entropy

MDPI

Article

Exponential Strong Converse for Successive Refinement with Causal Decoder Side Information †

Lin Zhou * and Alfred Hero *

Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, USA
* Correspondence: linzhou@umich.edu (L.Z.); hero@eecs.umich.edu (A.H.)
† Part of this paper has been accepted to ISIT 2019, Paris, France.

Received: 27 February 2019; Accepted: 13 April 2019; Published: 17 April 2019

Abstract: We consider the k-user successive refinement problem with causal decoder side information and derive an exponential strong converse theorem. The rate-distortion region for the problem can be derived as a straightforward extension of the two-user case by Maor and Merhav (2008). We show that for any rate-distortion tuple outside the rate-distortion region of the k-user successive refinement problem with causal decoder side information, the joint excess-distortion probability approaches one exponentially fast. Our proof follows by judiciously adapting the recently proposed strong converse technique by Oohama using the information spectrum method, the variational form of the rate-distortion region and Hölder's inequality. The lossy source coding problem with causal decoder side information considered by El Gamal and Weissman is a special case ($k = 1$) of the current problem. Therefore, the exponential strong converse theorem for the El Gamal and Weissman problem follows as a corollary of our result.

Keywords: exponential strong converse; information spectrum method; successive refinement; causal side information

1. Introduction

We consider the k-user successive refinement problem with causal decoder side information shown in Figure 1, which we refer to as the k-user causal successive refinement problem. The decoders aim to recover the source sequence based on the encoded symbols and causally available private side information sequences. Specifically, given the source sequence X^n, each encoder f_j where $j \in \{1, \ldots, k\}$ compresses X^n into a codeword S_j. At time $i \in \{1, \ldots, n\}$, for each $j \in \{1, \ldots, k\}$, the j-th user aims to recover the i-th source symbol using the codewords from encoders $(f_1, \ldots, f_j)$, the side information up to time i and a decoding function $\phi_{j,i}$, i.e., $\hat{X}_{j,i} = \phi_{j,i}(S_1, \ldots, S_j, Y_{j,1}, \ldots, Y_{j,i})$. Finally, at time n, for all $j \in \{1, \ldots, k\}$, the j-th user outputs the source estimate $\hat{X}_j^n$ which, under a distortion measure d_j, is required to be less than or equal to a specified distortion level D_j.

The causal successive refinement problem was first considered by Maor and Merhav in [1] who fully characterized the rate-distortion region for the two-user version. Maor and Merhav showed that, unlike the case with non-causal side information [2,3], no special structure e.g., degradedness, is required between the side information Y_1^n and Y_2^n. Furthermore, Maor and Merhav discussed the performance loss due to causal decoder side information compared with non-causal side information [2,3]. In general, for the k-user successive refinement problem, the loss of performance due to causal decoder side information can be derived using Theorem 1 of the present paper and the results in [2,3] for the k-user case, under certain conditions on the degradedness of the side information in [2,3].

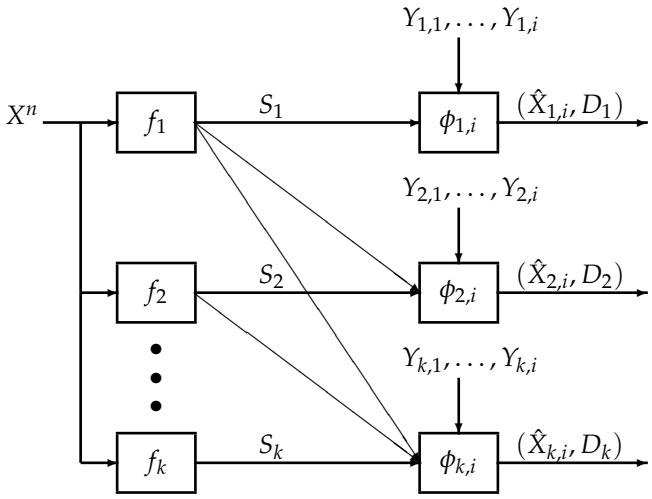

Figure 1. Encoder-decoder system model for the k-user successive refinement problem with causal decoder side information at time $i \in [n]$. Each encoder f_j where $j \in [k]$ compresses the source information into codewords S_j. Given accumulated side information $(Y_{j,1}, \ldots, Y_{j,i})$ and the codewords $(S_1, \ldots, S_j)$, decoder $\phi_{j,i}$ reproduces the i-th source symbol as $\hat{X}_{j,i}$. At time n, for $j \in [k]$, the estimate $\hat{X}_j^n$ for user j is required to satisfy distortion constraint D_j under a distortion measure d_j.

