime} \triangleq\left(U_{0}, U_{j}\right)$ for $j=1,2$. Using the Markov chains $U_{0} \rightarrow T \rightarrow\left(X_{1}, X_{2}, Z\right), Y \rightarrow\left(U_{1}, U_{2}, Z, T\right) \rightarrow$ $\left(X_{1}, X_{2}, U_{0}\right),\left(U_{1}, U_{2}\right) \rightarrow\left(X_{1}, X_{2}, U_{0}, T\right) \rightarrow Z, U_{1} \rightarrow$ $\left(X_{1}, U_{0}, T\right) \rightarrow\left(X_{2}, Z\right)$, and $U_{2} \rightarrow\left(X_{2}, U_{0}, T\right) \rightarrow\left(X_{1}, Z\right)$, the region can be simplified as the set of $\left(R_{1}, R_{2}, R_{01}, R_{02}\right)$ such that

$$
\begin{aligned}
& R_{1} \geq \max \left\{I\left(U_{1}^{\prime} ; X_{1} \mid Z, T\right)\right. \\
& \left.\quad I\left(U_{1}^{\prime} ; X_{1} \mid U_{2}^{\prime}, X_{2}, Z, T\right)\right\}
\end{aligned}
$$

$$
\begin{array}{r}
R_{2} \geq \max \left\{I\left(U_{2}^{\prime} ; X_{2} \mid Z, T\right),\right. \\
\left.I\left(U_{2}^{\prime} ; X_{2} \mid U_{1}^{\prime}, X_{1}, Z, T\right)\right\} \\
R_{1}+R_{2} \geq I\left(U_{1}^{\prime}, U_{2}^{\prime} ; X_{1}, X_{2} \mid Z, T\right) \\
R_{1}+R_{01} \geq I\left(U_{1}^{\prime} ; X_{1}, X_{2}, Y \mid Z, T\right) \\
R_{2}+R_{02} \geq I\left(U_{2}^{\prime} ; X_{1}, X_{2}, Y \mid Z, T\right) \\
R_{1}+R_{2}+R_{01}+R_{02} \geq I\left(U_{1}^{\prime}, U_{2}^{\prime} ; X_{1}, X_{2}, Y \mid Z, T\right),
\end{array}
$$

for some p.m.f.

$$
\begin{aligned}
& p\left(x_{1}, x_{2}, z, t, u_{1}^{\prime}, u_{2}^{\prime}, y\right)= \\
& \quad p\left(x_{1}, x_{2}, z\right) p(t) p\left(u_{1}^{\prime}, u_{2}^{\prime} \mid x_{1}, x_{2}, t\right) p\left(y \mid u_{1}^{\prime}, u_{2}^{\prime}, z, t\right)
\end{aligned}
$$

such that

$$
\begin{aligned}
p\left(u_{1}^{\prime} \mid x_{1}, x_{2}, z, t\right) & =p\left(u_{1}^{\prime} \mid x_{1}, t\right) \\
p\left(u_{2}^{\prime} \mid x_{1}, x_{2}, z, t\right) & =p\left(u_{2}^{\prime} \mid x_{2}, t\right) \\
\left\|p\left(x_{1}, x_{2}, y, z \mid t\right)-q\left(x_{1}, x_{2}, y, z\right)\right\|_{1} & \leq \epsilon \text { for all } t
\end{aligned}
$$

3) Converse Proof of Theorem 3: Proof: We prove the following lemma later.

Lemma 4: Consider a p.m.f. $q_{X_{1} X_{2} Z Y}$ such that the random variables $X_{1}$ and $X_{2}$ are conditionally independent given $Z$. Then any rate tuple $\left(R_{1}, R_{2}, R_{00}, R_{01}, R_{02}\right)$ in $\mathcal{R}_{\text {MAC-coord }}$ satisfies, for every $\epsilon \in\left(0, \frac{1}{4}\right]$,

$$
\begin{align*}
R_{1} & \geq I\left(U_{1} ; X_{1} \mid Z, T\right)  \tag{64}\\
R_{2} & \geq I\left(U_{2} ; X_{2} \mid Z, T\right)  \tag{65}\\
R_{1}+R_{01} & \geq I\left(U_{1} ; X_{1}, Y \mid X_{2}, Z, T\right)-g(\epsilon) \\
R_{2}+R_{02} & \geq I\left(U_{2} ; X_{2}, Y \mid X_{1}, Z, T\right)-g(\epsilon)  \tag{66}\\
R_{1}+R_{2}+R_{01}+R_{02} & \geq I\left(U_{1}, U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right)-g(\epsilon), \tag{67}
\end{align*}
$$

with $g(\epsilon)=2 \sqrt{\epsilon}\left(H_{q}\left(X_{1}, X_{2}, Y, Z\right)+\log \frac{\left(\left|\mathcal{X}_{1}\left\|\mathcal{X}_{2}\right\| \mathcal{Y} \| \mathcal{Z}\right|\right)}{\epsilon}\right)$ (which tends to 0 as $\epsilon \rightarrow 0$ ), for some p.m.f.

$$
\begin{align*}
& p\left(x_{1}, x_{2}, z, t, u_{1}, u_{2}, y\right)= \\
& p(z) p\left(x_{1} \mid z\right) p\left(x_{2} \mid z\right) p(t) p\left(u_{1} \mid x_{1}, t\right) p\left(u_{2} \mid x_{2}, t\right) p\left(y \mid u_{1}, u_{2}, z, t\right) \tag{69}
\end{align*}
$$

such that $\left\|p\left(x_{1}, x_{2}, y, z \mid t\right)-q\left(x_{1}, x_{2}, y, z\right)\right\|_{1} \leq \epsilon$ for all $t$.
Notice that Lemma 4 gives an almost matching outer bound to the inner bound in Theorem 1 for $X_{1}$ and $X_{2}$ conditionally independent given $Z$ (see (41)), in the sense that continuity at $\epsilon=0$ is the only difference between them. When the shared randomness rate $R_{02}$ is sufficiently large, we can in fact prove that such a continuity argument holds. We argue this below.

First we prove the following lemma concerning the cardinality bounds on the auxiliary random variables $U_{1}$ and $U_{2}$ under sufficiently large shared randomness rate $R_{02}$. Let $S_{\epsilon}$ denote the set of rate tuples $\left(R_{1}, R_{2}, R_{01}\right)$ that satisfy, for every $\epsilon>0$,

$$
\begin{aligned}
R_{1} & \geq I\left(U_{1} ; X_{1} \mid Z, T\right) \\
R_{2} & \geq I\left(U_{2} ; X_{2} \mid Z, T\right) \\
R_{1}+R_{01} & \geq I\left(U_{1} ; X_{1}, Y \mid X_{2}, Z, T\right)-g(\epsilon)
\end{aligned}
$$

with $g(\epsilon)=2 \sqrt{\epsilon}\left(H_{q}\left(X_{1}, X_{2}, Y, Z\right)+\log \frac{\left(\left|\mathcal{X}_{1}\left\|\mathcal{X}_{2}\right\| \mathcal{Y} \| \mathcal{Z}\right|\right)}{\epsilon}\right)$ (which tends to 0 as $\epsilon \rightarrow 0$ ), for some p.m.f.

$$
\begin{align*}
& p\left(x_{1}, x_{2}, z, t, u_{1}, u_{2}, y\right)= \\
& p(z) p\left(x_{1} \mid z\right) p\left(x_{2} \mid z\right) p(t) p\left(u_{1} \mid x_{1}, t\right) p\left(u_{2} \mid x_{2}, t\right) p\left(y \mid u_{1}, u_{2}, z, t\right) \tag{70}
\end{align*}
$$

such that $\left\|p\left(x_{1}, x_{2}, y, z \mid t\right)-q\left(x_{1}, x_{2}, y, z\right)\right\|_{1} \leq \epsilon$ for all $t$.
Lemma 5: The size of the auxiliary random variable alphabets in $S_{\epsilon}$ can be restricted to:

$$
\begin{aligned}
\left|\mathcal{U}_{1}\right| & \leq\left|\mathcal{X}_{1}\left\|\mathcal{X}_{2}| | \mathcal{Y}\right\| \mathcal{Z}\right| \\
\left|\mathcal{U}_{2}\right| & \leq\left|\mathcal{U}_{1}\right|\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\|\mathcal{Y}\| \mathcal{Z}\right|, \\
|\mathcal{T}| & \leq 3
\end{aligned}
$$

Proof: See Appendix E.
From Lemmas 4 and 5, the converse for Theorem 3 follows using the continuity of total variation distance and mutual information in the probability simplex along the same lines as [5, Lemma VI.5] and [8, Lemma 6]. It remains to prove Lemma 4.

Proof of Lemma 4: Consider a code that induces a joint distribution on $\left(X_{1}^{n}, X_{2}^{n}, Z^{n}, Y^{n}\right)$ such that

$$
\begin{equation*}
\left\|p_{X_{1}^{n}, X_{2}^{n}, Z^{n}, Y^{n}}-q_{X_{1} X_{2} Z Y}^{(n)}\right\|_{1}<\epsilon . \tag{71}
\end{equation*}
$$

Let us first prove the lower bound on $R_{j}$ for $j \in\{1,2\}$ in (64)-(65).

$$
\begin{align*}
n R_{j} & \geq H\left(M_{j}\right) \\
& \geq H\left(M_{j} \mid S_{j}, Z^{n}\right) \\
& \geq I\left(M_{j} ; X_{j}^{n} \mid S_{j}, Z^{n}\right) \\
& \stackrel{(a)}{=} I\left(M_{j}, S_{j} ; X_{j}^{n} \mid Z^{n}\right) \\
& =\sum_{i=1}^{n} I\left(M_{j}, S_{j} ; X_{j i} \mid X_{j, i+1}^{n}, Z^{n}\right) \\
& \stackrel{(b)}{=} \sum_{i=1}^{n} I\left(M_{j}, S_{j}, Z_{\sim i}, X_{j, i+1}^{n} ; X_{j i} \mid Z_{i}\right) \\
& \geq \sum_{i=1}^{n} I\left(M_{j}, S_{j}, X_{j, i+1}^{n}, Z^{i-1} ; X_{j i} \mid Z_{i}\right) \\
& \stackrel{(c)}{\geq} \sum_{i=1}^{n} I\left(U_{j i} ; X_{j i} \mid Z_{i}\right) \\
& \stackrel{(d)}{=} n I\left(U_{j T} ; X_{j T} \mid Z_{T}, T\right) \\
& \stackrel{(e)}{=} n I\left(U_{j} ; X_{j} \mid Z, T\right), \tag{72}
\end{align*}
$$

where (a) follows since $S_{j}$ is independent of $\left(X_{j}^{n}, Z^{n}\right)$, (b) follows since $\left(X_{j i}, Z_{i}\right), i=1, \ldots, n$, are jointly i.i.d., (c) follows by defining $U_{1 i}=\left(M_{1}, S_{1}, X_{1, i+1}^{n}, Z^{i-1}\right)$ and $U_{2 i}=\left(M_{2}, S_{2}, X_{2, i+1}^{n}\right)$, (d) follows by introducing a uniform time-sharing random variable $T \in[1: n]$ that is independent of everything else, while (e) follows by defining $U_{1}:=U_{1 T}$, $U_{2}:=U_{2 T}, X_{1}:=X_{1 T}, X_{2}:=X_{2 T}, Y:=Y_{T}$ and $Z:=Z_{T}$.

We next derive the lower bound on $\left(R_{1}+R_{01}\right)$ in (66).

$$
\begin{aligned}
& n\left(R_{1}+R_{01}\right) \\
& \geq H\left(M_{1}, S_{1}\right) \\
& \geq H\left(M_{1}, S_{1} \mid X_{2}^{n}, Z^{n}\right)
\end{aligned}
$$

$$
\begin{align*}
& \geq I\left(M_{1}, S_{1} ; X_{1}^{n}, Y^{n} \mid X_{2}^{n}, Z^{n}\right) \\
&= \sum_{i=1}^{n} I\left(M_{1}, S_{1} ; X_{1 i}, Y_{i} \mid X_{1, i+1}^{n}, Y_{i+1}^{n}, X_{2}^{n}, Z^{n}\right) \\
&= \sum_{i=1}^{n} I\left(M_{1}, S_{1}, X_{1, i+1}^{n}, Y_{i+1}^{n}, X_{2 \sim i}, Z_{\sim i} ; X_{1 i}, Y_{i} \mid X_{2 i}, Z_{i}\right) \\
& \quad-\sum_{i=1}^{n} I\left(X_{1, i+1}^{n}, Y_{i+1}^{n}, X_{2 \sim i}, Z_{\sim i} ; X_{1 i}, Y_{i} \mid X_{2 i}, Z_{i}\right) \\
&\left(\begin{array}{l}
(a) \\
\geq \\
\sum_{i=1}^{n} I\left(M_{1}, S_{1}, X_{1, i+1}^{n}, Z^{i-1} ; X_{1 i}, Y_{i} \mid X_{2 i}, Z_{i}\right)-n g(\epsilon) \\
= \\
\sum_{i=1}^{n} I\left(U_{1 i} ; X_{1 i}, Y_{i} \mid X_{2 i}, Z_{i}\right)-n g(\epsilon) \\
= \\
= \\
= \\
\\
n I\left(U_{1 T} ; X_{1 T}, Y_{1} ; X_{1}, Y \mid X_{2 T}, Z, T\right)-n g(\epsilon)
\end{array}\right.
\end{align*}
$$

where (a) follows since

$$
\begin{aligned}
& \sum_{i=1}^{n} I\left(X_{1, i+1}^{n}, Y_{i+1}^{n}, X_{2 \sim i}, Z_{\sim i} ; X_{1 i}, Y_{i} \mid X_{2 i}, Z_{i}\right) \\
& \quad \leq \sum_{i=1}^{n} I\left(X_{1 \sim i}, X_{2 \sim i}, Y_{\sim i}, Z_{\sim i} ; X_{1 i}, X_{2 i}, Y_{i}, Z_{i}\right) \\
& \quad \leq n g(\epsilon)
\end{aligned}
$$

by (71) and Lemma 2. The bound $n\left(R_{2}+R_{02}\right) \geq$ $I\left(U_{2} ; X_{2}, Y \mid X_{1}, Z, T\right)$ follows in a similar manner.

For the lower bound on $\left(R_{1}+R_{2}+R_{01}+R_{02}\right)$ in (68), we proceed as follows.

$$
\begin{align*}
& n\left(R_{1}+R_{2}+R_{01}+R_{02}\right) \\
& \geq H\left(M_{[1: 2]}, S_{[1: 2]}\right) \\
& \geq H\left(M_{[1: 2]}, S_{[1: 2]} \mid Z^{n}\right) \\
& \geq I\left(M_{[1: 2]}, S_{[1: 2]} ; X_{1}^{n}, X_{2}^{n}, Y^{n} \mid Z^{n}\right) \\
& =\sum_{i=1}^{n} I\left(M_{[1: 2]}, S_{[1: 2]} ; X_{1 i}, X_{2 i}, Y_{i} \mid X_{1, i+1}^{n}, X_{2, i+1}^{n}, Y_{i+1}^{n}, Z^{n}\right) \\
& =\sum_{i=1}^{n} I\left(M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Y_{i+1}^{n}, Z_{\sim i} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right) \\
& \quad-\sum_{i=1}^{n} I\left(X_{1, i+1}^{n}, X_{2, i+1}^{n}, Y_{i+1}^{n}, Z_{\sim i} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right) \\
& \begin{array}{l}
\geq \\
\geq
\end{array} \sum_{i=1}^{n} I\left(M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{i-1} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right) \\
& \quad-n g(\epsilon) \\
& = \\
& =\sum_{i=1}^{n} I\left(U_{1 i}, U_{2 i} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right)-n g(\epsilon) \\
& =n I\left(U_{1 T}, U_{2 T} ; X_{1 T}, X_{2 T}, Y_{T} \mid Z_{T}, T\right)-n g(\epsilon)  \tag{74}\\
& =n I\left(U_{1}, U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right)-n g(\epsilon),
\end{align*}
$$

where (a) follows since

$$
\begin{array}{rl}
\sum_{i=1}^{n} & I\left(X_{1, i+1}^{n}, X_{2, i+1}^{n}, Y_{i+1}^{n}, Z_{\sim i} ; X_{1 i}, X_{2 i}, Y_{i} \mid Z_{i}\right) \\
& \leq \sum_{i=1}^{n} I\left(X_{1 \sim i}, X_{2 \sim i}, Y_{\sim i}, Z_{\sim i} ; X_{1 i}, X_{2 i}, Y_{i}, Z_{i}\right) \\
\quad \leq n g(\epsilon)
\end{array}
$$

by (71) and Lemma 2.

We now prove the Markov chains $U_{1 i} \rightarrow X_{1 i} \rightarrow\left(X_{2 i}, Z_{i}\right)$, $U_{2 i} \rightarrow X_{2 i} \rightarrow\left(U_{1 i}, X_{1 i}, Z_{i}\right),\left(U_{1 i}, U_{2 i}\right)-\left(X_{1 i}, X_{2 i}\right)-Z_{i}$, and $Y_{i} \rightarrow\left(U_{1 i}, U_{2 i}, Z_{i}\right) \rightarrow\left(X_{1 i}, X_{2 i}\right)$. Note that this implies that the joint p.m.f. satisfies (70). Recall the auxiliary random variable identifications

$$
\begin{align*}
U_{1 i} & =\left(M_{1}, S_{1}, X_{1, i+1}^{n}, Z^{i-1}\right)  \tag{75}\\
U_{2 i} & =\left(M_{2}, S_{2}, X_{2, i+1}^{n}\right) \tag{76}
\end{align*}
$$

Let us first show that $\left(U_{1 i}, U_{2 i}\right) \rightarrow\left(X_{1 i}, X_{2 i}\right) \rightarrow Z_{i}$.

$$
\begin{align*}
& I\left(U_{1 i}, U_{2 i} ; Z_{i} \mid X_{1 i}, X_{2 i}\right) \\
& \quad=I\left(M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{i-1} ; Z_{i} \mid X_{1 i}, X_{2 i}\right) \\
& \quad \leq I\left(M_{[1: 2]}, S_{[1: 2]}, X_{1 \sim i}, X_{2 \sim i}, Z^{i-1} ; Z_{i} \mid X_{1 i}, X_{2 i}\right) \\
& \quad=I\left(S_{[1: 2]}, X_{1 \sim i}, X_{2 \sim i}, Z^{i-1} ; Z_{i} \mid X_{1 i}, X_{2 i}\right) \\
& \quad+I\left(M_{[1: 2]} ; Z_{i} \mid X_{1}^{n}, X_{2}^{n}, S_{[1: 2]}, Z^{i-1}\right) \\
& \quad \stackrel{(a)}{=} 0+0=0 \tag{77}
\end{align*}
$$

where in (a), the first term is zero since $S_{[1: 2]}$ is independent of $\left(X_{1}^{n}, X_{2}^{n}, Z^{n}\right)$ and $\left(X_{1 i}, X_{2 i}, Z_{i}\right), i=1, \ldots, n$, are jointly i.i.d., while the second term is zero because of the Markov chain $M_{[1: 2]} \rightarrow\left(X_{1}^{n}, X_{2}^{n}, S_{[1: 2]}\right) \rightarrow Z^{n}$. We now show that $U_{1 i} \rightarrow X_{1 i} \rightarrow\left(X_{2 i}, Z_{i}\right)$ is a Markov chain.

$$
\begin{align*}
& I\left(U_{1 i} ; X_{2 i}, Z_{i} \mid X_{1 i}\right) \\
& \quad=I\left(M_{1}, S_{1}, X_{1, i+1}^{n}, Z^{i-1} ; X_{2 i}, Z_{i} \mid X_{1 i}\right) \\
& \quad \leq I\left(M_{1}, S_{1}, X_{1 \sim i}, Z^{i-1} ; X_{2 i}, Z_{i} \mid X_{1 i}\right) \\
& \quad=I\left(S_{1}, X_{1 \sim i}, Z^{i-1} ; X_{2 i}, Z_{i} \mid X_{1 i}\right) \\
& \quad+I\left(M_{1} ; X_{2 i}, Z_{i} \mid X_{1}^{n}, S_{1}, Z^{i-1}\right) \\
& \quad \stackrel{(a)}{=} 0+0=0, \tag{78}
\end{align*}
$$

where in (a), the first term is zero since $S_{1}$ is independent of $\left(X_{1}^{n}, X_{2}^{n}, Z^{n}\right)$ and $\left(X_{1 i}, X_{2 i}, Z_{i}\right), i=1, \ldots, n$, are jointly i.i.d., while the second term is zero because of the Markov chain $M_{1} \rightarrow\left(X_{1}^{n}, S_{1}\right) \rightarrow\left(X_{2}^{n}, Z^{n}\right)$. Next, we show that $Y_{i} \rightarrow\left(U_{1 i}, U_{2 i}, Z_{i}\right) \rightarrow\left(X_{1 i}, X_{2 i}\right)$ is a Markov chain.

$$
\begin{aligned}
& I\left(Y_{i} ; X_{1 i}, X_{2 i} \mid U_{1 i}, U_{2 i}, Z_{i}\right) \\
& \quad=I\left(Y_{i} ; X_{1 i}, X_{2 i} \mid M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{i-1}, Z_{i}\right) \\
& \quad \leq I\left(Y_{i}, Z_{i+1}^{n} ; X_{1 i}, X_{2 i} \mid M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{i}\right) \\
& \quad=I\left(Z_{i+1}^{n} ; X_{1 i}, X_{2 i} \mid M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{i}\right) \\
& \quad+I\left(Y_{i} ; X_{1 i}, X_{2 i} \mid M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{n}\right) \\
& \quad \stackrel{(a)}{=} I\left(Z_{i+1}^{n} ; X_{1 i}, X_{2 i} \mid M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{i}\right) \\
& \quad \leq I\left(Z_{i+1}^{n} ; X_{1}^{i}, X_{2}^{i} \mid M_{[1: 2]}, S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{i}\right) \\
& \quad \leq I\left(Z_{i+1}^{n} ; X_{1}^{i}, X_{2}^{i}, M_{[1: 2]} \mid S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{i}\right) \\
& \quad=I\left(Z_{i+1}^{n} ; X_{1}^{i}, X_{2}^{i} \mid S_{[1: 2]}, X_{1, i+1}^{n}, X_{2, i+1}^{n}, Z^{i}\right) \\
& \quad+I\left(Z_{i+1}^{n} ; M_{[1: 2]} \mid S_{[1: 2]}, X_{1}^{n}, X_{2}^{n}, Z^{i}\right) \\
& \quad \stackrel{(a)}{=} 0+0=0,
\end{aligned}
$$

where (a) follows from the Markov chain $Y^{n} \rightarrow$ $\left(M_{[1: 2]}, S_{[1: 2]}, Z^{n}\right) \rightarrow\left(X_{1}^{n}, X_{2}^{n}\right)$ and in (b), the first term is zero since $S_{[1: 2]}$ is independent of $\left(X_{1}^{n}, X_{2}^{n}, Z^{n}\right)$ and $\left(X_{1 i}, X_{2 i}, Z_{i}\right), i=1, \ldots, n$, are jointly i.i.d., while the
second term is zero because of the Markov chain $M_{[1: 2]} \rightarrow$ $\left(X_{1}^{n}, X_{2}^{n}, S_{[1: 2]}\right) \rightarrow Z^{n}$.

