Hierarchical Identification with Pre-processing

Minh Thanh Vu, Tobias J. Oechtering *Senior Member, IEEE*, Mikael Skoglund *Senior Member, IEEE*

Abstract

We study a two-stage identification problem with pre-processing to enable efficient data retrieval and reconstruction. In the enrollment phase, users' data are stored into the database in two layers. In the identification phase an observer obtains an observation, which is originated from an unknown user in the enrolled database through a memoryless channel. The observation is sent for processing in two stages. In the first stage, the observation is pre-processed, and the result is then used in combination with the stored first layer information in the database to output a list of compatible users to the second stage. Then the second step uses the information of users contained in the list from both layers and the original observation sequence to return the exact user identity and a corresponding reconstruction sequence. The rate-distortion regions are characterized for both discrete and Gaussian scenarios. Specifically, for a fixed list size and distortion level, the compression-identification trade-off in the Gaussian scenario results in three different operating cases characterized by three auxiliary functions. While the choice of the auxiliary random variable for the first layer information is essentially unchanged when the identification rate is varied, the second one is selected based on the dominant function within those three. Due to the presence of a mixture of discrete and continuous random variables, the proof for the Gaussian case is highly non-trivial, which makes a careful measure theoretic analysis necessary. In addition, we study a connection of the previous setting to a two observer identification and a related problem with a lower bound for the list size, where the latter is motivated from privacy concerns.

Index Terms

Identification systems, list decoding, pre-processing, Gaussian distribution, rate-distortion trade-off.

I. Introduction

The blooming numbers of smart devices and services lead to an increase in high-dimensional contents such as videos or audios. Because of the large amount of data, efficient data storage and compression mechanisms are necessary. Given an observed sequence, e.g. an image, reliable identification of an related user inside a database is crucial in many image or video processing applications in eHealth, IoT, etc.. However, using high-dimensional observations directly puts a heavy toll on the system. We propose a pre-processing procedure, e.g. a letter-wise quantization, to reduce the search complexity.

The identification problem was first studied by Willems in [1], where he characterized the identification capacity for biometric systems. The compression and distortion aspects of users' data were taken into account in [2] and [3], where the trade-offs between compression and identification rates, and compression-distortion-capacity, respectively, were provided. In [4] the authors additionally considered compressing the observation and sending it to the processing center. Clustering was studied in [5], [6], and [7] as a method to improve the search speed, where in the enrollment phase users were distributed into clusters (groups) based on their data sequences. Each user could appear in several clusters.

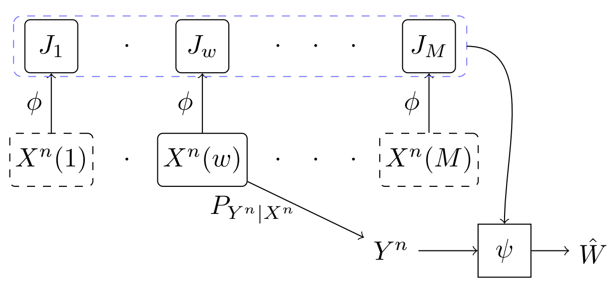


Fig. 1: A simplified model of identification systems studied in [1] and [2].

Generally speaking, The identification problem consists of two phases. In the first phase, *the enrollment phase*, data from M users $(x^n(i))_{i=1}^M$ are enrolled into a database as $(j_i)_{i=1}^M$, cf. Fig. 1, in a *compressed* or an *uncompressed* format. The users' data are not available after the enrollment phase. In the second phase, the *identification phase*, an observation y^n, which is related to $x^n(w)$ where w is chosen uniformly at random from M users, is available to the system. The task of the system is to identify the correct user w based on the observation y^n and the stored information in the database $(j_i)_{i=1}^M$. The identification capacity corresponds to the maximum number of users M such that the probability of correct identification approaches one.

In some areas, for example, in forensics or surveilance, one would like to identify the suspects as quickly as possible and view their criminal records. In these scenarios, we can also reduce the search complexity by first providing a list of possible

The authors are with the Department of Information Science and Engineering, KTH-Royal Institute of Technology, Stockholm, Sweden. This paper was presented in part at ISIT 2017 and at DCC 2018.

suspects and refining the search inside the given list. The records are stored in a second node which might be even a legal requirement, e.g. if only further details about suspicious people are stored.

Motivated by these examples, in this work we study an identification problem in which we assume only a single cluster of users and two storage nodes to store their data sequences in the system. The focus is, hence, to study the *capacity-list-compression-distortion* trade-off for the given cluster. In our setup each data sequence $x^n(i)$, which corresponds to the i-th user is compressed and stored in two layers, cf. Fig. 2. We note that a *user* is actually represented by a data entry, which does not have to be an actual person. The first layer stores some representative features of the sequence as in [8]. The second layer contains refinement information. This information layer helps to identify the user exactly and reconstruct the corresponding data sequence. This approach becomes interesting when querying information in the second layer is costly and therefore the system needs to limit the number of queries. An observation y^n is provided to the processing unit which needs to return the correct user's identity and its corresponding reconstruction sequence. To facilitate the processing time and power, the observation is first passed through a fixed channel $P_{Z|Y}$, which can be thought of as a feature extraction operation, or a fixed observation compression scheme. The processed observation is then compared with the information in the first layer, which results in a list of compatible users, $\mathcal{L}$. Then, the processing unit retrieves the information contained in both layers for all users in the list, $(j_{\mathcal{L}}, k_{\mathcal{L}})$. The retrieved information is then combined with observation y^n to identity the correct user. Finally, the system outputs the corresponding reconstruction sequence of the identified user.

We summarize our contribution as follows.

- Complete characterizations are provided for both discrete and Gaussian scenarios. The discrete case serves as the backbone for deriving the rate-distortion trade-off. The Gaussian setting not only provides an explicit illustration of the rate-distortion trade-off but also is interesting by itself for practical reasons. Specifically, we provide a complete achievability proof for the Gaussian setting that combines ideas in [9], [10] with an interesting tweak in the error analysis. In the converse direction, we estimate the distortion of the exact information aided by a genie and show that using the optimal estimator the estimated distortion level is also below the target distortion level. As a result, we are able to derive and simplify an outerbound using standard steps which we also provide some careful measure theoretic justifications for completeness of the arguments.
- In addition, from the identification process' perspective there is no difference between z^n obtained from the pre-processed procedure and z^n obtained from another weaker user in the sense of degradedness. Therefore, we extend our consideration in the discrete case to another related problem where two observers participate in the identification process. We provide a tight characterization in the scenario where reconstruction is not required.
- We study a spin-off problem where the processing unit needs to publish the list of users after the first processing stage. To guarantee some privacy, e.g. according to k-anonymity criterion[1], we require that the publishable list which contains the original list, must have a lower bound on its size. From this restriction, we see that our original setting literally provides some additional privacy guarantee at a small cost.

The paper is organized as follows. In Section II we consider the scenario where the users' data, the observation and the pre-processed information are discrete. Section III is devoted for the case where the users' data, observation, and pre-processed information are jointly Gaussian. The main proofs are provided at the end of each section. Additionally, some technical proofs are given in the Appendix for further justification.

II. The discrete identification problem

A. Notation

We begin by introducing some notations. Random variables, their realizations and alphabets are denoted by uppercase, lowercase and calligraphic letters, respectively, unless otherwise stated. In this section, we consider finite alphabets where the i-th user sequence $x^n(i)$ is generated iid from the probability distribution P_X on $\mathcal{X}$. The random reconstruction symbol and its alphabet are denoted by $\hat{X}$ and $\hat{\mathcal{X}}$, respectively. The letter-wise distortion measure is a bounded mapping of the form $d\colon \mathcal{X} \times \hat{\mathcal{X}} \to [0, d_{\max}]$. With abuse of notation, the sequence distortion measure is defined as

$$d(x^n, \hat{x}^n) = \frac{1}{n}\sum_{i=1}^{n} d(x_i, \hat{x}_i).$$

The strongly typical set is denoted by $\mathcal{T}_\epsilon^n$. For a set $\mathcal{A}$, $|\mathcal{A}|$ and $\mathcal{A}^c$ denote its cardinality and complement, respectively.

B. Formal Problem Formulation & Result

The (big) data $x^n(i) \in \mathcal{X}^n$, where[2] $i \in \mathcal{W} = [1:M]$ with $M = |\mathcal{W}|$, is compressed and stored hierarchically in two layers. The enrollment can be described by (possibly stochastic) mappings

$$\phi_{kn}\colon \mathcal{X}^n \to \mathcal{M}_k,\ k = 1, 2. \tag{1}$$

[1]Informally, k-anonymity is a property of a data table that for a given set of attributes each tuple of values appears at least k times in the table. For a formal definition, the reader is referred to [11].

[2]For $a \in \mathbb{Z}$ we use the shorthand notation $[1 : a]$ for the set $\{1, \dots, a\}$.

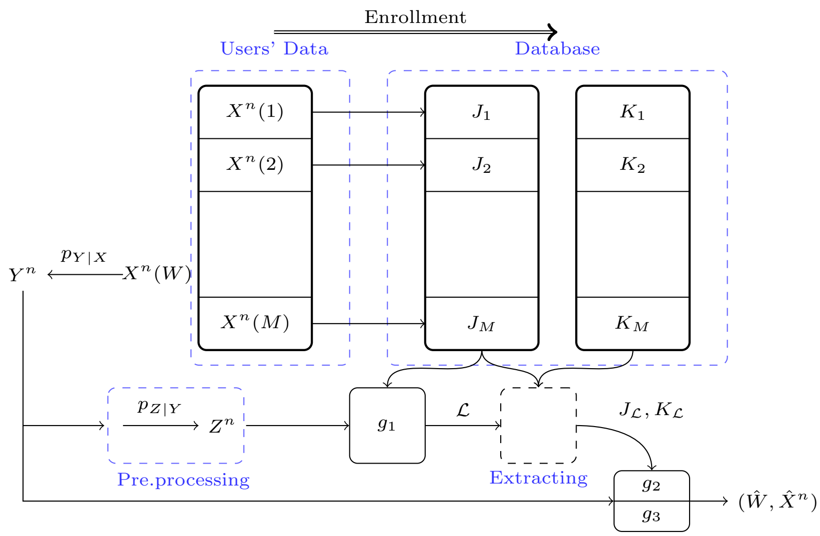


Fig. 2: An overview of the two stage identification system. We assume that there always exists a user W which has been enrolled previously and to which the observation Y^n is the output of a memoryless channel $P_{Y|X}$ with the input $X^n(W)$. Furthermore, W is uniformly distributed over $[1:M]$ and independent of users' data. The first and second layer information are represented by the collections $(J_i)_{i=1}^M$ and $(K_i)_{i=1}^M$, respectively.

We denote database indices $\phi_{1n}(x^n(i))$ and $\phi_{2n}(x^n(i))$ as $j_i \in \mathcal{M}_1$ and $k_i \in \mathcal{M}_2$ for all $i \in \mathcal{W}$.

An observer obtains information y^n about a user in the database from the output of the memoryless channel $P_{Y|X}$ with input $x^n(w)$, where w is an instance of a uniformly distributed random variable W over the set $\mathcal{W}$, which is independent of the users' data. The observer sends y^n to a processing unit to identify w and obtain a reconstruction $\hat{x}^n$ of $x^n(w)$ within the distortion D.

In the processing unit, the observation y^n is first pre-processed. The pre-processing is modeled by a *fixed* channel $P_{Z|Y}$ to produce a noisy version z^n, which can be linked to a quantization or a feature extraction process. Then, based on z^n and the first layer database $(j_i)_{i=1}^M$, a list $\mathcal{L} \in \mathfrak{L}$ of at most $2^{n\Delta}$ possible matching indices of a given size, is produced. This action can be described by a processing mapping

$$\begin{aligned} g_1 \colon \mathcal{Z}^n \times \mathcal{M}_1^M &\to \mathfrak{L}, \\ g_1(z^n, (j_i)_{i=1}^M) &\mapsto \mathcal{L}, \end{aligned} \tag{2}$$

where

$$\mathfrak{L} = \left\{ \mathcal{S} \middle| \mathcal{S} \subseteq \mathcal{W},\ |\mathcal{S}| \leq 2^{n\Delta} \right\} \cup \left\{ \{e\} \right\}$$

is the set of subsets of users in $\mathcal{W}$ with cardinality at most $2^{n\Delta}$ and the set $\{e\}$, which describes an error event. This means that we allow the mapping g_1 to declare an error. The *extracting* action, which takes as its inputs the index list $\mathcal{L}$ and the stored information of all users in both layers to return the information of all chosen users inside the list along with the list, can be described by a projection mapping

$$\begin{aligned} &\pi \colon \mathcal{M}_1^M \times \mathcal{M}_2^M \times \mathfrak{L} \to \mathfrak{M}_{12}, \\ &\pi((j_i)_{i=1}^M, (k_i)_{i=1}^M, \mathcal{L}) \mapsto \begin{cases} ((j_i)_{i\in\mathcal{L}}, (k_i)_{i\in\mathcal{L}}, \mathcal{L}) & \text{if } \mathcal{L} \neq \{e\} \\ (1, 1, \{e\}) & \text{otherwise} \end{cases}, \end{aligned} \tag{3}$$

where

$$\mathfrak{M}_{12} = \bigcup_{\mathcal{L}\neq\{e\}} \left\{ ((j_i)_{i\in\mathcal{L}}, (k_i)_{i\in\mathcal{L}}, \mathcal{L}) \middle| (j_i)_{i\in\mathcal{L}} \in \mathcal{M}_1^{|\mathcal{L}|},\ (k_i)_{i\in\mathcal{L}} \in \mathcal{M}_2^{|\mathcal{L}|},\ \mathcal{L} \in \mathfrak{L} \right\} \cup \{(1, 1, \{e\})\}.$$

It should be clear from the definition of $\mathfrak{M}_{12}$ that the vectors $(j_i)_{i\in\mathcal{L}}$ and $(k_i)_{i\in\mathcal{L}}$ can contain repeated elements. Therefore, the inclusion of $\mathcal{L}$ at the output of π helps to pinpoint which combination of users the output information belongs to[3]. For brevity, elements of $\mathfrak{M}_{12}$ are denoted by $(j_{\mathcal{L}}, k_{\mathcal{L}})$. In the second stage the processing unit returns an estimate of the index $\hat{w}$ which is the output of a deterministic processing mapping $g_2(\cdot)$ where

$$g_2 \colon \mathcal{Y}^n \times \mathfrak{M}_{12} \to \mathcal{W} \cup \{e\},$$

[3]The choice of $(1, 1)$ as the output information when $\mathcal{L} = \{e\}$ is inconsequential.

$$g_2(y^n,(j_{\mathcal{L}},k_{\mathcal{L}}))\mapsto \hat{w}, \tag{4}$$

i.e., g_2 can declare an error event as well. Furthermore, the processing unit needs to output a reconstruction sequence $\hat{x}^n$ of the data sequence $x^n(w)$. To describe the reconstruction processing mapping, first define a second projection mapping

$$\begin{aligned} \hat{\pi}\colon \mathfrak{M}_{12}\times\Big(\mathcal{W}\cup\{e\}\Big) &\to \hat{\mathfrak{M}}_{12}=\mathcal{M}_1\times\mathcal{M}_2\times\Big(\mathcal{W}\cup\{e\}\Big) \\ \hat{\pi}((j_{\mathcal{L}},k_{\mathcal{L}}),\hat{w}) &\mapsto \begin{cases}(j_{\hat{w}},k_{\hat{w}},\hat{w}) & \text{if } \ \hat{w}\in\mathcal{L}\neq\{e\} \\ (1,1,e) & \text{otherwise}\end{cases}. \end{aligned} \tag{5}$$

Similarly, we denote elements of $\hat{\mathfrak{M}}_{12}$ as $(j_{\hat{w}},k_{\hat{w}})$, then the reconstruction mapping is given by

$$\begin{aligned} g_3\colon \mathcal{Y}^n\times\hat{\mathfrak{M}}_{12} &\to \hat{\mathcal{X}}^n \\ g_3(y^n,(j_{\hat{w}},k_{\hat{w}})) &\mapsto \hat{x}^n. \end{aligned} \tag{6}$$

The two projection mappings π and $\hat{\pi}$ are inherent, hence need not to be designed explicitly.

Definition 1. *For a given pre-processing channel $P_{Z|Y}$, an identification scheme of length n consists of two enrollment mappings $\{\phi_{kn}\}_{k=1}^{2}$ and three processing mappings $\{g_k\}_{k=1}^{3}$.*

Definition 2. *For a given pre-processing scheme $P_{Z|Y}$, a rate-distortion tuple (R,R_1,R_2,R_L,D) is achievable if for every $\epsilon>0$, there exists an identification scheme of length n such that*

$$\begin{gathered} \frac{1}{n}\log M > R-\epsilon, \quad \frac{1}{n}\log|\mathcal{M}_1| < R_1+\epsilon \\ \frac{1}{n}\log|\mathcal{M}_2| < R_2+\epsilon, \quad \Delta < R_L+\epsilon, \quad \Pr(W\notin\boldsymbol{\mathcal{L}})<\epsilon, \\ \Pr(W\neq\hat{W})<\epsilon, \quad \mathbb{E}[d(X^n(W),\hat{X}^n)] < D+\epsilon, \end{gathered} \tag{7}$$

for all sufficiently large n. The set of all achievable tuples is denoted by $\mathcal{R}$.

Note that given $\Pr\{W\neq\hat{W}\}$ in the finite alphabet case our constraint $\mathbb{E}[d(X^n(W),\hat{X}^n)] < D+\epsilon$ is equivalent to the constraint $\mathbb{E}[d(X^n(W),\hat{X}^n)|\hat{W}=W]\leq D+\epsilon$, which is considered in [3], since the distortion measure is bounded.

Definition 3. *Let $\mathcal{R}^\star$ be the collection of tuples (R,R_1,R_2,R_L,D) such that there exist random variables U and V defined on finite alphabets $\mathcal{U}$ and $\mathcal{V}$ which satisfy*

$$|\mathcal{U}|\leq|\mathcal{X}|+5, |\mathcal{V}|\leq(|\mathcal{X}|+5)(|\mathcal{X}|+2) \tag{8}$$

and a deterministic reconstruction mapping $f\colon\mathcal{U}\times\mathcal{V}\times\mathcal{Y}\to\hat{\mathcal{X}}$ such that the followings expressions are fulfilled:

$$U-V-X-Y-Z,$$

$$R_1\geq I(X;U), \tag{9a}$$

$$R_1+R_2\geq I(X;U)+I(X;V|U,Y), \tag{9b}$$

$$R_1+R_2-R\geq I(X;U,V|Y), \tag{9c}$$

$$R\leq\min\{R_L+I(Z;U),I(Y;U,V)\}, \tag{9d}$$

$$D\geq\mathbb{E}[d(X,f(U,V,Y))]. \tag{9e}$$

The above definitions imply that both $\mathcal{R}$ and $\mathcal{R}^\star$ are closed subsets of $\mathbb{R}^5$ w.r.t. ℓ_1 metric. Furthermore, $\mathcal{R}^\star$ is not empty since it contains $(0,0,0,0,d_{\max})$. We state our first result in the following theorem.

Theorem 1. *For a given pre-processing strategy $P_{Z|Y}$, memoryless data source P_X, and observation model $P_{Y|X}$, the rate-distortion region for our setting is given by*

$$\mathcal{R}=\mathcal{R}^\star. \tag{10}$$

The proof of Theorem 1 is given in Subsection II-D.

Remark 1. *For a given choice of auxiliary random variables U, V such that the distortion constraint* (9e) *is fulfilled, the first inequality* (9a) *shows the minimum compression rate for the first layer. The second inequality* (9b) *indicates the trade-off between total compression rate in both layers. One notices that the second term on the right-hand side of* (9b) *suggest the use of binning for the stored data on the second storage node.* (9c) *shows the trade-off between the total compression rate and the identification rate. Namely, the identification rate is strictly smaller than the total compression rate if the right-hand* *side of* (9c) *is positive. Lastly, the first term on the right-hand side of* (9d) *is the maximum identification rate resulting from the first layer information and pre-processed information. The second term in* (9d) *is the maximum identification rate when the identification process is performed jointly, i.e., with the original observation and information from both layers.*

Remark 2. *In the special case where $R_L = R$ we notice that U can be set to a deterministic value, e.g. $U = \varnothing$, so that the rate-distortion region* (9) *reduces to the one given in [3, Theorem 1].*

C. Related problems

1) The identification problem: When the distortion level $D = d_{\max}$, i.e., the distortion constraint can be removed, then binning for the second layer, described by V, is not necessary. We obtain the following corollary.

Corollary 1. *For a fixed $P_{Z|Y}$, the rate region of our identification setting, i.e., $D = d_{\max}$, is given by the set of tuples (R, R_1, R_2, R_L) such that*

$$\begin{aligned} &U - V - X - Y - Z, \\ &R_1 \geq I(X;U), \quad R_1 + R_2 \geq I(X;U,V), \\ &R \leq \min\{R_L + I(Z;U), I(Y;U,V)\}, \end{aligned} \tag{11}$$

where U and V are random variables taking values on alphabets $\mathcal{U}$ and $\mathcal{V}$, respectively, with $|\mathcal{U}| \leq |\mathcal{X}| + 4$ and $|\mathcal{V}| \leq (|\mathcal{X}| + 4)(|\mathcal{X}| + 1)$.

The proof of Corollary 1 is given in Appendix C.

2) A two observer problem: A related problem is stated in the following. The data sequence $x^n(w)$ is observed through the channel $P_{ZY|X}$ by two Observers 1 and 2, which obtain y^n and z^n, respectively. Moreover, Observer 2 has only access to the information stored in the first layer and is interested in obtaining a list of users in the database only, for instance due to complexity or due to privilege restriction. Accordingly, the decoding mapping and the requirement for the second observer are given by

$$\mathcal{L} = g_2(z^n, (j_i)_{i=1}^M), \text{ and } \Pr(W \notin \boldsymbol{\mathcal{L}}) < \epsilon, \tag{12}$$

where $|\mathcal{L}| \leq 2^{n\Delta}$. Observer 2 in the current setting corresponds to the first processing stage in the previous settings. In contrast, Observer 1 has access to both layers and wants to identify the user correctly, i.e., the decoding mapping and the requirement of the first observer are

$$\hat{w} = g_1(y^n, (j_i)_{i=1}^M, (k_i)_{i=1}^M), \text{ and } \Pr(W \neq \hat{W}) < \epsilon. \tag{13}$$

In other words, the identification processes for two observers are separated. Note that there is no reconstruction requirement in the current problem. The rate region for this problem can be described by the following proposition.

Proposition 1. *The optimal rate region for the stated problem is the set of tuples (R, R_1, R_2, R_L) such that*

$$\begin{aligned} &U - V - X - (Y, Z), \\ &R_1 \geq I(X;U), \quad R_1 + R_2 \geq I(X;U,V), \\ &R \leq \min\{R_L + I(Z;U), I(Y;U,V)\}, \end{aligned} \tag{14}$$

where U and V are random variables taking values on finite alphabets $\mathcal{U}$ and $\mathcal{V}$, respectively, with $|\mathcal{U}| \leq |\mathcal{X}| + 4$ and $|\mathcal{V}| \leq (|\mathcal{X}| + 4)(|\mathcal{X}| + 1)$.

Note that the Markov condition $X - Y - Z$ is not needed since the two identification processes work independently. This means that our original problem can be viewed as a sequential cooperation scheme between two identification processes. The proof of Proposition 1 is given in Appendix D.

3) A lower bound on the list size: We consider the following related setting. The original setting is considered, namely the identification scheme involves $\{\phi_{kn}\}_{k=1}^2$ and $\{g_k\}_{k=1}^3$. Additionally, the processing unit needs to release a list $\hat{\mathcal{L}}$, which contains the original list $\mathcal{L}$ when $\mathcal{L} \neq \{e\}$, i.e., $\mathcal{L} \subseteq \hat{\mathcal{L}}$, such that

$$2^{n\Delta'} \leq |\hat{\mathcal{L}}| \leq 2^{n\Delta}, \tag{15}$$

which means that the list $\hat{\mathcal{L}}$ is always bounded. This restriction is motivated from the privacy criterion such as k-anonymity when the processing unit needs to release a list of possible users. For this problem we have the following result.

Corollary 2. *The trade-off among (R, R_1, R_2, R_L, D) is the same as in* (9)*, while additionally $R_L \geq \Delta'$ due to the presence of the constraint* (15).

Sketch of Proof: The achievability part follows the one given in Theorem 1 with a modification. The processing unit publishes the list $\hat{\mathcal{L}}$ if it fulfills the two-side constraint (15). As before we denote $\mathcal{L}_1$ the list of suitable indices resulting from the first

stage processing, i.e., satisfying the condition (107). If $|\mathcal{L}_1| \leq 2^{n\Delta'}$ then we randomly pick up $\lceil 2^{n\Delta'} \rceil - |\mathcal{L}_1| + 1$ indices from $\mathcal{W} \backslash \mathcal{L}_1$ and append them to $\mathcal{L}_1$ before returning as $\hat{\mathcal{L}}$. If the list of suitable indices violates the upper bound constraint, i.e., $|\mathcal{L}_1| > 2^{n\Delta}$, then the processing unit publishes the list $\hat{\mathcal{L}}$ of the first $\lceil 2^{n\Delta'} \rceil + 1$ indices from $\mathcal{L}_1$.
The converse follows immediately since the constraint (15) is more restrictive than $|\boldsymbol{\mathcal{L}}| \leq 2^{n\Delta}$. ■

D. Proof of Theorem 1

The achievability proof is standard and presented in Appendix B. We provide herein the converse proof, which is relevant for the proof of Theorem 3.

1) Cardinality bounding of $\mathcal{U}$ and $\mathcal{V}$: It is sufficient to preserve the following quantities $H(X|U)$, $H(X|U,Y)$, $H(X|U,V,Y)$, $H(Z|U)$, $H(Y|U,V)$, the distortion constraint, and $p(x)$ for all but one $x \in \mathcal{X}$. By the support lemma [12, Appendix C] the cardinality of $\mathcal{U}$ and $\mathcal{V}$ can be bounded by

$$\begin{aligned} &|\mathcal{U}| \leq |\mathcal{X}| + 5, \\ &|\mathcal{V}| \leq (|\mathcal{X}| + 5)(|\mathcal{X}| + 2). \end{aligned} \tag{16}$$

This implies that $\mathcal{R}^\star$ is a closed region.