It remains to prove the Markov chain $U_{2 i} \rightarrow X_{2 i} \rightarrow$ ( $\left.U_{1 i}, X_{1 i}, Z_{i}\right)$. Consider the following chain of inequalities.

$$
\begin{align*}
& I\left(U_{2 i} ;\right.\left.U_{1 i}, X_{1 i}, Z_{i} \mid X_{2 i}\right) \\
&= I\left(M_{2}, S_{2}, X_{2, i+1}^{n} ; M_{1}, S_{1}, X_{1, i}^{n}, Z^{i} \mid X_{2 i}\right) \\
&= I\left(S_{2}, X_{2, i+1}^{n} ; M_{1}, S_{1}, X_{1, i}^{n}, Z^{i} \mid X_{2 i}\right) \\
&+I\left(M_{2} ; M_{1}, S_{1}, X_{1, i}^{n}, Z^{i} \mid S_{2}, X_{2, i}^{n}\right) \\
&= I\left(S_{2}, X_{2, i+1}^{n} ; S_{1}, X_{1, i}^{n}, Z^{i} \mid X_{2 i}\right) \\
&+I\left(S_{2}, X_{2, i+1}^{n} ; M_{1} \mid S_{1}, X_{1, i}^{n}, Z^{i}, X_{2 i}\right) \\
&+I\left(M_{2} ; S_{1}, X_{1, i}^{n}, Z^{i} \mid S_{2}, X_{2, i}^{n}\right) \\
&+I\left(M_{2} ; M_{1} \mid S_{1}, S_{2}, X_{1, i}^{n}, X_{2, i}^{n}, Z^{i}\right) \\
& \text { (a) } \\
& \leq 0+I\left(S_{2}, X_{2, i+1}^{n} ; M_{1}, X_{1}^{i-1}, Z^{i-1} \mid S_{1}, X_{1, i}^{n}, Z^{i}, X_{2 i}\right) \\
&+I\left(M_{2}, X_{2}^{i-1} ; S_{1}, X_{1, i}^{n}, Z^{i} \mid S_{2}, X_{2, i}^{n}\right) \\
&+I\left(M_{2}, X_{2}^{i-1} ; M_{1} \mid S_{1}, S_{2}, X_{1, i}^{n}, X_{2, i}^{n}, Z^{i}\right) \\
&= I\left(S_{2}, X_{2, i+1}^{n} ; X_{1}^{i-1}, Z^{i-1} \mid S_{1}, X_{1, i}^{n}, Z^{i}, X_{2 i}\right) \\
&+I\left(S_{2}, X_{2, i+1}^{n} ; M_{1} \mid S_{1}, X_{1}^{n}, Z^{n}, X_{2 i}\right) \\
&+I\left(X_{2}^{i-1} ; S_{1}, X_{1, i}^{n}, Z^{i} \mid S_{2}, X_{2, i}^{n}\right) \\
&+I\left(M_{2} ; S_{1}, X_{1, i}^{n}, Z^{i} \mid S_{2}, X_{2}^{n}\right) \\
&+I\left(X_{2}^{i-1} ; M_{1} \mid S_{1}, S_{2}, X_{1, i}^{n}, X_{2, i}^{n}, Z^{i}\right) \\
&+I\left(M_{2} ; M_{1} \mid S_{1}, S_{2}, X_{1, i}^{n}, X_{2}^{n}, Z^{i}\right) \\
& \stackrel{(b)}{=} 0+0+0+0+I\left(X_{2}^{i-1} ; M_{1} \mid S_{1}, S_{2}, X_{1, i}^{n}, X_{2, i}^{n}, Z^{i}\right)+0 \\
& \leq I\left(X_{2}^{i-1} ; M_{1}, X_{1}^{i-1} \mid S_{1}, S_{2}, X_{1, i}^{n}, X_{2, i}^{n}, Z^{i}\right) \\
&= I\left(X_{2}^{i-1} ; X_{1}^{i-1} \mid S_{1}, S_{2}, X_{1, i}^{n}, X_{2, i}^{n}, Z^{i}\right) \\
&+I\left(X_{2}^{i-1} ; M_{1} \mid S_{1}, S_{2}, X_{1}^{n}, X_{2, i}^{n}, Z^{i}\right) \\
& \stackrel{(c)}{=} I\left(X_{2}^{i-1} ; X_{1}^{i-1} \mid Z^{i-1}\right) \\
&+I\left(X_{2}^{i-1} ; M_{1} \mid S_{1}, S_{2}, X_{1}^{n}, X_{2, i}^{n}, Z^{i}\right)  \tag{79}\\
&(\stackrel{d}{=} 0+0=0, \\
&+09)
\end{align*}
$$

where the fact that the first term in (a) and the first and third terms in (b) are zeros, as well as (c), follow since $S_{[1: 2]}$ is independent of $\left(X_{1}^{n}, X_{2}^{n}, Z^{n}\right)$ and $\left(X_{1 i}, X_{2 i}, Z_{i}\right), i=$ $1, \ldots, n$, are jointly i.i.d. The second, fourth and sixth terms in (b) and the second term in (d) are zeros because of the Markov chains $M_{1} \rightarrow\left(X_{1}^{n}, S_{1}\right) \rightarrow\left(S_{2}, X_{2}^{n}, Z^{n}\right)$ and $M_{2} \rightarrow$ $\left(X_{2}^{n}, S_{2}\right) \rightarrow\left(S_{1}, M_{1}, X_{1}^{n}, Z^{n}\right)$. The first term in (d) is zero because $\left(X_{1 i}, X_{2 i}, Z_{i}\right), i=1, \ldots, n$, are jointly i.i.d. with $q_{X_{1} X_{2}, Z}$ and $I\left(X_{1} ; X_{2} \mid Z\right)=0$. Notice that we also have $\left\|p_{X_{1}, X_{2}, Z, Y \mid T=t}-q_{X_{1}, X_{2}, Z, Y \mid T=t}\right\|_{1} \leq \epsilon$ along the same lines as in the proof of Theorem 5.
4) Proof of Theorem 6: Proof: We prove the following lemma later.

Lemma 6: Consider a p.m.f. $q_{X_{1} X_{2} Z Y}$ such that the random variables $X_{1}$ and $X_{2}$ are conditionally independent given $Z$. Then any rate tuple $\left(R_{1}, R_{2}, R_{00}, R_{01}, R_{02}\right)$ in $\mathcal{R}_{\mathrm{MAC}-\text { coord }}$ satisfies, for every $\epsilon \in\left(0, \frac{1}{4}\right]$,

$$
\begin{aligned}
& R_{1} \geq I\left(U_{0}, U_{1} ; X_{1} \mid Z, T\right) \\
& R_{2} \geq I\left(U_{0}, U_{2} ; X_{2} \mid Z, T\right)
\end{aligned}
$$

$$
\begin{aligned}
& R_{1}+R_{2} \geq I\left(U_{0}, U_{1}, U_{2} ; X_{1}, X_{2} \mid Z, T\right) \\
& R_{1}+R_{01} \geq I\left(U_{1} ; X_{1}, Y \mid X_{2}, Z, T\right)-g(\epsilon) \\
& R_{2}+R_{02} \geq I\left(U_{2} ; X_{2}, Y \mid X_{1}, Z, T\right)-g(\epsilon) \\
& R_{00}+R_{1}+R_{01} \geq I\left(U_{0}, U_{1} ; X_{1}, Y \mid X_{2}, Z, T\right)-g(\epsilon) \\
& R_{00}+R_{2}+R_{02} \geq I\left(U_{0}, U_{2} ; X_{2}, Y \mid X_{1}, Z, T\right)-g(\epsilon) \\
& R_{1}+R_{2}+R_{01}+R_{02} \geq I\left(U_{1}, U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right)-g(\epsilon) \\
& R_{00}+R_{1}+R_{2}+R_{01}+R_{02} \\
& \geq I\left(U_{0}, U_{1}, U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right)-g(\epsilon),
\end{aligned}
$$

with $g(\epsilon)=2 \sqrt{\epsilon}\left(H_{q}\left(X_{1}, X_{2}, Y, Z\right)+\log \frac{\left(\left|\mathcal{X}_{1}\left\|\mathcal{X}_{2}\right\| \mathcal{Y} \| \mathcal{Z}\right|\right)}{\epsilon}\right)$ (which tends to 0 as $\epsilon \rightarrow 0$ ), for some p.m.f.

$$
\begin{aligned}
& p\left(x_{1}, x_{2}, z, t, u_{0}, u_{1}, u_{2}, y\right)=p(z) p\left(x_{1} \mid z\right) p\left(x_{2} \mid z\right) p(t) \\
& \quad \times p\left(u_{0} \mid t\right) p\left(u_{1} \mid x_{1}, u_{0}, t\right) p\left(u_{2} \mid x_{2}, u_{0}, t\right) p\left(y \mid u_{1}, u_{2}, z, t\right)
\end{aligned}
$$

such that $\left\|p\left(x_{1}, x_{2}, y, z \mid t\right)-q\left(x_{1}, x_{2}, y, z\right)\right\|_{1} \leq \epsilon$ for all $t$.
Notice that Lemma 6 gives an epsilon rate region, whose continuity at $\epsilon=0$ is unknown. When the shared randomness rates $\left(R_{00}, R_{01}, R_{02}\right)$ are sufficiently large, we can in fact prove that such a continuity argument holds. We argue this below.

First we prove the following lemma concerning the cardinality bounds on the auxiliary random variables $U_{0}$, $U_{1}$ and $U_{2}$ under sufficiently large shared randomness rates ( $R_{00}, R_{01}, R_{02}$ ). Let $S_{\epsilon}^{\prime}$ denote the set of rate pairs $\left(R_{1}, R_{2}\right)$ that satisfy, for every $\epsilon>0$,

$$
\begin{aligned}
R_{1} & \geq I\left(U_{0}, U_{1} ; X_{1} \mid Z, T\right) \\
R_{2} & \geq I\left(U_{0}, U_{2} ; X_{2} \mid Z, T\right) \\
R_{1}+R_{2} & \geq I\left(U_{0}, U_{1}, U_{2} ; X_{1}, X_{2} \mid Z, T\right)
\end{aligned}
$$

for some p.m.f.

$$
\begin{aligned}
& p\left(x_{1}, x_{2}, z, t, u_{0}, u_{1}, u_{2}, y\right)=p(z) p\left(x_{1} \mid z\right) p\left(x_{2} \mid z\right) p(t) \\
& \quad \times p\left(u_{0} \mid t\right) p\left(u_{1} \mid x_{1}, u_{0}, t\right) p\left(u_{2} \mid x_{2}, u_{0}, t\right) p\left(y \mid u_{1}, u_{2}, z, t\right)
\end{aligned}
$$

such that $\left\|p\left(x_{1}, x_{2}, y, z \mid t\right)-q\left(x_{1}, x_{2}, y, z\right)\right\|_{1} \leq \epsilon$ for all $t$.
Lemma 7: The size of the auxiliary random variable alphabets in $S_{\epsilon}^{\prime}$ can be restricted to:

$$
\begin{aligned}
\left|\mathcal{U}_{0}\right| & \leq\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\right||\mathcal{Y}||\mathcal{Z}|, \\
\left|\mathcal{U}_{1}\right| & \leq\left|\mathcal{U}_{0}\right|\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\|\mathcal{Y}\| \mathcal{Z}\right|, \\
\left|\mathcal{U}_{2}\right| & \leq\left|\mathcal{U}_{0}\right|\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\|\mathcal{Y}\| \mathcal{Z}\right|, \\
|\mathcal{T}| & \leq 3 .
\end{aligned}
$$

## Proof: See Appendix F.

From Lemmas 6 and 7, the outer bound in Theorem 6 follows using the continuity of total variation distance and mutual information in the probability simplex along the same lines as [5, Lemma VI.5] and [8, Lemma 6]. It remains to prove Lemma 6.

Proof of Lemma 6: With the same choice of auxiliary random variables as in the proof of Theorem 5, i.e., $U_{0 i}=S_{0}$, $U_{1 i}=\left(M_{1}, S_{1}, X_{1, i+1}^{n}, Z^{i-1}\right)$, and $U_{2 i}=\left(M_{2}, S_{2}, X_{2, i+1}^{n}\right)$, we will show that the Markov chain $U_{2} \rightarrow\left(X_{2}, U_{0}, T\right) \rightarrow$ $\left(U_{1}, X_{1}, Z\right)$ holds when $X_{1} \rightarrow Z \rightarrow X_{2}$. Then in addition to the Markov chains $U_{0} \rightarrow T \rightarrow\left(X_{1}, X_{2}, Z\right)$,
$U_{1} \rightarrow\left(X_{1}, U_{0}, T\right) \rightarrow\left(X_{2}, Z\right),\left(U_{1}, U_{2}\right) \rightarrow\left(X_{1}, X_{2}\right.$, $\left.U_{0}, T\right) \rightarrow Z$ and $Y \rightarrow\left(U_{1}, U_{2}, Z, T\right) \rightarrow\left(X_{1}, X_{2}, U_{0}\right)$, we note the following simplifications to the rate constraints in Theorem 5:

$$
\begin{aligned}
& R_{1} \geq \max \left\{I\left(U_{1} ; X_{1} \mid U_{0}, Z, T\right), I\left(U_{1} ; X_{1} \mid U_{0}, U_{2}, X_{2}, Z, T\right)\right\} \\
&=\max \left\{I\left(U_{1} ; X_{1} \mid U_{0}, Z, T\right), I\left(U_{1} ; X_{1} \mid U_{0}, Z, T\right)\right\} \\
&=I\left(U_{0}, U_{1} ; X_{1} \mid Z, T\right) \\
& R_{2} \geq \max \left\{I\left(U_{2} ; X_{2} \mid U_{0}, Z, T\right), I\left(U_{2} ; X_{2} \mid U_{0}, U_{1}, X_{1}, Z, T\right)\right\} \\
&=\max \left\{I\left(U_{2} ; X_{2} \mid U_{0}, Z, T\right), I\left(U_{2} ; X_{2} \mid U_{0}, Z, T\right)\right\} \\
&=I\left(U_{0}, U_{2} ; X_{2} \mid Z, T\right) \\
& R_{1}+R_{2} \geq I\left(U_{1}, U_{2} ; X_{1}, X_{2} \mid U_{0}, Z, T\right) \\
&=I\left(U_{0}, U_{1}, U_{2} ; X_{1}, X_{2} \mid Z, T\right) \\
& R_{1}+R_{01} \geq I\left(U_{1} ; X_{1}, X_{2}, Y \mid Z, T\right) \\
&=I\left(U_{1} ; X_{1}, Y \mid X_{2}, Z, T\right), \\
& \\
& R_{2}+R_{02} \geq I\left(U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right) \\
&=I\left(U_{2} ; X_{2}, Y \mid X_{1}, Z, T\right), \\
& R_{00}+R_{1}+R_{01} \geq I\left(U_{0}, U_{1} ; X_{1}, X_{2}, Y \mid Z, T\right) \\
&=I\left(U_{0}, U_{1} ; X_{1}, Y \mid X_{2}, Z, T\right) \\
& R_{00}+R_{2}+R_{02} \geq I\left(U_{0}, U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right) \\
&=I\left(U_{0}, U_{2} ; X_{2}, Y \mid X_{1}, Z, T\right), \\
& R_{1}+R_{2}+R_{01}+R_{02} \geq I\left(U_{1}, U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right), \\
& R_{00}+R_{1}+R_{2}+R_{01}+ R_{02} \\
& \geq I\left(U_{0}, U_{1}, U_{2} ; X_{1}, X_{2}, Y \mid Z, T\right) .
\end{aligned}
$$

It remains to prove the Markov chain $U_{2 i} \rightarrow\left(X_{2 i}, U_{0 i}\right) \rightarrow$ ( $U_{1 i}, X_{1 i}, Z_{i}$ ). Recall the choice of auxiliary random variables.

$$
\begin{aligned}
& U_{0 i}=S_{0} \\
& U_{1 i}=\left(M_{1}, S_{1}, X_{1, i+1}^{n}, Z^{i-1}\right) \\
& U_{2 i}=\left(M_{2}, S_{2}, X_{2, i+1}^{n}\right)
\end{aligned}
$$

Consider the following chain of inequalities.

$$
\begin{aligned}
& I\left(U_{2 i} ; U_{1 i}, X_{1 i}, Z_{i} \mid X_{2 i}, U_{0 i}\right) \\
& =I\left(M_{2}, S_{2}, X_{2, i+1}^{n} ; M_{1}, S_{1}, X_{1, i}^{n}, Z^{i} \mid X_{2 i}, S_{0}\right) \\
& =I\left(S_{2}, X_{2, i+1}^{n} ; M_{1}, S_{1}, X_{1, i}^{n}, Z^{i} \mid X_{2 i}, S_{0}\right) \\
& \quad+I\left(M_{2} ; M_{1}, S_{1}, X_{1, i}^{n}, Z^{i} \mid S_{0}, S_{2}, X_{2, i}^{n}\right) \\
& =I\left(S_{2}, X_{2, i+1}^{n} ; S_{1}, X_{1, i}^{n}, Z^{i} \mid X_{2 i}, S_{0}\right) \\
& \quad+I\left(S_{2}, X_{2, i+1}^{n} ; M_{1} \mid S_{1}, X_{1, i}^{n}, Z^{i}, X_{2 i}, S_{0}\right) \\
& \quad+I\left(M_{2} ; S_{1}, X_{1, i}^{n}, Z^{i} \mid S_{0}, S_{2}, X_{2, i}^{n}\right) \\
& \quad+I\left(M_{2} ; M_{1} \mid S_{0}, S_{1}, S_{2}, X_{1, i}^{n}, X_{2, i}^{n}, Z^{i}\right) \\
& \quad(a) \\
& \leq 0 \\
& \quad+I\left(S_{2}, X_{2, i+1}^{n} ; M_{1}, X_{1}^{i-1}, Z^{i-1} \mid S_{0}, S_{1}, X_{1, i}^{n}, Z^{i}, X_{2 i}\right) \\
& \quad+I\left(M_{2}, X_{2}^{i-1} ; S_{1}, X_{1, i}^{n}, Z^{i} \mid S_{0}, S_{2}, X_{2, i}^{n}\right) \\
& \quad+I\left(M_{2}, X_{2}^{i-1} ; M_{1} \mid S_{0}, S_{1}, S_{2}, X_{1, i}^{n}, X_{2, i}^{n}, Z^{i}\right) \\
& =I\left(S_{2}, X_{2, i+1}^{n} ; X_{1}^{i-1}, Z^{i-1} \mid S_{0}, S_{1}, X_{1, i}^{n}, Z^{i}, X_{2 i}\right) \\
& \quad+I\left(S_{2}, X_{2, i+1}^{n} ; M_{1} \mid S_{0}, S_{1}, X_{1}^{n}, Z^{n}, X_{2 i}\right) \\
& \quad+I\left(X_{2}^{i-1} ; S_{1}, X_{1, i}^{n}, Z^{i} \mid S_{0}, S_{2}, X_{2, i}^{n}\right) \\
& \quad+I\left(M_{2} ; S_{1}, X_{1, i}^{n}, Z^{i} \mid S_{0}, S_{2}, X_{2}^{n}\right) \\
& \quad+I\left(X_{2}^{i-1} ; M_{1} \mid S_{0}, S_{1}, S_{2}, X_{1, i}^{n}, X_{2, i}^{n}, Z^{i}\right) \\
& \quad+I\left(M_{2} ; M_{1} \mid S_{0}, S_{1}, S_{2}, X_{1, i}^{n}, X_{2}^{n}, Z^{i}\right)
\end{aligned}
$$

$$
\begin{align*}
& \stackrel{(b)}{=} 0+0+0+0 \\
& \quad+I\left(X_{2}^{i-1} ; M_{1} \mid S_{0}, S_{1}, S_{2}, X_{1, i}^{n}, X_{2, i}^{n}, Z^{i}\right)+0 \\
& \leq \\
& \quad I\left(X_{2}^{i-1} ; M_{1}, X_{1}^{i-1} \mid S_{0}, S_{1}, S_{2}, X_{1, i}^{n}, X_{2, i}^{n}, Z^{i}\right) \\
& =I\left(X_{2}^{i-1} ; X_{1}^{i-1} \mid S_{0}, S_{1}, S_{2}, X_{1, i}^{n}, X_{2, i}^{n}, Z^{i}\right) \\
& \quad+I\left(X_{2}^{i-1} ; M_{1} \mid S_{0}, S_{1}, S_{2}, X_{1}^{n}, X_{2, i}^{n}, Z^{i}\right) \\
& \stackrel{(c)}{=} I\left(X_{2}^{i-1} ; X_{1}^{i-1} \mid Z^{i-1}\right) \\
& \quad+I\left(X_{2}^{i-1} ; M_{1} \mid S_{0}, S_{1}, S_{2}, X_{1}^{n}, X_{2, i}^{n}, Z^{i}\right)  \tag{80}\\
& \stackrel{(d)}{=} 0 \\
& \\
& \quad+0=0
\end{align*}
$$

where the first term in (a), the first and third terms in (b) are zeros and (c) follows since $S_{[0: 2]}$ is independent of $\left(X_{1}^{n}, X_{2}^{n}, Z^{n}\right)$ and $\left(X_{1 i}, X_{2 i}, Z_{i}\right), i=1, \ldots, n$, are jointly i.i.d. The second, fourth and sixth terms in (b) and the second term in (d) are zeros because of the Markov chains $M_{1} \rightarrow$ $\left(X_{1}^{n}, S_{0}, S_{1}\right) \rightarrow\left(S_{2}, X_{2}^{n}, Z^{n}\right)$ and $M_{2} \rightarrow\left(X_{2}^{n}, S_{0}, S_{2}\right) \rightarrow$ $\left(S_{1}, M_{1}, X_{1}^{n}, Z^{n}\right)$. The first term in (d) is zero because $\left(X_{1 i}, X_{2 i}, Z_{i}\right), i=1, \ldots, n$, are jointly i.i.d. with $q_{X_{1} X_{2}, Z}$ and $I\left(X_{1} ; X_{2} \mid Z\right)=0$.

## Appendix A <br> Proof of Theorem 1

The proof employs the Output Statistics of Random Binning (OSRB) framework developed by Yassaee et al. [53]. In the sequel, we use capital letters (like $P_{X}$ ) to denote random p.m.f.'s (see, e.g., [5], [53]) and lower-case letters (like $p_{X}$ ) to denote non-random p.m.f.'s. We use $p_{\mathcal{A}}^{\mathrm{U}}$ to denote the uniform distribution over the set $\mathcal{A}$. The notation $\approx$ for pmf approximations is adopted from [53] - for two random pmfs $P_{X}$ and $Q_{X}$ on the same alphabet $\mathcal{X}$, we say that $P_{X} \stackrel{\epsilon}{\approx} Q_{X}$ provided $\mathbb{E}\left[\left\|P_{X}-Q_{X}\right\|_{1}\right] \leq \epsilon$. For any two sequences of random p.m.f.'s $P_{X^{(n)}}$ and $Q_{X^{(n)}}$ on $\mathcal{X}^{(n)}$, we write $P_{X^{(n)}} \approx Q_{X^{(n)}}$ if $\lim _{n \rightarrow \infty} \mathbb{E}\left\|P_{X^{(n)}}-Q_{X^{(n)}}\right\|_{1}=0$. Likewise, we use $p_{X} \stackrel{\epsilon}{\approx} q_{X}$ for two (non-random) p.m.f.'s provided $\left\|p_{X}-q_{X}\right\|_{1} \leq \epsilon$. For any two sequences of random pmfs $P_{X^{n}}$ and $Q_{X^{n}}$ on $\mathcal{X}^{n}$, we write $P_{X^{n}} \approx Q_{X^{n}}$ if $\lim _{n \rightarrow \infty} \mathbb{E}\left[\left\|P_{X^{n}}-Q_{X^{n}}\right\|_{1}\right]=0$. Similarly, we write $p_{X^{n}} \approx$ $q_{X^{n}}$ for two sequences of (non-random) p.m.f.'s. We also quote the following results that will prove useful in the proof of Theorem 1, where the first one is a restatement of the OSRB result [54, Th. 1].

Theorem 7: [53, Th. 1] Given a discrete memoryless source $\left(A^{n}, B_{1}^{n}, \cdots, B_{L}^{n}\right) \sim$ i.i.d. with $p_{A, B_{1}, \cdots, B_{L}}$ on $\mathcal{A} \times \prod_{j=1}^{L} \mathcal{B}_{i}$. For $j \in[1: L]$, let $\phi_{j}: \mathcal{B}_{j}^{n} \rightarrow\left[1: 2^{n R_{j}}\right]$ be a random binning in which $\phi_{j}$ maps each sequence of $\mathcal{B}_{j}^{n}$ uniformly and independently to the set $\left[1: 2^{n R_{j}}\right]$. Let $K_{j}=\phi_{j}\left(B_{j}^{n}\right), j \in$ $\{1, \cdots, L\}$. If for each $\mathcal{S} \subseteq[1: L]$, the following constraint

$$
\begin{equation*}
\sum_{t \in \mathcal{S}} R_{t} \leq H\left(B_{\mathcal{S}} \mid A\right) \tag{81}
\end{equation*}
$$

holds, then we have

$$
\begin{equation*}
\lim _{n \rightarrow \infty} \mathbb{E}_{\phi_{1}, \cdots, \phi_{L}}\left[\left\|P_{A^{n}, K_{1}, \cdots, K_{L}}-p_{A^{n}} \prod_{j=1}^{L} p_{K_{j}}^{U}\right\|_{1}\right]=0 \tag{82}
\end{equation*}
$$

where $\mathbb{E}_{\phi_{1}, \cdots, \phi_{L}}$ denotes the expectation over the random binnings, $P_{A^{n}, K_{1}, \cdots, K_{L}}$ is a random p.m.f., and $p_{K_{j}}^{U}$ is the uniform distribution over $\left[1: 2^{n R_{j}}\right]$.

Lemma 8: [53, Lemma 4]

1) If $P_{X^{n}} \approx Q_{X^{n}}$, then $P_{X^{n}} P_{Y^{n} \mid X^{n}} \approx Q_{X^{n}} P_{Y^{n} \mid X^{n}}$. Also if $P_{X^{n}} P_{Y^{n} \mid X^{n}} \approx Q_{X^{n}} Q_{Y^{n} \mid X^{n}}$, then $P_{X^{n}} \approx Q_{X^{n}}$.
2) If $p_{X^{n}} p_{Y^{n} \mid X^{n}} \approx q_{X^{n}} q_{Y^{n} \mid X^{n}}$, then there exists a sequence $x^{n} \in \mathcal{X}^{n}$ such that $p_{Y^{n} \mid X^{n}=x^{n}} \approx q_{Y^{n} \mid X^{n}=x^{n}}$.
3) If $P_{X^{n}} \approx Q_{X^{n}}$ and $P_{X^{n}} P_{Y^{n} \mid X^{n}} \approx P_{X^{n}} Q_{Y^{n} \mid X^{n}}$, then $P_{X^{n}} P_{Y^{n} \mid X^{n}} \approx Q_{X^{n}} Q_{Y^{n} \mid X^{n}}$.
For simplicity, we prove the achievability for $|\mathcal{T}|=1$. The proof with the general time-sharing random variable $T$ then follows using standard time sharing argument outlined below.

Suppose the rate tuples $\boldsymbol{R}^{\prime}:=\left(R_{1}^{\prime}, R_{2}^{\prime}, R_{01}^{\prime}, R_{02}^{\prime}\right)$ and $\boldsymbol{R}^{\prime \prime}:=\left(R_{1}^{\prime \prime}, R_{2}^{\prime \prime}, R_{01}^{\prime \prime}, R_{02}^{\prime \prime}\right)$ are achievable for $q_{X_{1} X_{2} Z Y}$, i.e., there exist p.m.f.'s $p^{\prime}\left(x_{1}, x_{2}, z, u_{1}, u_{2}, y\right)$ and $p^{\prime \prime}\left(x_{1}, x_{2}\right.$, $\left.z, u_{1}, u_{2}, y\right)$ which together with the rate tuples $\boldsymbol{R}^{\prime}$ and $\boldsymbol{R}^{\prime \prime}$, respectively, satisfy the constraints in Theorem 1 (with $T=\emptyset$ ). For each blocklength $n$, we use the code corresponding to $\boldsymbol{R}^{\prime}$ for the first $\alpha n$ (with $\alpha \in[0,1]$, where $\alpha n$ is an integer) transmissions and use the code corresponding to $\boldsymbol{R}^{\prime \prime}$ for the rest of the transmissions. Let $p_{X_{1}^{n}, X_{2}^{n}, Z^{n}, Y^{n}}$ be the induced distribution with this new code. For ease of notation, let $D \triangleq$ $\left(X_{1}, X_{2}, Z, Y\right)$. Note that the induced distribution decomposes as $p_{D^{n}}=p_{D^{\alpha n}} \cdot p_{D_{\alpha n+1}^{n}}$, where $D_{\alpha n+1}^{n}=\left(D_{\alpha n+1}, \ldots, D_{n}\right)$, because of the independence of the two segments. Then, the overall total variation distance of interest with this new code is given by

$$
\begin{align*}
& \left\|p_{D^{n}}-\prod q_{D}\right\| \\
& \begin{aligned}
&=\left\|p_{D^{\alpha n}} \cdot p_{D_{\alpha n+1}^{n}}-q_{D}^{(n)}\right\| \\
&=\sum_{d^{n}} \mid \mid p_{D^{\alpha n}}\left(d^{\alpha n}\right) p_{D_{\alpha n+1}^{n}}\left(d_{\alpha n+1}^{n}\right) \\
& \quad-q_{D}^{(\alpha n)}\left(d^{\alpha n}\right) q_{D}^{(\alpha n+1: n)}\left(d_{\alpha n+1}^{n}\right) \mid \\
& \leq \sum_{d^{n}} \mid p_{D^{\alpha n}}\left(d^{\alpha n}\right) p_{D_{\alpha n+1}^{n}}\left(d_{\alpha n+1}^{n}\right) \\
& \quad-p_{D^{\alpha n}}\left(d^{\alpha n}\right) q_{D}^{(\alpha n+1: n)}\left(d_{\alpha n+1}^{n}\right) \mid \\
& \quad+\sum_{d^{n}} \mid p_{D^{\alpha n}}\left(d^{\alpha n}\right) q_{D}^{(\alpha n+1: n)}\left(d_{\alpha n+1}^{n}\right) \\
& \quad-q_{D}^{(\alpha n)}\left(d^{\alpha n}\right) q_{D}^{(\alpha n+1: n)}\left(d_{\alpha n+1}^{n}\right) \mid \\
&= \sum_{d_{\alpha n+1}^{n}}\left|p_{D_{\alpha n+1}^{n}}\left(d_{\alpha n+1}^{n}\right)-q_{D}^{(\alpha n+1: n)}\left(d_{\alpha n+1}^{n}\right)\right| \\
& \quad+\sum_{d^{n \alpha}}\left|p_{D^{\alpha n}}\left(d^{\alpha n}\right)-q_{D}^{(\alpha n)}\left(d^{\alpha n}\right)\right| \\
& \rightarrow 0 \text { as } n \rightarrow \infty,
\end{aligned}
\end{align*}
$$

where (83) follows by defining $q_{D}^{(\alpha n+1: n)}\left(d_{\alpha n+1}^{n}\right)=$ $\prod_{i=\alpha n+1}^{n} q_{D}\left(d_{i}\right)$, (84) follows by adding and subtracting the term $p_{D^{\alpha n}}\left(d^{\alpha n}\right) q_{D}^{(\alpha n+1: n)}\left(d_{\alpha n+1}^{n}\right)$ for each $d^{n}$ inside the absolute value in (83) and then applying the triangle inequality, and (85) follows because the codes in each of the segments
satisfy (1) with the respective blocklengths. So, the rate tuple $\alpha \boldsymbol{R}^{\prime}+(1-\alpha) \boldsymbol{R}^{\prime \prime}$ is achievable. This handles the case of $|\mathcal{T}|=2$ with $\mathcal{T}=\{1,2\}$ and $p_{T}(1)=1-p_{T}(2)=\alpha$. The generalization to larger alphabets for $T$ follows along similar lines.

We now follow the standard structure of an achievability proof via OSRB. This involves defining two protocols, one each based on random coding and random binning, that induce a joint distribution on the random variables defined during the protocols.

## A. Random Binning Scheme

Let $\left(U_{1}^{n}, U_{2}^{n}, X_{1}^{n}, X_{2}^{n}, Z^{n}, Y^{n}\right)$ be drawn i.i.d. with the joint distribution

$$
p\left(x_{1}, x_{2}, z\right) p\left(u_{1} \mid x_{1}\right) p\left(u_{2} \mid x_{2}\right) p\left(y \mid u_{1}, u_{2}, z\right)
$$

such that $p\left(x_{1}, x_{2}, z, y\right)=q\left(x_{1}, x_{2}, z, y\right)$. Now we employ the following random binning:

- Generate $\left(S_{1}, M_{1}, F_{1}\right)$ as three uniform binnings of $U_{1}^{n}$ independently, i.e. $S_{1}=\phi_{11}\left(U_{1}^{n}\right) \in\left[1: 2^{n R_{01}}\right], M_{1}=$ $\phi_{12}\left(U_{1}^{n}\right) \in\left[1: 2^{n R_{1}}\right]$ and $F_{1}=\phi_{13}\left(U_{1}^{n}\right) \in\left[1: 2^{n \tilde{R}_{1}}\right]$. Here, $S_{1}$ stands for the pairwise shared randomness between the first encoder and the decoder, $M_{1}$ stands for a message communicated over the noiseless link, while $F_{1}$ is additional shared randomness assumed in OSRB to be eliminated later without disturbing the i.i.d. distribution.
- Likewise, generate $\left(S_{2}, M_{2}, F_{2}\right)$ as three uniform binnings of $U_{2}^{n}$ independently, i.e. $S_{2}=\phi_{21}\left(U_{2}^{n}\right) \in[1$ : $\left.2^{n R_{02}}\right], M_{2}=\phi_{22}\left(U_{2}^{n}\right) \in\left[1: 2^{n R_{2}}\right]$ and $F_{2}=$ $\phi_{23}\left(U_{2}^{n}\right) \in\left[1: 2^{n \tilde{R}_{2}}\right]$.
The receiver uses a Slepian-Wolf decoder to estimate $\left(\hat{u}_{1}^{n}, \hat{u}_{2}^{n}\right)$ from $\left(s_{1}, s_{2}, f_{1}, f_{2}, m_{1}, m_{2}, z^{n}\right)$. The corresponding random p.m.f. induced is (the randomness is due to the binning)

$$
\begin{align*}
& P\left(x_{1}^{n}, x_{2}^{n}, z^{n}, y^{n}, u_{1}^{n}, u_{2}^{n}, s_{1}, f_{1}, m_{1}, s_{2}, f_{2}, m_{2}, \hat{u}_{1}^{n}, \hat{u}_{2}^{n}\right) \\
& =p\left(x_{1}^{n}, x_{2}^{n}, z^{n}\right) p\left(u_{1}^{n} \mid x_{1}^{n}\right) p\left(u_{2}^{n} \mid x_{2}^{n}\right) p\left(y^{n} \mid u_{1}^{n}, u_{2}^{n}, z^{n}\right) \\
& \quad \times P\left(s_{1}, f_{1}, m_{1} \mid u_{1}^{n}\right) P\left(s_{2}, f_{2}, m_{2} \mid u_{2}^{n}\right) \\
& \quad \times P^{S W}\left(\hat{u}_{1}^{n}, \hat{u}_{2}^{n} \mid s_{1}, f_{1}, m_{1}, s_{2}, f_{2}, m_{2}, z^{n}\right) \\
& =p\left(x_{1}^{n}, x_{2}^{n}, z^{n}\right) P\left(s_{1}, f_{1}, m_{1}, u_{1}^{n} \mid x_{1}^{n}\right) P\left(s_{2}, f_{2}, m_{2}, u_{2}^{n} \mid x_{2}^{n}\right) \\
& \quad \times P^{S W}\left(\hat{u}_{1}^{n}, \hat{u}_{2}^{n} \mid s_{1}, f_{1}, m_{1}, s_{2}, f_{2}, m_{2}, z^{n}\right) p\left(y^{n} \mid u_{1}^{n}, u_{2}^{n}, z^{n}\right) \\
& = \\
& \quad p\left(x_{1}^{n}, x_{2}^{n}, z^{n}\right) P\left(s_{1}, f_{1} \mid x_{1}^{n}\right) P\left(u_{1}^{n} \mid s_{1}, f_{1}, x_{1}^{n}\right) P\left(m_{1} \mid u_{1}^{n}\right) \\
& \quad \times P\left(s_{2}, f_{2} \mid x_{2}^{n}\right) P\left(u_{2}^{n} \mid s_{2}, f_{2}, x_{2}^{n}\right) P\left(m_{2} \mid u_{2}^{n}\right) \\
& \quad \times P^{S W}\left(\hat{u}_{1}^{n}, \hat{u}_{2}^{n} \mid s_{1}, f_{1}, m_{1}, s_{2}, f_{2}, m_{2}, z^{n}\right) p\left(y^{n} \mid u_{1}^{n}, u_{2}^{n}, z^{n}\right) \\
& =P\left(x_{1}^{n}, x_{2}^{n}, z^{n}, s_{1}, s_{2}, f_{1}, f_{2}\right) P\left(u_{1}^{n} \mid s_{1}, f_{1}, x_{1}^{n}\right) P\left(m_{1} \mid u_{1}^{n}\right) \\
& \quad \times P\left(u_{2}^{n} \mid s_{2}, f_{2}, x_{2}^{n}\right) P\left(m_{2} \mid u_{2}^{n}\right)  \tag{87}\\
& \quad \times P^{S W}\left(\hat{u}_{1}^{n}, \hat{u}_{2}^{n} \mid s_{1}, f_{1}, m_{1}, s_{2}, f_{2}, m_{2}, z^{n}\right) p\left(y^{n} \mid u_{1}^{n}, u_{2}^{n}, z^{n}\right) .
\end{align*}
$$

where (86) uses the Markov chains $\left(U_{1}, U_{2}\right) \rightarrow\left(X_{1}, X_{2}\right) \rightarrow$ $Z, U_{1} \rightarrow X_{1} \rightarrow X_{2} \rightarrow U_{2}, Y \rightarrow\left(U_{1}, U_{2}, Z\right) \rightarrow\left(X_{1}, X_{2}\right)$, and the binning construction.

## B. Random Coding Scheme

We assume that additional shared randomness $F_{j}$ of rate $\tilde{R}_{j}, j \in\{1,2\}$ are available between the respective encoders and the decoder in the main problem. Encoder $j \in\{1,2\}$, knowing $\left(s_{j}, f_{j}, x_{j}^{n}\right)$, generates $u_{j}^{n}$ according to the p.m.f. $P\left(u_{j}^{n} \mid s_{j}, f_{j}, x_{j}^{n}\right)$ (from the previous protocol) and sends the bin index of $u_{j}^{n}$ corresponding to the binning $\phi_{j 2}$ in the previous protocol over the noiseless link to the decoder. The decoder obtains $\left(s_{1}, s_{2}, f_{1}, f_{2}, m_{1}, m_{2}, z^{n}\right)$, and employs the Slepian-Wolf decoder from the previous protocol, i.e. $P^{S W}\left(\hat{u}_{1}^{n}, \hat{u}_{2}^{n} \mid s_{1}, f_{1}, m_{1}, s_{2}, f_{2}, m_{2}, z^{n}\right)$, to estimate $\left(u_{1}^{n}, u_{2}^{n}\right)$. Then it constructs $y^{n}$ according to the distribution $p_{Y^{n} \mid U_{1}^{n}, U_{2}^{n}, Z^{n}}\left(y^{n} \mid \hat{u}_{1}^{n}, \hat{u}_{2}^{n}, z^{n}\right)$. The induced random p.m.f. from this protocol is given by

$$
\begin{align*}
& \hat{P}\left(x_{1}^{n}, x_{2}^{n}, z^{n}, y^{n}, u_{1}^{n}, u_{2}^{n}, s_{1}, f_{1}, m_{1}, s_{2}, f_{2}, m_{2}, \hat{u}_{1}^{n}, \hat{u}_{2}^{n}\right) \\
& =p^{\mathrm{U}}\left(s_{1}\right) p^{\mathrm{U}}\left(f_{1}\right) p^{\mathrm{U}}\left(s_{2}\right) p^{\mathrm{U}}\left(f_{2}\right) p\left(x_{1}^{n}, x_{2}^{n}, z^{n}\right) \\
& \quad \times P\left(u_{1}^{n} \mid s_{1}, f_{1}, x_{1}^{n}\right) P\left(m_{1} \mid u_{1}^{n}\right) \\
& \quad \times P\left(u_{2}^{n} \mid s_{2}, f_{2}, x_{2}^{n}\right) P\left(m_{2} \mid u_{2}^{n}\right) \\
& \quad \times P^{S W}\left(\hat{u}_{1}^{n}, \hat{u}_{2}^{n} \mid s_{1}, f_{1}, m_{1}, s_{2}, f_{2}, m_{2}, z^{n}\right) p\left(y^{n} \mid \hat{u}_{1}^{n}, \hat{u}_{2}^{n}, z^{n}\right) . \tag{88}
\end{align*}
$$

Next we find constraints that imply that the induced p.m.f.'s from the two protocols are almost identical. Then one can restrict attention to the source coding side of the problem (related to the random binning protocol) and investigate the desired properties like vanishing total variation distance.

## C. Analysis of Rate Constraints

We now derive sufficient conditions for the joint statistics of the random variables from the two protocols to be identical. Since $\left(s_{j}, f_{j}\right)$ are the bin indices of $u_{j}^{n}$ for $j \in\{1,2\}$, if we ensure that

$$
\begin{equation*}
R_{01}+\tilde{R}_{1} \leq H\left(U_{1} \mid X_{1}, X_{2}, Z\right)=H\left(U_{1} \mid X_{1}\right) \tag{89}
\end{equation*}
$$

$$
\begin{align*}
R_{01}+\tilde{R}_{1}+R_{02}+\tilde{R}_{2} & \leq H\left(U_{1}, U_{2} \mid X_{1}, X_{2}, Z\right)  \tag{90}\\
& =H\left(U_{1}, U_{2} \mid X_{1}, X_{2}\right) \tag{91}
\end{align*}
$$

(where the equalities in (89)-(91) follow from the Markov chains $U_{1} \rightarrow X_{1} \rightarrow\left(X_{2}, Z\right), U_{2} \rightarrow X_{2} \rightarrow\left(X_{1}, Z\right)$, and $\left(U_{1}, U_{2}\right) \rightarrow\left(X_{1}, X_{2}\right) \rightarrow Z$, respectively), then by Theorem 7, we obtain

$$
\begin{align*}
& P\left(x_{1}^{n}, x_{2}^{n}, z^{n}, s_{1}, s_{2}, f_{1}, f_{2}\right) \\
& \quad \approx p^{\mathrm{U}}\left(s_{1}\right) p^{\mathrm{U}}\left(f_{1}\right) p^{\mathrm{U}}\left(s_{2}\right) p^{\mathrm{U}}\left(f_{2}\right) p\left(x_{1}^{n}, x_{2}^{n}, z^{n}\right) \\
& \quad=\hat{P}\left(x_{1}^{n}, x_{2}^{n}, z^{n}, s_{1}, s_{2}, f_{1}, f_{2}\right) . \tag{92}
\end{align*}
$$

This in turn results in

$$
\begin{align*}
& P\left(x_{1}^{n}, x_{2}^{n}, z^{n}, u_{1}^{n}, u_{2}^{n}, s_{1}, f_{1}, m_{1}, s_{2}, f_{2}, m_{2}, \hat{u}_{1}^{n}, \hat{u}_{2}^{n}\right) \\
& \quad \approx \hat{P}\left(x_{1}^{n}, x_{2}^{n}, z^{n}, u_{1}^{n}, u_{2}^{n}, s_{1}, f_{1}, m_{1}, s_{2}, f_{2}, m_{2}, \hat{u}_{1}^{n}, \hat{u}_{2}^{n}\right) \tag{93}
\end{align*}
$$

Note that the condition (91) above is redundant because $H\left(U_{1}, U_{2} \mid X_{1}, X_{2}\right)=H\left(U_{1} \mid X_{1}\right)+H\left(U_{2} \mid X_{2}\right)$ using the Markov chain $U_{1} \rightarrow X_{1} \rightarrow X_{2} \rightarrow U_{2}$.