2) Converse: Given $\epsilon > 0$ small enough, assume that there exist mappings such that all the conditions are fulfilled for *all* sufficiently large n. Furthermore by taking n large enough, the condition $\frac{1}{n} < \epsilon$ is valid. For notation brevity we abbreviate $(J_i)_{i=1}^M$ as $\boldsymbol{J}$ and $(K_i)_{i=1}^M$ as $\boldsymbol{K}$. The corresponding realizations are denoted by $\boldsymbol{j}$, and $\boldsymbol{k}$. Since $\Pr(\hat{W} \neq W) < \epsilon$, Fano's inequality for the second stage implies

$$\begin{aligned} H(W|Y^n, (J_{\boldsymbol{\mathcal{L}}}, K_{\boldsymbol{\mathcal{L}}})) &< 1 + \Pr(\hat{W} \neq W) \log_2 M \\ &< 1 + \epsilon \log_2 M. \end{aligned} \tag{17}$$

We also establish a variant of Fano's inequality for the first stage. Define

$$E = \chi_{\left\{W \in g_1(Z^n, \boldsymbol{J})\right\}}, \tag{18}$$

where χ_B is the indicator function of the set B. Since W is in the list when $E = 1$, the error probability $P_e = \Pr(E = 0)$ is bounded by ϵ. We obtain the following inequality

$$\begin{aligned} H(W|Z^n, \boldsymbol{J}) &\leq h_b(P_e) + P_e \log_2 M + n(R_L + \epsilon) \\ &= n(R_L + \epsilon + \frac{1}{n}(h_b(P_e) + P_e \log_2 M)) \leq n(R_L + \epsilon_n), \end{aligned} \tag{19}$$

where $\epsilon_n = 2\epsilon + \frac{1}{n}\epsilon \log_2 M$ and $h_b(\cdot)$ is the binary entropy function. The detailed derivation is given in Appendix H-B, more specifically in (249). Define random variables

$$\begin{aligned} U_i &= (W, J_W, Y^{i-1}), \\ V_i &= (U_i, K_W, Y_{i+1}^n), \quad i \in [1:n]. \end{aligned} \tag{20}$$

Observe that $U_i - V_i - X_i(W) - Y_i - Z_i$ for all $i \in [1:n]$, due to the memoryless property of the observation and pre-processing channels and the source. The identification rate can be bounded firstly as

$$\begin{aligned} n(R - \epsilon) &\leq \log_2 M = H(W) \\ &= I(W; Z^n, \boldsymbol{J}) + H(W|Z^n, \boldsymbol{J}) \\ &\overset{(\star)}{\leq} I(W; Z^n|\boldsymbol{J}) + n(R_L + \epsilon_n) \\ &\leq I(W, \boldsymbol{J}; Z^n) + n(R_L + \epsilon_n) \\ &= I(W, J_W; Z^n) + n(R_L + \epsilon_n) \\ &= \sum_{i=1}^n I(W, J_W, Z^{i-1}; Z_i) + n(R_L + \epsilon_n), \\ &\overset{(a)}{\leq} \sum_{i=1}^n I(W, J_W, Y^{i-1}; Z_i) + n(R_L + \epsilon_n), \end{aligned} \tag{21}$$

where $(\star)$ holds due to (19) and since W is independent of $\boldsymbol{J}$. (a) holds due to the Markov chain $Z^{i-1} - Y^{i-1} - (Z_i, W, J_W)$ for all $i \in [1:n]$, due to the memoryless property of the pre-processing. This implies that

$$(R - \epsilon)(1 - \epsilon) \leq \frac{1}{n} \sum_{i=1}^n I(U_i; Z_i) + R_L + 2\epsilon. \tag{22}$$

Secondly,

$$
\begin{aligned}
n(R-\epsilon) &\le \log_2 M = H(W) \\
&= I(W; Y^n, (J_{\boldsymbol{\mathcal{L}}}, K_{\boldsymbol{\mathcal{L}}})) + H(W|Y^n, (J_{\boldsymbol{\mathcal{L}}}, K_{\boldsymbol{\mathcal{L}}})) \\
&\overset{(b)}{\le} I(W; Y^n, Z^n, \boldsymbol{J}, \boldsymbol{K}) + 1 + \epsilon \log_2 M \\
&\overset{(c)}{=} I(W; Y^n, \boldsymbol{J}, \boldsymbol{K}) + 1 + \epsilon \log_2 M \\
&\overset{(\star\star)}{\le} I(W, \boldsymbol{J}, \boldsymbol{K}; Y^n) + 1 + \epsilon \log_2 M \\
&= I(W, J_W, K_W; Y^n) + 1 + \epsilon \log_2 M \\
&\le \sum_{i=1}^{n} I(W, J_W, K_W, Y^{n\setminus i}; Y_i) + 1 + \epsilon \log_2 M,
\end{aligned} \tag{23}
$$

where (b) holds since by eqs. (2) and (3), $\boldsymbol{\mathcal{L}} = g_1(Z^n, \boldsymbol{J})$, and

$$
(J_{\boldsymbol{\mathcal{L}}}, K_{\boldsymbol{\mathcal{L}}}) = ((J_i)_{i\in\boldsymbol{\mathcal{L}}}, (K_i)_{i\in\boldsymbol{\mathcal{L}}}, \boldsymbol{\mathcal{L}}) = \pi(\boldsymbol{J}, \boldsymbol{K}, \boldsymbol{\mathcal{L}}),
$$

hold. We also use the inequality (17) in (b). (c) is valid due to the Markov chain $Z^n - Y^n - (W, \boldsymbol{J}, \boldsymbol{K})$. $(\star\star)$ holds since W is independent of both $\boldsymbol{J}$ and $\boldsymbol{K}$. Using (20) and (23) gives us

$$
(R-\epsilon)(1-\epsilon) \le \frac{1}{n}\sum_{i=1}^{n} I(U_i, V_i; Y_i) + \epsilon. \tag{24}
$$

Furthermore, the sum of the compressed rates can be bounded as

$$
\begin{aligned}
n(R_1 + R_2 + \epsilon) &\ge H(J_W, K_W|W) \\
&\ge I(X^n(W), Y^n; J_W, K_W|W) \\
&\ge I(Y^n; J_W|W) + I(X^n(W); J_W, K_W|W, Y^n) \\
&\overset{(d)}{=} I(Y^n; J_W, W) + I(X^n(W); J_W, K_W, W|Y^n) \\
&\overset{(e)}{=} \sum_{i=1}^{n} \Big(I(Y_i; W, J_W, Y^{i-1}) + I(X_i(W); W, J_W, K_W, Y^{n\setminus i}, X^{i-1}(W)|Y_i) \Big) \\
&\overset{(f)}{\ge} \sum_{i=1}^{n} \Big(I(X_i(W), Y_i; W, J_W, Y^{i-1}) + I(X_i(W); K_W, Y_{i+1}^n|Y_i, W, J_W, Y^{i-1}) \Big) \\
&\overset{(g)}{=} \sum_{i=1}^{n} I(X_i(W); U_i) + I(X_i(W); V_i|U_i, Y_i),
\end{aligned} \tag{25}
$$

where (d) is valid since W is independent of both $X^n(W)$ and Y^n. (e) is true due to the memoryless property of the observational channel. (f) holds since we drop the term $X^{i-1}(W)$ in the second term. (g) follows from the Markov chain $Y_i - X_i(W) - (W, J_W, Y^{i-1})$ for all $i \in [1:n]$. Similarly, we can show that

$$
\begin{aligned}
n(R_1 + \epsilon) &\ge H(J_W|W) \\
&\ge I(X^n(W); J_W|W) = I(X^n(W); J_W, W) \\
&= \sum_{i=1}^{n} I(X_i(W); W, J_W, X^{i-1}(W)) \\
&\ge \sum_{i=1}^{n} I(X_i(W); U_i).
\end{aligned} \tag{26}
$$

In addition, using the first line of (25) and the second last line in (23) we obtain

$$
\begin{aligned}
& n(R_1 + R_2 + \epsilon) - \log_2 M \\
&\ge H(J_W, K_W|W) - I(W, J_W, K_W; Y^n) - (1 + \epsilon \log_2 M) \\
&\ge I(X^n(W); J_W, K_W, W) - I(Y^n; J_W, K_W, W) - (1 + \epsilon \log_2 M) \\
&\overset{(*)}{=} I(X^n(W); J_W, K_W, W|Y^n) - (1 + \epsilon \log_2 M) \\
&= \sum_{i=1}^{n} I(X_i(W); J_W, K_W, W|Y^n, X^{i-1}(W)) - (1 + \epsilon \log_2 M)
\end{aligned}
$$

$$\stackrel{(e)}{=} \sum_{i=1}^{n} I(X_i(W); J_W, K_W, W, Y^{n\setminus i}, X^{i-1}(W)|Y_i) - (1+\epsilon\log_2 M)$$
$$\geq \sum_{i=1}^{n} I(X_i(W); U_i, V_i|Y_i) - (1+\epsilon\log_2 M), \tag{27}$$

where $(*)$ holds due to the memoryless property of the observation channel, i.e., $Y^n - X^n(W) - (W, J_W, K_W)$ and (e) holds as before. This leads to

$$R_1 + R_2 + 2\epsilon - (R-\epsilon)(1-\epsilon) \geq \frac{1}{n}\sum_{i=1}^{n} I(X_i(W); U_i, V_i|Y_i). \tag{28}$$

Since $(J_{\hat{W}}, K_{\hat{W}})$ is an abbreviation of $(J_{\hat{W}}, K_{\hat{W}}, \hat{W})$ and

$$D + \epsilon > \mathbb{E}[d(X^n(W), g_3((J_{\hat{W}}, K_{\hat{W}}), Y^n))] = \mathbb{E}[d(X^n(W), g_3(J_{\hat{W}}, K_{\hat{W}}, \hat{W}, Y^n))]$$
$$> \Pr(\hat{W} = W) \times \mathbb{E}[d(X^n(W), g_3(J_W, K_W, W, Y^n))|\hat{W} = W], \tag{29}$$

the following chain of expressions holds

$$\mathbb{E}[d(X^n(W), g_3(W, J_W, K_W, Y^n))] \leq \mathbb{E}[d(X^n(W), g_3(J_W, K_W, W, Y^n))|\hat{W} = W]\Pr(\hat{W} = W)$$
$$+ \Pr(\hat{W} \neq W) d_{\max} < D + (1 + d_{\max})\epsilon. \tag{30}$$

Let Q be a random variable uniformly distributed on $[1:n]$ and independent of everything else. Define

$$U = (U_Q, Q), \quad V = (V_Q, Q), \tag{31}$$

and

$$f(U, V, Y_Q) = g_{3Q}(J_W, K_W, W, Y^n). \tag{32}$$

Note that $U - V - X_Q(W) - Y_Q - Z_Q$ still holds. Then the above constraints can be rewritten as

$$\begin{aligned}
(R-\epsilon)(1-\epsilon) &\leq I(U_Q; Z_Q|Q) + R_L + 2\epsilon = I(U; Z_Q) + R_L + 2\epsilon \\
(R-\epsilon)(1-\epsilon) &\leq I(U, V; Y_Q) + \epsilon \\
R_1 + R_2 + \epsilon &\geq I(X_Q(W); U) + I(X_Q(W); V|U, Y_Q) \\
R_1 + \epsilon &\geq I(X_Q(W); U) \\
R_1 + R_2 + 2\epsilon - (R-\epsilon)(1-\epsilon) &\geq I(X_Q(W); U, V|Y_Q) \\
D + (1 + d_{\max})\epsilon &> \mathbb{E}[d(X_Q(W), f(U, V, Y_Q))].
\end{aligned} \tag{33}$$

Since $(X_Q(W), Y_Q, Z_Q)$ has the joint distribution as $P_{XY} \times P_{Z|Y}$,

$$\Big((R-\epsilon)(1-\epsilon) - \epsilon, R_1 + \epsilon, R_2 + \epsilon, R_L + \epsilon, D + (1 + d_{\max})\epsilon\Big) \in \mathcal{R}^{\star}$$

by the cardinality bounding arguments presented in Subsection II-D1. Taking $\epsilon \to 0$ completes the backward direction since $\mathcal{R}^{\star}$ is closed. ■

III. The Gaussian Identification Problem

In this section we consider the setup where the users' data are Gaussian distributed, i.e., $X_i(w) \sim \mathcal{N}(0, \sigma_X^2)$, $\forall i, w$, and

$$Y_i = X_i(W) + N_{1i},\ Z_i = Y_i + N_{2i}$$

where $N_{1i} \sim \mathcal{N}(0, \sigma_{N_1}^2)$ and $N_{2i} \sim \mathcal{N}(0, \sigma_{N_2}^2)$ are iid random variables, which are also independent of the users' data and each other. In other words, the observation and pre-processing channels are iid Gaussian. The reconstruction set is the set of real numbers, i.e., $\hat{\mathcal{X}} = \mathbb{R}$. The distortion measure is the squared error distance

$$d(x^n, \hat{x}^n) = \frac{1}{n}\|x^n - \hat{x}^n\|_2^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2. \tag{34}$$

The definition of an identification scheme and achievability follows similarly as the ones given in Definitions 1 and 2 in which the processing mappings $\{g_i\}_{i=1}^3$ are measurable[4]. The enrollment mappings $\{\phi_{in}\}_{i=1}^2$ given by

$$\phi_{in} \colon \mathbb{R}^n \to \mathcal{M}_i,\ i = 1, 2,$$

[4]To be more accurate, they are of the form $f\colon (A_1, \mathcal{A}_1) \to (A_2, \mathcal{A}_2)$ where $\{(A_i, \mathcal{A}_i)\}_{i=1}^2$ are measurable spaces with $\mathcal{A}_i$ being the corresponding Borel σ-algebra. The Borel σ-algebra of $\mathbb{R}$ equipped with the Euclidean distance is $\mathcal{B}(\mathbb{R})$, while the Borel σ-algebra of a discrete set A equipped with the discrete metric is its power set 2^A. If a mapping takes multiple arguments as input or output, in which each argument's range can be either discrete or $\mathbb{R}$, then the corresponding (Borel) σ-algebra is the product of the (Borel) σ-algebra of each individual argument. This particular assumption is a consequence of [13, Lemma 1.2] which states that "for countable products of separable metric spaces, the product and Borel σ-fields agree".

are also measurable. Let us denote $(\Omega, \mathcal{A}, \mathbb{P})$ the underlying probability space. In the Appendix E we provide a proof of the following observation.

Theorem 2. *Let (R, R_1, R_2, R_L, D) be a rate-distortion tuple such that there exist random variables U and V with a joint conditional probability density*[5] *$p_{UV|X}$ and a measurable reconstruction mapping $g\colon \mathbb{R}^3 \to \mathbb{R}$ such that the following conditions are fulfilled.*

$$R_1 \geq I(X;U), \tag{35a}$$
$$R_1 + R_2 \geq I(X;U) + I(X;V|U,Y), \tag{35b}$$
$$R_1 + R_2 - R \geq I(X;U,V|Y), \tag{35c}$$
$$R \leq \min\{R_L + I(Z;U), I(Y;U,V)\}, \tag{35d}$$
$$D \geq \mathbb{E}[d(X, g(U,V,Y))]. \tag{35e}$$

Then (R, R_1, R_2, R_L, D) is achievable in the sense of Definition 2.

It will be clear from Appendix E that our proof for Theorem 2 can be transferred directly to the discrete case as the pmfs in the discrete case can be viewed as density functions w.r.t. the counting measure. Due to the formal analytical complexity of the Gaussian case, where we have a mixture of discrete and continuous random variables, we choose to present its proof separately for the sake of clarity. Theorem 2 allows us to derive the rate-distortion region for the Gaussian setting, denoted by $\mathcal{R}_{GS}$, which is given by the following theorem.

Theorem 3. *Assume that $0 \leq R_L \leq R$ and $0 < D \leq \frac{\sigma_X^2 \sigma_{N_1}^2}{\sigma_Y^2}$. Then the corresponding rate-distortion region $\mathcal{R}_{GS}$ is given by*

$$R < R_\gamma, \tag{36a}$$
$$R_1 \geq \frac{1}{2}\log_2\left(\frac{\sigma_X^2}{\sigma_Z^2 2^{-2(R-R_L)} - (\sigma_{N_1}^2 + \sigma_{N_2}^2)}\right), \tag{36b}$$
$$R_1 + R_2 \geq \frac{1}{2}\log_2 \frac{\sigma_Y^2}{\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2} + \Gamma, \tag{36c}$$
$$R_1 + R_2 - R \geq \Gamma, \tag{36d}$$

where

$$R_\gamma = \min\left\{\frac{1}{2}\log_2\left(\frac{\sigma_Z^2}{\sigma_{N_1}^2 + \sigma_{N_2}^2}\right) + R_L, \frac{1}{2}\log_2\left(\frac{\sigma_Y^2}{\sigma_{N_1}^2}\right)\right\},$$
$$\Gamma = \frac{1}{2}\max\left\{\log_2 \frac{\sigma_X^2\sigma_{N_1}^2}{\sigma_Y^2 D}, \log_2 \frac{\sigma_X^2 2^{-2R}}{\sigma_Y^2 2^{-2R} - \sigma_{N_1}^2}, \log_2 \frac{\sigma_X^2}{\sigma_Y^2}\frac{\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2}{\sigma_Z^2 2^{-2(R-R_L)} - (\sigma_{N_1}^2 + \sigma_{N_2}^2)}\right\}. \tag{37}$$

Remark 3. *The constraint* (36a) *corresponds to the constraint* (35d) *where R_γ can be seen as the supermum of the right-hand side of* (35d) *w.r.t any pair of auxiliary random variables U and V such that $YZ - X - UV$ holds and the mutual information terms are well-defined.*
The constraint $R_L \leq R$ is motivated from the fact that the first layer cannot reasonably output a list with size larger than the number of users in the system. As for the second restriction on the distortion on D, if we consider for each $i \in [1:n]$ estimating $X_i(W)$ using Y_i and the MMSE estimator, then, the distortion level is exactly $\frac{\sigma_X^2\sigma_{N_1}^2}{\sigma_Y^2}$. With additional information, the system in general can do better than this bound. If for some unknown reason, the target list size or the target distortion level is set above the corresponding thresholds, then the corresponding terms, related to D or R_L, in (36) *are omitted. For instance, the rate-distortion trade-off when $R_L > R$ and $0 < D \leq \sigma_X^2\sigma_{N_1}^2/\sigma_Y^2$ is given by*

$$R_1 + R_2 \geq R + \frac{1}{2}\max\left\{\log_2 \frac{\sigma_X^2\sigma_{N_1}^2}{\sigma_Y^2 D}, \log_2 \frac{\sigma_X^2 2^{-2R}}{\sigma_Y^2 2^{-2R} - \sigma_{N_1}^2}\right\}$$
$$0 \leq R < \frac{1}{2}\log_2\left(\frac{\sigma_Y^2}{\sigma_{N_1}^2}\right). \tag{38}$$

By definition the rate-distortion region $\mathcal{R}$ is closed in the finite dimensional metric space induced by the ℓ_1 distance. However, the constraint $R < R_L$ and $D > 0$ may lead to the impression that the region is not necessary closed. In Appendix F we show that $\mathcal{R}_{GS}$ is indeed closed.

The proof of Theorem 3 is divided into the following parts. We first establish an outer bound on the achievable rate-distortion region. Then, we discuss how to resolve the complicated outer bound into small subregions that can be achieved by different parameterized coding schemes. Finally, we show that each region can be achieved, hence implying that the complete outer

[5]The ranges of U and V are $\mathbb{R}$ and the joint density p_{XUV} is with respect to the (product) Lebesgue measure in $\mathbb{R}^3$.

bound is achievable. The crucial idea for deriving the outer bound is to minimize the term "related to" $I(X;U,V|Y)$ while all other parameters are fixed. The approach is particularly helpful in our scenario, since it does not create additional parameters for describing the region.

A. *Study of extreme cases*

We first consider extreme cases which provide some points and hints about the whole rate-distortion region.

- Our setup can be regarded as a blowing up of the Heegard-Berger [14] scheme without a constraint on the distortion in the first layer, i.e., the distortion constraint in the first layer is "viewed" as ∞. Hence when additionally $M=1$, which also reduces the setting to the Wyner-Ziv problem, the rate region collapses into
$$R_1 + R_2 \geq \frac{1}{2}\log_2 \frac{\sigma_X^2 \sigma_{N_1}^2}{\sigma_Y^2 D}. \tag{39}$$
- Similar to Remark 1, for a given R_L the identification capacity is the minimum of the first stage identification capacity $\frac{1}{2}\log_2 \frac{\sigma_Z^2}{\sigma_{N_1}^2+\sigma_{N_2}^2} + R_L$ and the identification capacity $\frac{1}{2}\log_2 \frac{\sigma_Y^2}{\sigma_{N_1}^2}$ when the processing unit has the full access to both storage nodes.

Assume that we want to design an identification scheme such that a given tuple (R, R_1, R_2, R_L, D) is achievable in which the list size R_L is large enough such that $R_\gamma = 1/2\log_2(\sigma_Y^2/\sigma_{N_1}^2)$. When the identification rate R is small, the distortion level D can be matched. However, when the identification rate R is close to the threshold R_γ, then the achieved distortion level by the identification scheme is likely to be lower than the requested distortion D. One can explain this observation as follows. In order for the identification rate to come close to the identification capacity R_γ, the compressed information must be *close* to the corresponding user's data, i.e., the distortion level for stored sequences will be extremely small and hence smaller than the requested level D. In other words, the distortion constraint in (35e) becomes inactive. This provides a hint that there will be a transition point from a region where the distortion constraint is active to a region where the distortion constraint is inactive when R increases. When the list size R_L is small or moderate, there exist additional transition points where the identification rate is limited at the first stage.

B. *An outerbound*

Suppose that the rate-distortion tuple (R, R_1, R_2, R_L, D) is achievable, i.e., for a given $\epsilon > 0$ there exists an identification scheme such that all conditions in Definition 2 are satisfied for *all* sufficiently large n. In the following we consider the case where we have $R_L \leq R$ and $D \leq \frac{\sigma_X^2 \sigma_{N_1}^2}{\sigma_Y^2}$ and derive an outerbound for the achievable rate-distortion region. Similarly, we denote by $\boldsymbol{J}$ and $\boldsymbol{K}$ the tuples $(J_i)_{i=1}^M$ and $(K_i)_{i=1}^M$. The distortion constraint implies that
$$\begin{aligned} D + \epsilon &> \mathbb{E}[d(X^n(W), g_3((J_{\hat{W}}, K_{\hat{W}}), Y^n))] \\ &\geq \inf_g \mathbb{E}[d(X^n(W), g(W, \boldsymbol{J}, \boldsymbol{K}, \hat{W}, J_{\hat{W}}, K_{\hat{W}}, Y^n))] \\ &\geq \frac{1}{n}\sum_{i=1}^n \inf_{g_i} \mathbb{E}[d(X_i(W), g_i(W, \boldsymbol{J}, \boldsymbol{K}, \hat{W}, J_{\hat{W}}, K_{\hat{W}}, Y^n))], \end{aligned} \tag{40}$$
where the infimum is taken over all possible measurable functions g_i on $\mathcal{W} \times \mathcal{M}_1^M \times \mathcal{M}_2^M \times (\mathcal{W} \cup \{e\}) \times \mathcal{M}_1 \times \mathcal{M}_2 \times \mathbb{R}^n$. In our identification scheme, $(\hat{W}, J_{\hat{W}}, K_{\hat{W}})$ are functions of $(\boldsymbol{J}, \boldsymbol{K}, Y^n, Z^n)$, which lead to the following relations
$$X^n(W) - (W, Y^n, Z^n, \boldsymbol{J}, \boldsymbol{K}) - (\hat{W}, J_{\hat{W}}, K_{\hat{W}}). \tag{41}$$
as well as
$$X^n(W) - (Y^n, W, J_W, K_W) - (Z^n, \boldsymbol{J}_{\setminus W}, \boldsymbol{K}_{\setminus W}), \tag{42}$$
where we use $\boldsymbol{J}_{\setminus W}$ as a shorthand notation of $(J_l)_{l=1,l\neq W}^M$ and similarly for $\boldsymbol{K}_{\setminus W}$. Furthermore, for notation simplicity we denote herein by T the tuple $(\hat{W}, J_{\hat{W}}, K_{\hat{W}}, \boldsymbol{J}_{\setminus W}, \boldsymbol{K}_{\setminus W})$ and by t a realization tuple $(\hat{w}, \hat{j}, \hat{k}, \boldsymbol{j}_{\setminus w}, \boldsymbol{k}_{\setminus w})$.
Thus, we have[6]
$$\begin{aligned} &\mathbb{E}[d(X_i(W), g_i(W, \boldsymbol{J}, \boldsymbol{K}, \hat{W}, J_{\hat{W}}, K_{\hat{W}}, Y^n))] \\ &= \mathbb{E}_{Y^n W K_W J_W T}\Big\{\mathbb{E}_{X_i(W)|Y^n W J_W K_W}\big[d(X_i(W), g_i(W, J_W, K_W, T, Y^n))\big]\Big\}. \end{aligned} \tag{43}$$

A detailed justification for this can be found in Appendix G-A. Let[7]
$$a^*(w, j_w, k_w, y^n) = \arg\min_t \mathbb{E}_{X_i(w)|y^n w j_w k_w}\big[d(X_i(w), g_i(w, j_w, k_w, t, y^n))\big], \tag{44}$$

[6]We restrict our attention to those functions g_i where the expectation is finite.

[7]If there exist multiple tuples $(\hat{w}, \hat{j}, \hat{k}, \boldsymbol{j}_{\setminus w}, \boldsymbol{j}_{\setminus w})$ which achieve the minimum we select the first according to the lexigraphical order.

where the minimum is attainable since the argument set is finite. Define the map g_i' as

$$g_i'(w, j_w, k_w, y^n) = g_i(w, j_w, k_w, a^*(w, j_w, k_w, y^n), y^n) \tag{45}$$

The measurability of a^* and g_i' are discussed in Appendix G-B. Then we have

$$\mathbb{E}[d(X_i(W), g_i(W, J_W, K_W, T, Y^n))] \geq \inf_{g_i'} \mathbb{E}[d(X_i(W), g_i'(W, J_W, K_W, Y^n))] \tag{46}$$

where the infimum is taken over all possible measurable functions g_i' on $\mathcal{W} \times \mathcal{M}_1 \times \mathcal{M}_2 \times \mathcal{Y}^n$ not just the one given in (45). Thus

$$\inf_{g_i} \mathbb{E}[d(X_i(W), g_i(W, J_W, K_W, T, Y^n))] = \inf_{g_i'} \mathbb{E}[d(X_i(W), g_i'(W, J_W, K_W, Y^n))], \tag{47}$$

which implies that

$$\begin{aligned} D + \epsilon &> \sum_{i=1}^{n} \frac{1}{n} \inf_{g_i'} \mathbb{E}[d(X_i(W), g_i'(W, J_W, K_W, Y^n))] \\ &= \sum_{i=1}^{n} \frac{1}{n} \mathbb{E}[d(X_i(W), \mathbb{E}[X_i(W)|W, J_W, K_W, Y^n])], \end{aligned} \tag{48}$$

since the distortion measure is the squared error. In Appendix G-C we present another route to arrive at (48), which is perhaps more formal. The constraint (48) can be interpreted in the following sense. A genie provides us the exact information (W, J_W, K_W). Then, we use the optimal estimator in the square error sense to reconstruct the orignal sequence using the aided information and the available information $(\hat{W}, J_{\hat{W}}, K_{\hat{W}}, Y^n)$. It turns out that the optimal estimator depends only on the exact information and the observation sequence.

As a standard step for a Gaussian setting, we next relate the distortion constraint (48) to a differential entropy term. For this we need to verify that the involved differential entropy term is well-defined. We establish that assertion in the following claim.

Claim 1. *There exists a conditional density function $p(x^n|w, j_w, k_w, y^n)$, which is jointly measurable on*

$$(\mathcal{W} \times \mathcal{M}_1 \times \mathcal{M}_2 \times \mathbb{R}^{2n}, 2^{\mathcal{W} \times \mathcal{M}_1 \times \mathcal{M}_2} \times \mathcal{B}(\mathbb{R}^{2n})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

such that $h(X^n(W)|W, J_W, K_W, Y^n)$ is well-defined according to the definition

$$h(X^n(W)|W, J_W, K_W, Y^n) = \mathbb{E}[-\log_2(p(X^n(W)|W, J_W, K_W, Y^n))].$$

The proof of Claim 1 and some consequences are given in Appendix H-A. Using the Claim 1 and the fact that the Gaussian distribution maximizes the conditional differential entropy subject to fixed error variance, (48) implies that

$$h(X^n(W)|J_W, K_W, W, Y^n) \leq \frac{n}{2} \log_2(2\pi e(D + \epsilon)). \tag{49}$$

Furthermore

$$h(X^n(W)|J_W, K_W, W, Y^n) \overset{(*)}{\leq} h(X^n(W)|Y^n) = \frac{n}{2} \log_2 \left(2\pi e \frac{\sigma_X^2 \sigma_{N_1}^2}{\sigma_Y^2}\right), \tag{50}$$

so that the assumption $D \leq \frac{\sigma_X^2 \sigma_{N_1}^2}{\sigma_Y^2}$ is to make the constraint (49) possibly active. $(*)$ is valid since conditioning reduces the entropy. In the next steps we need to verify the validity of the inequality (19) in our current Gaussian setting. To that end we need the following claim.