For the Slepian-Wolf decoder to succeed, we require (by Slepian-Wolf theorem [54], see also [53, Lemma 1])

$$
\begin{align*}
R_{01}+\tilde{R}_{1}+R_{1} & \geq H\left(U_{1} \mid U_{2}, Z\right),  \tag{94}\\
R_{02}+\tilde{R}_{2}+R_{2} & \geq H\left(U_{2} \mid U_{1}, Z\right),  \tag{95}\\
R_{01}+\tilde{R}_{1}+R_{1}+R_{02}+\tilde{R}_{2}+R_{2} & \geq H\left(U_{1}, U_{2} \mid Z\right) . \tag{96}
\end{align*}
$$

This ensures that

$$
\begin{gather*}
P\left(x_{1}^{n}, x_{2}^{n}, z^{n}, u_{1}^{n}, u_{2}^{n}, s_{1}, f_{1}, m_{1}, s_{2}, f_{2}, m_{2}, \hat{u}_{1}^{n}, \hat{u}_{2}^{n}\right) \\
\approx P\left(x_{1}^{n}, x_{2}^{n}, z^{n}, u_{1}^{n}, u_{2}^{n}, s_{1}, f_{1}, m_{1}, s_{2}, f_{2}, m_{2}\right) \\
\times \mathbb{1}\left\{\hat{u}_{1}^{n}=u_{1}^{n}, \hat{u}_{2}^{n}=u_{2}^{n}\right\} . \tag{97}
\end{gather*}
$$

Using (97), (93) and the first and third parts of Lemma 8, we can write the following for the joint probability distribution involving $y^{n}$

$$
\begin{align*}
& \hat{P}\left(x_{1}^{n}, x_{2}^{n}, z^{n}, u_{1}^{n}, u_{2}^{n}, s_{1}, f_{1}, m_{1}, s_{2}, f_{2}, m_{2}, \hat{u}_{1}^{n}, \hat{u}_{2}^{n}, y^{n}\right) \\
& =\hat{P}\left(x_{1}^{n}, x_{2}^{n}, z^{n}, u_{1}^{n}, u_{2}^{n}, s_{1}, f_{1}, m_{1}, s_{2}, f_{2}, m_{2}, \hat{u}_{1}^{n}, \hat{u}_{2}^{n}\right) \\
& \quad \times p\left(y^{n} \mid \hat{u}_{1}^{n}, \hat{u}_{2}^{n}, z^{n}\right) \\
& \approx P\left(x_{1}^{n}, x_{2}^{n}, z^{n}, u_{1}^{n}, u_{2}^{n}, s_{1}, f_{1}, m_{1}, s_{2}, f_{2}, m_{2}\right) \\
& \quad \times \mathbb{1}\left\{\hat{u}_{1}^{n}=u_{1}^{n}, \hat{u}_{2}^{n}=u_{2}^{n}\right\} p\left(y^{n} \mid \hat{u}_{1}^{n}, \hat{u}_{2}^{n}, z^{n}\right) \\
& =P\left(x_{1}^{n}, x_{2}^{n}, z^{n}, u_{1}^{n}, u_{2}^{n}, s_{1}, f_{1}, m_{1}, s_{2}, f_{2}, m_{2}\right) \\
& \quad \times \mathbb{1}\left\{\hat{u}_{1}^{n}=u_{1}^{n}, \hat{u}_{2}^{n}=u_{2}^{n}\right\} p\left(y^{n} \mid u_{1}^{n}, u_{2}^{n}, z^{n}\right) \\
& =P\left(x_{1}^{n}, x_{2}^{n}, z^{n}, u_{1}^{n}, u_{2}^{n}, s_{1}, f_{1}, m_{1}, s_{2}, f_{2}, m_{2}, y^{n}\right) \\
& \quad \times \mathbb{1}\left\{\hat{u}_{1}^{n}=u_{1}^{n}, \hat{u}_{2}^{n}=u_{2}^{n}\right\} . \tag{98}
\end{align*}
$$

Thus, using the first part of Lemma 8, we can conclude that

$$
\begin{equation*}
\hat{P}\left(x_{1}^{n}, x_{2}^{n}, z^{n}, y^{n}, f_{1}, f_{2}\right) \approx P\left(x_{1}^{n}, x_{2}^{n}, z^{n}, y^{n}, f_{1}, f_{2}\right) \tag{99}
\end{equation*}
$$

We require $\left(X_{1}^{n}, X_{2}^{n}, Z^{n}, Y^{n}\right)$ to be independent of the extra shared randomness $\left(F_{1}, F_{2}\right)$ to eliminate them without disturbing the desired i.i.d. distribution. This can be accomplished by imposing the following conditions according to Theorem 7.

$$
\begin{align*}
\tilde{R}_{1} & \leq H\left(U_{1} \mid X_{1}, X_{2}, Y, Z\right),  \tag{100}\\
\tilde{R}_{2} & \leq H\left(U_{2} \mid X_{1}, X_{2}, Y, Z\right),  \tag{101}\\
\tilde{R}_{1}+\tilde{R}_{2} & \leq H\left(U_{1}, U_{2} \mid X_{1}, X_{2}, Y, Z\right) . \tag{102}
\end{align*}
$$

This ensures that

$$
\begin{equation*}
P\left(x_{1}^{n}, x_{2}^{n}, z^{n}, y^{n}, f_{1}, f_{2}\right) \approx p^{\mathrm{U}}\left(f_{1}\right) p^{\mathrm{U}}\left(f_{2}\right) p\left(x_{1}^{n}, x_{2}^{n}, z^{n}, y^{n}\right) \tag{103}
\end{equation*}
$$

which along with (99) and the triangle inequality, implies that

$$
\begin{equation*}
\hat{P}\left(x_{1}^{n}, x_{2}^{n}, z^{n}, y^{n}, f_{1}, f_{2}\right) \approx p^{\mathrm{U}}\left(f_{1}\right) p^{\mathrm{U}}\left(f_{2}\right) p\left(x_{1}^{n}, x_{2}^{n}, z^{n}, y^{n}\right) \tag{104}
\end{equation*}
$$

Hence there exists a fixed binning with corresponding pmf $\tilde{p}$ such that if we replace $P$ by $\tilde{p}$ in (88) and denote the resulting pmf by $\hat{p}$, then

$$
\begin{equation*}
\hat{p}\left(x_{1}^{n}, x_{2}^{n}, z^{n}, y^{n}, f_{1}, f_{2}\right) \approx p^{\mathrm{U}}\left(f_{1}\right) p^{\mathrm{U}}\left(f_{2}\right) p\left(x_{1}^{n}, x_{2}^{n}, y^{n}, z^{n}\right) \tag{105}
\end{equation*}
$$

Now the second part of Lemma 8 allows us to conclude that there exist instances $F_{1}=f_{1}^{*}, F_{2}=f_{2}^{*}$ such that

$$
\begin{equation*}
\hat{p}\left(x_{1}^{n}, x_{2}^{n}, z^{n}, y^{n} \mid f_{1}^{*}, f_{2}^{*}\right) \approx p\left(x_{1}^{n}, x_{2}^{n}, z^{n}, y^{n}\right) \tag{106}
\end{equation*}
$$

Now along with the rate constraints imposed in equations (89) - (90), (94) - (96) and (100) - (102), we also need to impose the non-negativity constraints on all the rates. But it turns out that the constraints $\tilde{R}_{1} \geq 0$ and $\tilde{R}_{2} \geq 0$ are redundant, which can be shown along the lines of [53, Remark 4]. We prove that if ( $\tilde{R}_{1}, \tilde{R}_{2}$ ) are not necessarily all positive and satisfy (89) - (90), (94) - (96) and (100) - (102) along with $\left(R_{1}, R_{2}, R_{01}, R_{02}\right)$ for some $\left(U_{1}, U_{2}\right)$ such that $\left(U_{1}, U_{2}\right) \rightarrow\left(X_{1}, X_{2}\right) \rightarrow Z, U_{1} \rightarrow X_{1} \rightarrow X_{2} \rightarrow U_{2}$ and $Y \rightarrow\left(U_{1}, U_{2}, Z\right) \rightarrow\left(X_{1}, X_{2}\right)$, then there exists $\left(\bar{U}_{1}, \bar{U}_{2}\right)$ with $\left(\bar{U}_{1}, \bar{U}_{2}\right) \rightarrow\left(X_{1}, X_{2}\right) \rightarrow Z, \bar{U}_{1} \rightarrow X_{1} \rightarrow X_{2} \rightarrow \bar{U}_{2}$ and $Y \rightarrow\left(\bar{U}_{1}, \bar{U}_{2}, Z\right) \rightarrow\left(X_{1}, X_{2}\right)$ and $\bar{R}_{1} \geq 0, \bar{R}_{2} \geq 0$ such that $\left(\bar{R}_{1}, \bar{R}_{2}\right)$ along with $\left(R_{1}, R_{2}, R_{01}, R_{02}\right)$ satisfy (89) - (90), (94) - (96) and (100) - (102) for ( $\left.\bar{U}_{1}, \bar{U}_{2}\right)$ instead of $\left(U_{1}, U_{2}\right)$.

Towards this end, suppose $\tilde{R}_{1}<0$ and $\tilde{R}_{2}<0$. Let ( $W_{1}, W_{2}$ ) be random variables such that $H\left(W_{1}\right)>\left|\tilde{R}_{1}\right|$ and $H\left(W_{2}\right)>\left|\tilde{R}_{2}\right|$. We also assume that $W_{1}$ as well as $W_{2}$ are independent of all the other random variables. Define $\bar{R}_{1}=\tilde{R}_{1}+H\left(W_{1}\right), \bar{R}_{2}=\tilde{R}_{2}+H\left(W_{2}\right)$ and $\bar{U}_{1}=\left(U_{1}, W_{1}\right)$, $\bar{U}_{2}=\left(U_{2}, W_{2}\right)$. Clearly, we have $\bar{R}_{1}, \bar{R}_{2} \geq 0$ and it is easy to see that $\left(\bar{R}_{1}, \bar{R}_{2}\right)$ along with $\left(R_{1}, R_{2}, R_{01}, R_{02}\right)$ satisfy (89) - (90), (94) - (96) and (100) - (102) for $\left(\bar{U}_{1}, \bar{U}_{2}\right)$ using the independence of $W_{1}$ and $W_{2}$ from all other random variables and the fact that $\left(\tilde{R}_{1}, \tilde{R}_{2}\right)$ satisfy (89) - (90), (94) - (96) and (100) - (102) along with $\left(R_{1}, R_{2}, R_{01}, R_{02}\right)$. Next suppose $\tilde{R}_{1}<0, \tilde{R}_{2} \geq 0$ - here the proof follows by defining $W_{1}, \bar{R}_{1}$ and $\bar{U}_{1}$ as above and noting that ( $\bar{R}_{1}, \tilde{R}_{2}$ ) along with ( $R_{1}, R_{2}, R_{01}, R_{02}$ ) satisfy (89) - (90), (94) - (96) and (100) - (102) for $\left(\bar{U}_{1}, U_{2}\right)$. The remaining configuration $\tilde{R}_{1} \geq 0, \tilde{R}_{2}<0$ can be dealt similarly. Finally on eliminating $\left(\tilde{R}_{1}, \tilde{R}_{2}\right)$ from equations (89) - (90), (94) - (96) and (100) - (102) by Fourier-Motzkin elimination (FME), we obtain the rate constraints

$$
\begin{align*}
& R_{1} \geq H\left(U_{1} \mid U_{2}, Z\right)-H\left(U_{1} \mid X_{1}, U_{2}, Z\right)=I\left(U_{1} ; X_{1} \mid U_{2}, Z\right),  \tag{107}\\
& R_{2} \geq H\left(U_{2} \mid U_{1}, Z\right)-H\left(U_{2} \mid X_{2}, U_{1}, Z\right)=I\left(U_{2} ; X_{2} \mid U_{1}, Z\right),  \tag{108}\\
& R_{1}+R_{2} \geq H\left(U_{1}, U_{2} \mid Z\right)-H\left(U_{1}, U_{2} \mid X_{1}, X_{2}, Z\right) \\
& =I\left(U_{1}, U_{2} ; X_{1}, X_{2} \mid Z\right),  \tag{109}\\
& \begin{aligned}
R_{1}+R_{01} & \geq H\left(U_{1} \mid U_{2}, Z\right)-H\left(U_{1} \mid X_{1}, X_{2}, Y, Z\right) \\
& =I\left(U_{1} ; X_{1}, X_{2}, Y \mid Z\right)-I\left(U_{1} ; U_{2} \mid Z\right),
\end{aligned}  \tag{110}\\
& \begin{aligned}
R_{1}+R_{01} & \geq H\left(U_{1} \mid U_{2}, Z\right)-H\left(U_{1} \mid X_{1}, X_{2}, Y, Z\right) \\
& =I\left(U_{1} ; X_{1}, X_{2}, Y \mid Z\right)-I\left(U_{1} ; U_{2} \mid Z\right),
\end{aligned} \\
& R_{2}+R_{02} \geq H\left(U_{2} \mid U_{1}, Z\right)-H\left(U_{2} \mid X_{1}, X_{2}, Y, Z\right) \\
& =I\left(U_{2} ; X_{1}, X_{2}, Y \mid Z\right)-I\left(U_{1} ; U_{2} \mid Z\right),  \tag{111}\\
& R_{1}+R_{2}+R_{01} \\
& \geq H\left(U_{1}, U_{2} \mid Z\right)-H\left(U_{2} \mid X_{2}, Z\right)-H\left(U_{1} \mid X_{1}, X_{2}, Y, Z\right) \\
& =I\left(U_{1} ; X_{1}, X_{2}, Y \mid Z\right)+I\left(U_{2} ; X_{2} \mid U_{1}, Z\right) \text {, }  \tag{112}\\
& R_{1}+R_{2}+R_{02} \\
& \geq H\left(U_{1}, U_{2} \mid Z\right)-H\left(U_{1} \mid X_{1}, Z\right)-H\left(U_{2} \mid X_{1}, X_{2}, Y, Z\right) \\
& =I\left(U_{2} ; X_{1}, X_{2}, Y \mid Z\right)+I\left(U_{1} ; X_{1} \mid U_{2}, Z\right) \text {, } \tag{113}
\end{align*}
$$

$$
\begin{align*}
& R_{1}+R_{2}+R_{01}+R_{02} \\
& \quad \geq H\left(U_{1}, U_{2} \mid Z\right)-H\left(U_{1}, U_{2} \mid X_{1}, X_{2}, Y, Z\right) \\
& \quad=I\left(U_{1}, U_{2} ; X_{1}, X_{2}, Y \mid Z\right) \tag{114}
\end{align*}
$$

Thus when the rate constraints in (107) - (114) are met, there exists a sequence of $\left(2^{n R_{01}}, 2^{n R_{02}}, 2^{n R_{1}}, 2^{n R_{2}}, n\right)$ codes with encoders and decoders as described in the second protocol with the particular realization of binning along with the fixed instances $f_{1}^{*}, f_{2}^{*}$ resulting in vanishing total variation distance.

## Appendix B

## Specialization of Theorem 1 to Deterministic Function Computation

Let $|\mathcal{T}|=1$ for simplicity. When $Y$ is a deterministic function of ( $X_{1}, X_{2}, Z$ ), we note the following simplifications to the rate constraints of Theorem 1 .

$$
\begin{align*}
R_{1}+R_{01} & \geq I\left(U_{1} ; X_{1}, X_{2}, Y \mid Z\right)-I\left(U_{1} ; U_{2} \mid Z\right) \\
& \stackrel{(a)}{=} I\left(U_{1} ; X_{1}, X_{2} \mid Z\right)-I\left(U_{1} ; U_{2} \mid Z\right) \\
& =I\left(U_{1} ; X_{1}, X_{2} \mid U_{2}, Z\right)-I\left(U_{1} ; U_{2} \mid X_{1}, X_{2}, Z\right) \\
& \stackrel{(b)}{=} I\left(U_{1} ; X_{1} \mid U_{2}, Z\right) \tag{115}
\end{align*}
$$

where (a) follows since $Y$ is determined by $\left(X_{1}, X_{2}, Z\right)$ while (b) follows from the Markov chains $U_{1} \rightarrow X_{1} \rightarrow\left(U_{2}, X_{2}, Z\right)$ and $U_{2} \rightarrow X_{2} \rightarrow\left(U_{1}, X_{1}, Z\right)$. Similarly, it follows that

$$
\begin{equation*}
R_{2}+R_{02} \geq I\left(U_{2} ; X_{2} \mid U_{1}, Z\right) \tag{116}
\end{equation*}
$$

Furthermore, we note that

$$
\begin{align*}
R_{1}+R_{2}+R_{01} & \geq I\left(U_{1} ; X_{1}, X_{2}, Y \mid Z\right)+I\left(U_{2} ; X_{2} \mid U_{1}, Z\right) \\
& \stackrel{(a)}{=} I\left(U_{1} ; X_{1}, X_{2} \mid Z\right)+I\left(U_{2} ; X_{2} \mid U_{1}, Z\right) \\
& \stackrel{(b)}{=} I\left(U_{1} ; X_{1}, X_{2} \mid Z\right)+I\left(U_{2} ; X_{1}, X_{2} \mid U_{1}, Z\right) \\
& =I\left(U_{1}, U_{2} ; X_{1}, X_{2} \mid Z\right), \tag{117}
\end{align*}
$$

where (a) follows since $Y$ is determined by $\left(X_{1}, X_{2}, Z\right)$ while (b) follows from the Markov chain $U_{2} \rightarrow X_{2} \rightarrow\left(U_{1}, X_{1}, Z\right)$. Similarly, it follows that

$$
\begin{equation*}
R_{1}+R_{2}+R_{02} \geq I\left(U_{1}, U_{2} ; X_{1}, X_{2} \mid Z\right) \tag{118}
\end{equation*}
$$

Finally, we note that

$$
\begin{align*}
R_{1}+R_{2}+R_{01}+R_{02} & \geq I\left(U_{1}, U_{2} ; X_{1}, X_{2}, Y \mid Z\right) \\
& =I\left(U_{1}, U_{2} ; X_{1}, X_{2} \mid Z\right) \tag{119}
\end{align*}
$$

Thus when $Y$ is a deterministic function of $\left(X_{1}, X_{2}, Z\right)$, the rate constraints involving the shared randomness rates $R_{01}$ and $R_{02}$ become redundant in Theorem 1. Hence, the region simplifies to the set of rate pairs $\left(R_{1}, R_{2}\right)$ satisfying

$$
\begin{aligned}
R_{1} & \geq I\left(U_{1} ; X_{1} \mid U_{2}, Z\right) \\
R_{2} & \geq I\left(U_{2} ; X_{2} \mid U_{1}, Z\right) \\
R_{1}+R_{2} & \geq I\left(U_{1}, U_{2} ; X_{1}, X_{2} \mid Z\right),
\end{aligned}
$$

for some p.m.f.

$$
p\left(x_{1}, x_{2}, z, u_{1}, u_{2}\right)=p\left(x_{1}, x_{2}, z\right) p\left(u_{1} \mid x_{1}\right) p\left(u_{2} \mid x_{2}\right)
$$

such that $H\left(Y \mid U_{1}, U_{2}, Z\right)=0$. This is precisely the inner bound of [31, Th. 2] specialized to the multiple-access network where the condition $H\left(Y \mid U_{1}, U_{2}, Z\right)=0$, in conjunction with the structure of the p.m.f., is expressed in an alternate form involving graph entropy using [51, Lemma 3].

## Appendix C

## Specialization of Theorem 3 to Deterministic Function Computation

Let $|\mathcal{T}|=1$ again for simplicity. The joint distribution in (70) is such that $\left(U_{1}, X_{1}\right) \rightarrow Z \rightarrow\left(U_{2}, X_{2}\right)$ is a Markov chain. This leads to the following simplification to the rate constraints involving shared randomness when $Y$ is a deterministic function of $\left(X_{1}, X_{2}, Z\right)$, say, $Y=f\left(X_{1}, X_{2}, Z\right)$.

$$
\begin{aligned}
I\left(U_{1} ; X_{1}, Y \mid X_{2}, Z\right) & =I\left(U_{1} ; X_{1}, X_{2}, Y \mid Z\right) \\
& =I\left(U_{1} ; X_{1}, X_{2} \mid Z\right) \\
& =I\left(U_{1} ; X_{1} \mid Z\right) .
\end{aligned}
$$

This simplification renders the bound on $R_{1}+R_{01}$ in Theorem 3 redundant. Hence, the region simplifies to the set of rate pairs $\left(R_{1}, R_{2}\right)$ satisfying

$$
\begin{aligned}
& R_{1} \geq I\left(U_{1} ; X_{1} \mid Z\right) \\
& R_{2} \geq I\left(U_{2} ; X_{2} \mid Z\right)
\end{aligned}
$$

for some p.m.f.

$$
\begin{aligned}
& p\left(x_{1}, x_{2}, z, u_{1}, u_{2}, y\right)= \\
& \quad p(z) p\left(x_{1} \mid z\right) p\left(x_{2} \mid z\right) p\left(u_{1} \mid x_{1}\right) p\left(u_{2} \mid x_{2}\right) p\left(y \mid u_{1}, u_{2}, z\right)
\end{aligned}
$$

such that $H\left(Y \mid U_{1}, U_{2}, Z\right)=0$. This is the rate region in [31, Th. 3] specialized to the multiple-access network where it is expressed in an alternate form involving graph entropies, which follows by using [51, Proof of Th. 3].

## Appendix D <br> Proof of Claim 1

It suffices to restrict attention to realizations $u_{1}, u_{2}$ for which $H\left(X_{1} \mid U_{1}=u_{1}\right)>0$ and $H\left(X_{2} \mid U_{2}=u_{2}\right)>0$. We prove the claim in three steps.
Step 1: We prove that under Case 3 , for any $u_{1}$ with $P\left(U_{1}=u_{1}\right)>0$, there exists a $k \in\{1,2\}$ such that $H\left(X_{1 k} \mid U_{1}=u_{1}\right)=0$.
Step 2: Then for any $u_{2}$ with $P\left(U_{2}=u_{2}\right)>0$, for the $k$ from Step 1, we show that $H\left(X_{2 k} \mid U_{2}=u_{2}\right)=0$.
Step 3: Finally, for any $u_{1}^{\prime}$ with $u_{1}^{\prime} \neq u_{1}$ and $P\left(U_{1}=u_{1}^{\prime}\right)>0$, for the $k$ from Step 1, we prove that $H\left(X_{1 k} \mid U_{1}=u_{1}^{\prime}\right)=0$.

Step 1: We prove this by contradiction. Suppose $H\left(X_{1 i} \mid U_{1}=u_{1}\right)>0$, for $i=1,2$. This implies that the support of $p_{X_{1} \mid U_{1}=u_{1}}$ cannot be of the form $\left\{b_{1}, b_{2}\right\},\left\{\left(b_{1}, b_{2}\right),\left(b_{1}, 1-b_{2}\right)\right\},\left\{\left(b_{1}, b_{2}\right),\left(1-b_{1}, b_{2}\right)\right\}$, for $b_{1}, b_{2} \in\{0,1\}$. The remaining possibility is that the support has to be a superset of either $\{(0,1),(1,0)\}$ or $\{(0,0),(1,1)\}$. In the sequel, we use the independence of $\left(U_{1}, X_{1}\right)$ and $\left(U_{2}, X_{2}\right)$, and the Markov chain $Y \rightarrow\left(U_{1}, U_{2}\right) \rightarrow\left(X_{1}, X_{2}\right)$ repeatedly. Consider a $u_{2}$ such that $P\left(U_{2}=u_{2}\right)>0$. It turns out that the support cannot be a superset of $\{(0,0),(1,1)\}$. To see this, first notice that $P\left(X_{1}=(0,0) \mid U_{1}=u_{1}, U_{2}=\right.$ $\left.u_{2}\right), P\left(X_{1}=(1,1) \mid U_{1}=u_{1}, U_{2}=u_{2}\right)>0$. Now since
$X_{1 J}-\left(U_{1}, U_{2}\right)-X_{1}$, we have $P\left(X_{1 J}=0 \mid U_{1}=u_{1}, U_{2}=\right.$ $\left.u_{2}\right)=P\left(X_{1 J}=0 \mid U_{1}=u_{1}, U_{2}=u_{2}, X_{1}=(1,1)\right)=0$, where the last equality follows from the correctness of output $Y=\left(X_{1 J}, X_{2 J}\right)$. Similarly, $P\left(X_{1 J}=1 \mid U_{1}=u_{1}, U_{2}=\right.$ $\left.u_{2}\right)=P\left(X_{1 J}=1 \mid U_{1}=u_{1}, U_{2}=u_{2}, X_{1}=(0,0)\right)=0$. This is a contradiction since $p_{X_{1 J} \mid U_{1}=u_{1}, U_{2}=u_{2}}$ has to be a probability distribution. The only other possibility is that the support of $p_{X_{1} \mid U_{1}=u_{1}}$ is a superset of $\{(0,1),(1,0)\}$. Since $H\left(X_{2} \mid U_{2}\right)>0$, there exists a $u_{2}$ with $P\left(U_{2}=u_{2}\right)>$ 0 such that $H\left(X_{2} \mid U_{2}=u_{2}\right)>0$. So, either exactly one or both of $H\left(X_{21} \mid U_{2}=u_{2}\right)$ and $H\left(X_{22} \mid U_{2}=u_{2}\right)$ will be strictly positive. Suppose exactly one of them is strictly positive. Without loss of generality, suppose $H\left(X_{21} \mid U_{2}=\right.$ $\left.u_{2}\right)>0$, i.e., $P\left(X_{21}=0 \mid U_{2}=u_{2}\right), P\left(X_{21}=1 \mid U_{2}=\right.$ $\left.u_{2}\right)>0$ and $H\left(X_{22} \mid U_{2}=u_{2}\right)=0$. Also, assume that $P\left(X_{22}=1 \mid U_{2}=u_{2}\right)=1$. Consider the probability distribution $p_{Y \mid U_{1}=u_{1}, U_{2}=u_{2}}$. This is well defined because $P\left(U_{1}=u_{1}, U_{2}=u_{2}\right)>0$ as $U_{1}$ is independent of $U_{2}$ and $P\left(U_{1}=u_{1}\right), P\left(U_{2}=u_{2}\right)>0$. From the above we have the following.
$P\left(X_{1}=(0,1), X_{2}=(0,1) \mid U_{1}=u_{1}, U_{2}=u_{2}\right)>0$ which implies that $P\left(Y=(0,1) \mid U_{1}=u_{1}, U_{2}=u_{2}\right)=0$ and $P\left(Y=(1,0) \mid U_{1}=u_{1}, U_{2}=u_{2}\right)=0$.
$P\left(X_{1}=(1,0), X_{2}=(0,1) \mid U_{1}=u_{1}, U_{2}=u_{2}\right)>0$ which implies that $P\left(Y=(0,0) \mid U_{1}=u_{1}, U_{2}=u_{2}\right)=0$ and $P\left(Y=(1,1) \mid U_{1}=u_{1}, U_{2}=u_{2}\right)=0$.