Claim 2. $\Pr\{E = e, W = w, \boldsymbol{J} = \boldsymbol{j}|Z^n = z^n\}$*, which is a measurable function of z^n, is also a jointly measurable function of $(e, w, \{j_i\}_{i=1}^M, z^n)$ on the Borel σ-algebra $2^{\{0,1\} \times \mathcal{W} \times \mathcal{M}_1^M} \times \mathcal{B}(\mathbb{R}^n)$.*

Claim 2 allows us to show the following inequality which is the Gaussian counterpart of (19)

$$H(W|Z^n, \boldsymbol{J}) \leq h_b(P_e) + P_e \log_2 M + n(R_L + \epsilon). \tag{51}$$

The proofs of Claim 2 and (51) as well as further implications are given in Appendix H-B. Next, using the variant of Fano's inequality justified in (51) we arrive at the following expression, which corresponds to (21),

$$\begin{aligned} n(R - \epsilon) &\leq \log_2 M \leq I(W, J_W; Z^n) + n(R_L + \epsilon_n) \\ &= h(Z^n) - h(Z^n|W, J_W) + n(R_L + \epsilon_n) \\ &= \frac{n}{2} \log_2(2\pi e \sigma_Z^2) - h(Z^n|W, J_W) + n(R_L + \epsilon_n), \end{aligned} \tag{52a}$$

where[8] $\epsilon_n = 2\epsilon + \frac{1}{n}\epsilon \log_2 M$. This leads to

$$n(R-\epsilon)(1-\epsilon) \le \frac{n}{2}\log_2(2\pi e\sigma_Z^2) - h(Z^n|W,J_W) + n(R_L+2\epsilon) \tag{53}$$

which implies that

$$h(Z^n|W,J_W) \le \frac{n}{2}\log_2(2\pi e\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon}). \tag{54}$$

Claim 1 also shows that for a given (w,j_w) the conditional pdfs $p_{X^n(W)|WJ_W}(x^n|w,j_w)$ and $p_{Y^n|WJ_W}(y^n|w,j_w)$ are well defined. Furthermore, N_1^n and N_2^n are independent of (W,J_W,K_W). Due to the entropy power inequality [12, p. 22], cf. also [15, Eq. (20)], we obtain

$$\begin{aligned} 2^{\frac{2}{n}h(Z^n|W,J_W)} &\ge 2^{\frac{2}{n}h(Y^n|W,J_W)} + 2^{\frac{2}{n}h(N_2^n|W,J_W)}, \\ 2^{\frac{2}{n}h(Z^n|W,J_W)} &\ge 2^{\frac{2}{n}h(X^n(W)|W,J_W)} + 2^{\frac{2}{n}h(N_1^n|W,J_W)} + 2^{\frac{2}{n}h(N_2^n|W,J_W)}, \end{aligned} \tag{55}$$

which leads to

$$\begin{aligned} 2\pi e(\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon} - \sigma_{N_2}^2) &\ge 2^{\frac{2}{n}h(Y^n|W,J_W)} \\ 2\pi e(\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon} - (\sigma_{N_1}^2+\sigma_{N_2}^2)) &\ge 2^{\frac{2}{n}h(X^n|W,J_W)}. \end{aligned} \tag{56}$$

Since $h(X^n|W,J_W) > -\infty$, we therefore have the following condition

$$(R-\epsilon)(1-\epsilon) - R_L - 2\epsilon < \frac{1}{2}\log_2\left(\frac{\sigma_Z^2}{\sigma_{N_1}^2+\sigma_{N_2}^2}\right). \tag{57}$$

Inequality (57) further leads to, since by Definition 2 it must hold for every $\epsilon > 0$,

$$R \le \frac{1}{2}\log_2\left(1+\frac{\sigma_X^2}{\sigma_{N_1}^2+\sigma_{N_2}^2}\right) + R_L, \tag{58}$$

as we take $\epsilon \to 0$. Under the assumption $R_L \le R$, the constraints (57) and (58) are not ruled out as they are not obviously true. Additionally, corresponding to (23) we obtain

$$\begin{aligned} n(R-\epsilon) &\le I(W,J_W,K_W;Y^n) + 1 + \epsilon\log_2 M \\ &= h(Y^n) - h(Y^n|W,J_W,K_W) + 1 + \epsilon\log_2 M \\ &= \frac{n}{2}\log_2(2\pi e\sigma_Y^2) - h(Y^n|W,J_W,K_W) + 1 + \epsilon\log_2 M, \end{aligned} \tag{59}$$

which leads to

$$h(Y^n|W,J_W,K_W) \le \frac{n}{2}\log_2(2\pi e\sigma_Y^2 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon}). \tag{60}$$

Similarly, using the entropy power inequality results in that

$$\begin{aligned} 2\pi e\sigma_Y^2 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon} \ge 2^{\frac{2}{n}h(Y^n|W,J_W,K_W)} &\ge 2^{\frac{2}{n}h(X^n(W)|W,J_W,K_W)} + 2^{\frac{2}{n}h(N_1^n|W,J_W,K_W)} \\ &> 2\pi e\sigma_{N_1}^2, \end{aligned} \tag{61}$$

since $h(X^n|W,J_W,K_W) > -\infty$. Thus, there exists an α_1 with $0 \le \alpha_1 < 1$, *which depends on other parameters*, such that

$$h(Y^n|W,J_W,K_W) = \frac{n}{2}\log_2(2\pi e((1-\alpha_1)\sigma_Y^2 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon} + \alpha_1\sigma_{N_1}^2)), \tag{62}$$

and

$$h(X^n(W)|W,J_W,K_W) \le \frac{n}{2}\log_2(2\pi e(1-\alpha_1)(\sigma_Y^n 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon} - \sigma_{N_1}^2)). \tag{63}$$

From (61) we also obtain a constraint on the rate R, namely

$$R \le \frac{1}{2}\log_2\left(\frac{\sigma_Y^2}{\sigma_{N_1}^2}\right). \tag{64}$$

Thus (58) and (64) imply that

$$0 \le R_L \le R \le R_\gamma. \tag{65}$$

Using the second inequality in (56) we have

$$\begin{aligned} n(R_1+\epsilon) &\ge I(X^n(W);J_W,W) \\ &\ge \frac{n}{2}\Bigg(\log_2(2\pi e\sigma_X^2) - \log_2(2\pi e(\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon} - (\sigma_{N_1}^2+\sigma_{N_2}^2)))\Bigg) \end{aligned}$$

[8]We also provide a direct justification of the second equality (52a) in Appendix I for the interested readers.

$$= \frac{n}{2}\log_2\left(\frac{\sigma_X^2}{\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon} - (\sigma_{N_1}^2+\sigma_{N_2}^2)}\right). \tag{66}$$

Taking $\epsilon \to 0$ we obtain

$$R_1 \geq \frac{1}{2}\log_2\left(\frac{\sigma_X^2}{\sigma_Z^2 2^{-2(R-R_L)} - (\sigma_{N_1}^2+\sigma_{N_2}^2)}\right). \tag{67}$$

Similarly, corresponding to (25) we obtain

$$\begin{aligned} n(R_1+R_2+\epsilon) &\geq I(Y^n; J_W, W) + I(X^n(W); J_W, K_W, W|Y^n) \\ &= \underbrace{h(Y^n) - h(Y^n|W, J_W)}_{\Delta_1} + \underbrace{h(X^n(W)|Y^n) - h(X^n(W)|J_W, K_W, W, Y^n)}_{\Delta_2}. \end{aligned} \tag{68}$$

The first term in (68) is bounded based on the first inequality in (56) as

$$\begin{aligned} \Delta_1 &\geq \frac{n}{2}\Bigg(\log_2(2\pi e\sigma_Y^2) - \log_2(2\pi e(\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon} - \sigma_{N_2}^2)\Bigg) \\ &= \frac{n}{2}\Bigg(\log_2\frac{\sigma_Y^2}{\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon} - \sigma_{N_2}^2}\Bigg). \end{aligned} \tag{69}$$

The second term is bounded in three different ways:

1) From (49) we obtain

$$\begin{aligned} \Delta_2 &\geq \frac{n}{2}\log_2\left(2\pi e\frac{\sigma_X^2\sigma_{N_1}^2}{\sigma_Y^2}\right) - \frac{n}{2}\log_2 2\pi e(D+\epsilon) \\ &= \frac{n}{2}\log_2\frac{\sigma_X^2\sigma_{N_1}^2}{\sigma_Y^2(D+\epsilon)}. \end{aligned} \tag{70}$$

This implies in combination with (69) that

$$R_1+R_2 \geq \frac{1}{2}\log_2\frac{\sigma_Y^2}{\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2} + \frac{1}{2}\log_2\frac{\sigma_X^2\sigma_{N_1}^2}{\sigma_Y^2 D}. \tag{71}$$

2) Secondly, the expressions in (62) and (63) lead to

$$\begin{aligned} \Delta_2 &= h(X^n(W)|Y^n) - h(X^n(W)|W, J_W, K_W) - h(Y^n|X^n(W)) + h(Y^n|W, J_W, K_W) \\ &\geq \frac{n}{2}\log_2\left(\frac{\sigma_X^2}{\sigma_Y^2}\frac{(1-\alpha_1)\sigma_Y^2 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon} + \alpha_1\sigma_{N_1}^2}{(1-\alpha_1)(\sigma_Y^2 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon} - \sigma_{N_1}^2)}\right) \\ &\overset{(a)}{\geq} \frac{n}{2}\log_2\left(\frac{\sigma_X^2}{\sigma_Y^2(\sigma_Y^2 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon} - \sigma_{N_1}^2)}\inf_{0\leq\alpha_1<1}(\sigma_Y^2 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon} + \frac{\alpha_1}{1-\alpha_1}\sigma_{N_1}^2)\right). \end{aligned} \tag{72}$$

We note that due to the inequality (61) the term $\frac{\sigma_X^2}{\sigma_Y^2(\sigma_Y^2 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon} - \sigma_{N_1}^2)}$ is positive hence (a) is valid. Note also that since α_1 might depend on ϵ and n, we should avoid taking the limit directly. Since $\frac{\alpha_1}{1-\alpha_1}$ is an increasing and positive function of α_1 on $[0,1)$, the infimum is attained at $\alpha_1 = 0$. Hence

$$\Delta_2 \geq \frac{n}{2}\log_2\left(\frac{\sigma_X^2 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon}}{\sigma_Y^2 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon} - \sigma_{N_1}^2}\right). \tag{73}$$

This implies by taking $\epsilon \to 0$ that

$$R_1+R_2 \geq \frac{1}{2}\log_2\frac{\sigma_Y^2}{\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2} + \frac{1}{2}\log_2\left(\frac{\sigma_X^2 2^{-2R}}{\sigma_Y^2 2^{-2R} - \sigma_{N_1}^2}\right), \tag{74}$$

3) Lastly, by applying a similar derivation we also observe that

$$R_1+R_2 \geq \frac{1}{2}\log_2\frac{\sigma_Y^2}{\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2} + \frac{1}{2}\log_2\frac{\sigma_X^2}{\sigma_Y^2}\frac{\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2}{\sigma_Z^2 2^{-2(R-R_L)} - (\sigma_{N_1}^2+\sigma_{N_2}^2)}. \tag{75}$$

The details are given in Appendix H-C.

Combining these three bounds we obtain

$$R_1+R_2 \geq \frac{1}{2}\log_2\frac{\sigma_Y^2}{\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2}$$

$$+\frac{1}{2}\max\Bigg\{\underbrace{\log_2\frac{\sigma_X^2\sigma_{N_1}^2}{\sigma_Y^2 D}}_{h_0(R)},\underbrace{\log_2\frac{\sigma_X^2 2^{-2R}}{\sigma_Y^2 2^{-2R}-\sigma_{N_1}^2}}_{h_1(R)},\underbrace{\log_2\frac{\sigma_X^2}{\sigma_Y^2}\frac{\sigma_Z^2 2^{-2(R-R_L)}-\sigma_{N_2}^2}{\sigma_Z^2 2^{-2(R-R_L)}-(\sigma_{N_1}^2+\sigma_{N_2}^2)}}_{h_2(R)}\Bigg\}. \tag{76}$$

Additionally, we have the following constraint which corresponds to (27)

$$\begin{aligned} n(R_1+R_2+\epsilon)-\log M &\geq I(X^n(W);W,J_W,K_W|Y^n)-(1+\epsilon\log M)\\ &= \Delta_2-(1+\epsilon\log M). \end{aligned} \tag{77}$$

which implies that

$$R_1+R_2-R\geq\Gamma. \tag{78}$$

In summary, we obtain the following outerbound region

$$\begin{aligned} 0<D\leq\frac{\sigma_X^2\sigma_{N_1}^2}{\sigma_Y^2},\ 0\leq R_L\leq R\leq R_\gamma,\\ R_1\geq\frac{1}{2}\log_2\left(\frac{\sigma_X^2}{\sigma_Z^2 2^{-2(R-R_L)}-(\sigma_{N_1}^2+\sigma_{N_2}^2)}\right),\\ R_1+R_2\geq\frac{1}{2}\log_2\frac{\sigma_Y^2}{\sigma_Z^2 2^{-2(R-R_L)}-\sigma_{N_2}^2}+\Gamma,\\ R_1+R_2-R\geq\Gamma. \end{aligned} \tag{79}$$

As $R\to R_\gamma$ either $h_1(R)$ or $h_2(R)$ goes to ∞. However, since both R_1 and R_2 are finite we must have

$$0\leq R_L\leq R<R_\gamma. \tag{80}$$

C. Analyzing the outerbound

An illustration of the following different situations is given in Fig. 3 and Fig. 4.

1) Phase transion points: The above outerbound matches some properties which are mentioned in Subsection III-A. We observe that for fixed D and R_L the three functions $h_0(R)$, $h_1(R)$ and $h_2(R)$ provide the key for the transition behavior from one extreme case to another since they are monotone in the identification rate R. We show in the following that there are three possible transition points $R_{cr_{12}}$, $R_{cr_{01}}$ and $R_{cr_{02}}$ as R varies, where the corresponding subscripts indicate which functions are involved. More specifically, we have

$$\begin{aligned} R_{cr_{12}} &= \frac{1}{2}\log_2\frac{\sigma_Z^2 2^{2R_L}-\sigma_Y^2}{\sigma_{N_2}^2},\quad R_{cr_{01}}=\frac{1}{2}\log_2\frac{\sigma_Y^2(1-\frac{D}{\sigma_{N_1}^2})}{\sigma_{N_1}^2}\\ R_{cr_{02}} &= R_L+\frac{1}{2}\log_2\frac{\sigma_Z^2}{\sigma_{N_2}^2+\frac{\sigma_{N_1}^2}{1-\frac{D}{\sigma_{N_1}^2}}}. \end{aligned} \tag{81}$$

To derive $R_{cr_{12}}$ we first notice that the function

$$\frac{\sigma_Y^2 2^{-2R}-\sigma_{N_1}^2}{\sigma_X^2 2^{-2R}}=\frac{\sigma_Y^2}{\sigma_X^2}-\frac{\sigma_{N_1}^2}{\sigma_X^2 2^{-2R}} \tag{82}$$

is a decreasing function w.r.t. R, which implies that $h_1(R)$ is an increasing one. Similarly, $h_2(R)$ is also an increasing function w.r.t. R. Hence by solving the following equation

$$h_1(R)=h_2(R) \tag{83}$$

we can find the (possibly) unique intersection point $R_{cr_{12}}$ if the equation has a solution. The above expression implies that

$$\begin{aligned} &\Rightarrow \sigma_Y^2-\frac{\sigma_{N_1}^2}{2^{-2R}}=\sigma_Y^2(1-\frac{\sigma_{N_1}^2}{\sigma_Z^2 2^{-2(R-R_L)}-\sigma_{N_2}^2})\\ &\Leftrightarrow R_{cr_{12}}=\frac{1}{2}\log_2\frac{\sigma_Z^2 2^{2R_L}-\sigma_Y^2}{\sigma_{N_2}^2}. \end{aligned} \tag{84}$$

Note that however $R_{cr_{12}}$ can lie outside the interval $[R_L, R_\gamma)$, i.e., $h_1(R) \neq h_2(R)$ for all $R \in [R_L.R_\gamma)$.
Next note that $h_1(0) = 0$ and $h_1(R) \to \infty$ as $R \to \frac{1}{2}\log_2 \frac{\sigma_Y^2}{\sigma_{N_1}^2}$. Since $h_1(R)$ is increasing, there is a unique point $R_{cr_{01}} \in [0, \frac{1}{2}\log_2 \frac{\sigma_Y^2}{\sigma_{N_1}^2})$ such that $h_1(R) = h_0(R)$, i.e.,

$$\frac{\sigma_{N_1}^2}{\sigma_Y^2 D} = \frac{2^{-2R_{cr_{01}}}}{\sigma_Y^2 2^{-2R_{cr_{01}}} - \sigma_{N_1}^2}, \tag{85}$$

above which the $h_1(R)$ dominates $h_0(R)$. Solving for $R_{cr_{01}}$ we obtain

$$R_{cr_{01}} = \frac{1}{2}\log_2 \frac{\sigma_Y^2(1 - \frac{D}{\sigma_{N_1}^2})}{\sigma_{N_1}^2}. \tag{86}$$

Analogously, $R_{cr_{02}} \in [R_L, R_L + \frac{1}{2}\log_2 \frac{\sigma_Z^2}{\sigma_{N_1}^2+\sigma_{N_2}^2})$ given as in (81) is the unique intersection point of $h_2(R)$ and $h_0(R)$, above which the $h_2(R)$ dominates $h_0(R)$, as $h_2(R_L) = 0$ while $h_2(R) \to \infty$ when $R \to R_L + \frac{1}{2}\log_2 \frac{\sigma_Z^2}{\sigma_{N_1}^2+\sigma_{N_2}^2}$. As the solution of the equation $h_2(R) = h_0(R)$, $R_{cr_{02}}$ also satisfies

$$D = \sigma_{N_1}^2\left(1 - \frac{\sigma_{N_1}^2}{\sigma_Z^2 2^{-2(R_{cr_{02}}-R_L)} - \sigma_{N_2}^2}\right). \tag{87}$$

2) Discussion: In this part we discuss some additional properties of the three functions and transitions points. We note that we have $R_{cr_{12}} \geq R_L$ because from $R_L \geq 0$ it follows that $\sigma_Z^2 2^{2R_L} - \sigma_Y^2 \geq 2^{2R_L}(\sigma_Z^2 - \sigma_Y^2) = 2^{2R_L}\sigma_{N_2}^2$. Additionally, again because we have $R_L \geq 0$ it follows that $h_1(0) = 0 \geq h_2(0)$ as

$$\sigma_Y^2(1 - \frac{\sigma_{N_1}^2}{\sigma_Z^2 2^{2R_L} - \sigma_{N_2}^2}) \geq \sigma_X^2. \tag{88}$$

Thus for $0 \leq R \leq \min\{R_\gamma, R_{cr_{12}}\}$, $h_1(R) \geq h_2(R)$. Furthermore, we observe that when $R \leq \min\{R_{cr_{12}}, R_\gamma\}$ the following holds

$$R \geq \frac{1}{2}\log_2 \frac{\sigma_Y^2}{\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2}. \tag{89}$$

Therefore, the constraint (36c), can be omitted in this case. If the interval $(R_{cr_{12}}, R_\gamma)$ is not empty then the reverse inequality holds on it and the constraint (36d) can be omitted.
The following relation is helpful to relate Case II and Case V in the later paragraph.

$$\sigma_Y^2(1 - 2^{-2R_{cr_{01}}}) = \sigma_Z^2(1 - 2^{-2(R_{cr_{02}}-R_L)}) = \sigma_Y^2 - \frac{\sigma_{N_1}^2}{1 - \frac{D}{\sigma_{N_1}^2}}. \tag{90}$$

Importantly, when $D \to 0$, we observe that as $R \to R_\gamma$ either $h_1(R)$ or $h_2(R)$ goes to ∞. Hence at least one of the point $R_{cr_{01}}$ or $R_{cr_{02}}$ lies in the interval $[R_L, R_\gamma)$. If $D \to \frac{\sigma_X^2\sigma_{N_1}^2}{\sigma_Y^2}$, then $R_{cr_{01}}$ goes to 0 and hence might lie outside the interval $R_L \leq R < R_\gamma$. In this case $R_{cr_{02}}$ is always inside.

D. Achievability

From Fig. 3 and Fig. 4 we see that different constraints will be active in the outer bound depending on the identification rate R. In the achievability we will therefore distinguish between different cases and select the parameter accordingly. In Table I we provide an overview about different cases as well as information about the marginal distributions of U and V that are encountered in the following.

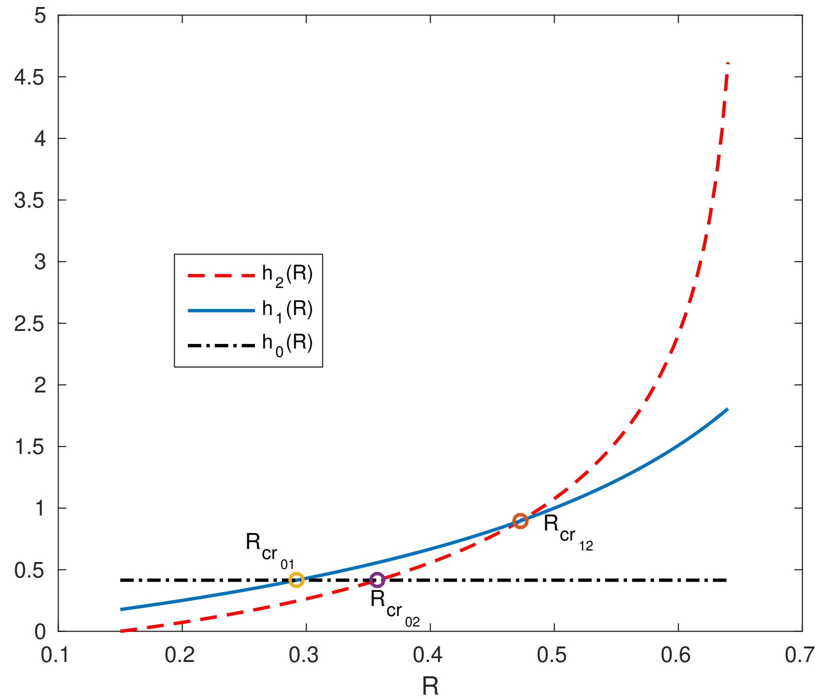
Fix a value of D and R_L where $0 \leq D \leq \frac{\sigma_X^2\sigma_{N_1}^2}{\sigma_Y^2}$ and $0 \leq R_L < R_\gamma(R_L)$.

1) The case $R_{cr_{12}} < R_\gamma$: We consider first that $R_{cr_{12}} < R_\gamma$ which implies that both $h_1(R_{cr_{12}})$ and $h_2(R_{cr_{12}})$ are defined, i.e., $R_{cr_{12}}$ lies in both domains of $h_1(R)$ and $h_2(R)$. Note also that $R_{cr_{12}} \geq R_L$ holds, cf. (84).

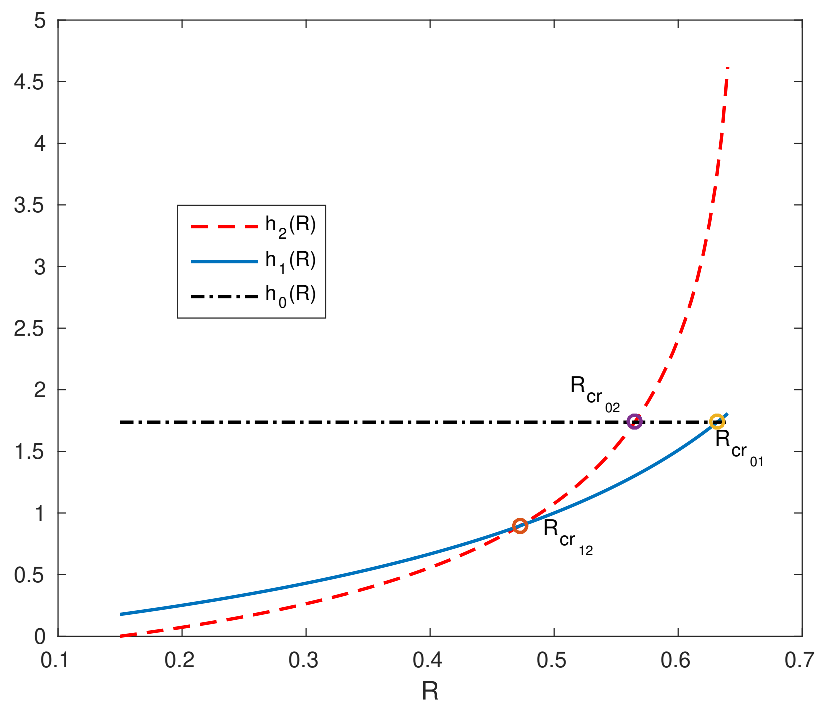
a) $R_{cr_{01}} \leq R_{cr_{12}}$. In cases I and II we need to truncate the corresponding interval if necessary so that the condition $R \geq R_L$ holds.

- Case I: $R_{cr_{01}} \leq R < R_{cr_{12}}$, i.e., $h_1(R)$ is the dominant component in the outerbound since $h_1(R) \geq h_0(R)$ when $R \geq R_{cr_{01}}$ and $h_1(R) > h_2(R)$ when $R < R_{cr_{12}}$. Let $X = V + N_0$ where V and N_0 are independent Gaussian random variables with $\sigma_V^2 = \sigma_Y^2(1-2^{-2R})$. Note that $\sigma_V^2 < \sigma_X^2$ since $R < R_\gamma$. V should be understood as the output of the *test channel* $p_{V|X}$, cf. [16, p. 311]. Then, let $V = U + N_0'$ where U and N_0' are independent Gaussian random variables such that $\sigma_U^2 = \sigma_Z^2(1 - 2^{-2(R-R_L)})$. We also observe that $\sigma_U^2 > 0$ if $R > R_L$. We note that

$$2^{-2R}(\sigma_Z^2 2^{2R_L} - \sigma_Y^2) > \sigma_{N_2}^2 \text{ or } \sigma_Y^2(1 - 2^{-2R}) > \sigma_Z^2(1 - 2^{-2(R-R_L)}), \tag{91}$$



(a) Case 1: $R_{cr_{12}} < R_\gamma$ and $R_L \leq R_{cr_{01}} \leq R_{cr_{12}}$. We can see that when $R_L \leq R \leq R_{cr_{01}}$, $h_0(R)$ dominates over $h_1(R)$ and $h_2(R)$. $h_1(R)$ is the dominant component when $R_{cr_{01}} < R \leq R_{cr_{12}}$. When $R_{cr_{12}} < R < R_\gamma$, then $h_2(R)$ dominates the other two functions.



(b) Case 2: $R_{cr_{12}} < R_\gamma$ and $R_{cr_{01}} \geq R_{cr_{12}}$. In this case $h_0(R)$ dominates over the other two functions when $R_L \leq R \leq R_{cr_{02}}$. For $R \in [R_{cr_{02}}, R_\gamma)$, $h_2(R)$ is the dominant component

Fig. 3: Two cases when D varies for fixed R_L.

Cases		Subcases	Dominating functions			Distributions of U and V
			$h_0(R)$	$h_1(R)$	$h_2(R)$	
$R_{cr_{12}} < R_\gamma$	$R_{cr_{01}} \leq R_{cr_{12}}$	I. $R_{cr_{01}} \leq R < R_{cr_{12}}$		✓		$U \sim P_U$, $V \sim P_V$
		II. $0 \leq R < R_{cr_{01}}$	✓			$U \sim P_U$, $V \sim P_V(R_{cr_{01}})$
		III. $R_{cr_{12}} \leq R < R_\gamma$			✓	$U \sim P_U$, V degenerate
	$R_{cr_{01}} > R_{cr_{12}}$	IV. $R_{cr_{02}} \leq R < R_\gamma$			✓	As in Case III
		V. $R_L \leq R < R_{cr_{02}}$	✓			$U \sim P_U$, $V \sim P_V(R_{cr_{01}})$
$R_{cr_{12}} \geq R_\gamma$	$R_{cr_{01}} > R_L$	VI. $R_{cr_{01}} \leq R < R_\gamma$		✓		As in Case I
		VII. $R_L \leq R < R_{cr_{01}}$	✓			As in Case II
	$R_{cr_{01}} \leq R_L$	VIII. $\forall R$		✓		As in Case I

TABLE I: Summary of optimal (marginal) distributions of the auxiliary random variables U and V for all possible cases specified by the relation among $R_{cr_{12}}$, $R_{cr_{01}}$, $R_{cr_{02}}$, R_L, R_γ and R where $P_U = \mathcal{N}\big(0, \sigma_Z^2(1-2^{-2(R-R_L)})\big)$ and $P_V = \mathcal{N}\big(0, \sigma_Y^2(1-2^{-2R})\big)$. Note that the marginal distribution of the auxiliary random variable U does not change. Additionally, due to the relation (90) the distribution of V in Case V is identical to the one in Case II.

since $R < R_{cr_{12}}$. This means that $\sigma_U^2 < \sigma_V^2$. Similarly, U is the output of the test channel $p_{U|V}$. By our choice of U and V the relation $U - V - X - Y - Z$ holds. We next examine whether the chosen random variables satisfy the constraints corresponding to the fixed parameters. The condition $I(Z;U) = R - R_L$ is satisfied by the chosen U. Furthermore, $I(Y;V) = R$ due to the choice of V. This means that the choice of U and V does not violate the constraint

$$R \leq \min\{I(Z;U) + R_L, I(Y;V)\}. \tag{92}$$

Next we calculate

$$\begin{aligned} h(X|V,Y) &= h(Y|X) + h(X|V) - h(Y|V) \\ &= \frac{1}{2}\log_2\left(2\pi e \frac{(\sigma_X^2 - \sigma_V^2)\sigma_{N_1}^2}{\sigma_Y^2 - \sigma_V^2}\right) \\ &= \frac{1}{2}\log_2\left(2\pi e \frac{\sigma_{N_1}^2}{\sigma_Y^2}\frac{\sigma_Y^2 2^{-2R} - \sigma_{N_1}^2}{2^{-2R}}\right) \\ &\overset{(\star)}{\leq} \frac{1}{2}\log_2(2\pi e D) \end{aligned} \tag{93}$$

where $(\star)$ is valid due to (85) as $R \geq R_{cr_{01}}$, i.e., the distortion level D is attainable using the MMSE decoder. Now, plugging the random variable U into the first expression in the achievable region we obtain

$$R_1 \geq I(X;U) = \frac{1}{2}\log_2 \frac{\sigma_X^2}{\sigma_Z^2 2^{-2(R-R_L)} - (\sigma_{N_1}^2 + \sigma_{N_2}^2)}. \tag{94}$$

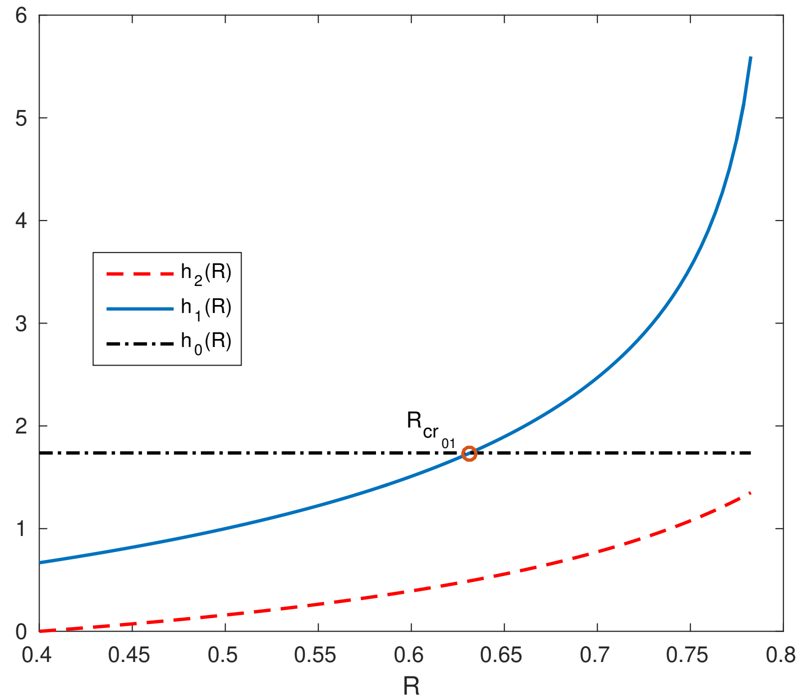


Fig. 4: Case 3: $R_{cr_{12}} \geq R_\gamma$ and $R_{cr_{01}} \geq R_L$. In this case for $R_L \leq R \leq R_{cr_{01}}$, $h_0(R)$ is the dominant component, while for $R_{cr_{01}} < R < R_\gamma$, $h_1(R)$ is the dominant component.

Moreover,

$$\begin{aligned} I(X;V|Y) &= \frac{1}{2}\log_2\left(\frac{\sigma_X^2\sigma_{N_1}^2}{\sigma_Y^2}\frac{\sigma_Y^2}{\sigma_{N_1}^2}\frac{2^{-2R}}{\sigma_Y^2 2^{-2R}-\sigma_{N_1}^2}\right) \\ &= \frac{1}{2}\log_2\frac{\sigma_X^2 2^{-2R}}{\sigma_Y^2 2^{-2R}-\sigma_{N_1}^2}. \end{aligned} \tag{95}$$

Since

$$\begin{aligned} I(X;U)+I(X;V|U,Y) &= I(X;U)+I(X;U,V|Y)-I(X;U|Y) \\ &= I(Y;U)+I(X;U,V|Y) \\ &= I(Y;U)+I(X;V|Y), \end{aligned} \tag{96}$$

where the second equality holds since $I(X;U|Y) = I(X;U) - I(Y;U)$ as $U - X - Y$, we obtain that

$$\begin{aligned} R_1+R_2 &\geq \frac{1}{2}\log_2\frac{\sigma_Y^2}{\sigma_Z^2 2^{-2(R-R_L)}-\sigma_{N_2}^2}+\frac{1}{2}\log_2\frac{\sigma_X^2 2^{-2R}}{\sigma_Y^2 2^{-2R}-\sigma_{N_1}^2} \\ R_1+R_2-R &\geq \frac{1}{2}\log_2\frac{\sigma_X^2 2^{-2R}}{\sigma_Y^2 2^{-2R}-\sigma_{N_1}^2}, \end{aligned} \tag{97}$$

which matches the outerbound. When $R = R_L$ we can simply choose U to be a Gaussian random variable which is independent of everything else. As discussed previously in (89), since $R \leq R_{cr_{12}}$ the first constraint in (97) can be omitted.

The other cases can be matched similarly using the same principle. Since this results in lengthy derivations with limited new insights, we provide the remaining proof in Appendix J.

Appendix A
Supporting Lemmata

In the following lemma and corollary we present useful properties of the conditional expectation over a σ-algebera (σ-field) which are used in later sections. The unfamilar reader is referred to, for instance, [17, Chapter 5] for a comprehensive introduction.

Lemma 1. *[13, Doob's conditional independence lemma, Proposition 5.6] For any σ-fields $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{H}$, we have $\mathcal{F}$ and $\mathcal{H}$ are conditionally independent given $\mathcal{G}$ iff*

$$P[H|\mathcal{F},\mathcal{G}] = P[H|\mathcal{G}],\ \mathbb{P}-a.s,\ H \in \mathcal{H}. \tag{98}$$

The conditioning on the left-hand side of (98) should be understood as w.r.t. the join σ-algebra $\sigma(\mathcal{F},\mathcal{G})$.

Corollary 3. *Assume that $\mathcal{F}$ and $\mathcal{H}$ are conditionally independent given $\mathcal{G}$. Let f be a nonnegative $\mathcal{H}$-measurable, integrable function. Then*

$$\mathbb{E}[f|\mathcal{F},\mathcal{G}] = \mathbb{E}[f|\mathcal{G}]\ ),\ \mathbb{P}-a.s.. \tag{99}$$

Proof: We note that for a given $H \in \mathcal{H}$, $\{\omega \mid P[H|\mathcal{F},\mathcal{G}](\omega) \neq P[H|\mathcal{G}](\omega)\} \in \sigma(\mathcal{F},\mathcal{G})$. Lemma 1 implies that for any positive simple function $\chi = \sum_{i=1}^{k} a_i \chi_{\mathcal{A}_i}$ where $\mathcal{A}_i \in \mathcal{H}$, and $a_i > 0$, $\forall i$,

$$\mathbb{E}[\chi|\mathcal{F},\mathcal{G}] = \mathbb{E}[\chi|\mathcal{G}] \text{ a.s..} \tag{100}$$

Since f is a nonnegative $\mathcal{H}$-measurable, integrable function, there is a sequence of increasing nonnegative, $\mathcal{H}$-measurable, simple functions χ_n that converges pointwise to f. We have by monotone convergence theorem

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}[\chi_n|\mathcal{F},\mathcal{G}] &= \mathbb{E}[f|\mathcal{F},\mathcal{G}] \text{ a.s.,}\\ \lim_{n\to\infty} \mathbb{E}[\chi_n|\mathcal{G}] &= \mathbb{E}[f|\mathcal{G}] \text{ a.s..}\end{aligned} \tag{101}$$

Denote $B_1 = \{\omega \mid \lim_{n\to\infty} \mathbb{E}[\chi_n|\mathcal{F},\mathcal{G}](\omega) \neq \mathbb{E}[f|\mathcal{F},\mathcal{G}](\omega)\}$, $B_2 = \{\omega \mid \lim_{n\to\infty} \mathbb{E}[\chi_n|\mathcal{G}](\omega) \neq \mathbb{E}[f|\mathcal{G}](\omega)\}$ and

$$C_i = \{\omega \mid \mathbb{E}[\chi_i|\mathcal{F},\mathcal{G}](\omega) \neq \mathbb{E}[\chi_i|\mathcal{G}](\omega)\},\ i = 1,\dots. \tag{102}$$

Define $B = B_1 \cup B_1 \bigcup_i C_i$. We observe that $B \in \sigma(\mathcal{F},\mathcal{G})$ and $\mathbb{P}(B) = 0$. For $\omega \in B^c$ then

$$\mathbb{E}[f|\mathcal{F},\mathcal{G}](\omega) = \lim_{n\to\infty} \mathbb{E}[\chi_n|\mathcal{F},\mathcal{G}](\omega) = \lim_{n\to\infty} \mathbb{E}[\chi_n|\mathcal{G}](\omega) = \mathbb{E}[f|\mathcal{G}](\omega). \tag{103}$$

Hence, the conclusion follows. ∎

Appendix B
Achievability of Theorem 1

Fix a conditional pmf $P_{UV|X}$ where $U - V - X$ and a deterministic reconstruction mapping f such that we have

$$\mathbb{E}[d(X, f(U,V,Y))] = D. \tag{104}$$

Additionally, for a fixed $\epsilon > 0$, we assume that the number of enrolled users is given by $M = 2^{n\hat{R}}$ where $\hat{R} = R - \epsilon/2$ and the actual list size is $\hat{\Delta} = R_L + \epsilon/2$. Also let $\hat{R}_U = R_1 + \epsilon/2$, $\hat{R}_V = R_2 + \epsilon/2$ and $\hat{R}'_V = R'_V - \epsilon/4$ be the actual code rates. The set of suitable tuples (R, R_L, R_1, R_2, R'_V) will be specified later in (124). We also select an $\bar{\epsilon} > 0$ for the strongly typical set, which depends on n and $\bar{\epsilon} \to 0$ as $n \to \infty$.

Codebook generation: The codebook used in the enrollment process is *identical* for all users and constructed as follows: Generate $2^{n\hat{R}_U}$ iid codewords $u^n(j)$ where $j \in [1 : 2^{n\hat{R}_U}]$ according to the marginal distribution P_U. For each j, we draw $2^{n(\hat{R}_V+\hat{R}'_V)}$ codewords $v^n(j,l)$ where $l \in [1 : 2^{n(\hat{R}_V+\hat{R}'_V)}]$ iid via the conditional distribution $P_{V|U}$. Each index l is parsed into a unique tuple $l = (k,k')$ where $k \in [1 : 2^{n\hat{R}_V}]$ and $k' \in [1 : 2^{n\hat{R}'_V}]$. Denote by

$$\mathfrak{B}(k) = \{l \mid l = (k,k'),\ \text{for some } k'\}, \tag{105}$$

the k-th bin, where $k \in [1 : 2^{n\hat{R}_V}]$. Additionally, we include a *fixed* pair of codewords (u^n_e, v^n_e) corresponding to the error event. The codebook is known in the whole system.

Enrollment: For each user index $i \in \mathcal{M}$, a codeword $u^n(j_i)$ is looked for such that $(x^n(i), u^n(j_i)) \in \mathcal{T}^n_{\bar{\epsilon}}(XU)$. The chosen j_i is stored in the first layer. Next, a codeword $v^n(j_i, l_i)$ is searched for such that

$$(x^n(i), u^n(j_i), v^n(j_i, l_i)) \in \mathcal{T}^n_{\bar{\epsilon}}(XUV). \tag{106}$$

The chosen bin index k_i is stored in the second layer. We note that in both steps if there is more than one suitable index, we select one of them uniformly at random. If there is none, an index is selected from the corresponding set of all indices uniformly at random.

Identification and Reconstruction: The observation y^n is first passed through the memoryless pre-processing channel given by $P_{Z|Y}$ to produce z^n which is used in the first stage of our identification and reconstruction process.

First stage: We look for all indices $i \in \mathcal{M}$ such that

$$(z^n, u^n(j_i)) \in \mathcal{T}^n_{\bar{\epsilon}}(ZU), \tag{107}$$

and put them into the list $\mathcal{L}$. If there are more than $2^{n\hat{\Delta}}$ suitable indices then an error is declared, i.e., we output the set $\mathcal{L}=\{e\}$. In this way, our list always meets the size constraint in (2).

Second stage: If $\mathcal{L}=\{e\}$, then we set $\hat{w}=e$. Otherwise, if $\mathcal{L}\neq\{e\}$, we find a unique index $\hat{w}$ in $\mathcal{L}$ such that

$$(y^n, u^n(j_{\hat{w}}), v^n(j_{\hat{w}}, \tilde{l})) \in \mathcal{T}_{\bar{\epsilon}}^n(YUV) \tag{108}$$

for some $\tilde{l}$, where $\tilde{l}\in\mathfrak{B}(k_{\hat{w}})$, and $j_{\hat{w}}$ and $k_{\hat{w}}$ are the stored information of the $\hat{w}$-th user. If there is no such $\hat{w}$ or there is more than one then we set $\hat{w}=e$. In the next step, if $\hat{w}\neq e$ then we search for a unique $\tilde{l}\in\mathfrak{B}(k_{\hat{w}})$ such that

$$(y^n, u^n(j_{\hat{w}}), v^n(j_{\hat{w}}, \tilde{l})) \in \mathcal{T}_{\bar{\epsilon}}^n. \tag{109}$$

If there is more than one $\tilde{l}$ or there is none then we set $\tilde{l}=e$. Moreover, if $\hat{w}=e$ or $\tilde{l}=e$ then we set $u^n(j_{\hat{w}})=u_e^n$ and $v^n(j_{\hat{w}},\tilde{l})=v_e^n$. The reconstruction sequence is given as $\hat{x}_\tau=f(u_\tau(j_{\hat{w}}), v_\tau(j_{\hat{w}},\tilde{l}), y_\tau)$ for all $\tau=[1:n]$.
Note that the search for the unique pair $(\hat{w},\tilde{l})$ could be done in a single step, however, to mitigate the complexity of describing (g_2,g_3) we choose the separate descriptions, cf. the Gaussian setting for more information.
Analysis: Let J_i and L_i, $i\in\mathcal{M}$, be the chosen indices for the i-th user. Furthermore, let $\boldsymbol{\mathcal{L}}_1$ be the list of indices $i\in\mathcal{M}$ that satisfy (107) in the first stage of the identification process, while the return list is denoted by $\boldsymbol{\mathcal{L}}$. Consider the following events

$$\begin{aligned}
&\mathcal{E}_u = \{(X^n(W), U^n(J_W)) \notin \mathcal{T}_{\bar{\epsilon}}^n\},\\
&\mathcal{E}_v = \{(X^n(W), U^n(J_W), V^n(J_W, L_W)) \notin \mathcal{T}_{\bar{\epsilon}}^n\},\\
&\mathcal{E}_{yz} = \left\{(Y^n, Z^n, X^n(W), U^n(J_W), V^n(J_W, L_W)) \notin \mathcal{T}_{\bar{\epsilon}}^n\right\},\\
&\mathcal{E}_1 = \{|\boldsymbol{\mathcal{L}}_1| > 2^{n\hat{\Delta}}\},\\
&\mathcal{E}_2 = \{\exists l\neq L_W, l\in\mathfrak{B}(K_W), (Y^n, U^n(J_W), V^n(J_W, l)) \in \mathcal{T}_{\bar{\epsilon}}^n\},\\
&\mathcal{E}_3 = \left\{\exists(w', l_{w'}), w'\neq W, w'\in\boldsymbol{\mathcal{L}}_1, (Y^n, U^n(J_{w'}), V^n(J_{w'}, l_{w'})) \in \mathcal{T}_{\bar{\epsilon}}^n,\ l_w'\in\mathfrak{B}(K_{w'})\right\}.
\end{aligned} \tag{110}$$

Define

$$\mathcal{E} = \mathcal{E}_u \cup \mathcal{E}_v \cup \mathcal{E}_{yz} \bigcup_{i=1}^{3} \mathcal{E}_i, \tag{111}$$

to be the event that summarizes all "errors." By the covering lemma [12, Lemma 3.3] we obtain

$$\Pr(\mathcal{E}_u) = \frac{1}{M}\sum_i \Pr\big((X^n(i), U^n(J_i)) \notin \mathcal{T}_{\bar{\epsilon}}^n\big) \to 0 \tag{112}$$

if $\hat{R}_U > I(X;U)+\gamma_n$, where $\gamma_n>0$ and $\gamma_n\to 0$ as $n\to\infty$, since W is independent of both $\big(X^n(i)\big)_{i=1}^M$ and the codebook. Similarly, we have $\Pr(\mathcal{E}_u^c\cap\mathcal{E}_v)\to 0$ if $\hat{R}_V+\hat{R}_V' > I(X;V|U)+\gamma_n$. Due to the Markov lemma [12, p.27], we have

$$\Pr(\mathcal{E}_u^c \cap \mathcal{E}_v^c \cap \mathcal{E}_{yz}) \to 0. \tag{113}$$

For the sake of simplicity in the analysis of the last three events we use the symmetric property of our problem. Due to symmetry, it is sufficient to condition on the event $\{W=1\}$. Following the analysis in [12, Section 11.3] we have

$$\Pr(\mathcal{E}_2|W=1) \to 0 \tag{114}$$

as $n\to\infty$ if $\hat{R}_V' < I(Y;V|U)-\gamma_n$. We focus on the two remaining events $\mathcal{E}_1$ and $\mathcal{E}_3$. For each $i\in\mathcal{M}$ define random variable

$$B_i = \chi_{\left\{(Z^n, U^n(J_i))\in\mathcal{T}_{\bar{\epsilon}}^n\right\}}. \tag{115}$$

Note that $\Pr(B_1=1|W=1)\to 1$ as $n\to\infty$. Hence, it is sufficient to consider the following probability

$$\begin{aligned}
&\Pr(B_1=1,\ |\boldsymbol{\mathcal{L}}_1|>2^{n\hat{\Delta}}|W=1) = \Pr\left\{B_1=1, \sum_{i=1}^{2^{n\hat{R}}} B_i > 2^{n\hat{\Delta}}\middle| W=1\right\}\\
&\le \Pr\left\{\sum_{i=2}^{2^{n\hat{R}}} B_i > 2^{n\hat{\Delta}}-1\middle| W=1\right\}\\
&\le \frac{\sum_{i=2}^{2^{n\hat{R}}}\mathbb{E}[B_i|W=1]}{2^{n\hat{\Delta}}-1} = \frac{\sum_{i=2}^{2^{n\hat{R}}}\Pr\{B_i|W=1\}}{2^{n\hat{\Delta}}-1} \overset{(\star)}{\le} \xi 2^{n(\hat{R}-\hat{\Delta})}2^{-n(I(Z;U)-\gamma_n)} \to 0
\end{aligned} \tag{116}$$

if $\hat{R} - \hat{\Delta} < I(Z;U) - \gamma_n$ where $\xi = (1 - 1/2^{n\hat{R}})/(1 - 1/2^{n\hat{\Delta}}) \to 1$ as $n \to \infty$. $(\star)$ is valid since conditioning on $W = 1$, Z^n is independent[9] of $U^n(J_i)$ for $i \in [2:M]$. Therefore, for $i \geq 2$

$$\begin{aligned}\Pr\{B_i|W=1\} &= \sum_{j_i}\sum_{u^n}\sum_{z^n\in\mathcal{T}_{\bar{\epsilon}}^n(Z|u^n)} P_{J_i}(j_i)P_{U^n(J_i)|J_i}(u^n|j_i)P(Z^n=z^n|W=1)\\ &\overset{(\star\star)}{\leq} \sum_{j_i}\sum_{u^n}\sum_{z^n\in\mathcal{T}_{\bar{\epsilon}}^n(Z|u^n)} P_{J_i}(j_i)P_{U^n(J_i)|J_i}(u^n|j_i)2^{-n(H(Z)-\gamma_n)}\\ &\leq 2^{-n(I(Z;U)-\gamma_n)}, \end{aligned} \tag{117}$$

where $(\star\star)$ holds since W is independent of Z^n and Z^n is iid according to the distribution P_Z. The expressions (113) and (116) imply that

$$\delta_{1,n} = \Pr(W \notin \boldsymbol{\mathcal{L}}) \to 0$$

as $n \to \infty$.

The probability of the last event can be bounded as

$$\begin{aligned}&\Pr(\mathcal{E}_3|W=1) = \Pr\Big\{\exists(w', l_{w'}),\ w'\neq 1,\ w'\in\boldsymbol{\mathcal{L}}_1, (Y^n, U^n(J_{w'}), V^n(J_{w'}, l_{w'}) \in \mathcal{T}_{\bar{\epsilon}}^n,\ l'_w \in \mathfrak{B}(K_{w'}))|W=1\Big\}\\ &\leq \Pr\Big\{\exists(w', l_{w'}),\ w'\neq 1, (Y^n, U^n(J_{w'}), V^n(J_{w'}, l_{w'})) \in \mathcal{T}_{\bar{\epsilon}}^n,\ l'_w\in\mathfrak{B}(K_{w'})|W=1\Big\}\\ &\leq \sum_{i=2}^{2^{n\hat{R}}} \Pr\{\exists l_i \in \mathfrak{B}(K_i), (Y^n, U^n(J_i), V^n(J_i, l_i)) \in \mathcal{T}_{\bar{\epsilon}}^n|W=1\}.\\ &\leq \sum_{i=2}^{2^{n\hat{R}}}\sum_{j_i,k_i}\sum_{l_i\in\mathfrak{B}(k_i)} P_{J_iK_i}(j_i,k_i)\Pr\{(Y^n, U^n(j_i), V^n(j_i,l_i)) \in \mathcal{T}_{\bar{\epsilon}}^n|W=1, J_i=j_i, K_i=k_i\}.\end{aligned}$$

Since for $i = 2, \ldots, M$

$$\begin{aligned}&\Pr\{(Y^n, U^n(j_i), V^n(j_i,l_i)) \in \mathcal{T}_{\bar{\epsilon}}^n|W=1, J_i=j_i, K_i=k_i\}\\ &\quad= \sum_{u^n,v^n}\sum_{y^n\in\mathcal{T}_{\bar{\epsilon}}^n(Y|u^n,v^n)} P_{U^n(j_i)V^n(j_i,l_i)|J_iK_i}(u^n,v^n|j_i,k_i)P(Y^n=y^n|W=1)\\ &\quad\overset{(a)}{\leq} 2^{-n(H(Y)-H(Y|U,V)-\gamma_n)} = 2^{-n(I(Y;U,V)-\gamma_n)},\end{aligned} \tag{118}$$

$\Pr(\mathcal{E}_3|W=1) \to 0$ if $\hat{R} + \hat{R}'_V < I(Y;U,V) - \gamma_n$, where (a) is valid due to the independence of Y^n and W. Since $\Pr(\mathcal{E}_1) \to 0$ and $\Pr(\mathcal{E}_3) \to 0$,

$$\delta_{2,n} = \Pr(\hat{W} \neq W) \to 0.$$

Moreover, due to the union bound

$$\Pr\{\mathcal{E}\} \to 0,\ \text{as } n \to \infty. \tag{119}$$

Given $\mathcal{E}^c$, we obtain

$$(1-\bar{\epsilon})\mathbb{E}[d(X, f(U,V,Y))] < d(X^n(W), \hat{X}^n) < (1+\bar{\epsilon})\mathbb{E}[d(X, f(U,V,Y))], \tag{120}$$

by the typical average lemma [12, p.26], which implies that $\left|d(X^n(W), \hat{X}^n) - D\right| < \bar{\epsilon}D$. Hence, choosing $\bar{\epsilon} \to 0$ as $n \to \infty$

$$\mathbb{E}\big[\left|d(X^n(W), \hat{X}^n) - D\right|\big] < \mathbb{E}\big[|d(X^n(W), \hat{X}^n) - D|\big|\mathcal{E}^c\big] + \Pr(\mathcal{E})(d_{\max} + D) = \delta_{3,n}. \tag{121}$$

Since $\Pr(\hat{W} \neq W) = \mathbb{E}[\chi_{\{\hat{W}\neq W\}}]$ and $\Pr(W \notin \boldsymbol{\mathcal{L}}) = \mathbb{E}[\chi_{\{W\notin\boldsymbol{\mathcal{L}}\}}]$, by the Selection Lemma [18, Lemma 2.2] there exists a codebook $\mathcal{H}_n$ such that

$$\begin{aligned}\Pr(\hat{W}\neq W|\mathcal{H}_n) < \delta_n, \qquad \Pr(W\notin\boldsymbol{\mathcal{L}}|\mathcal{H}_n) < \delta_n,\\ \mathbb{E}\big[|d(X^n(W), \hat{X}^n) - D|\big|\mathcal{H}_n\big] < \delta_n,\end{aligned} \tag{122}$$

where $\delta_n = 4\max(\{\delta_{i,n}\}_{i=1}^3)$. Since the space of codebooks is discrete,

$$\left|\mathbb{E}[d(X^n(W), \hat{X}^n)|\mathcal{H}_n] - D\right| \leq \mathbb{E}\big[|d(X^n(W), \hat{X}^n) - D|\big|\mathcal{H}_n\big],$$

[9]This can be explained in more details as follows. Conditioning on $W = 1$, Z^n is independent of $X^n(i)$ for $i \in [2:M]$ and the codebook, while $U^n(J_i)$ depends only on $X^n(i)$ and the codebook. Hence, we can use a single codebook for all users.

which implies that

$$\mathbb{E}[d(X^n(W), \hat{X}^n)|\mathcal{H}_n] < D + \delta_n. \tag{123}$$

In summary, given an $\epsilon > 0$ if the following conditions

$$\begin{aligned} R_1 > I(X;U), \quad R_2 + R'_V > I(X;V|U), \\ R'_V < I(Y;V|U), \\ R < R_L + I(Z;U), \\ R + R'_V < I(Y;U,V), \end{aligned} \tag{124}$$

hold, then there exists a data processing scheme that satisfies all the requirements of Definition 2 for sufficiently large n. By using Fourier-Motzkin elimination [12, Appendix D] to eliminate R'_V we obtain

$$\begin{aligned} R_1 > I(X;U), \quad R_2 > I(X;V|U,Y), \\ R_2 - R > I(X;V|U) - I(Y;U,V) \\ R < \min\{R_L + I(Z;U), I(Y;U,V)\}. \end{aligned} \tag{125}$$

In the next step we simplify the above region by a rate transfer argument. Assume that R'_1, R'_2, and Θ are positive numbers such that

$$\begin{aligned} &R'_1 - \Theta > I(X;U), \\ &R'_2 + \Theta > \max\Big\{I(X;V|U,Y), R + I(X;V|U) - I(Y;U,V)\Big\}. \end{aligned} \tag{126}$$

Herein, Θ is the rate transferred from storage Node 2 to storage Node 1. Since $I(X;U) \geq 0$, by (125) there exists an identification scheme such that $(R'_1 - \Theta, R'_2 + \Theta)$ is achievable for the given R. This implies the achievability[10] of (R'_1, R'_2) for the given R. Applying the Fourier-Motzkin approach for a second time to eliminate Θ, the achievable rate region is enlarged to

$$\begin{aligned} R'_1 \geq I(X;U), \\ R'_1 + R'_2 \geq I(X;U) + I(X;V|U,Y), \\ R'_1 + R'_2 - R \geq I(X;U,V|Y), \\ R \leq \min\{R_L + I(Z;U), I(Y;U,V)\}, \end{aligned} \tag{127}$$

since by definition, the achievable region is closed.

Appendix C
Proof of Corollary 1

Direct part: Rate tuples that fulfill the conditions given in (11) also satisfy the conditions given in (9) with $D = d_{\max}$ and an arbitrary deterministic mapping f. Hence they are achievable.
Converse part: Define

$$\begin{aligned} U_i &= (W, J_W, Z^{i-1}), \\ V_i &= (U_i, K_W, Y^{i-1}), \quad i \in [1:n]. \end{aligned} \tag{128}$$

Then $U_i - V_i - X_i(W) - Y_i - Z_i$ for all $i \in [1:n]$. The two first constraints on the compression rates can be derived shortly as

$$\begin{aligned} n(R_1 + \epsilon) &\geq \sum_{i=1}^{n} I(X_i(W); W, J_W, X(W)^{i-1}) \\ &\geq \sum_{i=1}^{n} I(X_i(W); U_i), \end{aligned} \tag{129}$$

and

$$n(R_1 + R_2 + \epsilon) \geq \sum_{i=1}^{n} I(X_i(W); W, J_W, K_W, X^{i-1}(W))$$

[10] For the given R and $\epsilon > 0$ assume that the corresponding scheme is $(\{\phi_{kn}\}_{k=1}^2, \{g_k\}_{k=1}^3)$, cf. Definition 1. Let $(\tilde{\mathcal{M}}_{21}, \tilde{\mathcal{M}}_{22})$ be a decomposition of $\mathcal{M}_2$ of the enrollment mapping ϕ_{2n} such that $\frac{1}{n}\log|\tilde{\mathcal{M}}_{21}| \leq \Theta + \epsilon/2$ and $\frac{1}{n}\log|\tilde{\mathcal{M}}_{22}| \leq R_2 + \epsilon/2$. For each user we can then store $(\mathcal{M}_1, \tilde{\mathcal{M}}_{21})$ in the first layer and $\tilde{\mathcal{M}}_{22}$ in the second layer. The processing mappings $\{\tilde{g}_k\}_{k=1}^3$ can be obtained from $\{g_k\}_{k=1}^3$ accordingly. Note that this process is possible since the second layer is always used in conjunction with the first layer, cf. (3) - (6).

$$= \sum_{i=1}^{n} I(X_i(W); W, J_W, K_W, X^{i-1}(W), Y^{i-1}, Z^{i-1})$$
$$\geq \sum_{i=1}^{n} I(X_i(W); U_i, V_i). \tag{130}$$

Following the same steps which lead to (23), we obtain

$$\begin{aligned} n(R-\epsilon) &\leq H(W) \\ &= I(W; Y^n, (J_{\mathcal{L}}, K_{\mathcal{L}})) + H(W|Y^n, (J_{\mathcal{L}}, K_{\mathcal{L}})) \\ &\overset{(*)}{\leq} I(W, J_W, K_W; Y^n) + 1 + \epsilon \log_2 M \\ &\leq \sum_{i=1}^{n} I(U_i, V_i; Y_i) + 1 + \epsilon \log_2 M, \end{aligned} \tag{131}$$

where $(*)$ holds due to the Markov chain $Z^n - Y^n - (W, \boldsymbol{J}, \boldsymbol{K})$. In addition, from (21) we obtain

$$\begin{aligned} n(R-\epsilon) &\leq \sum_{i=1}^{n} I(W, J_W, Z^{i-1}; Z_i) + n(R_L + \epsilon_n), \\ &= \sum_{i=1}^{n} I(Z_i; U_i) + n(R_L + \epsilon_n). \end{aligned} \tag{132}$$

The rest follows by defining a uniform random variable Q on the set $[1:n]$ and taking $\epsilon \to 0$ as in Theorem 1. The cardinality of $\mathcal{U}$ and $\mathcal{V}$ can be bounded similarly using the support lemma [12, Appendix C].