This is a contradiction since $p_{Y \mid U_{1}=u_{1}, U_{2}=u_{2}}$ has to be a probability distribution. Now suppose both $H\left(X_{21} \mid U_{2}=u_{2}\right)$ and $H\left(X_{22} \mid U_{2}=u_{2}\right)$ are strictly positive. Since $Y_{2}:=$ $X_{2 J} \in\left\{X_{21}, X_{22}\right\}$, the only possibility is that $p_{X_{2} \mid U_{2}=u_{2}}$ has a support that is a superset of $\{(0,1),(1,0)\}$. Then we have the following.
$P\left(X_{1}=(0,1), X_{2}=(0,1) \mid U_{1}=u_{1}, U_{2}=u_{2}\right)>0$ which implies that $P\left(Y=(0,1) \mid U_{1}=u_{1}, U_{2}=u_{2}\right)=0$ and $P\left(Y=(1,0) \mid U_{1}=u_{1}, U_{2}=u_{2}\right)=0$.
$P\left(X_{1}=(0,1), X_{2}=(1,0) \mid U_{1}=u_{1}, U_{2}=u_{2}\right)>0$ which implies that $P\left(Y=(0,0) \mid U_{1}=u_{1}, U_{2}=u_{2}\right)=0$ and $P\left(Y=(1,1) \mid U_{1}=u_{1}, U_{2}=u_{2}\right)=0$.

This is a contradiction since $p_{Y \mid U_{1}=u_{1}, U_{2}=u_{2}}$ has to be a probability distribution. Thus, we have, under Case 3, if $H\left(X_{1} \mid U_{1}=u_{1}\right)>0$, then there exists a $k \in\{1,2\}$ such that $H\left(X_{1 k} \mid U_{1}=u_{1}\right)=0$.

Step 2: Note that there exists $u_{1}$ with $P\left(U_{1}=u_{1}\right)>0$ such that $H\left(X_{1} \mid U_{1}=u_{1}\right)>0$ since $H\left(X_{1} \mid U_{1}\right)>0$. So, by the discussion in Step 1, there exists a $k$ such that $H\left(X_{1 k} \mid U_{1}=\right.$ $\left.u_{1}\right)=0$. Note that $H\left(X_{1 k^{\prime}} \mid U_{1}=u_{1}\right)>0$, where $k^{\prime}=3-k$, since $H\left(X_{1} \mid U_{1}=u_{1}\right)>0$. Similarly, there exists $u_{2}$ with $P\left(U_{2}=u_{2}\right)>0$ such that $H\left(X_{2} \mid U_{2}=u_{2}\right)>0$. Now we show that $H\left(X_{2 k} \mid U_{2}=u_{2}\right)=0$. We prove this by contradiction. Suppose $H\left(X_{2 k} \mid U_{2}=u_{2}\right)>0$. Then in view of the above discussion, we have $H\left(X_{2 k^{\prime}} \mid U_{2}=u_{2}\right)=0$. Without loss of generality, let $k=1$, i.e., $k^{\prime}=2$. Note that $p_{X_{12}, X_{21} \mid U_{1}=u_{1}, U_{2}=u_{2}}=p_{X_{12} \mid U_{1}=u_{1}} \cdot p_{X_{21} \mid U_{2}=u_{2}}$ has full support since $H\left(X_{12} \mid U_{1}=u_{1}\right), H\left(X_{21} \mid U_{2}=u_{2}\right)>0$. Also, assume that $P\left(X_{11}=0 \mid U_{1}=u_{1}\right)=1$ and $P\left(X_{22}=0 \mid U_{2}=\right.$ $\left.u_{2}\right)=1$ (other choices can be dealt similarly). Consider the probability distribution $p_{Y \mid U_{1}=u_{1}, U_{2}=u_{2}}$. Then we have the following.
$P\left(X_{1}=(0,1), X_{2}=(1,0) \mid U_{1}=u_{1}, U_{2}=u_{2}\right)=$ $P\left(X_{12}=1, X_{21}=1 \mid U_{1}=u_{1}, U_{2}=u_{2}\right)>0$, which implies that $P\left(Y=(0,0) \mid U_{1}=u_{1}, U_{2}=u_{2}\right)=0$ and $P\left(Y=(1,1) \mid U_{1}=u_{1}, U_{2}=u_{2}\right)=0$ since we have the Markov chain $Y-\left(U_{1}, U_{2}\right)-\left(X_{1}, X_{2}\right)$.
$P\left(X_{1}=(0,0), X_{2}=(0,0) \mid U_{1}=u_{1}, U_{2}=u_{2}\right)=$ $P\left(X_{12}=0, X_{21}=0 \mid U_{1}=u_{1}, U_{2}=u_{2}\right)>0$, which implies that $P\left(Y=(0,1) \mid U_{1}=u_{1}, U_{2}=u_{2}\right)=0$ and $P\left(Y=(1,0) \mid U_{1}=u_{1}, U_{2}=u_{2}\right)=0$ since we have the Markov chain $Y-\left(U_{1}, U_{2}\right)-\left(X_{1}, X_{2}\right)$.

This is a contradiction since $p_{Y \mid U_{1}=u_{1}, U_{2}=u_{2}}$ has to be a probability distribution.

Step 3: Now suppose that there exists a $u_{1}^{\prime}$ with $u_{1}^{\prime} \neq u_{1}$ and $P\left(U_{1}=u_{1}^{\prime}\right)>0$ such that $H\left(X_{1} \mid U_{1}=u_{1}^{\prime}\right)>0$. Since there is a $u_{2}$ such that $H\left(X_{2 k} \mid U_{2}=u_{2}\right)=0$, by the same argument as above (reversing the roles of 1 and 2 ), we have $H\left(X_{1 k} \mid U_{1}=u_{1}^{\prime}\right)=0$. This completes the proof of Claim 1.

## Appendix E

## Proof of Lemma 5

The cardinality bound $|\mathcal{T}| \leq 3$ on the time-sharing random variable $T$ can be obtained using standard arguments based on the support lemma [55, Appendix C]. ${ }^{3}$ We now focus on cardinality bounds for $U_{1}$ and $U_{2}$. Consider the region $S_{\epsilon}$ with $|\mathcal{T}|=1:$

$$
\begin{align*}
R_{1} & \geq I\left(U_{1} ; X_{1} \mid Z\right)  \tag{120}\\
R_{2} & \geq I\left(U_{2} ; X_{2} \mid Z\right)  \tag{121}\\
R_{1}+R_{01} & \geq I\left(U_{1} ; X_{1}, Y \mid X_{2}, Z\right)-g(\epsilon) \tag{122}
\end{align*}
$$

for some p.m.f.

$$
\begin{align*}
& p\left(x_{1}, x_{2}, z, u_{1}, u_{2}, y\right) \\
& \quad=p(z) p\left(x_{1} \mid z\right) p\left(x_{2} \mid z\right) p\left(u_{1} \mid x_{1}\right) p\left(u_{2} \mid x_{2}\right) p\left(y \mid u_{1}, u_{2}, z\right) \tag{123}
\end{align*}
$$

such that

$$
\begin{equation*}
\left\|\sum_{u_{1}, u_{2}} p\left(x_{1}, x_{2}, z, u_{1}, u_{2}, y\right)-q\left(x_{1}, x_{2}, z, y\right)\right\| \leq \epsilon \tag{124}
\end{equation*}
$$

Call this region $\mathcal{C}$. We now show that the auxiliary cardinalities can be restricted to $\left|\mathcal{U}_{1}\right| \leq\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\right||\mathcal{Y} \| \mathcal{Z}|$ and $\left|\mathcal{U}_{2}\right| \leq$ $\left|\mathcal{U}_{1}\left\|\mathcal{X}_{1}| | \mathcal{X}_{2}| | \mathcal{Y}\right\| \mathcal{Z}\right|$ via the perturbation argument of [56]. This is done in two steps.

- Step 1: We prove that

$$
\begin{equation*}
\mathcal{C}=\text { Closure }\left(\bigcup_{K_{1}, K_{2} \geq 0} \mathcal{C}^{K_{1}, K_{2}}\right) \tag{125}
\end{equation*}
$$

where the region $\mathcal{C}^{K_{1}, K_{2}}$ for positive integers $K_{1}, K_{2}$ is given by the union of rate triples ( $R_{1}, R_{2}, R_{01}$ ) satisfying (120)-(122) over random variables $\left(U_{1}, U_{2}, X_{1}, X_{2}, Y, Z\right)$ with cardinality bounds $\left|\mathcal{U}_{1}\right| \leq K_{1} \&\left|\mathcal{U}_{2}\right| \leq K_{2}$ and having a joint p.m.f.

[^2]$p(z) p\left(x_{1} \mid z\right) p\left(x_{2} \mid z\right) p\left(u_{1} \mid x_{1}\right) p\left(u_{2} \mid x_{2}\right) p\left(y \mid u_{1}, u_{2}, z\right)$ such that $\left\|p\left(x_{1}, x_{2}, z, y\right)-q\left(x_{1}, x_{2}, z, y\right)\right\| \leq \epsilon$.

- Step 2: If $\left|\mathcal{U}_{1}\right| \leq K_{1}$ and $\left|\mathcal{U}_{2}\right| \leq K_{2}$ for some constants $K_{1}, K_{2}$, we show that the auxiliary cardinalities can be brought down to $\left|\mathcal{U}_{1}\right| \leq\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\right||\mathcal{Y}||\mathcal{Z}|$ and $\left|\mathcal{U}_{2}\right| \leq$ $\left|\mathcal{U}_{1}\left\|\mathcal{X}_{1}\right\| \mathcal{X}_{2}\|\mathcal{Y}\| \mathcal{Z}\right|$.
Proof Step 1: It suffices to show that any rate triple $\left(R_{1}, R_{2}, R_{01}\right) \in \mathcal{C}$ is a limit point of the set $\underset{K_{1}, K_{2} \geq 0}{\bigcup} \mathcal{C}^{K_{1}, K_{2}}$. Firstly, if $\left(R_{1}, R_{2}, R_{01}\right) \in \mathcal{C}$, then random variables ( $\left.U_{1}, U_{2}, X_{1}, X_{2}, Y, Z\right)$ satisfying (120)-(122) exist such that their joint p.m.f. is of the form (123) and (124) holds. Assume that $\mathcal{U}_{j}=\{1,2, \cdots\}$ for $j=1,2$. Define modified versions $\left(U_{1}^{\prime}, U_{2}^{\prime}\right)$ of random variables $\left(U_{1}, U_{2}\right)$ taking values in $\{1,2, \cdots, m\} \cup \mathcal{X}_{1}$ and $\{1,2, \cdots, m\} \cup \mathcal{X}_{2}$ respectively, where $m$ is an integer. The alphabet cardinalities of $\left(U_{1}^{\prime}, U_{2}^{\prime}\right)$ are $m+\left|\mathcal{X}_{1}\right|$ and $m+\left|\mathcal{X}_{2}\right|$ respectively. Let the conditional p.m.f. of $U_{j}^{\prime}$ given $X_{j}$ for $j \in\{1,2\}, x_{j} \in \mathcal{X}_{j}$ be specified as follows:

$$
\begin{align*}
p_{U_{j}^{\prime} \mid X_{j}}\left(i \mid x_{j}\right) & =p_{U_{j} \mid X_{j}}\left(i \mid x_{j}\right), i=\{1,2, \cdots, m\}  \tag{126}\\
p_{U_{j}^{\prime} \mid X_{j}}\left(x_{j}^{\prime} \mid x_{j}\right) & =0 \forall x_{j}^{\prime} \neq x_{j}  \tag{127}\\
p_{U_{j}^{\prime} \mid X_{j}}\left(x_{j} \mid x_{j}\right) & =\operatorname{Pr}\left(U_{j}>m \mid X_{j}=x_{j}\right)=\sum_{i=m+1}^{\infty} p_{U_{j} \mid X_{j}}\left(i \mid x_{j}\right) . \tag{128}
\end{align*}
$$

From the definitions in (126)-(128) and the fact that the original random variables $\left(U_{1}, U_{2}, X_{1}, X_{2}, Y, Z\right)$ satisfy $U_{1} \rightarrow$ $X_{1} \rightarrow\left(U_{2}, X_{2}, Z\right)$ and $U_{2} \rightarrow X_{2} \rightarrow\left(U_{1}, X_{1}, Z\right)$, it follows that the Markov chains $U_{1}^{\prime} \rightarrow X_{1} \rightarrow\left(U_{2}^{\prime}, X_{2}, Z\right)$ and $U_{2}^{\prime} \rightarrow$ $X_{2} \rightarrow\left(U_{1}^{\prime}, X_{1}, Z\right)$ hold as well.

To define $p_{Y \mid U_{1}^{\prime}, U_{2}^{\prime}, Z}$, consider new auxiliary random variables $\left(U_{1}^{\prime \prime}, U_{2}^{\prime \prime}\right)$ whose alphabets are the same as that of ( $U_{1}, U_{2}$ ), and their conditional p.m.f. given $U_{j}^{\prime}$ for $j \in\{1,2\}$ is specified as follows:

- If $U_{j}^{\prime} \in\{1,2, \cdots, m\}$, then $U_{j}^{\prime \prime}=U_{j}^{\prime}$, i.e.,

$$
\begin{equation*}
p_{U_{j}^{\prime \prime} \mid U_{j}^{\prime}}^{\prime}(i \mid i)=1, i=1,2, \cdots, m . \tag{129}
\end{equation*}
$$

- Otherwise, let

$$
\begin{align*}
& p_{U_{j}^{\prime \prime} \mid U_{j}^{\prime}}\left(i \mid x_{j}\right)= \\
& \begin{cases}\operatorname{Pr}\left(U_{j}=i \mid X_{j}=x_{j}, U_{j}>m\right), & i=m+1, m+2, \cdots \\
0, & i=1,2, \cdots, m\end{cases} \tag{130}
\end{align*}
$$

Note that we have

$$
\begin{align*}
& \sum_{u_{j}^{\prime}} p_{U_{j}^{\prime} \mid X_{j}}\left(u_{j}^{\prime} \mid x_{j}\right) p_{U_{j}^{\prime \prime} \mid U_{j}^{\prime}}\left(u_{j}^{\prime \prime} \mid u_{j}^{\prime}\right) \\
&=p_{U_{j} \mid X_{j}}\left(u_{j}^{\prime \prime} \mid x_{j}\right), u_{j}^{\prime \prime} \in \mathcal{U}_{j}, x_{j} \in \mathcal{X}_{j}, j=1,2 \tag{131}
\end{align*}
$$

Further, let

$$
\begin{equation*}
p_{Y \mid U_{1}^{\prime \prime}, U_{2}^{\prime \prime}, Z}\left(y \mid u_{1}, u_{2}, z\right)=p_{Y \mid U_{1}, U_{2}, Z}\left(y \mid u_{1}, u_{2}, z\right) \tag{132}
\end{equation*}
$$

The conditional distribution $p_{Y \mid U_{1}^{\prime}, U_{2}^{\prime}, Z}$ is defined as

$$
\begin{align*}
& p_{Y \mid U_{1}^{\prime}, U_{2}^{\prime}, Z}\left(y \mid u_{1}^{\prime}, u_{2}^{\prime}, z\right) \\
& =\sum_{u_{1}^{\prime \prime}, u_{2}^{\prime \prime}} p_{U_{1}^{\prime \prime} \mid U_{1}^{\prime}}\left(u_{1}^{\prime \prime} \mid u_{1}^{\prime}\right) p_{U_{2}^{\prime \prime} \mid U_{2}^{\prime}}\left(u_{2}^{\prime \prime} \mid u_{2}^{\prime}\right) p_{Y \mid U_{1}^{\prime \prime}, U_{2}^{\prime \prime}, Z}\left(y \mid u_{1}^{\prime \prime}, u_{2}^{\prime \prime}, z\right) \\
& =\sum_{u_{1}^{\prime \prime}, u_{2}^{\prime \prime}} p_{U_{1}^{\prime \prime} \mid U_{1}^{\prime}}\left(u_{1}^{\prime \prime} \mid u_{1}^{\prime}\right) p_{U_{2}^{\prime \prime} \mid U_{2}^{\prime}}\left(u_{2}^{\prime \prime} \mid u_{2}^{\prime}\right) p_{Y \mid U_{1}, U_{2}, Z}\left(y \mid u_{1}^{\prime \prime}, u_{2}^{\prime \prime}, z\right) . \tag{133}
\end{align*}
$$

From the definitions in (130)-(133) and the fact that the original random variables $\left(U_{1}, U_{2}, X_{1}, X_{2}, Y, Z\right)$ satisfy $Y \rightarrow\left(U_{1}, U_{2}, Z\right) \rightarrow\left(X_{1}, X_{2}\right)$, it follows that the Markov chain $Y \rightarrow\left(U_{1}^{\prime}, U_{2}^{\prime}, Z\right) \rightarrow\left(X_{1}, X_{2}\right)$ holds as well. Due to these Markov constraints, the joint distribution of $\left(U_{1}^{\prime}, U_{2}^{\prime}, X_{1}, X_{2}, Y, Z\right)$ is $p_{U_{1}^{\prime}, U_{2}^{\prime}, X_{1}, X_{2}, Y, Z}=$ $p_{X_{1}, X_{2}, Z} p_{U_{1}^{\prime} \mid X_{1}} p_{U_{2}^{\prime} \mid X_{2}} p_{Y \mid U_{1}^{\prime}, U_{2}^{\prime}, Z}$. On marginalizing away $\left(U_{1}^{\prime}, U_{2}^{\prime}\right)$ from $p_{U_{1}^{\prime}, U_{2}^{\prime}, X_{1}, X_{2}, Y, Z}$, we obtain the p.m.f. $p_{X_{1}, X_{2}, Y, Z}$, which can be seen as follows.

$$
\begin{align*}
& \sum_{u_{1}^{\prime}, u_{2}^{\prime}}\left(p_{X_{1}, X_{2}, Z}\left(x_{1}, x_{2}, z\right) p_{U_{1}^{\prime} \mid X_{1}}\left(u_{1}^{\prime} \mid x_{1}\right) p_{U_{2}^{\prime} \mid X_{2}}\left(u_{2}^{\prime} \mid x_{2}\right)\right. \\
& \left.\quad \times p_{Y \mid U_{1}^{\prime}, U_{2}^{\prime}, Z}\left(y \mid u_{1}^{\prime}, u_{2}^{\prime}, z\right)\right) \\
& =\sum_{u_{1}^{\prime}, u_{2}^{\prime}} p_{X_{1}, X_{2}, Z}\left(x_{1}, x_{2}, z\right) p_{U_{1}^{\prime} \mid X_{1}}\left(u_{1}^{\prime} \mid x_{1}\right) p_{U_{2}^{\prime} \mid X_{2}}\left(u_{2}^{\prime} \mid x_{2}\right) \\
& \left(\sum_{u_{1}^{\prime \prime}, u_{2}^{\prime \prime}} p_{U_{1}^{\prime \prime} \mid U_{1}^{\prime}}\left(u_{1}^{\prime \prime} \mid u_{1}^{\prime}\right) p_{U_{2}^{\prime \prime} \mid U_{2}^{\prime}}\left(u_{2}^{\prime \prime} \mid u_{2}^{\prime}\right) p_{Y \mid U_{1}^{\prime \prime}, U_{2}^{\prime \prime}, Z}\left(y \mid u_{1}^{\prime \prime}, u_{2}^{\prime \prime}, z\right)\right) \\
& =\sum_{u_{1}^{\prime \prime}, u_{2}^{\prime \prime}} p_{X_{1}, X_{2}, Z}\left(x_{1}, x_{2}, z\right)\left(\sum_{u_{1}^{\prime}} p_{U_{1}^{\prime} \mid X_{1}}\left(u_{1}^{\prime} \mid x_{1}\right) p_{U_{1}^{\prime \prime} \mid U_{1}^{\prime}}\left(u_{1}^{\prime \prime} \mid u_{1}^{\prime}\right)\right) \\
& \left(\sum_{u_{2}^{\prime}} p_{U_{2}^{\prime} \mid X_{2}}\left(u_{2}^{\prime} \mid x_{2}\right) p_{U_{2}^{\prime \prime} \mid U_{2}^{\prime}}\left(u_{2}^{\prime \prime} \mid u_{2}^{\prime}\right)\right) p_{Y \mid U_{1}, U_{2}, Z}\left(y \mid u_{1}^{\prime \prime}, u_{2}^{\prime \prime}, z\right) \\
& \stackrel{(a)}{=} \sum_{u_{1}^{\prime \prime}, u_{2}^{\prime \prime}} p_{X_{1}, X_{2}, Z}\left(x_{1}, x_{2}, z\right) p_{U_{1} \mid X_{1}}\left(u_{1}^{\prime \prime} \mid x_{1}\right) p_{U_{2} \mid X_{2}}\left(u_{2}^{\prime \prime} \mid x_{2}\right) \\
& \quad p_{Y \mid U_{1}, U_{2}, Z}\left(y \mid u_{1}^{\prime \prime}, u_{2}^{\prime \prime}, z\right)  \tag{135}\\
& =p_{X_{1}, X_{2}, Y, Z}\left(x_{1}, x_{2}, y, z\right),
\end{align*}
$$

where (a) follows from (131). Hence $p_{U_{1}^{\prime} U_{2}^{\prime} X_{1} X_{2} Y Z}$ satisfies (124).

It follows that the joint distribution of the random variables ( $U_{1}^{\prime}, U_{2}^{\prime}, X_{1}, X_{2}, Y, Z$ ) converges to the the joint distribution of $\left(U_{1}, U_{2}, X_{1}, X_{2}, Y, Z\right)$ in the limit $m \rightarrow \infty$. As a result, the mutual information terms $I\left(U_{1}^{\prime} ; X_{1} \mid Z\right), I\left(U_{2}^{\prime} ; X_{2} \mid Z\right)$ and $I\left(U_{1}^{\prime} ; X_{1}, Y \mid X_{2}, Z\right)$ converge to $I\left(U_{1} ; X_{1} \mid Z\right), I\left(U_{2} ; X_{2} \mid Z\right)$ and $I\left(U_{1} ; X_{1}, Y \mid X_{2}, Z\right)$ respectively. Hence, we conclude that the given rate triple $\left(R_{1}, R_{2}, R_{01}\right)$ is a limit point of the set $\bigcup_{K_{1}, K_{2} \geq 0} \mathcal{C}^{K_{1}, K_{2}}$.

Proof Step 2: It suffices to consider an optimization of the weighted sum term $\lambda_{1} I\left(U_{1} ; X_{1} \mid Z\right)+\lambda_{2} I\left(U_{2} ; X_{2} \mid Z\right)+$ $\lambda_{3} I\left(U_{1} ; X_{1}, Y \mid X_{2}, Z\right)$ for non-negative reals $\lambda_{1}, \lambda_{2}, \lambda_{3}$, and find new auxiliary random variables whose cardinalities are bounded while not increasing the weighted sum and preserving the conditions (123) and on (124) on $p\left(x_{1}, x_{2}, u_{1}, u_{2}, y, z\right)$.