Appendix D
Proof of Proposition 1

The proof follows closely the one of Theorem 1 with some modifications.

Achievability: $2^{n\hat{R}_U}$ codewords $u^n(j)$ are generated as before. For each m we draw $2^{n\hat{R}_V}$ codewords $v^n(j,k)$ iid via the marginal $p_{V|U}$, i.e., no binning is used. The enrollment process follows accordingly. The identification process corresponding to Observer 2 works identically as the first stage in (107) while for Observer 1 the processing unit searches through *all* users to find the unique $\hat{w}$ such that

$$(y^n, u^n(j_{\hat{w}}), v^n(j_{\hat{w}}, k_{\hat{w}})) \in \mathcal{T}_\epsilon^n, \tag{133}$$

which leads to the following event in the analysis

$$\mathcal{E}_3' = \left\{ \exists w', w' \neq W, (Y^n, U^n(J_{w'}), V^n(J_{w'}, K_{w'})) \in \mathcal{T}_\epsilon^n \right\}.$$

Similarly, we have $\Pr(\mathcal{E}_3'|W=1) \to 0$ if $\hat{R} < I(Y; U, V) - \gamma_n$.

One might notice that the condition $X - Y - Z$ is not used in the achievability proof of Theorem 1. Hence, it can be concluded that the two stage processing in the achievability of Theorem 1 achieves the rate region of Proposition 1.

Converse: Define the random variables U_i and V_i as in (128). We also obtain the constraints as in (129), (130), and (132). To arrive at (131) we need the following modification

$$\begin{aligned} n(R-\epsilon) &\leq H(W) \\ &= I(W; Y^n, \boldsymbol{J}, \boldsymbol{K}) + H(W|Y^n, \boldsymbol{J}, \boldsymbol{K}) \\ &\overset{(\star)}{\leq} I(W, J_W, K_W; Y^n) + 1 + \epsilon \log_2 M, \end{aligned} \tag{134}$$

where $(\star)$ follows from the Fano's inequality and the requirement in (13).

Appendix E
Justification of the Gaussian setting

We provide a justification for Theorem 2 in several steps. In the first step we establish a supporting covering lemma, which bypasses the need of a Markov lemma for weak typicality[11]. In the next step we provide a coding scheme which is based on the adapted covering lemma. The analysis only highlights the important parts. Our approach resembles the one given in [9], [10] with a tweak in the "error" analysis. In more detail, in our coding approach we use weak typicality with an adapted covering lemma, the quantization of the reconstruction mapping and a distortion analysis as in Wyner-Ziv approach. However, it is interesting to note that we do not quantize the auxiliary random variables.

[11] Markov lemmas for continuous alphabets can be found in the works [15, Lemma 5] in the context of weak typicality and [19] in the sense of weak*-typicality. However, it is not obvious to extend Lemma 5 in [15] to multiple layers of auxiliary random variables used in our superposition coding scheme. Bounding the distortion level using the approach in [19] is difficult since the distortion measure is unbounded.

A. Preliminary

To differentiate between weak and strong typicality, given $0 < \delta < 1$ we denote the weakly typical set by $\mathcal{A}_\delta^n$ whose definition for a tuple of random variables $(X_1, \dots, X_k)$ with a joint probability density function $p_{X_1 X_2 \dots X_k}$ is given by [16, p. 521], [15, Lemma 3]

$$\mathcal{A}_\delta^n(X_1 \dots X_k) = \left\{ (x_1^n, \dots, x_k^n) \middle| \left| -\frac{1}{n} \log p_{X_\mathcal{S}}^n(x_\mathcal{S}^n) - h(X_\mathcal{S}) \right| < \delta, \ \forall \mathcal{S} \subseteq [1:k] \right\} \tag{135}$$

where $h(\cdot)$ denotes the differential entropy[12]. Some important properties of weakly typical sequences are given in the following:

- If $x_\mathcal{S}^n \in \mathcal{A}_\delta^n(X_\mathcal{S})$ then
$$2^{-n(h(X_\mathcal{S})+\delta)} \leq p_{X_\mathcal{S}}^n(x_\mathcal{S}^n) \leq 2^{-n(h(X_\mathcal{S})-\delta)}. \tag{136}$$
- If $\mathcal{S}_1 \cap \mathcal{S}_2 = \varnothing$ and $(x_{\mathcal{S}_1}^n, x_{\mathcal{S}_2}^n) \in \mathcal{A}_\delta^n(X_{\mathcal{S}_1 \cup \mathcal{S}_2})$ where $\mathcal{S}_1, \mathcal{S}_2 \subset [1:k]$ then
$$2^{-n(h(X_{\mathcal{S}_1}|X_{\mathcal{S}_2})+2\delta)} \leq p_{X_{\mathcal{S}_1}|X_{\mathcal{S}_2}}^n(x_{\mathcal{S}_1}^n|x_{\mathcal{S}_2}^n) \leq 2^{-n(h(X_{\mathcal{S}_1}|X_{\mathcal{S}_2})-2\delta)}. \tag{137}$$
- For $x_{\mathcal{S}_1}^n \in \mathcal{A}_\delta^n(X_{\mathcal{S}_1})$ then
$$\mathrm{Vol}(\mathcal{A}_\delta^n(X_{\mathcal{S}_2}|x_{\mathcal{S}_1}^n)) \leq 2^{n(h(X_{\mathcal{S}_2}|X_{\mathcal{S}_1})+2\delta)}, \tag{138}$$
where $\mathcal{A}_\delta^n(X_{\mathcal{S}_2}|x_{\mathcal{S}_1}^n)$ is the conditional typical set. Note that the left-hand side is zero if $x_{\mathcal{S}_1}^n \notin \mathcal{A}_\delta^n(X_{\mathcal{S}_1})$ as the set $\mathcal{A}_\delta^n(X_{\mathcal{S}_2}|x_{\mathcal{S}_1}^n)$ is empty in this case.

Assume that the tuple $(X^n, Y^n, Z^n, U^n, V^n)$ is generated iid from the joint density p_{XYZUV}. Then due to the weak law of large numbers we have the following properties

$$\Pr\{(Y^n, Z^n, U^n, V^n) \notin \mathcal{A}_\delta^n(YZUV)\} \to 0 \text{ as } n \to \infty, \tag{139}$$

as well as

$$\Pr\left\{ \left| d(X^n, g(U^n, V^n, Y^n)) - D \right| > \delta \right\} \to 0, \tag{140}$$

when we assume that $D = \mathbb{E}[d(X, g(U, V, Y))] < \infty$. As in [9] we define the following indicator function

$$\psi_n(x^n, y^n, z^n, u^n, v^n) = \begin{cases} 1 & \text{if } \left| d(x^n, g(u^n, v^n, y^n)) - D \right| > \delta, \text{ or } (y^n, z^n, v^n, u^n) \notin \mathcal{A}_\delta^n(YZUV) \\ 0 & \text{otherwise} \end{cases}. \tag{141}$$

Let $\delta_n = \mathbb{E}[\psi_n(X^n, Y^n, Z^n, U^n, V^n)]$, then due to the union bound, (139) and (140) we have

$$\delta_n \to 0, \quad \text{as } n \to \infty. \tag{142}$$

For brevity define[13]

$$\mathcal{S}_n^\delta = \{(x^n, u^n, v^n) \colon \eta_{XUV}(x^n, u^n, v^n) \leq \delta_n^{1/2}\}, \tag{143}$$

where

$$\begin{aligned} \eta_{XUV}(x^n, u^n, v^n) &= \mathbb{E}[\psi_n(x^n, Y^n, Z^n, u^n, v^n)|X^n = x^n, V^n = v^n, U^n = u^n] \\ &= \mathbb{E}[\psi_n(x^n, Y^n, Z^n, u^n, v^n)|X^n = x^n]. \end{aligned} \tag{144}$$

Due to the Markov inequality we have

$$\Pr\{(X^n, U^n, V^n) \notin \mathcal{S}_n^\delta\} \leq \frac{\mathbb{E}[\psi_n(X^n, Y^n, Z^n, U^n, V^n)]}{\delta_n^{1/2}} = \delta_n^{1/2}. \tag{145}$$

Finally, define[14]

$$\mathcal{B}_n^\delta = \mathcal{A}_\delta^n(UVX) \cap \mathcal{S}_n^\delta, \tag{146}$$

and $\mathcal{B}_n^\delta(x^n)$, $\mathcal{B}_n^\delta(x^n, u^n)$ as sections of $\mathcal{B}_n^\delta$ corresponding to the sequence x^n and on the pair (x^n, u^n), respectively. Note that $\mathcal{B}_n^\delta(x^n)$ can be the empty set. A similar statement can be made about $\mathcal{B}_n^\delta(x^n, u^n)$. The following lemma is useful for analyzing the coding scheme that is presented in the next subsection.

[12] Note that $\mathcal{A}_\delta^n(X_1 \dots X_k)$ is a Borel-measurable set.
[13] Since ψ_n is a Borel measurable mapping, η_{XUV} is also Borel measurable. Hence, the set $\mathcal{S}_n^\delta$ is Borel measurable.
[14] Hence $\mathcal{B}_n^\delta$ is a Borel measurable set as it is the intersection of two measurable sets, cf. Footnotes 12 and 13.

Lemma 2. *Assume that $X^n \sim p_X^n$. Generate M codewords $u^n(j)$ iid according to the marginal density p_U, where $M \geq 2^{nR_U}$. For each $u^n(j)$ draw L codewords $v^n(j,l)$ via the conditional density $p_{V|U}$, where $L \geq 2^{nR_V}$. Then for a given δ, where $0 < \delta < 1$,*

$$\Pr\Big\{(X^n, U^n(j), V^n(j,l)) \notin \mathcal{B}_n^\delta,\ \forall j,l\Big\} \to 0 \tag{147}$$

as $n \to \infty$ if $R_U \geq I(X;U) + 4\delta$ and $R_V \geq I(X;V|U) + 5\delta$.

Proof: For notational brevity we suppress the superscript δ in $\mathcal{B}_n^\delta$ in the rest of this subsection. It is sufficient to prove the lemma for $R_U = I(X;U) + 4\delta$, $R_V = I(X;V|U) + 5\delta$ and $2^{nR_U} \leq M \leq \hat{M} = 2^{2nR_U}$. We first expand the left-hand side of (147) as[15]

$$\Pr\Big\{(X^n, U^n(j), V^n(j,l)) \notin \mathcal{B}_n,\ \forall j,l\Big\} = \int p_X(x^n)\Pr\Big\{(U^n(j), V^n(j,l)) \notin \mathcal{B}_n(x^n),\ \forall j,l\Big\} dx^n. \tag{148}$$

The second term inside the integral can be decomposed as

$$\begin{aligned}
&\Pr\Big\{(U^n(j), V^n(j,l)) \notin \mathcal{B}_n(x^n),\ \forall j,l\Big\}\\
&\overset{(a)}{=} \prod_{j=1}^{M} \Pr\{(U^n(j), V^n(j,l)) \notin \mathcal{B}_n(x^n),\ \forall l\}\\
&\overset{(b)}{=} \Big\{\Pr\{(U^n(1), V^n(1,l)) \notin \mathcal{B}_n(x^n),\ \forall l\}\Big\}^M,
\end{aligned} \tag{149}$$

where (a) is valid due to the independence of tuples $\Big(U^n(j), \big(V^n(j,l)\big)_l\Big)_j$ for all j. (b) holds due to the iid of the codebook. Note that

$$\mathcal{B}_n(x^n) = \varnothing \implies \Pr\{(U^n(1), V^n(1,l)) \notin \mathcal{B}_n(x^n),\ \forall l\} = 1. \tag{150}$$

Otherwise, we define

$$\begin{aligned}
&\mathcal{C}_n(x^n) = \{u^n \mid u^n \in \mathcal{A}_\delta^n(U|x^n),\ \text{and}\ \{v^n \colon v^n \in \mathcal{B}_n(x^n, u^n)\} \neq \varnothing\}\\
&\mathcal{C}_n^c(x^n) = \mathcal{U}^n \backslash \mathcal{C}_n(x^n) = \mathbb{R}^n \backslash \mathcal{C}_n(x^n).
\end{aligned} \tag{151}$$

Then, for each x^n such that $\mathcal{B}_n(x^n) \neq \varnothing$ the following holds

$$\begin{aligned}
&\Pr\{(U^n(1), V^n(1,l)) \notin \mathcal{B}_n(x^n),\ \forall l\}\\
&= \int_{\mathcal{C}_n^c(x^n)} p_U^n(u^n) du^n + \int_{\mathcal{C}_n(x^n)} p_U^n(u^n) \Pr\{V^n(1,l) \notin \mathcal{B}_n(x^n, u^n),\ \forall l | U^n(1) = u^n\} du^n\\
&= \int_{\mathcal{C}_n^c(x^n)} p_U^n(u^n) du^n + \int_{\mathcal{C}_n(x^n)} p_U^n(u^n) \prod_{l=1}^{L} \Pr\{V^n(1,l) \notin \mathcal{B}_n(x^n, u^n) | U^n(1) = u^n\} du^n\\
&\overset{(b)}{=} \int_{\mathcal{C}_n^c(x^n)} p_U^n(u^n) du^n + \int_{\mathcal{C}_n(x^n)} p_U^n(u^n) \Big\{\Pr\{V^n(1,1) \notin \mathcal{B}_n(x^n, u^n) | U^n(1) = u^n\}\Big\}^L du^n,
\end{aligned} \tag{152}$$

where (b) holds due to the iid of the codebook. Moreover, for $u^n \in \mathcal{C}_n(x^n)$,

$$\begin{aligned}
&\Pr\{V^n(1,1) \notin \mathcal{B}_n(x^n, u^n) | U^n(1) = u^n\}\\
&= 1 - \Pr\{V^n(1,1) \in \mathcal{B}_n(x^n, u^n) | U^n(1) = u^n\}\\
&= 1 - \int_{\mathcal{B}_n(x^n, u^n)} p_{V|U}^n(v^n|u^n) dv^n.
\end{aligned} \tag{153}$$

From the definition of $\mathcal{B}_n$ for each $v^n \in \mathcal{B}_n(x^n, u^n)$ we have

$$\begin{aligned}
\frac{p_{V|U}^n(v^n|u^n)}{p_{V|UX}^n(v^n|u^n, x^n)} &\geq \frac{2^{-n(h(V|U)+2\delta)}}{2^{-n(h(V|U,X)-2\delta)}}\\
&= 2^{-n(I(X;V|U)+4\delta)}.
\end{aligned} \tag{154}$$

This implies that for $u^n \in \mathcal{C}_n(x^n)$ we have the following inequality

$$\Pr\{V^n(1,1) \notin \mathcal{B}_n(x^n, u^n) | U^n(1) = u^n\}$$

[15]Note that herein dx^n is a friendly notation for $d\lambda^{\otimes n}$, i.e., we are considering the product of Lebesgue measures.

$$\leq 1 - 2^{-n(I(X;V|U)+4\delta)} \int_{\mathcal{B}_n(x^n,u^n)} p^n_{V|UX}(v^n|u^n,x^n)dv^n. \tag{155}$$

Therefore, for $u^n \in \mathcal{C}_n(x^n)$ the second integrand of the second integral in (152) is bounded as

$$\begin{aligned}
&\{\Pr\{V^n(1,1) \notin \mathcal{B}_n(x^n,u^n)|U^n(1)=u^n\}^L \\
&\leq \left(1 - 2^{-n(I(X;V|U)+4\delta)} \int_{\mathcal{B}_n(x^n,u^n)} p^n_{V|UX}(v^n|u^n,x^n)dv^n\right)^L \\
&\overset{(*)}{\leq} 1 - \int_{\mathcal{B}_n(x^n,u^n)} p^n_{V|UX}(v^n|u^n,x^n)dv^n + \exp\left(-L2^{-n(I(X;V|U)+4\delta)}\right) \\
&\overset{(c)}{\leq} 1 - \int_{\mathcal{B}_n(x^n,u^n)} p^n_{V|UX}(v^n|u^n,x^n)dv^n + \exp(-2^{n\delta}),
\end{aligned} \tag{156}$$

where (c) follows from the definition of L. In $(*)$ we use the following inequality [16, Lemma 10.5.3]

$$(1-xy)^n \leq 1 - x + e^{-yn}, \tag{157}$$

where $0 \leq x, y \leq 1$, and $n > 0$. Thus, when $\mathcal{B}_n(x^n) \neq \varnothing$,

$$\begin{aligned}
&\Pr\{(U^n(1), V^n(1,l)) \notin \mathcal{B}_n(x^n),\ \forall l\} \\
&\leq \int_{\mathcal{C}^c_n(x^n)} p^n_U(u^n)du^n + \int_{\mathcal{C}_n(x^n)} p^n_U(u^n)\left(1 - \int_{\mathcal{B}_n(x^n,u^n)} p^n_{V|UX}(v^n|u^n,x^n)dv^n + \exp(-2^{n\delta})\right)du^n \\
&= 1 + \exp(-2^{n\delta}) \int_{\mathcal{C}_n(x^n)} p^n_U(u^n)du^n - \int_{\mathcal{C}_n(x^n)} p^n_U(u^n) \int_{\mathcal{B}_n(x^n,u^n)} p^n_{V|UX}(v^n|u^n,x^n)dv^n du^n \\
&\overset{(d)}{\leq} 1 + \exp(-2^{n\delta}) - 2^{-n(I(X;U)+3\delta)} \int_{\mathcal{B}_n(x^n)} p^n_{VU|X}(v^n,u^n|x^n)du^n dv^n \\
&= 1 + \exp(-2^{n\delta}) - 2^{-n(I(X;U)+3\delta)}\Pr\{(U^n,V^n) \in \mathcal{B}_n(x^n)|X^n = x^n\},
\end{aligned} \tag{158}$$

where (d) follows since for $u^n \in \mathcal{C}_n(x^n)$ we have

$$\begin{aligned}
\frac{p^n_U(u^n)}{p^n_{U|X}(u^n|x^n)} &\geq \frac{2^{-n(h(U)+\delta)}}{2^{-n(h(U|X)-2\delta)}} \\
&= 2^{-n(I(X;U)+3\delta)}.
\end{aligned} \tag{159}$$

Finally,

$$\begin{aligned}
&\Pr\left\{(U^n(j), V^n(j,l)) \notin \mathcal{B}_n(x^n),\ \forall j,l\right\} \\
&\leq \left(1 + \exp(-2^{n\delta}) - 2^{-n(I(X;U)+3\delta)}\Pr\{(U^n,V^n) \in \mathcal{B}_n(x^n)|X^n = x^n\}\right)^M \\
&= (1+\exp(-2^{n\delta}))^M\left(1 - 2^{-n(I(X;U)+3\delta)}\frac{\Pr\{(U^n,V^n) \in \mathcal{B}_n(x^n)|X^n = x^n\}}{1+\exp(-2^{n\delta})}\right)^M \\
&\overset{(*)}{\leq} (1+\exp(-2^{n\delta}))^M\left(1 - \frac{\Pr\{(U^n,V^n) \in \mathcal{B}_n(x^n)|X^n = x^n\}}{1+\exp(-2^{n\delta})} + \exp(-M2^{-n(I(X;U)+3\delta)})\right) \\
&\leq (1+\exp(-2^{n\delta}))^{\hat{M}}\left(1 - \frac{\Pr\{(U^n,V^n) \in \mathcal{B}_n(x^n)|X^n = x^n\}}{1+\exp(-2^{n\delta})} + \exp(-2^{n\delta})\right),
\end{aligned} \tag{160}$$

where $(*)$ has the same explanation as before. From equation (150) we observe that the bound in (160) holds as well for the case $\mathcal{B}_n(x^n) = \varnothing$. Furthermore, note that

$$(1+\exp(-2^{n\delta}))^{\hat{M}} \to 1 \tag{161}$$

as $n \to \infty$ which will be pointed out in the following. Define $\beta = 2^{n\delta}$ which implies that $\hat{M} = 2^{2n(I(X;U)+3\delta)} = \beta^\alpha$ where $\alpha = 2\frac{I(X;U)+3\delta}{\delta} > 0$. Also as $n \to \infty$, $\beta \to \infty$. It suffices to show that

$$\lim_{\beta\to\infty} \beta^\alpha \ln(1+e^{-\beta}) = 0, \tag{162}$$

which can be concluded from L'Hospital's rule. Next, we average over x^n which gives us

$$\Pr\left\{(X^n, U^n(j), V^n(j,l)) \notin \mathcal{B}_n,\ \forall j,l\right\}$$

$$\leq (1+\exp(-2^{n\delta}))^{\hat{M}}\left(1+\exp(-2^{n\delta})-\frac{\Pr\{(U^n,V^n,X^n)\in\mathcal{B}_n\}}{1+\exp(-2^{n\delta})}\right). \tag{163}$$

The fact that

$$\Pr\{(U^n,V^n,X^n)\in\mathcal{B}_n\}\to 1, \tag{164}$$

follows from

$$\Pr\{(U^n,V^n,X^n)\in\mathcal{A}_\delta^n(UVX)\}\to 1,\ \text{as } n\to\infty, \tag{165}$$

and (145). In conclusion we obtain

$$\Pr\left\{(X^n,U^n(j),V^n(j,l))\notin\mathcal{B}_n,\ \forall j,l\right\}\to 0,\ \text{as } n\to\infty. \tag{166}$$

■

B. A coding scheme

As in the discrete case we begin with the codebook construction. Given $0<\delta<1$, whose value is determined later, fix a conditional density $p_{UV|X}$ and a measurable mapping $g\colon\mathbb{R}^3\to\mathbb{R}$ such that

$$\mathbb{E}[d(X,g(U,V,Y))]=D<\infty,\ \text{and } I(Y;V|U)>0. \tag{167}$$

We will discuss the degenerate case where $I(Y;V|U)=0$ at the end of this subsection. We note that since the distortion measure is the squared error distance, there exists a measurable quantization mapping $f\colon\hat{\mathcal{X}}\to\{\hat{x}_i\}_{i=1}^N\subset\hat{\mathcal{X}}$, with N sufficiently large and $\hat{\mathcal{X}}=\mathbb{R}$ such that [10, Eq. 2.11]

$$\hat{D}=\mathbb{E}[d(X,f(g(U,V,Y)))]\leq(1+\delta)D. \tag{168}$$

Define[16]

$$\hat{g}=f\circ g. \tag{169}$$

With abuse of notation, we define $\mathcal{B}_n^\delta$ as before with $\hat{g}$ in place of g and $\hat{D}$ in place of D.
Additionally[17], we show in the following that there exist a deterministic mapping and an auxiliary random variable which produce the same effect as drawing an element from a set uniformly at random. We use the mapping and random variable in our formal coding scheme to show that the resulting mappings are measurable. Let $\mathcal{T}$ be the set of of all pairs (i,j) where $i\in[1:2^{nR_U}]$ and $j\in[1:2^{n\bar{R}_V}]$. The corresponding power set is $2^{\mathcal{T}}$. For each set $\mathcal{E}\in 2^{\mathcal{T}}$, we select one element of $\mathcal{E}$ uniformly at random if $\mathcal{E}\neq\varnothing$. Otherwise, we select one element of $\mathcal{T}$ uniformly at random. The corresponding conditional pmf is given by $\{P_{\mathcal{E}}(t)\mid t\in\mathcal{T}\}$. For each n by the functional representation lemma [12, Appendix B] there exists a discrete random variable $\hat{T}$, defined on the corresponding finite alphabet $\hat{\mathcal{T}}$, and a function $\hat{\psi}\colon 2^{\mathcal{T}}\times\hat{\mathcal{T}}\to\mathcal{T}$ such that

$$\hat{\psi}(\mathcal{E},\hat{T})\sim P_{\mathcal{E}},\ \forall\mathcal{E}\in 2^{\mathcal{T}}. \tag{170}$$

Codebook generation: We generate a *single* codebook for all users which consists of 2^{nR_U} iid sequence $u^n(j)$ from the marginal pdf p_U. For each j, $2^{n\bar{R}_V}$ codewords $v^n(j,l)$ are drawn iid from the conditional pdf $p_{V|U}$. Each index l is parsed into a unique pair (k,k'), where $k\in[1:2^{nR_V}]$, $k'\in[1:2^{nR'_V}]$ and $\bar{R}_V=R_V+R'_V$, i.e., k is the corresponding bin index of l where the bin is given as in (105). We also fix two sequences u_e^n and v_e^n corresponding to the error message $\{e\}$.
Enrollment: Given $x^n(i)$ where $i\in\mathcal{M}$, we search for the set $\mathcal{I}_i$ which is determined as

$$\mathcal{I}_i=\left\{(j_i,l_i)\mid(x^n(i),u^n(j_i),v^n(j_i,l_i))\in\mathcal{B}_n^\delta,\ j_i\in[1:2^{nR_U}],\ l_i\in[1:2^{n\bar{R}_V}]\right\}. \tag{171}$$

If the set $\mathcal{I}_i$ is not empty then we select a tuple (j_i,l_i) uniformly at random from $\mathcal{I}_i$. Otherwise, (j_i,l_i) is selected uniformly from the set of all pairs. Formally the action is described by $\hat{\psi}(\mathcal{I}_i,\hat{t})$ as in (170) where $\hat{t}$ is the corresponding realization of $\hat{T}$. We store j_i in the first layer and the bin index k_i in the second layer[18].
Identification and Reconstruction: The two stage identification works similarly as in the discrete case with the following modification. Condition (107) is replaced by

$$(z^n,u^n(j_i))\in\mathcal{A}_\delta^n(ZU). \tag{172}$$

[16] Note that $\hat{g}$ is a measurable mapping since it is a composition of two measurable mappings.

[17] By our restrictions, all mappings are required to be deterministic and measurable. However, in our proof we use randomization in the encoding step to simplify the analysis. Hence, the existence of the mapping and the auxiliary random variable allow us to perfom derandomization in the last step. Moreover, the output sequence is also a random vector since it is the ouput of the combination of deterministic transformations whose inputs are random vectors.

[18] We note that this encoding scheme is different from the one in Section II since the first layer message j_i is chosen after searching through codeword sequences in all layers. In contrast, in the discrete case the stored index in the first layer of the i-th user is chosen based only on the codewords in the first layer $\left(u^n(j)\right)_{j=1}^{2^{nR_U}}$.

Condition (108) is replaced by searching for a unique $\hat{w}$ such that

$$(y^n, u^n(j_{\hat{w}}), v^n(j_{\hat{w}}, \tilde{l})) \in \mathcal{A}_\delta^n(YUV). \tag{173}$$

for some $\tilde{l} \in \mathfrak{B}(k_{\hat{w}})$. Condition (109) is changed to searching for a unique $\tilde{l} \in \mathfrak{B}(k_{\hat{w}})$ when $\hat{w} \neq e$ such that

$$(y^n, u^n(j_{\hat{w}}), v^n(j_{\hat{w}}, \tilde{l})) \in \mathcal{A}_\delta^n(YUV). \tag{174}$$

If $\hat{w} = e$ we set $\tilde{l} = 1$. When $\hat{w} = e$ or $\tilde{l} = e$, we set $u^n(j_{\hat{w}}) = u_e^n$ and $v^n(j_{\hat{w}}, \tilde{l}) = v_e^n$. Then the processing center outputs the corresponding sequence $\hat{x}_\tau = \hat{g}(u_\tau(j_{\hat{w}}), v_\tau(j_{\hat{w}}, \tilde{l}), y_\tau)$ for all $\tau = [1:n]$ where $\hat{g}$ is defined in (169).

Properness of our coding scheme:

Roughly speaking, in each of the aforementioned steps the action consists of a combination of mappings whose pre-image of a Borel set is a finite intersections, or/and unions, of Borel sets. Hence the resulting mappings are measurable. The details are given in the following. We only need to show that mappings whose input arguments contain elements of $\mathbb{R}$ are measurable[19]. For notation brevity we define $\Xi = 1 + 2^{n(R_U + \bar{R}_V)}$, $\Upsilon = 1 + 2^{nR_U}$, $\boldsymbol{u}^n = (u^n(1), \ldots, u^n(2^{nR_U}))$ and $\boldsymbol{v}^n = (v^n(1,1), \ldots, v^n(2^{nR_U}, 2^{n\bar{R}_V}))$.

- We first show that the mappings from the users' data sequences and codebook to the stored indices are jointly measurable. For the sake of clarity, we focus on the first user. Consider the set of mappings $\{\psi_{i,j}\}$ where $i \in [1 : 2^{nR_U}]$ and $j \in [1 : 2^{n\bar{R}_V}]$ each is defined as

$$\begin{aligned} \psi_{i,j} &\colon \mathbb{R}^{n\times\Xi} \to \{*, (i,j)\} \\ \psi_{i,j}(x^n(1), \boldsymbol{u}^n, \boldsymbol{v}^n) &\mapsto \begin{cases} (i,j) & \text{if } (x^n(1), u^n(i), v^n(i,j)) \in \mathcal{B}_n^\delta \\ * & \text{otherwise} \end{cases} \end{aligned}$$

  where $*$ is a dummy symbol. Then each $\psi_{i,j}$ is a measurable mapping since the pre-image

$$\psi_{i,j}^{-1}((i,j)) = \{(x^n(1), \boldsymbol{u}^n, \boldsymbol{v}^n) \mid (x^n(1), u^n(i), v^n(i,j)) \in \mathcal{B}_n^\delta, \text{ other codewords take values in } \mathbb{R}^n\}, \tag{175}$$

  is a Borel set. Hence the map

$$\begin{aligned} \psi = (\psi_{1,1}, \ldots, \psi_{2^{nR_U}, 2^{n\bar{R}_V}}) &\colon \mathbb{R}^{n\times\Xi} \to \bigtimes_{i,j} \{*, (i,j)\} \\ \psi(x^n(1), \boldsymbol{u}^n, \boldsymbol{v}^n) &\mapsto \hat{\mathcal{I}}_1, \end{aligned} \tag{176}$$

  is a measurable mapping. The one-to-one correspondence π_0 between the vector $\hat{\mathcal{I}}_1$ and the set of suitable pairs $\mathcal{I}_1$ given in (171) by eliminating all $*$, e.g., for $\hat{\mathcal{I}}_1 = \big(*, (1,2), *, (1,4), *, \ldots, *\big)$

$$\mathcal{I}_1 = \pi_0(\hat{\mathcal{I}}_1) = \{(1,2), (1,4)\}, \tag{177}$$

  is obviously measurable. Let $\mathcal{T}$, $\hat{\psi}$ and $\hat{\mathcal{T}}$ be defined as in (170) then the map

$$\begin{aligned} \phi &\colon \mathbb{R}^{n\times\Xi} \times \hat{\mathcal{T}} \to \mathcal{T} \\ \phi(x^n(1), \boldsymbol{u}^n, \boldsymbol{v}^n, \hat{t}) = \hat{\psi}(\pi_0(\psi(x^n(1), \boldsymbol{u}^n, \boldsymbol{v}^n)), \hat{t}) &\mapsto (j_1, l_1), \end{aligned} \tag{178}$$

  which is our selection map, is

$$\big(\mathbb{R}^{n\times\Xi} \times \hat{\mathcal{T}}, \mathcal{B}(\mathbb{R}^{n\times\Xi}) \times 2^{\hat{\mathcal{T}}}\big) \to (\mathcal{T}, 2^{\mathcal{T}})$$

  measurable. We note that the mappings from the chosen pair to the stored pair are projections, hence measurable. In summary we show that the encoding mappings are measurable.

- To show that forming the list induces a measurable mapping, consider the following set of mappings $\{\hat{g}_{1i}\}_{i=1}^M$ where for each i, $\hat{g}_{1i}$ is defined as

$$\begin{aligned} \hat{g}_{1i} &\colon \mathbb{R}^{n\times\Upsilon} \times \mathcal{M}_1^M \to \{*, i\} \\ \hat{g}_{1i}(z^n, \boldsymbol{u}^n, \boldsymbol{j}) &\mapsto \begin{cases} i & \text{if } (z^n, u^n(j_i)) \in \mathcal{A}_\delta^n(ZU) \\ * & \text{otherwise} \end{cases}. \end{aligned} \tag{179}$$

  Since $\mathcal{A}_n^\delta(ZU)$ is a Borel set, it can be seen that the map

$$\begin{aligned} \hat{g}_1 = (\hat{g}_{11}, \ldots, \hat{g}_{1M}) &\colon \mathbb{R}^{n\times\Upsilon} \times \mathcal{M}_1^M \to \hat{\mathfrak{L}} = \bigtimes_{i=1}^M \{*, i\} \\ \hat{g}_1(z^n, \boldsymbol{u}^n, \boldsymbol{j}) &\mapsto \hat{\mathcal{L}}. \end{aligned} \tag{180}$$

[19]Mappings which map finite input alphabets to finite output alphabets are obviously measurable since the corresponding Borel σ-algebras are power sets.

is jointly measurable. Next, let π_1 be defined as

$$\pi_1\colon \hat{\mathfrak{L}} \to \mathfrak{L}$$
$$\pi_1(\hat{\mathcal{L}}) \mapsto \begin{cases} \mathcal{L} & \text{by eliminating all } * \text{ and if } 1 \le |\mathcal{L}| \le 2^{n\Delta} \\ \{e\} & \text{otherwise} \end{cases}.$$

Since π_1 is a mapping from a discrete set to another discrete set, it is measurable w.r.t. the power set σ-algebra. Hence the map $\bar{g}_1 = \pi_1 \circ \hat{g}_1$ is a jointly measurable on

$$(\mathbb{R}^{n\times\Upsilon} \times \mathcal{M}_1^M, \mathcal{B}(\mathbb{R}^{n\times\Upsilon}) \times 2^{\mathcal{M}_1^M}) \to (\mathfrak{L}, 2^{\mathfrak{L}}).$$

Our first stage processing map g_1 can be obtain from $\bar{g}_1$ once a set of codewords is fixed.

- Similarly, for user identification in the second stage we look at the following set mappings $\{\hat{g}_{2i}\}_{i=1}^M$, whereas each is defined as

$$\hat{g}_{2i}\colon \mathbb{R}^{n\times\Xi} \times \mathfrak{M}_{12} \to \{*, i\}$$
$$\hat{g}_{2i}(y^n, \boldsymbol{u}^n, \boldsymbol{v}^n, (j_{\mathcal{L}}, k_{\mathcal{L}})) \mapsto \begin{cases} i & \text{if } i \in \mathcal{L} \text{ and } (y^n, u^n(j_i), v^n(j_i, l)) \in \mathcal{A}_\delta^n(YUV) \\ & \text{for some } l \in \mathfrak{B}(k_i) \\ * & \text{otherwise} \end{cases}.$$

We observe that for each i the mapping $\hat{g}_{2i}$ is jointly measurable. Next we need the mapping

$$\pi_2\colon \bigtimes_{i=1}^{M} \{*, i\} \to \mathcal{W} \cup \{e\}$$
$$\pi_2(\boldsymbol{\alpha}) \mapsto \begin{cases} \hat{w} & \text{if it is the only non-}* \text{ element in } \boldsymbol{\alpha} \\ e & \text{otherwise} \end{cases}.$$

The second stage identification mapping g_2 can be obtained from $\bar{g}_2 = \pi_2 \circ \left((\hat{g}_{2i})_{i=1}^M\right)$ once a set of codewords is fixed.

- Finally, to describe the reconstruction mapping g_3 we need mappings $\hat{g}_3$ and π_3 which are defined in the following. Let $\hat{\mathfrak{M}}'_{12} = \mathcal{M}_1 \times \left\{[1:2^{n\bar{R}_V}] \cup \{e\}\right\} \times \left(\mathcal{W} \cup \{e\}\right)$. The mapping $\hat{g}_3$ searches for the unique second layer index $\tilde{l}$ of the chosen user $\hat{w}$, which has the bin index $k_{\hat{w}}$, and is defined formally as

$$\hat{g}_3\colon \mathbb{R}^{n\times\Xi} \times \hat{\mathfrak{M}}_{12} \to \hat{\mathfrak{M}}'_{12}$$
$$\hat{g}_3(y^n, \boldsymbol{u}^n, \boldsymbol{v}^n, (j_{\hat{w}}, k_{\hat{w}}, \hat{w})) \mapsto \begin{cases} (j_{\hat{w}}, \tilde{l}, \hat{w}) & \text{if } \hat{w} \ne e \text{ and } \tilde{l} \text{ is unique such that} \\ & (y^n, u^n(j_{\hat{w}}), v^n(j_{\hat{w}}, \tilde{l})) \in \mathcal{A}_\delta^n(YUV) \\ & \text{as well as } \tilde{l} \in \mathfrak{B}(k_{\hat{w}}) \\ (1, e, \hat{w}) & \text{if } \hat{w} \ne e \\ (1, 1, e) & \text{if } \hat{w} = e \end{cases}$$

where $\hat{\mathfrak{M}}_{12}$ is defined in (5). The mapping π_3 outputs the corresponding codeword pair $(u^n(j_{\hat{w}}), v^n(j_{\hat{w}}, \tilde{l}))$ given the input tuple $(j_{\hat{w}}, \tilde{l}, \hat{w})$ and the codebook. It is defined as

$$\pi_3\colon \mathbb{R}^{n\times(\Xi-1)} \times \hat{\mathfrak{M}}'_{12} \to \mathbb{R}^{2n}$$
$$\pi_3(\boldsymbol{u}^n, \boldsymbol{v}^n, (j_{\hat{w}}, \tilde{l}, \hat{w})) \mapsto \begin{cases} (u^n(j_{\hat{w}}), v^n(j_{\hat{w}}, \tilde{l})) & \text{if } \hat{w} \ne e \text{ and } \tilde{l} \ne e \\ (u_e^n, v_e^n) & \text{otherwise} \end{cases}. \tag{181}$$

The measurable properties of $\hat{g}_3$ and π_3 can be shown similarly as the ones of $\hat{g}_1$ and $\hat{g}_2$. The reconstruction mapping g_3 can be obtained from $\bar{g}_3(\cdot, y^n) = \hat{g}(\pi_3(\cdot, \hat{g}_3(\cdot)), y^n)$, where $\hat{g}$, which has a finite output alphabet and is defined in (169), is applied symbolwisely.

Analysis: Let J_i and L_i, $i \in \mathcal{W}$, be the chosen indices for the i-th user. Furthermore, let $\boldsymbol{\mathcal{L}}_1$ be the list of indices $i \in \mathcal{W}$ that satisfy (172) in the first stage of the identification process.
Denoted by $\boldsymbol{\mathcal{H}}$ the random variable which represents the randomly generated codebook, i.e.,

$$\boldsymbol{\mathcal{H}} = \left\{(U^n(j), V^n(j,l)) \mid i \in \mathcal{M}, j \in [1:2^{nR_U}], l \in [1:2^{n\bar{R}_V}]\right\}, \tag{182}$$

and its realization by $\mathcal{H}$. The Markov relation

$$(Y^n, Z^n) - X^n(W) - (W, J_W, L_W, \boldsymbol{\mathcal{H}})$$

follows by our coding scheme. However, for the error analysis we need the Markov relation in form of density terms.

Claim 3. *For each triple (w, j_w, l_w), the function*

$$\begin{aligned}&p_{X^n(W)Y^n\boldsymbol{\mathcal{H}}|J_WL_WW}(x^n, y^n, \mathcal{H}|j_w, l_w, w)\\&= \frac{\Pr\{J_w = j_w, L_w = l_w, W = w|X^n(W) = x^n, \boldsymbol{\mathcal{H}} = \mathcal{H}\}}{P_{J_WL_WW}(j_w, l_w, w)} \times p^n_X(x^n)p^n_{Y|X}(y^n|x^n)p(\boldsymbol{\mathcal{H}} = \mathcal{H})\end{aligned} \tag{183a}$$

is a conditional density function of the distribution $\mu(B, w, j_w, l_w) = \Pr\big\{(X^n(W), Y^n, \boldsymbol{\mathcal{H}}) \in B|J_w = j_w, L_w = l_w, W = w\big\}$ w.r.t. the product of Lebesgue measures $\lambda^{\otimes n(2+2^{n(R_U+\bar{R}_V)})}$, where $B \in \mathcal{B}\big(\mathbb{R}^{n\times(2+2^{n(R_U+\bar{R}_V)})}\big)$ is a Borel set. It can also be argued that this function is jointly measurable in $(x^n, y^n, \mathcal{H}, j_w, l_w, w)$.

Proof: It is immediate from the definition of $p_{X^n(W)Y^n\boldsymbol{\mathcal{H}}|J_WL_WW}$ in (183a) that it is a jointly measurable function in $(x^n, y^n, \mathcal{H})$. Lemma 1 implies the following relation

$$\begin{aligned}&\Pr\{J_w = j_w, L_w = l_w, W = w|X^n(W) = x^n, \boldsymbol{\mathcal{H}} = \mathcal{H}\}\\&\qquad= \Pr\{J_w = j_w, L_w = l_w, W = w|X^n(W) = x^n, Y^n = y^n, \boldsymbol{\mathcal{H}} = \mathcal{H}\}\\&\qquad P_{X^n(W)Y^n\boldsymbol{\mathcal{H}}} - \text{a.s.}\end{aligned} \tag{184}$$

Hence by integrating $p_{X^n(W)Y^n\boldsymbol{\mathcal{H}}|J_WL_WW}$, defined as in (183a), on each set Borel set B and using the relation (184) as well as the definition of conditional probability we obtain the conclusion. We further note that since our encoding procedure is identical among users and W is independent of users' data and the encoding process, we obtain

$$\begin{aligned}&p_{X^n(W)Y^n\boldsymbol{\mathcal{H}}|J_WL_WW}(x^n, y^n, \mathcal{H}|j_w, l_w, w)\\&= \frac{\Pr\{J_1 = j_w, L_1 = l_w|X^n(1) = x^n, \boldsymbol{\mathcal{H}} = \mathcal{H}\}}{P_{J_1L_1}(j_w, l_w)} \times p^n_X(x^n)p^n_{Y|X}(y^n|x^n)p(\boldsymbol{\mathcal{H}} = \mathcal{H}), .\end{aligned} \tag{185}$$

■

Note further that as $\Pr\{J_w = j, L_w = l|X^n(w) = x^n, \boldsymbol{\mathcal{H}} = \mathcal{H}\} = 0$ for some combinations of data sequence, observation and codebook for the w-th user, the corresponding density value is zero.

The following piggy-back's trick [20, Lemma 4.3] facilitates the need of the Markov lemma for the continuous alphabet. For brevity, we denote herein by $\tilde{Y}^n$ the pair (Y^n, Z^n), by $\boldsymbol{\psi}_n$ the random variable $\psi_n(X^n(W), \tilde{Y}^n, U^n(J_W), V^n(J_W, L_W))$ and by P_{cp} the distribution $P_{X^n(W)U^n(J_W)V^n(J_W,L_W)WJ_WL_W}$. Additionally, we define

$$\begin{aligned}\boldsymbol{\chi}_{\mathcal{B}_n} &= \chi_{\mathcal{B}_n}(X^n(W), U^n(J_W), V^n(J_W, L_W))\\\boldsymbol{\chi}_{\mathcal{B}^c_n} &= 1 - \boldsymbol{\chi}_{\mathcal{B}_n},\end{aligned} \tag{186}$$

where herein $\mathcal{B}_n$ is also a short notation for $\mathcal{B}^\delta_n$. We first notice that since $0 \le \psi_n(\cdot) \le 1$,

$$\begin{aligned}\mathbb{E}[\boldsymbol{\psi}_n] &= \mathbb{E}[\boldsymbol{\chi}_{\mathcal{B}^c_n}\boldsymbol{\psi}_n] + \mathbb{E}[\boldsymbol{\chi}_{\mathcal{B}_n}\boldsymbol{\psi}_n]\\&\le \Pr\{(X^n(W), U^n(J_W), V^n(J_W, L_W)) \notin \mathcal{B}_n\} + \mathbb{E}[\boldsymbol{\chi}_{\mathcal{B}_n}\boldsymbol{\psi}_n],\end{aligned} \tag{187}$$

With the help[20] of (183a) the second term can be bounded as

$$\begin{aligned}\mathbb{E}[\boldsymbol{\chi}_{\mathcal{B}_n}\boldsymbol{\psi}_n] &= \int \chi_{\mathcal{B}_n}(x^n, u^n, v^n) \times \mathbb{E}\big[\psi_n(x^n, u^n, v^n, \tilde{Y}^n)|X^n(w) = x^n, U^n(j_w) = u^n, V^n(j_w, l_w) = v^n,\\&\qquad W = w, J_w = j_w, L_w = l_w\big]dP_{\text{cp}}\\&= \int \chi_{\mathcal{B}_n}(x^n, u^n, v^n)\mathbb{E}[\psi_n(x^n, u^n, v^n, \tilde{Y}^n)|X^n(w) = x^n, W = w]dP_{\text{cp}}\\&= \int \chi_{\mathcal{B}_n}(x^n, u^n, v^n)\eta_{XUV}(x^n, u^n, v^n)dP_{\text{cp}}\\&\overset{(a)}{\le} \delta^{1/2}_n,\end{aligned} \tag{188}$$

where (a) holds since given $(x^n, u^n, v^n) \in \mathcal{B}_n$, we have $\eta_{XUV}(x^n, u^n, v^n) \le \delta^{1/2}_n$.
Due to the symmetry of the problem we obtain

$$\Pr\{(X^n(W), U^n(J_W), V^n(J_W, L_W)) \notin \mathcal{B}_n\} = \Pr\{(X^n(1), U^n(J_1), V^n(J_1, L_1)) \notin \mathcal{B}_n\} \tag{189}$$

as W is independent of the enrollment process. Moreover, by our encoding rule we have

$$\{\omega \in \Omega \mid (X^n(1), U^n(J_1), V^n(J_1, L_1)) \notin \mathcal{B}_n\} = \{\omega \in \Omega \mid (X^n(1), U^n(j_1), V^n(j_1, l_1)) \notin \mathcal{B}_n,\ \forall j_1, l_1\} \tag{190}$$

Then by Lemma 2

$$\Pr\{(X^n(W), U^n(J_W), V^n(J_W, L_W)) \notin \mathcal{B}_n\} \to 0, \tag{191}$$

[20]See also the disintegration arguments in Appendix A.

as $n \to \infty$ if

$$R_U \geq I(X;U) + 4\delta, \quad R_V + R'_V \geq I(X;V|U) + 5\delta. \tag{192}$$

Hence

$$\mathbb{E}[\boldsymbol{\psi}_n] \to 0, \text{ as } n \to \infty. \tag{193}$$

This implies that $(\tilde{Y}^n, U^n(J_W), V^n(J_W, L_W)) \in \mathcal{A}_\delta^n(YUV)$ with high probability, i.e.,

$$\Pr\{W \notin \boldsymbol{\mathcal{L}_1}\} \to 0, \text{ as } n \to \infty. \tag{194}$$

As in the discrete case we consider the following events

$$\begin{aligned}
&\mathcal{E}_1 = \{|\boldsymbol{\mathcal{L}}_1| > 2^{n\Delta}\}, \\
&\mathcal{E}_2 = \left\{(U^n(J_W), V^n(J_W, \tilde{l}), Y^n) \in \mathcal{A}_\delta^n(UVY), \text{ for some } \tilde{l} \neq L_W, \ \tilde{l} \in \mathfrak{B}(K_W)\right\}, \\
&\mathcal{E}_3 = \left\{\exists(w', \tilde{l}), w' \neq W, w' \in \boldsymbol{\mathcal{L}}_1, (Y^n, U^n(J_{w'}), V^n(J_{w'}, \tilde{l})) \in \mathcal{A}_\delta^n(YUV), \ \tilde{l} \in \mathfrak{B}(K_{w'})\right\}.
\end{aligned} \tag{195}$$

To bound the probability of the event $\mathcal{E}_1$ we only need to verify (117) for $i \geq 2$, which is expressed in our case as

$$\begin{aligned}
\Pr\{B_i|W=1\} &= \int_{u^n, j_i} \int_{\mathcal{A}_\delta^n(Z|u^n)} p_{Z^n|W}(z^n|1) dz^n dP_{U^n(J_i)J_i}(u^n, j_i) \\
&\leq \int_{u^n, j_i} \int_{\mathcal{A}_\delta^n(Z|u^n)} 2^{-n(h(Z)-\delta)} dz^n dP_{U^n(J_i)J_i}(u^n, j_i) \\
&\leq 2^{-n(I(Z;U)-3\delta)}.
\end{aligned} \tag{196}$$

Therefore as in the discrete case $\Pr\{\mathcal{E}_1\} \to 0$ if $R - \Delta < I(Z;U) - 3\delta$. The analysis in (118) can be carried out similarly and we obtain the condition $R + R'_V < I(Y;U,V) - \delta$, which is needed for $\Pr\{\mathcal{E}_3\} \to 0$. This further leads to

$$\Pr\{\hat{W} \neq W\} \to 0. \tag{197}$$

Hence, we only need to bound the probability of the second event $\mathcal{E}_2$.

We use the same technique as the one in [21, Lemma 1]. Due to the symmetry of the codebook construction and the encoding process, it is sufficient to condition on the following event[21]

$$\{J_1 = 1, L_1 = 1, W = 1\}. \tag{198}$$

We also assume that $l_1 = 1$ belongs to $\mathfrak{B}(1)$. Then due to the union bound and symmetry

$$\begin{aligned}
\Pr\{\mathcal{E}_2|J_1 = 1, L_1 = 1, W = 1\} &\leq \sum_{\tilde{l} \in \mathfrak{B}(1), \ \tilde{l} \neq 1} \Pr\{(U^n(1), V^n(1, \tilde{l}), Y^n) \in \mathcal{A}_\delta^n | J_1 = 1, L_1 = 1, W = 1\} \\
&\leq 2^{nR'_V} \Pr\{(U^n(1), V^n(1,2), Y^n) \in \mathcal{A}_\delta^n | J_1 = 1, L_1 = 1, W = 1\}.
\end{aligned} \tag{199}$$

The probability term in the right-hand side of (199) can be factorized[22]

$$\begin{aligned}
&\Pr\{(U^n(1), V^n(1,2), Y^n) \in \mathcal{A}_\delta^n | J_1 = 1, L_1 = 1, W = 1\} \\
&= \int_{\mathcal{A}_\delta^n(UVY)} p_{U^n(1)V^n(1,2)Y^n|J_1L_1W}(u^n, v^n, y^n|1,1,1) du^n dv^n dy^n \\
&= \int_{\mathcal{A}_\delta^n(UVY)} \bigg(\int p(U^n(1) = u^n, V^n(1,2) = v^n, Y^n = y^n, \\
&\qquad\qquad X^n(1) = x^n, V^n(1,1) = \tilde{v}^n | J_1 = 1, L_1 = 1, W = 1) dx^n d\tilde{v}^n \bigg) du^n dv^n dy^n \\
&\overset{(\star)}{=} \int_{\mathcal{A}_\delta^n(UVY)} \left(\int p^n_{Y|X}(y^n|x^n) p_{U^n(1)V^n(1,2)X^n(1)V^n(1,1)|J_1L_1W}(u^n, v^n, x^n, \tilde{v}^n|1,1,1) dx^n d\tilde{v}^n \right) du^n dv^n dy^n \\
&\overset{(\star\star)}{=} \int p^n_{Y|X}(y^n|x^n) p(X^n(1) = x^n, U^n(1) = u^n, V^n(1,1) = \tilde{v}^n | J_1 = 1, L_1 = 1)
\end{aligned}$$

[21]For simplicity we drop the subscript for the index of the first user, i.e., the notation (j_1, l_1) is simplified as (j, l).

[22]Since $\mathcal{A}_\delta^n(UVY)$ is a Borel measurable set, we do not need to consider the complete measure space. We also use the notation $p_{X|Y}(x|y)$ and $p(X = x|Y = y)$ for probability density function interchangeably where the latter is handy when a long tuple of random variables is present in the expression.

$$\times \left(\int_{\mathcal{A}_\delta^n(V|u^n,y^n)} p(V^n(1,2) = v^n | U^n(1) = u^n, V^n(1,1) = \tilde{v}^n, X^n(1) = x^n, \right.$$
$$\left. J_1 = 1, L_1 = 1) dv^n \right) dx^n dy^n du^n d\tilde{v}^n. \tag{200}$$

The equality in $(\star)$ holds according to the relation (183a). Since densities are non-negative, $(\star\star)$ holds due to Fubini's theorem and (185). For brevity, we denote $\mathcal{F} = \{U^n(1) = u^n, V^n(1,1) = \tilde{v}^n, X^n(1) = x^n\}$ and

$$\boldsymbol{\mathcal{C}} = \{V^n(1,l) \mid l \geq 3\} \bigcup \left\{U^n(j), V^n(j,l)\right\}_{j,l\big|j\geq 2}, \tag{201}$$

as the rest of the codebook[23]. For given $\boldsymbol{\mathcal{C}} = \mathcal{C}$ define

$$\begin{aligned} n(\mathcal{C},\mathcal{F}) &= \big|\{l \mid v^n(1,l) \in \mathcal{C},\ (u^n, v^n(1,l), x^n) \in \mathcal{B}_n\}\big| \\ &\quad + \big|\{(j,l) \mid j \geq 2, (u^n(j), v^n(j,l)) \in \mathcal{C},\ (u^n(j), v^n(j,l), x^n) \in \mathcal{B}_n\}\big|, \end{aligned} \tag{202}$$

which is a Borel measurable function, and

$$i(\mathcal{C},\mathcal{F}) = \begin{cases} 1 & \text{if } (x^n, u^n, \tilde{v}^n) \notin \mathcal{B}_n \text{ and } n(\mathcal{C},\mathcal{F}) = 0 \\ 0 & \text{otherwise} \end{cases}. \tag{203}$$

As a standard step, we further define[24] $\mathfrak{G} = \{\mathcal{C} \colon p(\boldsymbol{\mathcal{C}} = \mathcal{C} | \mathcal{F}, J_1 = 1, L_1 = 1) = 0\}$. Then

$$\Pr\{\boldsymbol{\mathcal{C}} \in \mathfrak{G} | \mathcal{F}, J_1 = 1, L_1 = 1\} = 0, \tag{204}$$

which implies that

$$\begin{aligned} &\int_{\mathcal{A}_\delta^n(V|u^n,y^n)} p(V^n(1,2) = v^n | \mathcal{F}, J_1 = 1, L_1 = 1) dv^n \\ &= \int_{\mathcal{A}_\delta^n(V|u^n,y^n)} \int_{\mathfrak{G}} p(V^n(1,2) = v^n, \boldsymbol{\mathcal{C}} = \mathcal{C} | \mathcal{F}, J_1 = 1, L_1 = 1) d\mathcal{C} dv^n \\ &\quad + \int_{\mathcal{A}_\delta^n(V|u^n,y^n)} \int_{\mathfrak{G}^c} p(V^n(1,2) = v^n, \boldsymbol{\mathcal{C}} = \mathcal{C} | \mathcal{F}, J_1 = 1, L_1 = 1) d\mathcal{C} dv^n \\ &\overset{(b)}{=} \int_{\mathcal{A}_\delta^n(V|u^n,y^n)} \int_{\mathfrak{G}^c} p(V^n(1,2) = v^n, \boldsymbol{\mathcal{C}} = \mathcal{C} | \mathcal{F}, J_1 = 1, L_1 = 1) d\mathcal{C} dv^n, \end{aligned} \tag{205}$$

where (b) is valid since (204) can be seen as the integration of $p(V^n(1,2) = v^n, \boldsymbol{\mathcal{C}} = \mathcal{C} | \mathcal{F}, J_1 = 1, L_1 = 1)$ over $\mathbb{R}^n \times \mathfrak{G}$, which implies that the first term in the above sum is zero. A similar line of reasoning can be applied to resolve the case where

$$p(X^n(1) = x^n, U^n(1) = u^n, V^n(1,1) = \tilde{v}^n | J_1 = 1, L_1 = 1) = 0, \tag{206}$$

in (200).
Additionally, consider the case that $(x^n, u^n, \tilde{v}^n) \notin \mathcal{B}_n$. Define[25]

$$\mathfrak{D} = \{\mathcal{C} : n(\mathcal{C},\mathcal{F}) > 0\}, \tag{207}$$

which is a Borel set. Then due to our encoding rule

$$\Pr\{\boldsymbol{\mathcal{C}} \in \mathfrak{D} | \mathcal{F}, J_1 = 1, L_1 = 1\} = 0, \tag{208}$$

which leads to

$$\begin{aligned} &\int_{\mathcal{A}_\delta^n(V|u^n,y^n)} p(V^n(1,2) = v^n | \mathcal{F}, J_1 = 1, L_1 = 1) dv^n \\ &= \int_{\mathcal{A}_\delta^n(V|u^n,y^n)} \int_{(\mathfrak{D}\cup\mathfrak{G})^c} p(V^n(1,2) = v^n, \boldsymbol{\mathcal{C}} = \mathcal{C} | \mathcal{F}, J_1 = 1, L_1 = 1) d\mathcal{C} dv^n. \end{aligned} \tag{209}$$

Therefore, to upper bound (200), by combining the arguments in (205) and (209), it is sufficient to consider the following inner integral

$$\int_{\mathfrak{C}} p(V^n(1,2) = v^n, \boldsymbol{\mathcal{C}} = \mathcal{C} | \mathcal{F}, J_1 = 1, L_1 = 1) d\mathcal{C}$$

[23]More precisely, $\boldsymbol{\mathcal{C}}$ is a random vector in which components are $V^n(1,l)$ where $l \geq 3$ and $(U^n(j), V^n(j,l))$ for $j \geq 2$ arranged in the presented order.

[24]From its definition $\mathfrak{G}$ is a Borel measurable set. In more details, due to the restriction (206) $\mathfrak{G}$ is the $(x^n, u^n, \tilde{v}^n)$-section of the measurable set $\bar{\mathfrak{G}} = \{(\mathcal{C}, x^n, u^n, \tilde{v}^n) \mid p(\mathcal{F}|J_1 = 1, L_1 = 1) > 0 \text{ and } p(\boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F} | J_1 = 1, L_1 = 1) = 0\}$. This implies that the inner integral over $\mathfrak{G}$ in (205) produces a measurable function in $(x^n, u^n, \tilde{v}^n, v^n)$.