For a given $p\left(u_{1}, u_{2}, x_{1}, x_{2}, y, z\right)$, consider the perturbation defined by
$p_{\epsilon}\left(u_{1}, u_{2}, x_{1}, x_{2}, y, z\right)=p\left(u_{1}, u_{2}, x_{1}, x_{2}, y, z\right)\left(1+\epsilon \phi\left(u_{1}\right)\right)$.
For $p_{\epsilon}\left(u_{1}, u_{2}, x_{1}, x_{2}, y, z\right)$ to be a valid p.m.f., we require that $\left(1+\epsilon \phi\left(u_{1}\right)\right) \geq 0$ for all $u_{1}$, and $\sum_{u_{1}} p\left(u_{1}\right) \phi\left(u_{1}\right)=0$. Furthermore, we will consider perturbations $\phi\left(u_{1}\right)$ such that

$$
\begin{align*}
& \mathbb{E}\left[\phi\left(U_{1}\right) \mid X_{1}=x_{1}, X_{2}=x_{2}, Y=y, Z=z\right] \\
& =\sum_{u_{1}} p\left(u_{1} \mid x_{1}, x_{2}, y, z\right) \phi\left(u_{1}\right)=0, \forall x_{1}, x_{2}, y, z \tag{136}
\end{align*}
$$

Observe that such a non-zero perturbation satisfying (136) (which also implies $\sum_{u_{1}} p\left(u_{1}\right) \phi\left(u_{1}\right)=0$ ) exists as long as $\left|\mathcal{U}_{1}\right|>\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\right||\mathcal{Y}||\mathcal{Z}|$, since the null-space of the constraints has rank at most $\left|\mathcal{X}_{1}\left\|\mathcal{X}_{2}\right\| \mathcal{Y} \| \mathcal{Z}\right|$. For sufficiently small values of $|\epsilon|$, we also have $\left(1+\epsilon \phi\left(u_{1}\right)\right) \geq 0$ for all $u_{1}$. Note that this perturbation preserves the distribution of $\left(X_{1}, X_{2}, Y, Z\right)$, i.e., $p\left(x_{1}, x_{2}, y, z\right)$. This follows since

$$
\begin{align*}
& p_{\epsilon}\left(x_{1}, x_{2}, y, z\right)=\sum_{u_{1}, u_{2}} p_{\epsilon}\left(u_{1}, u_{2}, x_{1}, x_{2}, y, z\right) \\
& =\sum_{u_{1}, u_{2}} p\left(u_{1}, u_{2}, x_{1}, x_{2}, y, z\right)\left(1+\epsilon \phi\left(u_{1}\right)\right) \\
& =p\left(x_{1}, x_{2}, y, z\right) \\
& \quad\left\{\sum_{u_{1}} p\left(u_{1} \mid x_{1}, x_{2}, y, z\right)+\epsilon \sum_{u_{1}} p\left(u_{1} \mid x_{1}, x_{2}, y, z\right) \phi\left(u_{1}\right)\right\} \\
& =p\left(x_{1}, x_{2}, y, z\right) \tag{137}
\end{align*}
$$

where the last step follows from (136).
We now show that the perturbed distribution $p_{\epsilon}(\cdot)$ preserves the structure of the p.m.f. in (123). To see this, note that

$$
\begin{align*}
& p_{\epsilon}\left(x_{1}, x_{2}, z, u_{1}, u_{2}, y\right) \\
& =p\left(x_{1}, x_{2}, z, u_{1}, u_{2}, y\right)\left(1+\epsilon \phi\left(u_{1}\right)\right) \\
& =p\left(x_{1}\right) p\left(z \mid x_{1}\right)\left\{p\left(u_{1} \mid x_{1}\right)\left(1+\epsilon \phi\left(u_{1}\right)\right)\right\} p\left(x_{2} \mid z\right) p\left(u_{2} \mid x_{2}\right) \\
& \quad p\left(y \mid u_{1}, u_{2}, z\right) \tag{138}
\end{align*}
$$

Marginalizing away $Z, X_{2}, U_{2}, Y$, we have

$$
\begin{equation*}
p_{\epsilon}\left(x_{1}, u_{1}\right)=p\left(x_{1}\right) p\left(u_{1} \mid x_{1}\right)\left(1+\epsilon \phi\left(u_{1}\right)\right) \tag{139}
\end{equation*}
$$

Since $p_{\epsilon}\left(x_{1}\right)=p\left(x_{1}\right)$ (by (137)), we have $p_{\epsilon}\left(u_{1} \mid x_{1}\right)=$ $p\left(u_{1} \mid x_{1}\right)\left(1+\epsilon \phi\left(u_{1}\right)\right)$. Thus,

$$
\begin{align*}
& p_{\epsilon}\left(x_{1}, x_{2}, z, u_{1}, u_{2}, y\right) \\
& \quad=p(z) p\left(x_{1} \mid z\right) p\left(x_{2} \mid z\right) p_{\epsilon}\left(u_{1} \mid x_{1}\right) p\left(u_{2} \mid x_{2}\right) p\left(y \mid u_{1}, u_{2}, z\right) \tag{140}
\end{align*}
$$

which is of the form (123).
Now if the distribution $p\left(u_{1}, u_{2}, x_{1}, x_{2}, y, z\right)$ minimizes $\lambda_{1} I\left(U_{1} ; X_{1} \mid Z\right)+\lambda_{2} I\left(U_{2} ; X_{2} \mid Z\right)+\lambda_{3} I\left(U_{1} ; X_{1}, Y \mid X_{2}, Z\right)$, then for any valid perturbation, we must have the extremality condition:

$$
\begin{align*}
& \frac{d}{d \epsilon}\left(\lambda_{1} I_{\epsilon}\left(U_{1} ; X_{1} \mid Z\right)+\lambda_{2} I_{\epsilon}\left(U_{2} ; X_{2} \mid Z\right)\right. \\
& \left.\quad+\lambda_{3} I_{\epsilon}\left(U_{1} ; X_{1}, Y \mid X_{2}, Z\right)\right)\left.\right|_{\epsilon=0}=0 \tag{141}
\end{align*}
$$

where the subscript $\epsilon$ in the mutual information terms is used to denote that these are evaluated under the perturbed
distribution $p_{\epsilon}\left(u_{1}, u_{2}, x_{1}, x_{2}, y, z\right)$. We examine the weighted sum term under the perturbed distribution $p_{\epsilon}(\cdot)$.

$$
\begin{align*}
& \lambda_{1} I_{\epsilon}\left(U_{1} ; X_{1} \mid Z\right)+\lambda_{2} I_{\epsilon}\left(U_{2} ; X_{2} \mid Z\right)+\lambda_{3} I_{\epsilon}\left(U_{1} ; X_{1}, Y \mid X_{2}, Z\right) \\
& =\lambda_{1}\left(H_{\epsilon}\left(X_{1}, Z\right)-H_{\epsilon}(Z)+H_{\epsilon}\left(U_{1}, Z\right)-H_{\epsilon}\left(U_{1}, X_{1}, Z\right)\right) \\
& +\lambda_{2}\left(H_{\epsilon}\left(X_{2}, Z\right)-H_{\epsilon}(Z)+H_{\epsilon}\left(U_{2}, Z\right)-H_{\epsilon}\left(U_{2}, X_{2}, Z\right)\right) \\
& +\lambda_{3}\left(H_{\epsilon}\left(X_{1}, X_{2}, Y, Z\right)-H_{\epsilon}\left(X_{2}, Z\right)+H_{\epsilon}\left(U_{1}, X_{2}, Z\right)\right. \\
& \text { - } \left.H_{\epsilon}\left(U_{1}, X_{1}, X_{2}, Y, Z\right)\right) \\
& \stackrel{(a)}{=} \lambda_{1}\left(H\left(X_{1}, Z\right)-H(Z)+H_{\epsilon}\left(U_{1}, Z\right)-H_{\epsilon}\left(U_{1}, X_{1}, Z\right)\right) \\
& +\lambda_{2}\left(H\left(X_{2}, Z\right)-H(Z)+H_{\epsilon}\left(U_{2}, Z\right)-H_{\epsilon}\left(U_{2}, X_{2}, Z\right)\right) \\
& +\lambda_{3}\left(H\left(X_{1}, X_{2}, Y, Z\right)-H\left(X_{2}, Z\right)+H_{\epsilon}\left(U_{1}, X_{2}, Z\right)\right. \\
& \left.-H_{\epsilon}\left(U_{1}, X_{1}, X_{2}, Y, Z\right)\right) \\
& \stackrel{(b)}{=} \lambda_{1}\left(H\left(X_{1}, Z\right)-H(Z)+H\left(U_{1}, Z\right)+\epsilon H_{\phi}\left(U_{1}, Z\right)\right. \\
& -\left[\sum_{u_{1}, z} p\left(u_{1}, z\right)\left(1+\epsilon \phi\left(u_{1}\right)\right) \log \left(1+\epsilon \phi\left(u_{1}\right)\right)\right] \\
& \text { - } H\left(U_{1}, X_{1}, Z\right)-\epsilon H_{\phi}\left(U_{1}, X_{1}, Z\right) \\
& \left.+\left[\sum_{u_{1}, x_{1}, z} p\left(u_{1}, x_{1}, z\right)\left(1+\epsilon \phi\left(u_{1}\right)\right) \log \left(1+\epsilon \phi\left(u_{1}\right)\right)\right]\right) \\
& +\lambda_{2}\left(H\left(X_{2}, Z\right)-H(Z)+H_{\epsilon}\left(U_{2}, Z\right)-H_{\epsilon}\left(U_{2}, X_{2}, Z\right)\right) \\
& +\lambda_{3}\left(H\left(X_{1}, X_{2}, Y, Z\right)-H\left(X_{2}, Z\right)+H\left(U_{1}, X_{2}, Z\right)+\right. \\
& \epsilon H_{\phi}\left(U_{1}, X_{2}, Z\right)- \\
& {\left[\sum_{u_{1}, x_{2}, z} p\left(u_{1}, x_{2}, z\right)\left(1+\epsilon \phi\left(u_{1}\right)\right) \log \left(1+\epsilon \phi\left(u_{1}\right)\right)\right]} \\
& -H\left(U_{1}, X_{1}, X_{2}, Y, Z\right)-\epsilon H_{\phi}\left(U_{1}, X_{1}, X_{2}, Y, Z\right)+ \\
& \left.\left[\sum_{u_{1}, x_{1}, x_{2}, y, z} p\left(u_{1}, x_{1}, x_{2}, y, z\right)\left(1+\epsilon \phi\left(u_{1}\right)\right) \log \left(1+\epsilon \phi\left(u_{1}\right)\right)\right]\right) \\
& \stackrel{(c)}{=} \lambda_{1}\left(I\left(U_{1} ; X_{1} \mid Z\right)+\epsilon H_{\phi}\left(U_{1}, Z\right)-\epsilon H_{\phi}\left(U_{1}, X_{1}, Z\right)\right) \\
& +\lambda_{2}\left(H\left(X_{2}, Z\right)-H(Z)+H_{\epsilon}\left(U_{2}, Z\right)-H_{\epsilon}\left(U_{2}, X_{2}, Z\right)\right) \\
& +\lambda_{3}\left(I\left(U_{1} ; X_{1}, Y \mid X_{2}, Z\right)+\epsilon H_{\phi}\left(U_{1}, X_{2}, Z\right)\right. \\
& \left.-\epsilon H_{\phi}\left(U_{1}, X_{1}, X_{2}, Y, Z\right)\right), \tag{142}
\end{align*}
$$

where (a) follows since the joint distribution of $\left(X_{1}, X_{2}, Y, Z\right)$ is preserved from (137), while in (b) we have defined

$$
\begin{aligned}
& H_{\phi}\left(U_{1}, Z\right)=-\sum_{u_{1}, z} p\left(u_{1}, z\right) \phi\left(u_{1}\right) \log p\left(u_{1}, z\right) \\
& H_{\phi}\left(U_{1}, X_{1}, Z\right)=-\sum_{u_{1}, x_{1}, z} p\left(u_{1}, x_{1}, z\right) \phi\left(u_{1}\right) \log p\left(u_{1}, x_{1}, z\right) \\
& H_{\phi}\left(U_{1}, X_{1}, X_{2}, Y, Z\right) \\
& =-\sum_{u_{1}, x_{1}, x_{2}, y, z} p\left(u_{1}, x_{1}, x_{2}, y, z\right) \phi\left(u_{1}\right) \log p\left(u_{1}, x_{1}, x_{2}, y, z\right),
\end{aligned}
$$

and (c) follows from the fact that all the sums in the square brackets are equal.

Note that from (140),

$$
\begin{equation*}
p_{\epsilon}\left(u_{2}, x_{2}, z\right)=p\left(u_{2}, x_{2}, z\right) \tag{143}
\end{equation*}
$$

Hence it follows that

$$
\begin{align*}
H_{\epsilon}\left(U_{2}, X_{2}, Z\right) & =H\left(U_{2}, X_{2}, Z\right)  \tag{144}\\
H_{\epsilon}\left(U_{2}, Z\right) & =H\left(U_{2}, Z\right) \tag{145}
\end{align*}
$$

Substituting (144) and (145) into (142), the weighted sum becomes

$$
\begin{align*}
& \lambda_{1} I_{\epsilon}\left(U_{1} ; X_{1} \mid Z\right)+\lambda_{2} I_{\epsilon}\left(U_{2} ; X_{2} \mid Z\right)+\lambda_{3} I_{\epsilon}\left(U_{1} ; X_{1}, Y \mid X_{2}, Z\right) \\
& =\lambda_{1}\left(I\left(U_{1} ; X_{1} \mid Z\right)+\epsilon H_{\phi}\left(U_{1}, Z\right)-\epsilon H_{\phi}\left(U_{1}, X_{1}, Z\right)\right) \\
& \quad+\lambda_{2} I\left(U_{2} ; X_{2} \mid Z\right)+\lambda_{3}\left(I\left(U_{1} ; X_{1}, Y \mid X_{2}, Z\right)\right. \\
& \left.\quad+\epsilon H_{\phi}\left(U_{1}, X_{2}, Z\right)-\epsilon H_{\phi}\left(U_{1}, X_{1}, X_{2}, Y, Z\right)\right) . \tag{146}
\end{align*}
$$

Now we apply the first derivative condition (141) to (146). This yields

$$
\begin{align*}
& \lambda_{1}\left(H_{\phi}\left(U_{1}, Z\right)-H_{\phi}\left(U_{1}, X_{1}, Z\right)\right) \\
& \quad+\lambda_{3}\left(H_{\phi}\left(U_{1}, X_{2}, Z\right)-H_{\phi}\left(U_{1}, X_{1}, X_{2}, Y, Z\right)\right)=0 \tag{147}
\end{align*}
$$

Substituting (147) into (146), we obtain

$$
\begin{aligned}
& \lambda_{1} I_{\epsilon}\left(U_{1} ; X_{1} \mid Z\right)+\lambda_{2} I_{\epsilon}\left(U_{2} ; X_{2} \mid Z\right)+\lambda_{3} I_{\epsilon}\left(U_{1} ; X_{1}, Y \mid X_{2}, Z\right) \\
& =\lambda_{1} I\left(U_{1} ; X_{1} \mid Z\right)+\lambda_{2} I\left(U_{2} ; X_{2} \mid Z\right)+\lambda_{3} I\left(U_{1} ; X_{1}, Y \mid X_{2}, Z\right) .
\end{aligned}
$$

Thus if $p\left(u_{1}, u_{2}, x_{1}, x_{2}, y, z\right)$ attains the minimum of the weighted sum rate, then the latter is preserved for any valid perturbation $p_{\epsilon}\left(u_{1}, u_{2}, x_{1}, x_{2}, y, z\right)$ that satisfies (136). Now we choose $\epsilon$ such that

$$
\min _{u_{1}}\left(1+\epsilon \phi\left(u_{1}\right)\right)=0
$$

and let $u_{1}=u_{1}^{*}$ attain this minimum. Clearly, $p_{\epsilon}\left(u_{1}^{*}\right)=0$, and hence there exists an $U_{1}$ with cardinality at most $\left|\mathcal{U}_{1}\right|-1$ such that $\lambda_{1} I\left(U_{1} ; X_{1} \mid Z\right)+\lambda_{2} I\left(U_{2} ; X_{2} \mid Z\right)+\lambda_{3} I\left(U_{1} ; X_{1}, Y \mid X_{2}, Z\right)$ is preserved. We can proceed by induction until $\left|\mathcal{U}_{1}\right|=$ $\left|\mathcal{X}_{1}\left\|\mathcal{X}_{2}|\|\mathcal{Y}\| \mathcal{Z}|\right.\right.$. When this happens, we are no longer guaranteed the existence of a non-trivial $\phi\left(u_{1}\right)$ satisfying (136). Hence we can restrict the cardinality to $\left|\mathcal{U}_{1}\right| \leq\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\right||\mathcal{Y}||\mathcal{Z}|$.

We next perturb $U_{2}$. For a given $p\left(x_{1}, x_{2}, u_{1}, u_{2}, y, z\right)$, consider another perturbation defined by
$p_{\epsilon}^{\prime}\left(x_{1}, x_{2}, u_{1}, u_{2}, y, z\right)=p\left(x_{1}, x_{2}, u_{1}, u_{2}, y, z\right)\left(1+\epsilon \phi^{\prime}\left(u_{2}\right)\right)$.
We require that $\left(1+\epsilon \phi^{\prime}\left(u_{2}\right)\right) \geq 0$ for all $u_{2}$, and $\sum_{u_{2}} p\left(u_{2}\right) \phi^{\prime}\left(u_{2}\right)=0$. Furthermore, let $\phi^{\prime}\left(u_{2}\right)$ be such that

$$
\begin{align*}
& \mathbb{E}\left[\phi^{\prime}\left(U_{2}\right) \mid U_{1}=u_{1}, X_{1}=x_{1}, X_{2}=x_{2}, Y=y, Z=z\right] \\
& =\sum_{u_{2}} p\left(u_{2} \mid u_{1}, x_{1}, x_{2}, y, z\right) \phi^{\prime}\left(u_{2}\right)=0, \forall u_{1}, x_{1}, x_{2}, y, z \tag{148}
\end{align*}
$$

Such a non-zero perturbation satisfying (148) exists as long as $\left|\mathcal{U}_{2}\right|>\left|\mathcal{U}_{1}\right|\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\right||\mathcal{Y} \| \mathcal{Z}|$. Again, it can be verified that $p_{\epsilon}^{\prime}(\cdot)$ preserves $p\left(x_{1}, x_{2}, y, z\right)$ as well as the structure of the p.m.f. in (123). We examine the weighted sum term under the perturbed distribution $p_{\epsilon}^{\prime}(\cdot)$.

$$
\begin{aligned}
& \lambda_{1} I_{\epsilon}\left(U_{1} ; X_{1} \mid Z\right)+\lambda_{2} I_{\epsilon}\left(U_{2} ; X_{2} \mid Z\right)+\lambda_{3} I_{\epsilon}\left(U_{1} ; X_{1}, Y \mid X_{2}, Z\right) \\
& =\lambda_{1}\left(H\left(X_{1}, Z\right)-H(Z)+H_{\epsilon}\left(U_{1}, Z\right)-H_{\epsilon}\left(U_{1}, X_{1}, Z\right)\right) \\
& +\lambda_{2}\left(H\left(X_{2}, Z\right)-H(Z)+H_{\epsilon}\left(U_{2}, Z\right)-H_{\epsilon}\left(U_{2}, X_{2}, Z\right)\right) \\
& +\lambda_{3}\left(H\left(X_{1}, X_{2}, Y, Z\right)-H\left(X_{2}, Z\right)+H_{\epsilon}\left(U_{1}, X_{2}, Z\right)\right. \\
& \left.\quad-H_{\epsilon}\left(U_{1}, X_{1}, X_{2}, Y, Z\right)\right) \\
& \stackrel{(a)}{=} \lambda_{1}\left(H\left(X_{1}, Z\right)-H(Z)+H_{\epsilon}\left(U_{1}, Z\right)-H_{\epsilon}\left(U_{1}, X_{1}, Z\right)\right) \\
& +\lambda_{2}\left(I\left(U_{2} ; X_{2} \mid Z\right)+\epsilon H_{\phi^{\prime}}\left(U_{2}, Z\right)-\epsilon H_{\phi^{\prime}}\left(U_{2}, X_{2}, Z\right)\right)
\end{aligned}
$$

$$
\begin{align*}
& +\lambda_{3}\left(H\left(X_{1}, X_{2}, Y, Z\right)-H\left(X_{2}, Z\right)+H_{\epsilon}\left(U_{1}, X_{2}, Z\right)\right. \\
& \left.\quad-H_{\epsilon}\left(U_{1}, X_{1}, X_{2}, Y, Z\right)\right) \tag{149}
\end{align*}
$$

where in (a) we have defined $H_{\phi^{\prime}}\left(U_{2}, X_{2}, Z\right)=$ $-\sum_{u_{2}, x_{2}, z} p\left(u_{2}, x_{2}, z\right) \phi^{\prime}\left(u_{2}\right) \log p\left(u_{2}, x_{2}, z\right)$. Consider the p.m.f. of $\left(U_{1}, X_{1}, X_{2}, Y, Z\right)$ under the perturbation $p_{\epsilon}^{\prime}(\cdot)$.
$p_{\epsilon}^{\prime}\left(u_{1}, x_{1}, x_{2}, y, z\right)=\sum_{u_{2}} p_{\epsilon}^{\prime}\left(u_{1}, u_{2}, x_{1}, x_{2}, y, z\right)$
$=\sum_{u_{2}} p\left(u_{1}, u_{2}, x_{1}, x_{2}, y, z\right)\left(1+\epsilon \phi^{\prime}\left(u_{2}\right)\right)$
$=p\left(u_{1}, x_{1}, x_{2}, y, z\right)\left\{1+\epsilon \sum_{u_{2}} p\left(u_{2} \mid u_{1}, x_{1}, x_{2}, y, z\right) \phi^{\prime}\left(u_{2}\right)\right\}$
$\stackrel{(a)}{=} p\left(u_{1}, x_{1}, x_{2}, y, z\right)$,
where (a) follows from (148). Hence it follows that the terms $H_{\epsilon}\left(U_{1}, X_{1}, X_{2}, Y, Z\right), H_{\epsilon}\left(U_{1}, X_{2}, Z\right), H_{\epsilon}\left(U_{1}, Z\right)$ and $H_{\epsilon}\left(U_{1}, X_{1}, Z\right)$ are preserved under $p_{\epsilon}^{\prime}(\cdot)$. Hence, the weighted sum in (149) becomes

$$
\begin{align*}
& \lambda_{1} I_{\epsilon}\left(U_{1} ; X_{1} \mid Z\right)+\lambda_{2} I_{\epsilon}\left(U_{2} ; X_{2} \mid Z\right)+\lambda_{3} I_{\epsilon}\left(U_{1} ; X_{1}, Y \mid X_{2}, Z\right) \\
& =\lambda_{1} I\left(U_{1} ; X_{1} \mid Z\right)+\lambda_{2}\left(I\left(U_{2} ; X_{2} \mid Z\right)+\epsilon H_{\phi^{\prime}}\left(U_{2}, Z\right)\right. \\
& \left.\quad-\epsilon H_{\phi^{\prime}}\left(U_{2}, X_{2}, Z\right)\right)+\lambda_{3} I\left(U_{1} ; X_{1}, Y \mid X_{2}, Z\right) . \tag{150}
\end{align*}
$$

Now we apply the first derivative condition to (150). This yields

$$
\begin{equation*}
\lambda_{2}\left(H_{\phi^{\prime}}\left(U_{2}, Z\right)-H_{\phi^{\prime}}\left(U_{2}, X_{2}, Z\right)\right)=0 . \tag{151}
\end{equation*}
$$

Substituting (151) into (150), it follows that the weighted sum is preserved. Thus if $p\left(u_{1}, u_{2}, x_{1}, x_{2}, y, z\right)$ attains the minimum of the weighted sum rate, then the latter is preserved for any valid perturbation $p_{\epsilon}^{\prime}\left(u_{1}, u_{2}, x_{1}, x_{2}, y, z\right)$ that satisfies (148). Now the proof is completed by choosing $\epsilon$ such that $\min _{u_{2}}\left(1+\epsilon \phi^{\prime}\left(u_{2}\right)\right)=0$ and the cardinality of $U_{2}$ drops by 1 . We can proceed by induction until $\left|\mathcal{U}_{2}\right|=\left|\mathcal{U}_{1}\right|\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\right||\mathcal{Y}||\mathcal{Z}|$.