[25]More specifically, $\mathfrak{D}$ is the $(x^n, u^n, \tilde{v}^n)$-section of the measurable set $\bar{\mathfrak{D}} = \{(\mathcal{C}, x^n, u^n, \tilde{v}^n) \mid (x^n, u^n, \tilde{v}^n) \notin \mathcal{B}_n \text{ and } n(\mathcal{C},\mathcal{F}) > 0\}$.

$$= \int_{\mathfrak{C}} p(\boldsymbol{\mathcal{C}}=\mathcal{C}|\mathcal{F}, J_1=1, L_1=1) p(V^n(1,2)=v^n|\boldsymbol{\mathcal{C}}=\mathcal{C}, \mathcal{F}, J_1=1, L_1=1) d\mathcal{C} \tag{210a}$$

$$\overset{(c)}{=} \int_{\mathfrak{C}} p(\boldsymbol{\mathcal{C}}=\mathcal{C}|\mathcal{F}, J_1=1, L_1=1) p(V^n(1,2)=v^n|\boldsymbol{\mathcal{C}}=\mathcal{C}, \mathcal{F}) \times \frac{\Pr\{J_1=1, L_1=1|V^n(1,2)=v^n, \boldsymbol{\mathcal{C}}=\mathcal{C}, \mathcal{F}\}}{\Pr\{J_1=1, L_1=1|\boldsymbol{\mathcal{C}}=\mathcal{C}, \mathcal{F}\}} d\mathcal{C}, \tag{210b}$$

where[26]

$$\mathfrak{C} = \begin{cases} (\mathfrak{D} \cup \mathfrak{G})^c & \text{if } (x^n, u^n, \tilde{v}^n) \notin \mathcal{B}_n \\ \mathfrak{G}^c & \text{if } (x^n, u^n, \tilde{v}^n) \in \mathcal{B}_n \end{cases}. \tag{211}$$

Note that in both cases, $\Pr\{J_1=1, L_1=1|\boldsymbol{\mathcal{C}}=\mathcal{C}, \mathcal{F}\} > 0$. In Appendix E-C we provide an argument to verify (c) in (210) independently for interested readers.

Next, we have

$$p(V^n(1,2)=v^n|\boldsymbol{\mathcal{C}}=\mathcal{C}, \mathcal{F}) = \prod_{i=1}^n p_{V|U}(v_i|u_i) \tag{212}$$

due to our codebook generation. In addition, we bound the numerator term in (210b) as follows:

$$\underbrace{\Pr\{J_1=1, L_1=1|V^n(1,2)=v^n, \boldsymbol{\mathcal{C}}=\mathcal{C}, \mathcal{F}\}}_{\gamma} \leq \underbrace{\frac{1}{2^{nR}} i(\mathcal{C}, \mathcal{F}) + \frac{1}{n(\mathcal{C}, \mathcal{F})+1}(1-i(\mathcal{C}, \mathcal{F}))}_{\gamma'}, \tag{213}$$

where $R = R_U + \bar{R}_V$. We verify the above inequality by the following cases:

- $(x^n, u^n, \tilde{v}^n) \notin \mathcal{B}_n$ then $n(\mathcal{C}, \mathcal{F}) = 0$ by our restriction which implies that $i(\mathcal{C}, \mathcal{F}) = 1$. We have
$$\gamma \leq \gamma' = \frac{1}{2^{nR}}, \tag{214}$$
with the equality when $(x^n, u^n, v^n) \notin \mathcal{B}_n$.
- $(x^n, u^n, \tilde{v}^n) \in \mathcal{B}_n$, i.e., $i(\mathcal{C}, \mathcal{F}) = 0$, we always have
$$\gamma \leq \gamma' = \frac{1}{n(\mathcal{C}, \mathcal{F})+1}, \tag{215}$$
with the equality when $(x^n, u^n, v^n) \notin \mathcal{B}_n$. The "+1" term in the denominator is due to the event $(x^n, u^n, \tilde{v}^n) \in \mathcal{B}_n$.

Moreover, the denominator in (210b) can be lower bounded as

$$\begin{aligned} &\Pr\{J_1=1, L_1=1|\boldsymbol{\mathcal{C}}=\mathcal{C}, \mathcal{F}\} \\ &\geq \int_{\mathcal{B}_n^c(x^n, u^n)} \Pr\{J_1=1, L_1=1|\boldsymbol{\mathcal{C}}=\mathcal{C}, \mathcal{F}, V^n(1,2)=v^n\} \\ &\quad \times p(V^n(1,2)=v^n|\boldsymbol{\mathcal{C}}=\mathcal{C}, \mathcal{F}) dv^n \\ &= \left(\frac{1}{2^{nR}} i(\mathcal{C}, \mathcal{F}) + \frac{1}{n(\mathcal{C}, \mathcal{F})+1}(1-i(\mathcal{C}, \mathcal{F}))\right) \\ &\quad \times \Pr\{V^n(1,2) \notin \mathcal{B}_n(x^n, u^n)|U^n(1)=u^n\} \\ &\geq \left(\frac{1}{2^{nR}} i(\mathcal{C}, \mathcal{F}) + \frac{1}{n(\mathcal{C}, \mathcal{F})+1}(1-i(\mathcal{C}, \mathcal{F}))\right) \\ &\quad \times \Pr\{V^n(1,2) \notin \mathcal{A}_\delta^n(V|u^n, x^n)|U^n(1)=u^n\}. \end{aligned} \tag{216}$$

Now for sufficiently large n,

$$\begin{aligned} \Pr\{V^n(1,2) \notin \mathcal{A}_\delta^n(V|u^n, x^n)|U^n(1)=u^n\} &= 1 - \int_{\mathcal{A}_\delta^n(V|u^n, x^n)} p_{V|U}^n(v^n|u^n) dv^n \\ &\geq 1 - 2^{-n(h(V|U)-2\delta)} \int_{\mathcal{A}_\delta^n(V|u^n, x^n)} dv^n \\ &\geq 1 - 2^{-n(h(V|U)-2\delta)} 2^{n(h(V|U,X)+2\delta)} \\ &= 1 - 2^{-n(I(X;V|U)-4\delta)}. \end{aligned} \tag{217}$$

[26]It can be seen that $\mathfrak{C}$ is the $(x^n, u^n, \tilde{v}^n)$-section of $(\bar{\mathfrak{G}} \cup \bar{\mathfrak{D}})^c$.

This analysis implies that when $\delta < I(X;V|U)/4$ and for sufficiently large n

$$\begin{aligned} p(V^n(1,2) = v^n|\boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F}) \times \frac{\Pr\{J_1 = 1, L_1 = 1|V^n(1,2) = v^n, \boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F}\}}{\Pr\{J_1 = 1, L_1 = 1|\boldsymbol{\mathcal{C}} = \mathcal{C}\}} \\ \leq \frac{1}{1 - 2^{-n(I(X;V|U)-4\delta)}} \prod_{i=1}^{n} p_{V|U}(v_i|u_i) \\ \overset{(e)}{\leq} (1+\hat{\epsilon})2^{-n(h(V|U)-2\delta)}, \end{aligned} \tag{218}$$

where $\hat{\epsilon}$ is a fixed positive number. (e) holds since $v^n \in \mathcal{A}_\delta^n(V|u^n, y^n)$. Combining (205), (210b) and (218) we obtain the following upper bound

$$\begin{aligned} &\int_{\mathcal{A}_\delta^n(V|u^n,y^n)} p(V^n(1,2) = v^n|\mathcal{F}, J_1 = 1, L_1 = 1) dv^n \\ &\leq \int_{\mathcal{A}_\delta^n(V|u^n,y^n)} (1+\hat{\epsilon})2^{-n(h(V|U)-2\delta)} dv^n \int_{\mathfrak{C}} p(\boldsymbol{\mathcal{C}} = \mathcal{C}|\mathcal{F}, J_1 = 1, L_1 = 1) d\mathcal{C} \\ &\leq (1+\hat{\epsilon})2^{-n(h(V|U)-2\delta)} 2^{n(h(V|U,Y)+2\delta)} \\ &= (1+\hat{\epsilon})2^{-n(I(V;Y|U)-4\delta)}. \end{aligned} \tag{219}$$

Hence, inserting the above inequality in (200) we obtain

$$\Pr\{(U^n(1), V^n(1,2), Y^n) \in \mathcal{A}_\delta^n | J_1 = 1, L_1 = 1, W = 1\} \leq (1+\hat{\epsilon})2^{-n(I(V;Y|U)-4\delta)}$$

and

$$\Pr\{\mathcal{E}_2\} \to 0 \text{ as } n \to \infty, \tag{220}$$

if $R_V' < I(V;Y|U) - 4\delta$ and $\delta < I(V;Y|U)/4$.
Lastly, we bound now the distortion level of the reconstruction sequence. Define

$$\phi_n = (1 - \boldsymbol{\psi}_n)(1 - \chi_{\mathcal{E}_1})(1 - \chi_{\mathcal{E}_2})(1 - \chi_{\mathcal{E}_3}), \tag{221}$$

and $\bar{\phi}_n = (1 - \phi_n)$. We have the following simple inequality, which is actually the union bound,

$$\bar{\phi}_n \leq 1 - (1 - \boldsymbol{\psi}_n)(1 - \chi_{\mathcal{E}_1})(1 - \chi_{\mathcal{E}_2}) + \chi_{\mathcal{E}_3} \leq \cdots \leq \boldsymbol{\psi}_n + \chi_{\mathcal{E}_1} + \chi_{\mathcal{E}_2} + \chi_{\mathcal{E}_3}. \tag{222}$$

Then $\mathbb{E}[\bar{\phi}_n] \to 0$ as $n \to \infty$. We notice that

$$\phi_n = 1 \implies \left\{ \left| d(X^n(W), \hat{X}^n) - \hat{D} \right| \leq \delta \right\}, \tag{223}$$

where $\hat{X}^n = \hat{g}(U^n(J_{\hat{W}}), V^n(J_{\hat{W}}, \tilde{L}), Y^n)$. Therefore the distortion level can be upperbounded as

$$\begin{aligned} \mathbb{E}\left[\left|d(X^n(W), \hat{X}^n) - \hat{D}\right|\right] &= \mathbb{E}\left[\phi_n \left|d(X^n, \hat{X}^n) - \hat{D}\right|\right] + \mathbb{E}\left[\bar{\phi}_n \left|d(X^n, \hat{X}^n) - \hat{D}\right|\right] \\ &\leq \delta + \mathbb{E}[\bar{\phi}_n \hat{D}] + \mathbb{E}[\bar{\phi}_n d(X^n, \hat{X}^n)]. \end{aligned} \tag{224}$$

The last term in (224) can be bounded using similar techniques as in [10, Lemma 5.1]. First note that

$$\mathbb{E}[\bar{\phi}_n d(X^n(W), \hat{X}^n)] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\bar{\phi}_n d(X_i(W), \hat{X}_i)] \leq \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\bar{\phi}_n \zeta(X_i(W))], \tag{225}$$

where $\zeta(X_i(W)) = \max_{\{\hat{x}\}_{k=1}^N} d(X_i(W), \hat{x}_k)$. We further observe that $\{\zeta(X_i(W))\}_{i=1}^n$ are iid $\sim P_{\zeta(X)}$ and integrable random variables. The latter statement holds due to the property of the square distortion measure and P_X. Then for all $i \in [1:n]$ the following is valid for any $a > 0$

$$\mathbb{E}[\bar{\phi}_n \zeta(X_i(W))] \leq a\mathbb{E}[\bar{\phi}_n] + \mathbb{E}[\zeta(X_i(W))\chi_{\{\zeta(X_i(W)) \geq a\}}]. \tag{226}$$

Due to the monotone convergence theorem and the iid property we have

$$\mathbb{E}[\zeta(X_i(W))\chi_{\{\zeta(X_i(W)) \geq a\}}] = \mathbb{E}[\zeta(X)\chi_{\{\zeta(X) \geq a\}}] \leq \delta, \ \forall i \tag{227}$$

for sufficiently large $a \geq a_0$ where a_0 depends only on $(d, P_X, \{\hat{x}_k\}_{k=1}^N)$. This implies that when $a \geq a_0$

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\bar{\phi}_n \zeta(X_i(W))] \leq a\mathbb{E}[\bar{\phi}_n] + \delta \leq 2\delta \tag{228}$$

when $n \to \infty$. In conclusion we obtain

$$\mathbb{E}\left[\left|d(X^n(W), \hat{X}^n) - \hat{D}\right|\right] \leq 4\delta. \tag{229}$$

for sufficiently large n. Recall that the discrete random variable $\hat{T}$ is used to select a pair of indices (j_i, l_i) randomly, cf. (170), and and $\boldsymbol{\mathcal{H}}$ is the random codebook. Put $\hat{\delta} = 4\delta$, by using Markov's inequality with the threshold $4\hat{\delta}$, as in the proof of [18, Lemma 2.2], we have for all sufficiently large n

$$\begin{aligned} &\left|\mathbb{E}[d(X^n(W), \hat{X}^n)|\boldsymbol{\mathcal{H}}, \hat{T}] - \hat{D}\right| < \mathbb{E}\left[\left|d(X^n(W), \hat{X}^n) - \hat{D}\right| |\boldsymbol{\mathcal{H}}, \hat{T}\right],\ \mathbb{P}-\text{a.s}, \\ &\Pr\Big\{\mathbb{E}[\chi_{\{W\neq\hat{W}\}}|\boldsymbol{\mathcal{H}}, \hat{T}] < 4\hat{\delta},\ \mathbb{E}[\chi_{\{W\notin\boldsymbol{\mathcal{L}}\}}|\boldsymbol{\mathcal{H}}, \hat{T}] < 4\hat{\delta}, \\ &\qquad \mathbb{E}\left[\left|d(X^n(W), \hat{X}^n) - \hat{D}\right| |\boldsymbol{\mathcal{H}}, \hat{T}\right] < 4\hat{\delta}\Big\} > 1/4, \end{aligned} \tag{230}$$

which implies the existence of a codebook $\mathcal{H}$ and an instance of randomness $\hat{t}$. Choosing δ small enough, we therefore arrive at the conditions in (124). The rest follows immediately.
Finally, we discuss about the case when $I(Y;V|U) = 0$. We then have that

$$I(X;V|U,Y) = I(Y,X;V|U) = I(X;V|U), \tag{231}$$

as $Y - X - (U,V)$. The second constraint (35b) becomes

$$R_1 + R_2 \geq I(X;U,V).$$

The third constraint (35c) can be omitted as

$$R + I(X;U,V|Y) \leq I(Y,X;U,V) = I(X;U,V), \tag{232}$$

where the first inequality holds since $R \leq \min\{R_L + I(Z;U), I(Y;U,V)\}$. In summary we need to prove the following region is achievable

$$\begin{aligned} R_1 &\geq I(X;U) \\ R_1 + R_2 &\geq I(X;U,V) \\ R &\leq \min\{R_L + I(Z;U), I(Y;U,V)\} \\ D &\geq \mathbb{E}[d(X, g(U,V,Y))]. \end{aligned} \tag{233}$$

The achievability of the region (233) can be proceeded in a similar manner as the one when $I(Y;V|U) > 0$. Namely, we need two layers of codewords $\boldsymbol{u}^n$ and $\boldsymbol{v}^n$. However binning is not used for the second layer. The reconstruction sequence is given by $\hat{g}(u^n(j_{\hat{w}}), v^n(j_{\hat{w}}, l_{\hat{w}}), y^n)$. In the analysis we simply omit the event $\mathcal{E}_2$ since no binning is used.
The following sub-region, which is useful in a later discussion, can be obtained by choosing V and g such that V is independent of everything else and $g\colon \mathbb{R}^2 \to \mathbb{R}$ such that

$$\begin{aligned} R_1 &\geq I(X;U) \\ R &\leq \min\{R_L + I(Z;U), I(Y;U)\} \\ D &\geq \mathbb{E}[d(X, g(U,Y))]. \end{aligned} \tag{234}$$

C. A detailed justification of (210)

The skeptic reader might be wary of the validity of (c) in (210) which is ensured by the following analysis. Let

$$\mathfrak{E} = \big\{(\mathcal{C}, x^n, u^n, \tilde{v}^n) \mid p(\boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F}) > 0\big\}. \tag{235}$$

Then, we have $\Pr\big\{(\boldsymbol{\mathcal{C}}, X^n(1), U^n(1), V^n(1,1)) \in \mathfrak{E}^c\big\} = 0$. Therefore, we can modify the expression (200) as follows. The LHS of (200) is expanded to be an integral over $\boldsymbol{\mathcal{C}}$ and $\mathcal{A}_\delta^n(UVY)$ of the corresponding conditional density term. Then by restricting the integral on the set $\mathfrak{E}$ and following similar steps as in (200), the last integral in (200) is changed to $\int\int_{\mathcal{A}_\delta^n(V|u^n,y^n)}\int_{\mathfrak{E}(x^n,u^n,\tilde{v}^n)}$, where $\mathfrak{E}(x^n, u^n, \tilde{v}^n)$ is the corresponding section of $\mathfrak{E}$. The set $\mathfrak{C}$ in (210) can be modified to $\mathfrak{C}'$ which is the $(x^n, u^n, \tilde{v}^n)$-section of $(\bar{\mathfrak{G}} \cup \bar{\mathfrak{D}})^c \cap \mathfrak{E}$, cf. Footnote 24 and 25.

Claim 4. *Let $\mathbb{R}^\alpha$ be the product space of tuples $(\mathcal{C}, u^n, \tilde{v}^n, x^n, v^n)$ where $\alpha = n(4 + (2^{n\bar{R}_V} - 2 + (2^{nR_U} - 1)2^{n\bar{R}_V})))$ with the corresponding Borel σ-algebra $\mathcal{B}(\mathbb{R}^\alpha)$. Then*

$$\begin{aligned} &p(V^n(1,2) = v^n|\boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F}, J_1 = 1, L_1 = 1)\Pr\{J_1 = 1, L_1 = 1|\boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F}\} \\ &\qquad = p(V^n(1,2) = v^n|\boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F}) \times \Pr\{J_1 = 1, L_1 = 1|V^n(1,2) = v^n, \boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F}\} \end{aligned}$$

$\lambda^{\otimes\alpha}$-almost everywhere on $\{(\mathcal{C}, u^n, \tilde{v}^n, x^n, v^n)|p(\boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F}) > 0\} \in \mathcal{B}(\mathbb{R}^\alpha)$. As a corollary, the conclusion also holds when we restrict to $v^n \in \mathcal{A}_\delta^n(V|u^n, y^n)$.

Proof: We first notice that $\Pr\{J_1 = 1, L_1 = 1|\boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F}\}p(\boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F}) = \Pr\{J_1 = 1, L_1 = 1\}p(\boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F}|J_1 = 1, L_1 = 1)$, $\lambda^{\otimes(\alpha-n)}$-almost everywhere on $\mathbb{R}^{\alpha-n}$, hence also $\lambda^{\otimes\alpha}$-a.e. on $\mathbb{R}^{\alpha}$. This can be seen by integrating both sides w.r.t. $(\mathcal{C}, u^n, \tilde{v}^n, x^n)$ on any set $\hat{\mathcal{E}} \in \mathcal{B}(\mathbb{R}^{\alpha-n})$ and using the definition of conditional probability distribution. Then for any set $\mathcal{E} \in \mathcal{B}(\mathbb{R}^{\alpha})$

$$
\begin{aligned}
&\int_{\mathcal{E}} p(V^n(1,2) = v^n|\boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F}, J_1 = 1, L_1 = 1)\Pr\{J_1 = 1, L_1 = 1|\boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F}\}p(\boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F})d\mathcal{C}du^n d\tilde{v}^n dx^n dv^n \\
&= P_{J_1L_1}(1,1)\int_{\mathcal{E}} p(V^n(1,2) = v^n|\boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F}, J_1 = 1, L_1 = 1)p(\boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F}|J_1 = 1, L_1 = 1)d\mathcal{C}du^n d\tilde{v}^n dx^n dv^n \\
&\overset{(d)}{=} \Pr\{J_1 = 1, L_1 = 1, (\boldsymbol{\mathcal{C}}, U^n(1), V^n(1,1), X^n(1), V^n(1,2)) \in \mathcal{E}\} \\
&= \int_{\mathcal{E}} p(V^n(1,2) = v^n|\mathcal{C}, \mathcal{F})\Pr\{J_1 = 1, L_1 = 1|V^n(1,2) = v^n, \boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F}\}p(\boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F})d\mathcal{C}du^n d\tilde{v}^n dx^n dv^n.
\end{aligned}
$$

In (d) we use the expression $p(V^n(1,2) = v^n, \boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F}|J_1 = 1, L_1 = 1) = p(\boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F}|J_1 = 1, L_1 = 1)p(V^n(1,2) = v^n|\mathcal{F}, \boldsymbol{\mathcal{C}} = \mathcal{C}, J_1 = 1, L_1 = 1)$ which holds except on a zero probability set where $p(\boldsymbol{\mathcal{C}} = \mathcal{C}, \mathcal{F}|J_1 = 1, L_1 = 1) = 0$. The conclusion of the claim follows. ∎

Since, we are doing integration over $(u^n, y^n, v^n) \in \mathcal{A}_\delta^n(UVY)$ and $(\mathcal{C}, x^n, u^n, \tilde{v}^n) \in (\bar{\mathfrak{G}} \cup \bar{\mathfrak{D}})^c \cap \mathfrak{E}$, Claim 4 indicates that replacing (210a) by (210b) does not change the value of (199).

Appendix F
On the closedness of $\mathcal{R}_{GS}$

Assume that the sequence of tuples $(R_m, R_{1,m}, R_{2,m}, R_{L,m}, D_m)_{m\in\mathbb{N}} \in \mathcal{R}_{GS}$ tends to (R, R_1, R_2, R_L, D) as $m \to \infty$ w.r.t. ℓ_1-distance. This implicitly means that neither R_1 nor R_2 is ∞. From the definition of $\mathcal{R}_{GS}$ we only need to show that we always have $R < R_\gamma(R_L)$ and $D > 0$, where due to the definition $R_\gamma(R_L)$ depends on R_L. We show this by a proof by contradiction. Suppose that we have $R = R_\gamma(R_L)$ or $D = 0$. For a given $\epsilon > 0$, there exists $m_0(\epsilon) \in \mathbb{N}$ such that $\forall m > m_0(\epsilon)$

$$D_m < D + \epsilon,\ R_m > R - \epsilon, \text{ and } R_m - R_{L,m} > R - R_L - \epsilon. \tag{236}$$

This implies that

$$
\begin{aligned}
R_{1,m} + R_{2,m} \geq{} & \frac{1}{2}\log_2 \frac{\sigma_Y^2}{\sigma_Z^2 2^{-2(R_m - R_{L,m})} - \sigma_{N_2}^2}, \\
&+ \frac{1}{2}\max\left\{\log_2 \frac{\sigma_X^2\sigma_{N_1}^2}{\sigma_Y^2(D+\epsilon)}, \log_2 \frac{\sigma_X^2 2^{-2(R-\epsilon)}}{\sigma_Y^2 2^{-2(R-\epsilon)} - \sigma_{N_1}^2}, \log_2 \frac{\sigma_X^2}{\sigma_Y^2}\frac{\sigma_Z^2 2^{-2(R-R_L-\epsilon)} - \sigma_{N_2}^2}{\sigma_Z^2 2^{-2(R-R_L-\epsilon)} - (\sigma_{N_1}^2 + \sigma_{N_2}^2)}\right\}.
\end{aligned} \tag{237}
$$

Taking $m \to \infty$, substituting $D = 0$ or $R = R_\gamma(R_L)$ into the right-hand side of (237), and then taking $\epsilon \to 0$ we see the violation since if $D = 0$ the first term in the maximization tends to ∞ whereas if $R = R_\gamma(R_L)$ one of the two latter terms in the maximization goes to ∞ which contradicts $R_1, R_2 < \infty$. Therefore $(R, R_1, R_2, R_L, D) \in \mathcal{R}_{GS}$.

Appendix G

A. *Justification of* (43)

Due to the Markov chain (42) we know that, cf. Lemma 1,

$$P[H|Y^n, W, J_W, K_W, T] = P[H|Y^n, W, J_W, K_W]\ \mathbb{P} - a.s.,\ \forall H \in \sigma(X_i(W)). \tag{238}$$

When the terms on both sides of (238) are regular conditional distributions, which exist since $\mathbb{R}$ is a Polish space, then (238) implies that $P_{X_i(W)|Y^nWJ_WK_W}$ is a regular conditional distribution of $X_i(W)$ given (Y^n, W, J_W, K_W, T). Since $\mathbb{E}[d(X_i(W), g_i(\cdot))] < \infty$ by our restriction, cf. Footnote 6, the disintegration theorem [13, Theorem 5.4] gives us

$$
\begin{aligned}
\mathbb{E}[d(X_i(W),&g_i(\cdot))|Y^n, W, J_W, K_W, T] \\
&= \int d(x, g_i(\cdot))dP_{X_i(W)|Y^nWJ_WK_W}(x), \quad \mathbb{P} - a.s..
\end{aligned} \tag{239}
$$

Hence (43) is justified.

B. *Measurability of g' in* (45)

Let $\bar{g}(Y^n, W, J_W, K_W, T)$ to be the right-hand side of (239), where $\bar{g}$ is measurable. Let t_0 be an arbitrary value of t. Since $\mathbb{E}[\bar{g}(\cdot)] = \mathbb{E}[d(X_i(W), g_i(\cdot))] < \infty$, the set of (w, j_k, k_w, y^n, t) for which $\bar{g}(w, j_k, k_w, y^n, t) = \infty$ has probability zero. Hence, by for example setting these $\bar{g}(w, j_k, k_w, y^n, t)$ to zero, it also suffices to assume that for all tuples (w, j_k, k_w, y^n, t) we have $\bar{g}(w, j_k, k_w, y^n, t) < \infty$. Since a^* is the argmin of $\bar{g}$ according to the lexigraphical restriction, we have

$$\begin{aligned}(a^*)^{-1}(t_0) = &\bigcap_{t \prec t_0} \big\{(w, j_w, k_w, y^n) \mid \bar{g}(w, j_w, k_w, y^n, t) > \bar{g}(w, j_w, k_w, y^n, t_0)\big\} \\ &\bigcap_{t_0 \prec t} \big\{(w, j_w, k_w, y^n) \mid \bar{g}(w, j_w, k_w, y^n, t) \geq \bar{g}(w, j_w, k_w, y^n, t_0)\big\} \end{aligned} \tag{240}$$

where $\prec$ is herein the lexigraphical order. This implies that a^* is a measurable function of (w, j_w, k_w, y^n) since the space of all possible t_0 is finite. The measurability of g_i' follows similarly from the finiteness.

C. *A formal arrival at* (48)

From (40) we have

$$\begin{aligned} D + \epsilon &> \frac{1}{n}\sum_{i=1}^{n} \inf_{g_i} \mathbb{E}[d(X_i(W), g_i(W, J_W, K_W, T, Y^n))] \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[d(X_i(W), \mathbb{E}[X_i(W)|W, J_W, K_W, T, Y^n])] \\ &\overset{(\star)}{=} \sum_{i=1}^{n} \frac{1}{n}\mathbb{E}[d(X_i(W), \mathbb{E}[X_i(W)|W, J_W, K_W, Y^n])] \end{aligned} \tag{241}$$

where $(\star)$ is explained using Corollary 3 as follows. In our case $\mathcal{H} = \sigma(X_i(W))$, $\mathcal{G} = \sigma(Y^n, W, J_W, K_W)$, $\mathcal{F} = \sigma(T)$ and $\sigma(\mathcal{F}, \mathcal{G}) = \sigma(W, J_W, K_W, T, Y^n)$. Since $X_i(W)$ is integrable, eq. (99) implies that

$$\begin{aligned} &\mathbb{E}[X_i(W)^+|\mathcal{F}, \mathcal{G}] = \mathbb{E}[X_i(W)^+|\mathcal{G}]),\ \mathbb{P} - a.s., \\ &\mathbb{E}[X_i(W)^-|\mathcal{F}, \mathcal{G}] = \mathbb{E}[X_i(W)^-|\mathcal{G}]),\ \mathbb{P} - a.s., \end{aligned} \tag{242}$$

where $X_i(W)^+ = \max\{X_i(W), 0\}$ and $X_i(W)^- = \max\{-X_i(W), 0\}$. This leads to $(\star)$. In the light of explanations in Appendix G-A, we also observe that (99) holds due to the disintegration theorem [13, Theorem 5.4].