## Appendix F

Proof of Lemma 7
Since the arguments are along the same lines as Appendix E, we only outline the differences below. Again, the cardinality bound on $T$ can be obtained using the support lemma [55, Appendix C], and we focus on cardinality bounds for $U_{0}$, $U_{1}$ and $U_{2}$. Let $\mathcal{C}^{\prime}$ denote the region in $S_{\epsilon}^{\prime}$ with $|\mathcal{T}|=1$, i.e., the collection of $\left(R_{1}, R_{2}\right)$ such that:

$$
\begin{align*}
R_{1} & \geq I\left(U_{0}, U_{1} ; X_{1} \mid Z\right)  \tag{152}\\
R_{2} & \geq I\left(U_{0}, U_{2} ; X_{2} \mid Z\right)  \tag{153}\\
R_{1}+R_{2} & \geq I\left(U_{0}, U_{1}, U_{2} ; X_{1}, X_{2} \mid Z\right), \tag{154}
\end{align*}
$$

for some p.m.f.

$$
\begin{align*}
& p\left(x_{1}, x_{2}, u_{0}, u_{1}, u_{2}, y, z\right)=p(z) p\left(x_{1} \mid z\right) p\left(x_{2} \mid z\right) p\left(u_{0}\right) \\
& \quad \times p\left(u_{1} \mid x_{1}, u_{0}\right) p\left(u_{2} \mid x_{2}, u_{0}\right) p\left(y \mid u_{1}, u_{2}, z\right) \tag{155}
\end{align*}
$$

such that

$$
\begin{equation*}
\left\|\sum_{u_{0}, u_{1}, u_{2}} p\left(x_{1}, x_{2}, u_{0}, u_{1}, u_{2}, y, z\right)-q\left(x_{1}, x_{2}, y, z\right)\right\| \leq \epsilon \tag{156}
\end{equation*}
$$

As before, we use the perturbation argument of [56], and the proof is in two steps.
$\left.\begin{array}{l}\text { - Step 1: We prove that } \mathcal{C}^{\prime} \quad= \\ \text { Closure }\left(\underset{K_{0}, K_{1}, K_{2} \geq 0}{\bigcup} \mathcal{C}^{\prime} K_{0}, K_{1}, K_{2}\right.\end{array}\right)$, where the region $\mathcal{C}^{\prime K_{0}, K_{1}, K_{2}}$ for positive integers $K_{0}, K_{1}, K_{2}$ is given by the union of rate pairs $\left(R_{1}, R_{2}\right)$ satisfying (152)-(154) over $\left(U_{0}, U_{1}, U_{2}, X_{1}, X_{2}, Y, Z\right)$ with cardinality bounds $\left|\mathcal{U}_{0}\right| \leq K_{0},\left|\mathcal{U}_{1}\right| \leq K_{1} \&\left|\mathcal{U}_{2}\right| \leq K_{2}$ and having a joint p.m.f.

$$
\begin{aligned}
& p(z) p\left(x_{1}\right) p\left(x_{2}\right) p\left(u_{0}\right) p\left(u_{1} \mid x_{1}, u_{0}\right) p\left(u_{2} \mid x_{2}, u_{0}\right) \\
& \quad \times p\left(y \mid u_{1}, u_{2}, z\right)
\end{aligned}
$$

such that

$$
\left\|p\left(x_{1}, x_{2}, y, z\right)-q\left(x_{1}, x_{2}, y, z\right)\right\| \leq \epsilon
$$

- Step 2: If $\left|\mathcal{U}_{0}\right| \leq K_{0},\left|\mathcal{U}_{1}\right| \leq K_{1}$ and $\left|\mathcal{U}_{2}\right| \leq K_{2}$ for some constants $K_{0}, K_{1}, K_{2}$, we show that the auxiliary cardinalities can be brought down to $\left|\mathcal{U}_{0}\right| \leq$ $\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\right||\mathcal{Y}||\mathcal{Z}|,\left|\mathcal{U}_{1}\right| \leq\left|\mathcal{U}_{0}\right|\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\right||\mathcal{Y}||\mathcal{Z}|$ and $\left|\mathcal{U}_{2}\right| \leq$ $\left|\mathcal{U}_{0}\left\|\mathcal{X}_{1}\right\| \mathcal{X}_{2}\|\mathcal{Y}\| \mathcal{Z}\right|$.
Proof Step 1: We show that any rate pair $\left(R_{1}, R_{2}\right) \in$ $\mathcal{C}^{\prime}$ is a limit point of the set $\bigcup_{K_{0}, K_{1}, K_{2}>0} \mathcal{C}^{\prime K_{0}, K_{1}, K_{2}}$. Firstly, if $\left(R_{1}, R_{2}\right) \in \mathcal{C}^{\prime}$, then random variables $\left(U_{0}, U_{1}, U_{2}, X_{1}\right.$, $\left.X_{2}, Y, Z\right)$ satisfying (152)-(154) exist such that their joint p.m.f. is of the form (155) and (156) holds. Assume that $\mathcal{U}_{j}=$ $\{1,2, \cdots\}$ for $j=0,1,2$. Define a modified version $U_{0}^{\prime}$ of the random variable $U_{0}$ with p.m.f. specified as:

$$
p_{U_{0}^{\prime}}(i)= \begin{cases}p_{U_{0}}(i), & i=\{1,2, \cdots, m\}  \tag{157}\\ \operatorname{Pr}\left(U_{0}>m\right), & i=m+1\end{cases}
$$

where $m$ is an integer. The alphabet of $U_{0}^{\prime}$ is $\mathcal{U}_{0}^{\prime}=$ $\{1,2, \cdots, m+1\}$. Likewise, define modified versions $\left(U_{1}^{\prime}, U_{2}^{\prime}\right)$ of random variables $\left(U_{1}, U_{2}\right)$ taking values in $\{1,2, \cdots, m\} \cup$ $\mathcal{X}_{1} \cup\left(\mathcal{X}_{1} \times \mathcal{U}_{0}^{\prime}\right)$ and $\{1,2, \cdots, m\} \cup \mathcal{X}_{2} \cup\left(\mathcal{X}_{2} \times \mathcal{U}_{0}^{\prime}\right)$ respectively. The alphabet cardinalities of $\left(U_{1}^{\prime}, U_{2}^{\prime}\right)$ are $m+\left|\mathcal{X}_{1}\right|+(m+$ 1) $\left|\mathcal{X}_{1}\right|$ and $m+\left|\mathcal{X}_{2}\right|+(m+1)\left|\mathcal{X}_{2}\right|$ respectively. Let the conditional p.m.f. of $U_{j}^{\prime}$ given $\left(X_{j}, U_{0}^{\prime}\right)$ for $j \in\{1,2\}$, $x_{j} \in \mathcal{X}_{j}$ be specified as:

$$
\begin{align*}
& p_{U_{j}^{\prime} \mid X_{j}, U_{0}^{\prime}}\left(i \mid x_{j}, u_{0}^{\prime}\right) \\
& =p_{U_{j} \mid X_{j}, U_{0}}\left(i \mid x_{j}, u_{0}^{\prime}\right), u_{0}^{\prime} \in\{1,2, \cdots, m\}, i \in\{1,2, \cdots, m\} \tag{158}
\end{align*}
$$

$p_{U_{j}^{\prime} \mid X_{j}, U_{0}^{\prime}}\left(\left(x_{j}, u_{0}^{\prime}\right) \mid x_{j}, u_{0}^{\prime}\right)$
$=\sum_{i=m+1}^{\infty} p_{U_{j} \mid X_{j}, U_{0}}\left(i \mid x_{j}, u_{0}^{\prime}\right), u_{0}^{\prime} \in\{1,2, \cdots, m\}$,
$p_{U_{j}^{\prime} \mid X_{j}, U_{0}^{\prime}}\left(x_{j} \mid x_{j}, m+1\right)=1$,
$p_{U_{j}^{\prime} \mid X_{j}, U_{0}^{\prime}}(\cdot \mid \cdot, \cdot)=0$, otherwise.
As before, to define $p_{Y \mid U_{1}^{\prime}, U_{2}^{\prime}, Z}$, consider new auxiliary random variables $\left(U_{1}^{\prime \prime}, U_{2}^{\prime \prime}\right)$ whose alphabets are the same as that of $\left(U_{1}, U_{2}\right)$. Let the conditional p.m.f. $p_{U_{1}^{\prime \prime}, U_{2}^{\prime \prime} \mid U_{1}^{\prime}, U_{2}^{\prime}}$ be specified as follows:

- If $U_{j}^{\prime} \in\{1,2, \ldots, m\} \cup\left(\mathcal{X}_{j} \times \mathcal{U}_{0}\right), j=1,2$, then, we define

$$
\begin{align*}
P_{U_{1}^{\prime \prime}, U_{2}^{\prime \prime} \mid U_{1}^{\prime}, U_{2}^{\prime}} & \left(u_{1}^{\prime \prime}, u_{2}^{\prime \prime} \mid u_{1}^{\prime}, u_{2}^{\prime}\right) \\
& =P_{U_{1}^{\prime \prime} \mid U_{1}^{\prime}}\left(u_{1}^{\prime \prime} \mid u_{1}^{\prime}\right) \cdot P_{U_{2}^{\prime \prime} \mid U_{2}^{\prime}}\left(u_{2}^{\prime \prime} \mid u_{2}^{\prime}\right) \tag{162}
\end{align*}
$$

where

$$
\begin{gather*}
p_{U_{j}^{\prime \prime} \mid U_{j}^{\prime}}(i \mid i)=1, i=1,2, \cdots, m . \\
p_{U_{j}^{\prime \prime} \mid U_{j}^{\prime}}\left(k \mid\left(x_{j}, i\right)\right)= \\
\begin{cases}\operatorname{Pr}\left(U_{j}=k \mid X_{j}=x_{j}, U_{0}=i, U_{j}>m\right), & k \geq m+1 \\
0, & k=1,2, \cdots, m\end{cases} \tag{164}
\end{gather*}
$$

- Note that, otherwise, $U_{0}^{\prime}=m+1$ and hence, $U_{1}^{\prime} \in \mathcal{X}_{1}$ and $U_{2}^{\prime} \in \mathcal{X}_{2}$. For this case, we define

$$
\begin{align*}
& p_{U_{1}^{\prime \prime}, U_{2}^{\prime \prime} \mid U_{1}^{\prime}, U_{2}^{\prime}}\left(u_{1}, u_{2} \mid x_{1}, x_{2}\right)= \\
& \sum_{u_{0}} \operatorname{Pr}\left(U_{0}=u_{0} \mid U_{0}>m\right) p_{U_{1} \mid X_{1}, U_{0}}\left(u_{1} \mid x_{1}, u_{0}\right) \\
& \quad p_{U_{2} \mid X_{2}, U_{0}}\left(u_{2} \mid x_{2}, u_{0}\right) \tag{165}
\end{align*}
$$

It is easy to verify that

$$
\begin{equation*}
p_{U_{1}^{\prime \prime}, U_{2}^{\prime \prime} \mid X_{1}, X_{2}}\left(u_{1}, u_{2} \mid x_{1}, x_{2}\right)=p_{U_{1}, U_{2} \mid X_{1}, X_{2}}\left(u_{1}, u_{2} \mid x_{1}, x_{2}\right) \tag{166}
\end{equation*}
$$

Further, let

$$
p_{Y \mid U_{1}^{\prime \prime}, U_{2}^{\prime \prime}, Z}\left(y \mid u_{1}, u_{2}, z\right)=p_{Y \mid U_{1}, U_{2}, Z}\left(y \mid u_{1}, u_{2}, z\right)
$$

. The conditional distribution $p_{Y \mid U_{1}^{\prime}, U_{2}^{\prime}, Z}$ is defined as

$$
\begin{align*}
& p_{Y \mid U_{1}^{\prime}, U_{2}^{\prime}, Z}\left(y \mid u_{1}^{\prime}, u_{2}^{\prime}, z\right) \\
& =\sum_{u_{1}^{\prime \prime}, u_{2}^{\prime \prime}} p_{U_{1}^{\prime \prime}, U_{2}^{\prime \prime} \mid U_{1}^{\prime}, U_{2}^{\prime}}\left(u_{1}^{\prime \prime}, u_{2}^{\prime \prime} \mid u_{1}^{\prime}, u_{2}^{\prime}\right) p_{Y \mid U_{1}^{\prime \prime}, U_{2}^{\prime \prime}, Z}\left(y \mid u_{1}^{\prime \prime}, u_{2}^{\prime \prime}, z\right)  \tag{167}\\
& =\sum_{u_{1}^{\prime \prime}, u_{2}^{\prime \prime}} p_{U_{1}^{\prime \prime}, U_{2}^{\prime \prime} \mid U_{1}^{\prime}, U_{2}^{\prime}}\left(u_{1}^{\prime \prime}, u_{2}^{\prime \prime} \mid u_{1}^{\prime}, u_{2}^{\prime}\right) p_{Y \mid U_{1}, U_{2}, Z}\left(y \mid u_{1}^{\prime \prime}, u_{2}^{\prime \prime}, z\right) \tag{168}
\end{align*}
$$

The above definitions preserve all the Markov chains associated with the original p.m.f. As a result, the joint distribution of $\left(U_{0}^{\prime}, U_{1}^{\prime}, U_{2}^{\prime}, X_{1}, X_{2}, Y, Z\right)$ factors according to $p_{U_{0}^{\prime}, U_{1}^{\prime}, U_{2}^{\prime}, X_{1}, X_{2}, Y, Z}=p_{X_{1}, X_{2}, Z} p_{U_{0}^{\prime}} p_{U_{1}^{\prime} \mid X_{1}, U_{0}^{\prime}}$ $p_{U_{2}^{\prime} \mid X_{2}, U_{0}^{\prime}} p_{Y \mid U_{1}^{\prime}, U_{2}^{\prime}, Z}$. On marginalizing away $\left(U_{0}^{\prime}, U_{1}^{\prime}, U_{2}^{\prime}\right)$ from $p_{U_{0}^{\prime}, U_{1}^{\prime}, U_{2}^{\prime}, X_{1}, X_{2}, Y, Z}$, we may verify that we obtain the p.m.f. $p_{X_{1}, X_{2}, Y, Z}$.

It follows that the joint distribution of the random variables $\left(U_{0}^{\prime}, U_{1}^{\prime}, U_{2}^{\prime}, X_{1}, X_{2}, Y, Z\right)$ converges to the the joint distribution of $\left(U_{0}, U_{1}, U_{2}, X_{1}, X_{2}, Y, Z\right)$ in the limit $m \rightarrow \infty$. As a result, the mutual information terms $I\left(U_{0}^{\prime}, U_{1}^{\prime} ; X_{1} \mid Z\right), \quad I\left(U_{0}^{\prime}, U_{2}^{\prime} ; X_{2} \mid Z\right)$ and $I\left(U_{0}^{\prime}, U_{1}^{\prime}, U_{2}^{\prime}, ; X_{1}, X_{2} \mid Z\right)$ converge to $I\left(U_{0}, U_{1} ; X_{1} \mid Z\right)$, $I\left(U_{0}, U_{2} ; X_{2} \mid Z\right)$ and $I\left(U_{0}, U_{1}, U_{2}, ; X_{1}, X_{2} \mid Z\right)$ respectively.

Hence, we conclude that the given rate pair $\left(R_{1}, R_{2}\right)$ is a limit point of the set $\bigcup_{K_{0}, K_{1}, K_{2} \geq 0} \mathcal{C}^{\prime K_{0}, K_{1}, K_{2}}$.

Proof Step 2: As in Appendix E, it suffices to consider an optimization of the weighted sum term $\lambda_{1} I\left(U_{0}, U_{1} ; X_{1} \mid Z\right)+$ $\lambda_{2} I\left(U_{0}, U_{2} ; X_{2} \mid Z\right)+\lambda_{3} I\left(U_{0}, U_{1}, U_{2} ; X_{1}, X_{2} \mid Z\right)$ for nonnegative reals $\lambda_{1}, \lambda_{2}, \lambda_{3}$, and find new auxiliary random variables whose cardinalities are bounded while not increasing the weighted sum. For a given $p\left(x_{1}, x_{2}, u_{0}, u_{1}, u_{2}, y, z\right)$, consider the perturbation defined by

$$
\begin{aligned}
p_{\epsilon}\left(x_{1}, x_{2}, u_{0}, u_{1}, u_{2}, y, z\right)= & p\left(x_{1}, x_{2}, u_{0}, u_{1}, u_{2}, y, z\right) \\
& \times\left(1+\epsilon \phi\left(u_{0}\right)\right)
\end{aligned}
$$

For $p_{\epsilon}\left(x_{1}, x_{2}, u_{0}, u_{1}, u_{2}, y, z\right)$ to be a valid p.m.f., we require that $\left(1+\epsilon \phi\left(u_{0}\right)\right) \geq 0$ for all $u_{0}$, and $\sum_{u_{0}} p\left(u_{0}\right) \phi\left(u_{0}\right)=0$. Furthermore, we will consider perturbations $\phi\left(u_{0}\right)$ such that

$$
\begin{align*}
& \mathbb{E}\left[\phi\left(U_{0}\right) \mid X_{1}=x_{1}, X_{2}=x_{2}, Y=y, Z=z\right] \\
& =\sum_{u_{0}} p\left(u_{0} \mid x_{1}, x_{2}, y, z\right) \phi\left(u_{0}\right)=0, \forall x_{1}, x_{2}, y, z \tag{169}
\end{align*}
$$

Observe that such a non-zero perturbation satisfying (169) exists as long as $\left|\mathcal{U}_{0}\right|>\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\right||\mathcal{Y}||\mathcal{Z}|$. Similar to Appendix E, it can be shown verified that this perturbation preserves $p\left(x_{1}, x_{2}, y, z\right)$ as well as the structure of the p.m.f. in (155).

We examine the weighted sum term under $p_{\epsilon}(\cdot)$.

$$
\begin{align*}
& \lambda_{1} I_{\epsilon}\left(U_{0}, U_{1} ; X_{1} \mid Z\right)+\lambda_{2} I_{\epsilon}\left(U_{0}, U_{2} ; X_{2} \mid Z\right) \\
& \quad+\lambda_{3} I_{\epsilon}\left(U_{0}, U_{1}, U_{2} ; X_{1}, X_{2} \mid Z\right) \\
& = \\
& \lambda_{1}\left(H\left(X_{1} \mid Z\right)+H_{\epsilon}\left(U_{0}, U_{1}, Z\right)-H_{\epsilon}\left(U_{0}, U_{1}, X_{1}, Z\right)\right) \\
& + \\
& +\lambda_{2}\left(H\left(X_{2} \mid Z\right)+H_{\epsilon}\left(U_{0}, U_{2}, Z\right)-H_{\epsilon}\left(U_{0}, U_{2}, X_{2}, Z\right)\right) \\
& + \\
& \quad \lambda_{3}\left(H\left(X_{1}, X_{2} \mid Z\right)+H_{\epsilon}\left(U_{0}, U_{1}, U_{2}, Z\right)\right. \\
& \left.\quad-H_{\epsilon}\left(U_{0}, U_{1}, U_{2}, X_{1}, X_{2}, Z\right)\right) \\
& \stackrel{(a)}{=} \lambda_{1}\left(I\left(U_{0}, U_{1} ; X_{1} \mid Z\right)+\epsilon H_{\phi}\left(U_{0}, U_{1}, Z\right)\right. \\
& \left.\quad-\epsilon H_{\phi}\left(U_{0}, U_{1}, X_{1}, Z\right)\right)  \tag{170}\\
& + \\
& \quad \lambda_{2}\left(I\left(U_{0}, U_{2} ; X_{2} \mid Z\right)+\epsilon H_{\phi}\left(U_{0}, U_{2}, Z\right)\right. \\
& \left.\quad-\epsilon H_{\phi}\left(U_{0}, U_{2}, X_{2}, Z\right)\right)+\lambda_{3}\left(I\left(U_{0}, U_{1}, U_{2} ; X_{1}, X_{2} \mid Z\right)\right. \\
& \left.\quad+\epsilon H_{\phi}\left(U_{0}, U_{1}, U_{2}, Z\right)-\epsilon H_{\phi}\left(U_{0}, U_{1}, U_{2}, X_{1}, X_{2}, Z\right)\right)
\end{align*}
$$

where in (a) we have defined

$$
\begin{aligned}
& H_{\phi}\left(U_{0}, U_{1}, U_{2}, X_{1}, X_{2}, Z\right) \\
& \quad=-\sum_{u_{0}, u_{1}, u_{2}, x_{1}, x_{2}, z} p\left(u_{0}, u_{1}, u_{2}, x_{1}, x_{2}, z\right) \phi\left(u_{0}\right) \\
& \quad \log p\left(u_{0}, u_{1}, u_{2}, x_{1}, x_{2}, z\right)
\end{aligned}
$$

Now we apply the first derivative condition to (170). This yields

$$
\begin{align*}
& \lambda_{1}\left(H_{\phi}\left(U_{0}, U_{1}, Z\right)-H_{\phi}\left(U_{0}, U_{1}, X_{1}, Z\right)\right) \\
& +\lambda_{2}\left(H_{\phi}\left(U_{0}, U_{2}, Z\right)-H_{\phi}\left(U_{0}, U_{2}, X_{2}, Z\right)\right) \\
& +\lambda_{3}\left(H_{\phi}\left(U_{0}, U_{1}, U_{2}, Z\right)-H_{\phi}\left(U_{0}, U_{1}, U_{2}, X_{1}, X_{2}, Z\right)\right)=0 \tag{171}
\end{align*}
$$

Substituting (171) into (170), it follows that the weighted sum is preserved. Thus if $p\left(u_{0}, u_{1}, u_{2}, x_{1}, x_{2}, y, z\right)$ attains the minimum of the weighted sum rate, then the latter is preserved for any valid perturbation $p_{\epsilon}\left(u_{0}, u_{1}, u_{2}, x_{1}, x_{2}, y, z\right)$ that satisfies (169). Now we choose $\epsilon$ such that $\min _{u_{0}}\left(1+\epsilon \phi\left(u_{0}\right)\right)=0$ and let $u_{0}=u_{0}^{*}$ attain this minimum. This makes $p_{\epsilon}\left(u_{0}^{*}\right)=0$, and the cardinality of $U_{0}$ can be reduced by one. We can proceed by induction until $\left|\mathcal{U}_{0}\right|=\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\right||\mathcal{Y}||\mathcal{Z}|$. Hence we can restrict the cardinality to $\left|\mathcal{U}_{0}\right| \leq\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\right||\mathcal{Y}||\mathcal{Z}|$.

We next perturb $U_{1}$. For a given $p\left(x_{1}, x_{2}, u_{0}, u_{1}, u_{2}, y, z\right)$, consider another perturbation defined by

$$
\begin{aligned}
p_{\epsilon}^{\prime}\left(x_{1}, x_{2}, u_{0}, u_{1}, u_{2}, y, z\right)= & p\left(x_{1}, x_{2}, u_{0}, u_{1}, u_{2}, y, z\right) \\
& \times\left(1+\epsilon \phi^{\prime}\left(u_{1}\right)\right)
\end{aligned}
$$

We require that $\left(1+\epsilon \phi^{\prime}\left(u_{1}\right)\right) \geq 0$ for all $u_{1}$, and $\sum_{u_{1}} p\left(u_{1}\right) \phi^{\prime}\left(u_{1}\right)=0$. Furthermore, let $\phi^{\prime}\left(u_{1}\right)$ be such that

$$
\begin{align*}
& \mathbb{E}\left[\phi^{\prime}\left(U_{1}\right) \mid U_{0}=u_{0}, X_{1}=x_{1}, X_{2}=x_{2}, Y=y, Z=z\right] \\
& =\sum_{u_{1}} p\left(u_{1} \mid u_{0}, x_{1}, x_{2}, y, z\right) \phi^{\prime}\left(u_{1}\right)=0, \forall u_{0}, x_{1}, x_{2}, y, z \tag{172}
\end{align*}
$$

Such a non-zero perturbation satisfying (172) exists as long as $\left|\mathcal{U}_{1}\right|>\left|\mathcal{U}_{0}\right|\left|\mathcal{X}_{1}\left\|\mathcal{X}_{2}\right\| \mathcal{Y} \| \mathcal{Z}\right|$. Again, it can be verified that $p_{\epsilon}^{\prime}(\cdot)$ preserves $p\left(x_{1}, x_{2}, y, z\right)$ as well as the structure of the p.m.f. in (155).

We examine the weighted sum term under the perturbed distribution $p_{\epsilon}^{\prime}(\cdot)$.

$$
\begin{align*}
\lambda_{1} & I_{\epsilon}\left(U_{0}, U_{1} ; X_{1} \mid Z\right)+\lambda_{2} I_{\epsilon}\left(U_{0}, U_{2} ; X_{2} \mid Z\right) \\
\quad & +\lambda_{3} I_{\epsilon}\left(U_{0}, U_{1}, U_{2} ; X_{1}, X_{2} \mid Z\right) \\
= & \lambda_{1}\left(H\left(X_{1} \mid Z\right)+H_{\epsilon}\left(U_{0}, U_{1}, Z\right)-H_{\epsilon}\left(U_{0}, U_{1}, X_{1}, Z\right)\right) \\
+ & \lambda_{2}\left(H\left(X_{2} \mid Z\right)+H_{\epsilon}\left(U_{0}, U_{2}, Z\right)-H_{\epsilon}\left(U_{0}, U_{2}, X_{2}, Z\right)\right) \\
+ & \lambda_{3}\left(H\left(X_{1}, X_{2} \mid Z\right)+H_{\epsilon}\left(U_{0}, U_{1}, U_{2}, Z\right)\right. \\
\quad & \left.-H_{\epsilon}\left(U_{0}, U_{1}, U_{2}, X_{1}, X_{2}, Z\right)\right) \\
\stackrel{(a)}{=} & \lambda_{1}\left(I\left(U_{0}, U_{1} ; X_{1} \mid Z\right)+\epsilon H_{\phi^{\prime}}\left(U_{0}, U_{1}, Z\right)\right. \\
& \left.-\epsilon H_{\phi^{\prime}}\left(U_{0}, U_{1}, X_{1}, Z\right)\right) \\
+ & \lambda_{2}\left(H\left(X_{2} \mid Z\right)+H_{\epsilon}\left(U_{0}, U_{2}, Z\right)-H_{\epsilon}\left(U_{0}, U_{2}, X_{2}, Z\right)\right) \\
+ & \lambda_{3}\left(I\left(U_{0}, U_{1}, U_{2} ; X_{1}, X_{2} \mid Z\right)+\epsilon H_{\phi^{\prime}}\left(U_{0}, U_{1}, U_{2}, Z\right)\right. \\
\quad & \left.-\epsilon H_{\phi^{\prime}}\left(U_{0}, U_{1}, U_{2}, X_{1}, X_{2}, Z\right)\right) \tag{173}
\end{align*}
$$

where in (a) we have defined

$$
\begin{aligned}
& H_{\phi^{\prime}}\left(U_{0}, U_{1}, U_{2}, X_{1}, X_{2}, Z\right) \\
& \quad=-\sum_{u_{0}, u_{1}, u_{2}, x_{1}, x_{2}, z} p\left(u_{0}, u_{1}, u_{2}, x_{1}, x_{2}, z\right) \phi^{\prime}\left(u_{1}\right) \\
& \quad \log p\left(u_{0}, u_{1}, u_{2}, x_{1}, x_{2}, z\right)
\end{aligned}
$$

Consider the p.m.f. of $\left(U_{0}, U_{2}, X_{2}, Z\right)$ under the perturbation $p_{\epsilon}^{\prime}(\cdot)$.

$$
\begin{aligned}
& p_{\epsilon}^{\prime}\left(u_{0}, u_{2}, x_{2}, z\right)=\sum_{u_{1}, x_{1}, y} p_{\epsilon}^{\prime}\left(u_{0}, u_{1}, u_{2}, x_{1}, x_{2}, y, z\right) \\
& =\sum_{u_{1}, x_{1}, y} p\left(u_{0}, u_{1}, u_{2}, x_{1}, x_{2}, y, z\right)\left(1+\epsilon \phi^{\prime}\left(u_{1}\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
& =p\left(u_{0}, u_{2}, x_{2}, z\right)\left\{1+\epsilon \sum_{u_{1}} p\left(u_{1} \mid u_{0}, u_{2}, x_{2}, z\right) \phi^{\prime}\left(u_{1}\right)\right\} \\
& \stackrel{(a)}{=} p\left(u_{0}, u_{2}, x_{2}, z\right)\left\{1+\epsilon \sum_{u_{1}} p\left(u_{1} \mid x_{2}, u_{0}\right) \phi^{\prime}\left(u_{1}\right)\right\} \\
& \stackrel{(b)}{=} p\left(u_{0}, u_{2}, x_{2}, z\right)
\end{aligned}
$$

where (a) follows since $U_{2} \rightarrow\left(X_{2}, U_{0}\right) \rightarrow\left(U_{1}, X_{1}, Z\right)$, while (b) follows from (172). Hence it follows that

$$
\begin{align*}
H_{\epsilon}\left(U_{0}, U_{2}, X_{2}, Z\right) & =H\left(U_{0}, U_{2}, X_{2}, Z\right)  \tag{174}\\
H_{\epsilon}\left(U_{0}, U_{2}, Z\right) & =H\left(U_{0}, U_{2}, Z\right) \tag{175}
\end{align*}
$$

Substituting (174) and (175) into (173), the weighted sum becomes

$$
\begin{align*}
& \lambda_{1} I_{\epsilon}\left(U_{0}, U_{1} ; X_{1} \mid Z\right)+\lambda_{2} I_{\epsilon}\left(U_{0}, U_{2} ; X_{2} \mid Z\right) \\
& \quad+\lambda_{3} I_{\epsilon}\left(U_{0}, U_{1}, U_{2} ; X_{1}, X_{2} \mid Z\right) \\
&= \lambda_{1}\left(I\left(U_{0}, U_{1} ; X_{1} \mid Z\right)+\epsilon H_{\phi^{\prime}}\left(U_{0}, U_{1}, Z\right)\right. \\
&\left.\quad-\epsilon H_{\phi^{\prime}}\left(U_{0}, U_{1}, X_{1}, Z\right)\right) \\
&+ \lambda_{2} I\left(U_{0}, U_{2} ; X_{2} \mid Z\right)+\lambda_{3}\left(I\left(U_{0}, U_{1}, U_{2} ; X_{1}, X_{2} \mid Z\right)\right. \\
&+\left.\epsilon H_{\phi^{\prime}}\left(U_{0}, U_{1}, U_{2}, Z\right)-\epsilon H_{\phi^{\prime}}\left(U_{0}, U_{1}, U_{2}, X_{1}, X_{2}, Z\right)\right) \tag{176}
\end{align*}
$$

Now we apply the first derivative condition to (176). This yields

$$
\begin{align*}
& \lambda_{1}\left(H_{\phi^{\prime}}\left(U_{0}, U_{1}, Z\right)-H_{\phi^{\prime}}\left(U_{0}, U_{1}, X_{1}, Z\right)\right)+ \\
& \lambda_{3}\left(H_{\phi^{\prime}}\left(U_{0}, U_{1}, U_{2}, Z\right)-H_{\phi^{\prime}}\left(U_{0}, U_{1}, U_{2}, X_{1}, X_{2}, Z\right)\right)=0 . \tag{177}
\end{align*}
$$

Substituting (177) into (176), it follows that the weighted sum is preserved. Thus if $p\left(u_{0}, u_{1}, u_{2}, x_{1}, x_{2}, y, z\right)$ attains the minimum of the weighted sum rate, then the latter is preserved for any valid perturbation $p_{\epsilon}^{\prime}\left(u_{0}, u_{1}, u_{2}, x_{1}, x_{2}, y, z\right)$ that satisfies (172). Now the proof is completed by choosing $\epsilon$ such that $\min _{u_{1}}\left(1+\epsilon \phi^{\prime}\left(u_{1}\right)\right)=0$ and the cardinality of $U_{1}$ drops by 1 . We can proceed by induction until $\left|\mathcal{U}_{1}\right|=$ $\left|\mathcal{U}_{0}\left\|\mathcal{X}_{1}\right\| \mathcal{X}_{2}\|\mathcal{Y}\| \mathcal{Z}\right|$. The same argument can be repeated to make $\left|\mathcal{U}_{2}\right| \leq\left|\mathcal{U}_{0}\right|\left|\mathcal{X}_{1}\right|\left|\mathcal{X}_{2}\right||\mathcal{Y}||\mathcal{Z}|$ as well.

## Acknowledgement

The authors thank the Associate Editor and the anonymous reviewers for their insightful comments on the article. In particular, the authors would like to thank them for making aware of [37, Sec. IV]. In addition, they thank Sandeep Pradhan for alerting them of an error in an earlier version of the article.

## REFERENCES

[1] P. W. Cuff, H. H. Permuter, and T. M. Cover, "Coordination capacity," IEEE Trans. Inf. Theory, vol. 56, no. 9, pp. 4181-4206, Sep. 2010.
[2] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem," IEEE Trans. Inf. Theory, vol. 48, no. 10, pp. 2637-2655, Oct. 2002.
[3] P. Harsha, R. Jain, D. McAllester, and J. Radhakrishnan, "The communication complexity of correlation," IEEE Trans. Inf. Theory, vol. 56, no. 1, pp. 438-449, Jan. 2010.
[4] A. Winter, "Compression of sources of probability distributions and density operators," 2022, arXiv:quant-ph/0208131.
[5] P. Cuff, "Distributed channel synthesis," IEEE Trans. Inf. Theory, vol. 59, no. 11, pp. 7071-7096, Nov. 2013.
[6] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, "The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels," IEEE Trans. Inf. Theory, vol. 60, no. 3, pp. 2926-2959, May 2014.
[7] M. M. Wilde, P. M. Hayden, F. Buscemi, and M.-H. Hsieh, "The information-theoretic costs of simulating quantum measurements," J. Phys. A, Math. Theor., vol. 45, no. 45, p. 453001, Nov. 2012.
[8] M. H. Yassaee, A. Gohari, and M. R. Aref, "Channel simulation via interactive communications," IEEE Trans. Inf. Theory, vol. 61, no. 6, pp. 2964-2982, Jun. 2015.
[9] F. Haddadpour, M. H. Yassaee, S. Beigi, A. Gohari, and M. R. Aref, "Simulation of a channel with another channel," IEEE Trans. Inf. Theory, vol. 63, no. 5, pp. 2659-2677, May 2017.
[10] G. Cervia, L. Luzzi, M. Le Treust, and M. R. Bloch, "Strong coordination of signals and actions over noisy channels with two-sided state information," IEEE Trans. Inf. Theory, vol. 66, no. 8, pp. 4681-4708, Aug. 2020.
[11] G. Cervia, L. Luzzi, M. R. Bloch, and M. Le Treust, "Polar coding for empirical coordination of signals and actions over noisy channels," in Proc. IEEE Inf. Theory Workshop (ITW), Sep. 2016, pp. 81-85.
[12] M. L. Treust, "Joint empirical coordination of source and channel," IEEE Trans. Inf. Theory, vol. 63, no. 8, pp. 5087-5114, Aug. 2017.
[13] R. A. Chou, M. R. Bloch, and J. Kliewer, "Empirical and strong coordination via soft covering with polar codes," IEEE Trans. Inf. Theory, vol. 64, no. 7, pp. 5087-5100, Jul. 2018.
[14] M. Mylonakis, P. A. Stavrou, and M. Skoglund, "Empirical coordination with multiple descriptions," in Proc. 57th Annu. Allerton Conf. Commun., Control, Comput. (Allerton), Sep. 2019, pp. 1074-1081.
[15] E. Kushilevitz and N. Nisan, Communication Complexity. Cambridge, U.K.: Cambridge Univ. Press, 1996.
[16] M. Braverman and A. Rao, "Information equals amortized communication," IEEE Trans. Inf. Theory, vol. 60, no. 10, pp. 6058-6069, Oct. 2014.
[17] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources (Corresp.)," IEEE Trans. Inf. Theory, vol. 25, no. 2, pp. 219-221, Mar. 1979.
[18] H. Yamamoto, "Wyner-Ziv theory for a general function of the correlated sources (Corresp.)," IEEE Trans. Inf. Theory, vol. 28, no. 5, pp. 803-807, Sep. 1982.
[19] A. H. Kaspi, "Two-way source coding with a fidelity criterion," IEEE Trans. Inf. Theory, vol. 31, no. 6, pp. 735-740, Nov. 1985.
[20] T. Han and K. Kobayashi, "A dichotomy of functions F(X, Y) of correlated sources(X, Y)," IEEE Trans. Inf. Theory, vol. 33, no. 1, pp. 69-76, Jan. 1987.
[21] N. Alon and A. Orlitsky, "Source coding and graph entropies," IEEE Trans. Inf. Theory, vol. 42, no. 5, pp. 1329-1339, Sep. 1996.
[22] A. Orlitsky and J. R. Roche, "Coding for computing," IEEE Trans. Inf. Theory, vol. 47, no. 3, pp. 903-917, Mar. 2001.
[23] B. Nazer and M. Gastpar, "Computation over multiple-access channels," IEEE Trans. Inf. Theory, vol. 53, no. 10, pp. 3498-3516, Oct. 2007.
[24] B. Nazer and M. Gastpar, "Lattice coding increases multicast rates for Gaussian multiple-access networks," in Proc. 45th Annu. Allerton Conf., 2007, pp. 1089-1096.
[25] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian sensor network," IEEE Trans. Inf. Theory, vol. 54, no. 11, pp. 5247-5251, Nov. 2008.
[26] N. Ma and P. Ishwar, "Some results on distributed source coding for interactive function computation," IEEE Trans. Inf. Theory, vol. 57, no. 9, pp. 6180-6195, Sep. 2011.
[27] R. Appuswamy, M. Franceschetti, N. Karamchandani, and K. Zeger, "Network coding for computing: Cut-set bounds," IEEE Trans. Inf. Theory, vol. 57, no. 2, pp. 1015-1030, Feb. 2011.
[28] H. Kowshik and P. R. Kumar, "Optimal function computation in directed and undirected graphs," IEEE Trans. Inf. Theory, vol. 58, no. 6, pp. 3407-3418, Jun. 2012.
[29] S. Kuzuoka and S. Watanabe, "A dichotomy of functions in distributed coding: An information spectral approach," IEEE Trans. Inf. Theory, vol. 61, no. 9, pp. 5028-5041, Sep. 2015.
[30] M. Sefidgaran, A. Gohari, and M. Aref, "On Körner-Marton's sum modulo two problem," in Proc. Iran Workshop Commun. Inf. Theory (IWCIT), May 2015, pp. 1-6.
[31] M. Sefidgaran and A. Tchamkerten, "Distributed function computation over a rooted directed tree," IEEE Trans. Inf. Theory, vol. 62, no. 12, pp. 7135-7152, Dec. 2016.
[32] S. Watanabe, "A classification of functions in multiterminal distributed computing," IEEE Trans. Inf. Theory, vol. 66, no. 10, pp. 6169-6183, Oct. 2020.
[33] S. Satpathy and P. Cuff, "Secure cascade channel synthesis," IEEE Trans. Inf. Theory, vol. 62, no. 11, pp. 6081-6094, Nov. 2016.
[34] B. N. Vellambi, J. Kliewer, and M. R. Bloch, "Strong coordination over multi-hop line networks using channel resolvability codebooks," IEEE Trans. Inf. Theory, vol. 64, no. 2, pp. 1132-1162, Feb. 2018.
[35] A. Bereyhi, M. Bahrami, M. Mirmohseni, and M. R. Aref, "Empirical coordination in a triangular multiterminal network," in Proc. IEEE Int. Symp. Inf. Theory, Jul. 2013, pp. 2149-2153.
[36] M. Mylonakis, P. A. Stavrou, and M. Skoglund, "Empirical coordination subject to a fidelity criterion," in Proc. IEEE Inf. Theory Workshop (ITW), Aug. 2019, pp. 1-5.
[37] F. Haddadpour, M. H. Yassaee, M. R. Aref, and A. Gohari, "When is it possible to simulate a DMC channel from another?" in Proc. IEEE Inf. Theory Workshop (ITW), Sep. 2013, pp. 1-5.
[38] F. M. J. Willems, "The discrete memoryless multiple access channel with partially cooperating encoders (Corresp.)," IEEE Trans. Inf. Theory, vol. 29, no. 3, pp. 441-445, May 1983.
[39] S. I. Bross, A. Lapidoth, and M. Wigger, "Dirty-paper coding for the Gaussian multiaccess channel with conferencing," IEEE Trans. Inf. Theory, vol. 58, no. 9, pp. 5640-5668, Sep. 2012.
[40] P. Noorzad, M. Effros, and M. Langberg, "The unbounded benefit of encoder cooperation for the $K$-user MAC," IEEE Trans. Inf. Theory, vol. 64, no. 5, pp. 3655-3678, May 2018.
[41] T. Anwar Atif, A. Padakandla, and S. Sandeep Pradhan, "Source coding for synthesizing correlated randomness," in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jun. 2020, pp. 1576-1581.
[42] D. Mayer et al., "Nonequilibrium thermodynamics and optimal cooling of a dilute atomic gas," 2019, arXiv:1901.06188.
[43] T. Anwar Atif, A. Padakandla, and S. Sandeep Pradhan, "Source coding for synthesizing correlated randomness," 2020, arXiv:2004.03651.
[44] D. Krithivasan and S. S. Pradhan, "Distributed source coding using abelian group codes: A new achievable rate-distortion region," IEEE Trans. Inf. Theory, vol. 57, no. 3, pp. 1495-1519, Mar. 2011.
[45] T. Anwar Atif and S. Sandeep Pradhan, "Distributed quantum faithful simulation and function computation using algebraic structured measurements," 2021, arXiv:2101.02360.
[46] S.-Y. Tung, "Multiterminal source coding," Ph.D. dissertation, School Elect. Comput. Eng., Cornell Univ., Ithaca, NY, USA, 1978.
[47] T. Berger, "Multiterminal source coding," Inf. Theory Approach Commun., vol. 229, pp. 171-231, Jul. 1977.
[48] G. R. Kurri, V. Ramachandran, S. R. B. Pillai, and V. M. Prabhakaran, "Multiple access channel simulation," in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jul. 2021, pp. 2411-2416.
[49] M. Gastpar, "The Wyner-Ziv problem with multiple sources," IEEE Trans. Inf. Theory, vol. 50, no. 11, pp. 2762-2768, Nov. 2004.
[50] M. Sefidgaran and A. Tchamkerten, "Computing a function of correlated sources: A rate region," in Proc. IEEE Int. Symp. Inf. Theory, Jul. 2011, pp. 1856-1860.
[51] M. Sefidgaran and A. Tchamkerten, "On computing a function of correlated sources," 2011, arXiv:1107.5806.
[52] P. Gacs and J. Körner, "Common information is far less than mutual information," Problems Control Inf. Theory, vol. 2, no. 2, pp. 149-162, 1973.
[53] M. H. Yassaee, M. R. Aref, and A. Gohari, "Achievability proof via output statistics of random binning," IEEE Trans. Inf. Theory, vol. 60, no. 11, pp. 6760-6786, Nov. 2014.
[54] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," IEEE Trans. Inf. Theory, vol. 19, no. 4, pp. 471-480, Jul. 1973.
[55] A. El Gamal and Y.-H. Kim, Network Information Theory. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[56] A. A. Gohari and V. Anantharam, "Evaluation of Marton's inner bound for the general broadcast channel," IEEE Trans. Inf. Theory, vol. 58, no. 2, pp. 608-619, Feb. 2012.

Gowtham R. Kurri (Member, IEEE) received the B.Tech. degree in electronics and communication engineering from the International Institute of Information Technology, Hyderabad, India, in 2011, and the M.Sc. and Ph.D. degrees from the Tata Institute of Fundamental Research, Mumbai, India, in 2020. From 2011 to 2012, he worked as an Associate Engineer at Qualcomm India Private Ltd., Hyderabad, India. From July to October, 2019, he was a Research Intern at the Blockchain Technology Group, IBM Research, Bengaluru, India. Since 2020, he has been a Post-Doctoral Researcher with the School of Electrical, Computer and Energy Engineering, Arizona State University. His research interests are in information theory and statistical machine learning.

Viswanathan Ramachandran was born in Kerala, India. He received the Ph.D. degree in electrical engineering from the Indian Institute of Technology Bombay in 2020. He worked as a Visiting Researcher at the Tata Institute of Fundamental Research in 2020. Since 2021, he has been a Post-Doctoral Fellow with the Department of Electrical Engineering, Technical University of Eindhoven, The Netherlands. His research interests are in information theory and multi-user information theory with applications to wireless and optical channels. He was a recipient of the Naik and Rastogi Award for excellence in Ph.D. research from the Indian Institute of Technology Bombay in 2021 and the Best Paper Award from the 25th National Conference on Communications (NCC), Indian Institute of Science, Bengaluru, in 2019.

Sibi Raj B. Pillai (Member, IEEE) received the Ph.D. degree in computer science and communication systems from Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, in July 2007. From October 2007 to April 2009, he was a Research Fellow with the University of Melbourne. Since 2009, he has been a Faculty Member with the Department of Electrical Engineering, Indian Institute of Technology Bombay. His research interests are in network information theory, feedback communications, cross layer scheduling, biological information inheritance, and radar signal processing.

Vinod M. Prabhakaran (Member, IEEE) received the M.E. degree from the Indian Institute of Science in 2001 and the Ph.D. degree from the University of California at Berkeley, Berkeley, in 2007. He was a Post-Doctoral Researcher at the University of Illinois at Urbana-Champaign, from 2008 to 2010 and the Ecole Polytechnique Fédérale de Lausanne, Switzerland, in 2011. Since 2011, he has been with the School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai. His research interests are in information theory, communication, cryptography, and signal processing. He was an Associate Editor for the Special Issue on Shannon Theory of the IEEE Transactions on Information Theory from 2016 to 2019. He is currently an Associate Editor for the Special Issue on Security and Privacy of the IEEE Transactions on Information Theory.


[^0]:    ${ }^{1}$ Referring to Shannon's channel coding theorem as the simulation of a noiseless channel using a noisy channel.

[^1]:    ${ }^{2}$ It turns out that shared randomness does not aid in deterministic function computation - see Remark 1.

[^2]:    ${ }^{3}$ It suffices for the alphabet $\mathcal{T}$ to have 3 elements to preserve $I\left(U_{1} ; X_{1} \mid Z, T\right), I\left(U_{2} ; X_{2} ; \mid Z, T\right)$, and $I\left(U_{1} ; X_{1}, Y \mid X_{2}, Z, T\right)$, thereby preserving the rate region.