APPENDIX H
PROOF OF CLAIMS

A. *Proof of Claim 1*

For notation brevity, we denote by η_2 the distribution $P_{X^n(W)Y^nZ^n}$. Consider the following definition, which results in a Markov kernel,

$$\eta_1(B, w, j_w, k_w) = \begin{cases} \dfrac{\Pr\{(X^n(w), Y^n, Z^n) \in B, W = w, J_w = j_w, K_w = k_w\}}{P_{WJ_WK_W}(w, j_w, k_w)} & \text{if } P_{WJ_WK_W}(w, j_w, k_w) > 0 \\ \eta_2(B) & \text{else} \end{cases}, \tag{243}$$

where $B \in \mathcal{B}(\mathbb{R}^{3n})$. Then $\eta_1(B, W, J_W, K_W)$ is (a version of) the conditional probability $\Pr\{(X^n(W), Y^n, Z^n) \in B | WJ_WK_W\}$. Note that η_2 and $\eta_1(\cdot, w, j_w, k_w)$ for a given (w, j_w, k_w) are probability measures on $\mathcal{B}(\mathbb{R}^{3n})$. Additionally, for each (w, j_w, k_w), $\eta_1(\cdot, w, j_w, k_w) \ll \eta_2 \ll \lambda^{\otimes 3n}$ where λ is the Lebesgue measure on $\mathcal{B}(\mathbb{R})$, $\lambda^{\otimes 3n}$ is the product of Lebesgue measures on $\mathcal{B}(\mathbb{R}^{3n})$ of (x^n, y^n, z^n), and $\ll$ denotes the absolute continuous relation between measures [17, Section 2.2]. By the Radon-Nikodym theorem, for each (w, j_w, k_w) there exists a conditional density $p(x^n, y^n, z^n | w, j_w, k_w)$ given by

$$p(x^n, y^n, z^n | w, j_w, k_w) = \frac{d\eta_1(\cdot, w, j_w, k_w)}{d\lambda^{\otimes 3n}}, \tag{244}$$

which is jointly Borel measurable in (x^n, y^n, z^n). Next we will show that it is also jointly measurable w.r.t. $(x^n, y^n, z^n, w, j_w, k_w)$. For each $B \in \mathcal{B}(\mathbb{R})$ we have

$$\begin{aligned} &\Big\{(w, j_w, k_w, x^n, y^n, z^n) \mid p(x^n, y^n, z^n | w, j_w, k_w) \in B\Big\} \\ &= \bigcup_{w, j_w, k_w} \big\{(w, j_w, k_w)\big\} \times p(\cdot | w, j_w, k_w)^{-1}(B) \in 2^{\mathcal{W} \times \mathcal{M}_1 \times \mathcal{M}_2} \times \mathcal{B}(\mathbb{R}^{3n}), \end{aligned} \tag{245}$$

where $p(\cdot|w, j_w, k_w)^{-1}(B) \in \mathcal{B}(\mathbb{R}^{3n})$ is the pre-image of B under $p(\cdot|w, j_w, k_w)$. This implies the joint measurability of $p(x^n, y^n, z^n|w, j_w, k_w)$ as the right-hand side of (245) is a finite union of measurable sets and further

$$p(x^n, y^n, z^n|w, j_w, k_w) = \frac{dP_{X^n(W)Y^nZ^nWJ_WL_W}}{d(\lambda^{\otimes 3n} \times P_{WJ_WK_W})}. \tag{246}$$

Then the function $p(x^n|y^n, w, j_w, k_w) = p(x^n, y^n|w, j_w, k_w)/p(y^n|w, j_w, k_w)$, defined when $p(y^n|w, j_w, k_w) \neq 0$, is the seeking jointly measurable conditional density. We can perform further marginalization to obtain, for example

$$p(y^n|w, j_w) = \int_{\mathcal{M}_2} p(y^n|w, j_w, k_w)dP_{K_W|W=w,J_W=j_w} = \frac{dP_{Y^nWJ_W}}{d(\lambda^{\otimes n} \times P_{WJ_W})}, \text{ etc.} \tag{247}$$

B. Proof of Claim 2 and Implications

The proof of Claim 2 employs similar steps as the one of Claim 1. Note that for each tuple $(e, w, \boldsymbol{j})$, the conditional probability

$$\Pr\{E = e, W = w, \boldsymbol{J} = \boldsymbol{j}|Z^n = z^n\}$$

is a measurable function of z^n due to the definition of conditional probability. It is also clear that given z^n, $\Pr\{E = e, W = w, \boldsymbol{J} = \boldsymbol{j}|Z^n = z^n\}$ is a measurable function of $(e, w, \boldsymbol{j})$ since there are only finite number of tuples $(e, w, \boldsymbol{j})$. Hence, similar as in (245) the jointly measurable of the conditional probability w.r.t. $(e, w, \boldsymbol{j}, z^n)$ can be shown as follows. For each $B \in \mathcal{B}(\mathbb{R})$

$$\begin{aligned}
&\left\{(e, w, \boldsymbol{j}, z^n)\middle|\Pr\{E = e, W = w, \boldsymbol{J} = \boldsymbol{j}|Z^n = z^n\} \in B\right\}\\
&\quad= \bigcup_{e,w,\boldsymbol{j}} \{(e, w, \boldsymbol{j})\} \times \Big(\Pr\{E = e, W = w, \boldsymbol{J} = \boldsymbol{j}|\cdot\}\Big)^{-1}(B)\\
&\quad\in 2^{\{0,1\}\times\mathcal{W}\times\mathcal{M}_1^M} \times \mathcal{B}(\mathbb{R}^n).
\end{aligned} \tag{248}$$

Claim 2 allows us to show the measurablity of $\Pr(e, w|z^n, \boldsymbol{j})$ via $\Pr(e, w|z^n, \boldsymbol{j}) = \Pr(e, w, \boldsymbol{j}|z^n)/\Pr(\boldsymbol{j}|z^n)$ defined when $\Pr(\boldsymbol{j}|z^n) \neq 0$ and so on.

Therefore, the inequalities (19) and (51) are justified step by step as follows

$$\begin{aligned}
H(E, W|Z^n, \boldsymbol{J}) &= \mathbb{E}[-\log_2 \Pr(E, W|Z^n, \boldsymbol{J})] = \mathbb{E}[-\log_2 \Pr(W|Z^n, \boldsymbol{J})] + \mathbb{E}[-\log_2 \Pr(E|W, Z^n, \boldsymbol{J})]\\
&= H(W|Z^n, \boldsymbol{J})\\
&= \mathbb{E}[-\log_2 \Pr(E|Z^n, \boldsymbol{J})] + \mathbb{E}[-\log_2 \Pr(W|E, Z^n, \boldsymbol{J})]\\
&\overset{(**)}{\leq} H(E) + \mathbb{E}\big[\Pr(E = 0|Z^n, \boldsymbol{J})H(W|E = 0, Z^n, \boldsymbol{J})\\
&+ \Pr(E = 1|Z^n, \boldsymbol{J})H(W|E = 1, Z^n, \boldsymbol{J})\big]\\
&\leq h_b(P_e) + P_e \log_2 M + n(R_L + \epsilon),
\end{aligned} \tag{249}$$

where $(**)$ follows from the law of total expectation, i.e., the computing order is $\mathbb{E}_{Z^n\boldsymbol{J}}[\mathbb{E}_{E|Z^n\boldsymbol{J}}[\mathbb{E}_{W|E,Z^n,\boldsymbol{J}}(\cdot)]]$. The inequality

$$\begin{aligned}
H(W|Y^n,(J_{\boldsymbol{\mathcal{L}}}, K_{\boldsymbol{\mathcal{L}}})) &< 1 + \Pr(\hat{W} \neq W) \log_2 M\\
&< 1 + \epsilon \log_2 M,
\end{aligned} \tag{250}$$

can be verified similarly, as $(J_{\boldsymbol{\mathcal{L}}}, K_{\boldsymbol{\mathcal{L}}})$ takes values on a finite set, cf. (3). The step

$$H(W) = I(W; Z^n, \boldsymbol{J}) + H(W|Z^n, \boldsymbol{J}) \tag{251}$$

in (21) is also valid for the Gaussian scenario due to [22, Theorem 2.4], since W is discrete. The interested reader is referred to Appendix I for an alternative direct justification of (251). The other steps follow from the chain rule of mutual information and data processing inequality [22, Theorem 2.5].

C. Derivation of (75)

The entropy power entropy

$$2^{\frac{2}{n}h(Y^n|W,J_W,K_W)} \geq 2^{\frac{2}{n}h(X^n(W)|W,J_W,K_W)} + 2^{\frac{2}{n}h(N_1^n|W,J_W,K_W)}, \tag{252}$$

leads to

$$\frac{n}{2}\log_2(2\pi e\sigma^2_{N_1}) < h(Y^n|W, J_W, K_W) \leq h(Y^n|W, J_W)$$

$$\leq \frac{n}{2}\log_2(2\pi e(\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon}-\sigma_{N_2}^2)), \tag{253}$$

which implies that there exists an α_2 with $0\leq\alpha_2<1$, such that

$$h(Y^n|W,J_W,K_W)=\frac{n}{2}\log_2(2\pi e((1-\alpha_2)(\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon}-\sigma_{N_2}^2)+\alpha_2\sigma_{N_1}^2)). \tag{254}$$

This further leads to

$$\begin{aligned}\Delta_2&=h(X^n(W)|Y^n)-h(X^n(W)|W,J_W,K_W)-h(Y^n|X^n(W))+h(Y^n|W,J_W,K_W)\\&\geq\frac{n}{2}\log_2\left(\frac{\sigma_X^2}{\sigma_Y^2}\frac{(1-\alpha_2)(\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon}-\sigma_{N_2}^2)+\alpha_2\sigma_{N_1}^2}{(1-\alpha_2)(\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon}-(\sigma_{N_1}^2+\sigma_{N_2}^2))}\right)\\&\overset{(b)}{\geq}\frac{n}{2}\log_2\left(\frac{\sigma_X^2\inf_{0\leq\alpha_2<1}((\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon}-\sigma_{N_2}^2)+\frac{\alpha_2}{1-\alpha_2}\sigma_{N_1}^2)}{\sigma_Y^2(\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon}-(\sigma_{N_1}^2+\sigma_{N_2}^2))}\right)\\&=\frac{n}{2}\log_2\left(\frac{\sigma_X^2}{\sigma_Y^2}\frac{\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon}-\sigma_{N_2}^2}{\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon}-(\sigma_{N_1}^2+\sigma_{N_2}^2)}\right).\end{aligned} \tag{255}$$

We note that (57) implies the term $\frac{\sigma_X^2}{\sigma_Y^2(\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon}-(\sigma_{N_1}^2+\sigma_{N_2}^2))}$ is positive. Hence we can move this term outside of the $\inf$-operation which implies that (b) is valid. (75) follows by taking $\epsilon\to 0$

APPENDIX I
SOME EXTRA JUSTIFICATION

A. *A direct proof of* (251)

We also observe from Claim 2 that $\Pr\{W=w|\boldsymbol{J}=\boldsymbol{j},Z^n=z^n\}=\frac{dP_{W\boldsymbol{J}Z^n}}{d(\mu_c\times P_{\boldsymbol{J}Z^n})}$, where

$$\Pr\{W=w|\boldsymbol{J}=\boldsymbol{j},Z^n=z^n\}=\frac{\Pr\{W=w,\boldsymbol{J}=\boldsymbol{j}|Z^n=z^n\}}{\Pr\{\boldsymbol{J}=\boldsymbol{j}|Z^n=z^n\}}$$

when $\Pr\{\boldsymbol{J}=\boldsymbol{j}|Z^n=z^n\}\neq 0$, μ_c is the counting measure. This provides another way to validate (251). Specifically, $P_{W\boldsymbol{J}Z^n}\ll P_W\times P_{\boldsymbol{J}Z^n}\ll\mu_c\times P_{\boldsymbol{J}Z^n}$ holds. The former holds since if $P_W\times P_{\boldsymbol{J}Z^n}(\mathcal{E})=0$ where $\mathcal{E}\in 2^{\mathcal{W}\times\mathcal{M}_1^M}\times\mathcal{B}(\mathbb{R}^n)$ then as $\mathcal{E}=\bigcup_w\{w\}\times\mathcal{E}_w$, $P_{\boldsymbol{J}Z^n}(\mathcal{E}_w)=0,\ \forall w$. Herein $\mathcal{E}_w$ is the w-section of $\mathcal{E}$. This further implies that $P_{W\boldsymbol{J}Z^n}\big(\{w\}\times\mathcal{E}_w\big)=0$, $\forall w$ and hence $P_{W\boldsymbol{J}Z^n}(\mathcal{E})=0$. Thus

$$\begin{aligned}I(W;\boldsymbol{J},Z^n)&=\mathbb{E}_{P_{W\boldsymbol{J}Z^n}}\big[\log_2\frac{dP_{W\boldsymbol{J}Z^n}}{d(P_W\times P_{\boldsymbol{J}Z^n})}\big]\\&=\mathbb{E}_{P_{W\boldsymbol{J}Z^n}}\big[\log_2\frac{\Pr\{W=w|\boldsymbol{J}=\boldsymbol{j},Z^n=z^n\}}{\Pr\{W=w\}}\big].\end{aligned} \tag{256}$$

B. *A direct proof of* (52a)

For the second equality, cf. equation (52a), we need to verify that $P_{Z^nWJ_W}\ll P_{Z^n}\times P_{WJ_W}\ll\lambda^{\otimes n}\times P_{WJ_W}$ holds. The first $\ll$ assertion is valid since if $P_{Z^n}\times P_{WJ_W}(\mathcal{E})=0$ where $\mathcal{E}\in\mathcal{B}(\mathbb{R}^n)\times 2^{\mathcal{W}\times\mathcal{M}_1}$ then for all (w,j_w) either $P_{WJ_W}(w,j_w)=0$ or $P_{Z^n}\{\mathcal{E}_{(w,j_w)}\}=0$, where $\mathcal{E}_{(w,j_w)}$ is the corresponding (w,j_w)-section of $\mathcal{E}$, is valid as $\mathcal{E}=\bigcup_{w,j_w}\mathcal{E}_{(w,j_w)}\times\{(w,j_w)\}$. In both cases we have

$$P_{Z^nWJ_W}\big(\mathcal{E}_{(w,j_w)}\times(w,j_w)\big)\leq\min\big\{P_{WJ_W}(w,j_w),P_{Z^n}(\mathcal{E}_{(w,j_w)})\big\}=0,\ \forall(w,j_w). \tag{257}$$

Therefore

$$\begin{aligned}I(W,J_W;Z^n)&=\mathbb{E}_{P_{Z^nWJ_W}}\big[\log_2\frac{dP_{Z^nWJ_W}}{d(P_Z^n\times P_{WJ_W})}\big]\\&=\mathbb{E}_{P_{Z^nWJ_W}}[\log_2\frac{p(z^n|w,j_w)}{p(z^n)}].\end{aligned} \tag{258}$$

APPENDIX J
ACHIEVABILITY IN THEOREM 3

1) The case $R_{cr_{12}} < R_\gamma$:

a) $R_{cr_{01}} \leq R_{cr_{12}}$

- Case II: $0 \leq R < R_{cr_{01}}$, $h_0(R)$ dominates both $h_1(R)$ and $h_2(R)$. Let V and N_0 be independent Gaussian random variables such that $X = V + N_0$ where $\sigma_V^2 = \sigma_Y^2(1-2^{-2R_{cr_{01}}})$. Since $R_{cr_{12}} < R_\gamma$ we also have $\sigma_V^2 < \sigma_X^2$. Additionally, let U and N_0' be independent Gaussian random variables such that $V = U + N_0'$ where $\sigma_U^2 = \sigma_Z^2(1 - 2^{-2(R-R_L)})$. Note that $\sigma_U^2 > 0$, if $R > R_L$. Furthermore, we also observe that $\sigma_U^2 < \sigma_V^2$ since $\sigma_U^2(R)$ is a increasing function of R, and $\sigma_U^2(R_{cr_{01}}) \leq \sigma_V^2$ holds because $R_{cr_{01}} \leq R_{cr_{12}}$. We also have the Markov chain $U - V - X - Y - Z$. Moreover
$$R - R_L = I(Z;U),\ R < R_{cr_{01}} = I(Y;V). \tag{259}$$
Additionally
$$h(X|V,Y) = \frac{1}{2}\log_2\left(2\pi e \frac{\sigma_{N_1}^2}{\sigma_Y^2}\frac{\sigma_Y^2 2^{-2R_{cr_{01}}} - \sigma_{N_1}^2}{2^{-2R_{cr_{01}}}}\right) = \frac{1}{2}\log_2 2\pi e D, \tag{260}$$
which implies that the distortion level is matched. Hence the chosen random variables satisfy the constraints for fixed parameters. The rate constraint for R_1 is given as in (94). The other sum rate constraints can be calculated as
$$\begin{aligned} R_1 + R_2 &\geq \frac{1}{2}\log_2 \frac{\sigma_Y^2}{\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2} + \frac{1}{2}\log_2 \frac{\sigma_X^2\sigma_{N_1}^2}{\sigma_Y^2 D}, \\ R_1 + R_2 - R &\geq \frac{1}{2}\log_2 \frac{\sigma_X^2\sigma_{N_1}^2}{\sigma_Y^2 D}. \end{aligned} \tag{261}$$
which again match with the outerbound. If $R = R_L$ we can choose U to be a Gaussian and independent of everything else. Similarly, the first sum rate constraint in the above region can be removed, cf. (89).
- Case III: $R_\gamma > R \geq R_{cr_{12}} \geq R_L$ then $h_2(R)$ dominates the other functions since $h_2(R) \geq h_1(R) \geq h_0(R)$ on this interval. We also observe that $R_{cr_{12}} \geq R_{cr_{02}}$, i.e., $R_{cr_{02}}$ lies inside the interval $[R_L, R_\gamma)$. Assume otherwise that $R_{cr_{02}} > R_{cr_{12}}$, then we have the following chain
$$h_0(R_{cr_{02}}) = h_2(R_{cr_{02}}) > h_2(R_{cr_{12}}) = h_1(R_{cr_{12}}) \geq h_1(R_{cr_{01}}) = h_0(R_{cr_{01}}), \tag{262}$$
which is a contradiction. Furthermore, note that as $\Gamma = \frac{1}{2}h_2(R)$ the third constraint (36c) becomes redundant due to (36b) as
$$\begin{aligned} R_1 + R_2 &\geq \frac{1}{2}\log_2 \frac{\sigma_Y^2}{\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2} + \frac{1}{2}h_2(R) \\ &= \frac{1}{2}\log_2 \frac{\sigma_X^2}{\sigma_Z^2 2^{-2(R-R_L)} - (\sigma_{N_1}^2 + \sigma_{N_2}^2)}. \end{aligned} \tag{263}$$
Additionally, since $R \geq R_{cr_{12}}$ we have
$$\begin{aligned} &2^{2R}(\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2) \leq \sigma_Y^2, \\ &\Rightarrow R + \frac{1}{2}h_2(R) \leq \frac{1}{2}\log_2 \frac{\sigma_X^2}{\sigma_Z^2 2^{-2(R-R_L)} - (\sigma_{N_1}^2 + \sigma_{N_2}^2)}, \end{aligned} \tag{264}$$
which implies that the fourth constraint $R_1 + R_2 \geq R + \Gamma$ also becomes irrelevant, cf. also (89). Since this is a degenerate case, we use the region (234) for achieving the corresponding outer bound. Let $X = U + N_0$ where U and N_0 are independent Gaussian random variables, where $\sigma_U^2 = \sigma_Z^2(1 - 2^{-2(R-R_L)})$. We observe that $\sigma_U^2 < \sigma_X^2$ since $R < R_\gamma$. The Markov chain $U - X - Y - Z$ is satisfied. Additionally, the condition $I(Z;U) = R - R_L$ is satisfied by the chosen U. Since $R \geq R_{cr_{12}}$, we have $\sigma_U^2 \geq \sigma_Y^2(1 - 2^{-2R})$, cf. (91). This implies that $I(Y;U) \geq R$. Next,
$$\begin{aligned} h(X|U,Y) &= \frac{1}{2}\log_2(2\pi e\sigma_{N_1}^2) + \frac{1}{2}\log_2 2\pi e(\sigma_Z^2 2^{-2(R-R_L)} - (\sigma_{N_1}^2 + \sigma_{N_2}^2)) \\ &\quad - \frac{1}{2}\log_2 2\pi e(\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2) \\ &= \frac{1}{2}\log_2 2\pi e\sigma_{N_1}^2\left(1 - \frac{\sigma_{N_1}^2}{\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2}\right) \\ &\overset{(87)}{\leq} \frac{1}{2}\log_2 2\pi e D, \end{aligned} \tag{265}$$

since $R \geq R_{cr_{12}} \geq R_{cr_{02}}$, i.e., the distortion level D is achievable. The last constraint (36b) is justified straightforwardly. We note again that in this case, the second layer is not necessary, hence *binning* can be omitted. A similar behavior is observed in [3, Section IV].

b) If $R_{cr_{01}} > R_{cr_{12}}$ then $R_{cr_{01}} \geq R_{cr_{02}} \geq R_{cr_{12}} \geq R_L$. If $R_{cr_{02}} < R_{cr_{12}}$ then the following meaningful chain of expressions, i.e., all involving terms are defined, shows the contradiction

$$h_0(R_{cr_{02}}) = h_2(R_{cr_{02}}) < h_2(R_{cr_{12}}) = h_1(R_{cr_{12}}) \leq h_1(R_{cr_{01}}) = h_0(R_{cr_{01}}). \tag{266}$$

Combining with our discussion in Subsection III-C, we have $R_{cr_{02}} \in [R_{cr_{12}}, R_L + \frac{1}{2}\log_2 \frac{\sigma_Z^2}{\sigma_{N_1}^2+\sigma_{N_2}^2})$. For $R_\gamma > R \geq R_{cr_{12}}$, we have $h_2(R) \geq h_1(R)$. Additionally, as $R \to R_\gamma$ either $h_1(R)$ or $h_2(R)$ tend to ∞. This implies that $h_2(R)$ goes to ∞ and intersects $h_0(R)$ before $h_1(R)$. Hence, both relations $R_{cr_{01}} \geq R_{cr_{02}}$ and $R_{cr_{02}} \in [R_{cr_{12}}, R_\gamma)$ follow.

- Case IV: If $R_\gamma > R \geq R_{cr_{02}}$, then $h_2(R)$ dominates the outerbound. Since $R \geq R_{cr_{12}}$ the two constraints (36c) and (36d) are again redundant. U is selected as in Case III. We note that the requirements $I(Y;U) \geq R$ and $h(X|U,Y) \leq \frac{1}{2}\log_2 2\pi e D$ are still fulfilled since $R \geq R_{cr_{02}} \geq R_{cr_{12}}$.
- Case V: If $R_L \leq R < R_{cr_{02}}$, then $h_0(R)$ dominates the outerbound, since not only $h_0(R) = h_0(R_{cr_{02}}) = h_2(R_{cr_{02}}) \geq h_1(R_{cr_{02}})$, but also both $h_1(R_{cr_{02}}) \geq h_1(R)$ and $h_2(R_{cr_{02}}) \geq h_2(R)$ hold. Let V and N_0 be independent Gaussian random variables such that $X = V + N_0$ where $\sigma_V^2 = \sigma_Z^2(1-2^{-2(R_{cr_{02}}-R_L)}) = \sigma_Y^2(1-2^{-2R_{cr_{01}}})$, cf. (90). Additionally, let U and N_0' be independent Gaussian random variables such that $V = U + N_0'$ where $\sigma_U^2 = \sigma_Z^2(1-2^{-2(R-R_L)})$ and $\sigma_U^2 > 0$ if $R > R_L$. Note that $\sigma_U^2 < \sigma_V^2$ since $R < R_{cr_{02}}$. Again we have the relation $U - V - X - Y - Z$. Next,

$$h(X|V,Y) = \frac{1}{2}\log_2 2\pi e \sigma_{N_1}^2 \left(1 - \frac{\sigma_{N_1}^2}{\sigma_Z^2 2^{-2(R_{cr_{02}}-R_L)} - \sigma_{N_2}^2}\right) \overset{(87)}{=} \frac{1}{2}\log_2 2\pi e D, \tag{267}$$

which implies that the distortion level is matched. The choice of U and V also leads to $I(U;Z) = R - R_L$ and from (267), cf. also (90),

$$\begin{aligned} I(Y;V) &= \frac{1}{2}\log_2 \frac{\sigma_Y^2}{\sigma_Z^2 2^{-2(R_{cr_{02}}-R_L)} - \sigma_{N_2}^2} \\ &= \frac{1}{2}\log_2 \sigma_Y^2 \frac{1 - \frac{D}{\sigma_{N_1}^2}}{\sigma_{N_1}^2} = R_{cr_{01}} \geq R. \end{aligned} \tag{268}$$

Lastly, the other constraints are calculated as

$$\begin{aligned} R_1 + R_2 &\geq \frac{1}{2}\log_2 \frac{\sigma_Y^2}{\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2} + \frac{1}{2}\log_2 \frac{\sigma_X^2 \sigma_{N_1}^2}{\sigma_Y^2 D}, \\ R_1 + R_2 - R &\geq \frac{1}{2}\log_2 \frac{\sigma_X^2 \sigma_{N_1}^2}{\sigma_Y^2 D}, \end{aligned} \tag{269}$$

which matches the outerbound. If $R = R_L$ we select U to be independent of the other random variables. When $R_L \leq R < R_{cr_{12}}$, ($R_{cr_{12}} \leq R \leq R_{cr_{02}}$), the first (second) constraint in the above region can be removed.

2) The case $R_{cr_{12}} \geq R_\gamma$: Since $R_{cr_{12}} \geq R_\gamma$, $h_1(R) > h_2(R)$ holds for all $R_L \leq R < R_\gamma$.
If $R_{cr_{01}} > R_L$ then the following argument shows that $R_{cr_{01}}$ lies in the interval $[R_L, R_\gamma)$. If $R_{cr_{01}}$ is outside the interval $[R_L, R_\gamma)$ then both $h_1(R)$ and $h_2(R)$ lie below $h_0(R)$ in $[R_L, R_\gamma)$. However this is not possible since $h_1(R)$ or both $h_1(R)$ and $h_2(R)$ tend to ∞ as $R \to R_\gamma$. Therefore, we need to consider the following subcases:

- Case VI: If $R_{cr_{01}} \leq R < R_\gamma$, then $h_1(R)$ dominates the outerbound. We select U and V as in Case I the previous discussion. We note that (91) still holds since $R < R_\gamma \leq R_{cr_{12}}$.
- Case VII: If $R_L \leq R < R_{cr_{01}}$, then $h_0(R)$ is the dominating component in the outerbound. U and V are chosen identically as in Case II. $\sigma_U^2 < \sigma_V^2$ is valid since $R_{cr_{01}} < R_\gamma \leq R_{cr_{12}}$.

Case VIII: If $R_{cr_{01}} \leq R_L$, then $h_1(R)$ dominates the other functions on $[R_L, R_\gamma)$. So the construction can be done similarly as in Case I as $R < R_{cr_{12}}$ always holds.

References

[1] F. M. J. Willems, T. Kalker, and J.-P. Linnartz, "On the capacity of a biometrical identification system," in *2003 IEEE International Symposium on Information Theory*. IEEE, p. 82.

[2] E. Tuncel, "Capacity/storage tradeoff in high-dimensional identification systems," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2097–2106, 2009.

[3] E. Tuncel and D. Gündüz, "Identification and lossy reconstruction in noisy databases," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 822–831, 2014.

[4] M. B. Westover and J. A. O'Sullivan, "Achievable rates for pattern recognition," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 299–320, 2008.

[5] F. M. J. Willems, "Searching methods for biometric identification systems: Fundamental limits," in *2009 IEEE International Symposium on Information Theory*. IEEE, 2009, pp. 2241–2245.

[6] E. Tuncel, "Recognition capacity versus search speed in noisy databases," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*. IEEE, 2012, pp. 2566–2570.

[7] F. Farhadzadeh and F. M. J. Willems, "Identification Rate, Search and Memory Complexity Tradeoff: Fundamental Limits," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6173–6188, 2016.

[8] E. Tuncel, P. Koulgi, and K. Rose, "Rate-distortion approach to databases: Storage and content-based retrieval," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 953–967, 2004.

[9] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[10] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder-ii: General sources," *Information and control*, vol. 38, no. 1, pp. 60–80, 1978.

[11] L. Sweeney, "k-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, 2002.

[12] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.

[13] O. Kallenberg, *Foundations of modern probability*. Springer Science & Business Media, 2006.

[14] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 727–734, 1985.

[15] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1057–1070, 1998.

[16] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[17] R. B. Ash and C. Doléans-Dade, *Probability and Measure Theory*. Academic Press, 2000.

[18] M. Bloch and J. Barros, *Physical-layer security: from information theory to security engineering*. Cambridge University Press, 2011.

[19] P. Mitran, "On a markov lemma and typical sequences for polish alphabets," *IEEE Trans. Inf. Theory*, vol. 61, no. 10, pp. 5342–5356, 2015.

[20] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 21, no. 3, pp. 294–300, 1975.

[21] P. Minero, S. H. Lim, and Y.-H. Kim, "A unified approach to hybrid coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1509–1523, 2015.

[22] Y. Polyanskiy and Y. Wu, "Lecture notes on information theory," *MIT (6.441), UIUC (ECE 563)*, pp. 2012–2013, 2014.