However, Maor and Merhav only presented a *weak* converse in [1]. In this paper, we strengthen the result in [1] by providing an exponential strong converse theorem for the full k-user causal successive refinement problem, which states that the joint excess-distortion probability approaches one exponentially fast if the rate-distortion tuple falls outside the rate-distortion region.

1.1. Related Works

We first briefly summarize existing works on the successive refinement problem. The successive refinement problem was first considered by Equitz and Cover [4] and by Koshelev [5] who considered necessary and sufficient conditions for a source-distortion triple to be successively refinable. Rimoldi [6] fully characterized the rate-distortion region of the successive refinement problem under the joint excess-distortion probability criterion while Kanlis and Narayan [7] derived the excess-distortion exponent in the same setting. The second-order asymptotic analysis of No and Weissman [8], which provides approximations to finite blocklength performance and implies strong converse theorems, was derived under the marginal excess-distortion probabilities criteria. This analysis was extended to the joint excess-distortion probability criterion by Zhou, Tan and Motani [9]. Other frameworks for successive refinement decoding include [10–13].

The study of source coding with causal decoder side information was initiated by Weissman and El Gamal in [14] where they derived the rate-distortion function for the lossy source coding problem with causal side information at the decoders (i.e., $k = 1$, see also [15], Chapter 11.2). Subsequently, Timo and Vellambi [16] characterized the rate-distortion regions of the Gu-Effros two-hop network [17] and the Gray-Wyner problem [18] with causal decoder side information; Maor and Merhav [19] derived the rate-distortion region for the successive refinement of the Heegard-Berger problem [20] with causal side information available at the decoders; Chia and Weissman [21] considered the cascade and triangular source coding problem with causal decoder side information. In all the aforementioned works, the authors used Fano's inquality to prove a weak converse. The weak converse implies that as the blocklength tends to infinity, if the rate-distortion tuple falls outside the rate-distortion region, then the joint excess-distortion probability is bounded away from zero. However, in this paper, we prove an exponential strong converse theorem for the k-user causal successive refinement problem, which significantly strengthens the weak converse as it implies that the joint excess-distortion probability tends to one exponentially fast with respect to the blocklength if the rate-distortion tuple falls outside the rate-distortion region (cf. Theorem 3). As a corollary of our result, for any $\varepsilon \in [0, 1)$, the ε-rate-distortion region (cf. Definition 2) remains the same as the

rate-distortion region (cf. Equation (27)). Please note that with weak converse, one can only assert that the ε-rate-distortion region equals the rate-distortion region when $\varepsilon = 0$. See [22] for yet another justification for the utility of a strong converse compared to a weak converse theorem.

As the information spectrum method will be used in this paper to derive an exponential strong converse theorem for the causal successive refinement problem, we briefly summarize the previous applications of this method to network information theory problems. In [23–25], Oohama used this method to derive exponential strong converses for the lossless source coding problem with one-helper [26,27] (i.e., the Wyner-Ahlswede-Körner (WAK) problem), the asymmetric broadcast channel problem [28], and the Wyner-Ziv problem [29] respectively. Furthermore, Oohama's information spectrum method was also used to derive exponential strong converse theorems for content identification with lossy recovery [30] by Zhou, Tan, Yu and Motani [31] and for Wyner's common information problem under the total variation distance measure [32] by Yu and Tan [33].

1.2. Main Contribution and Challenges

We consider the k-user causal successive refinement problem and present an exponential strong converse theorem. For given rates and blocklength, define the joint excess-distortion probability as the probability that any decoder incurs a distortion level greater than the specified distortion level (see (3)) and define the probability of correct decoding as the probability that all decoders satisfy the specified distortion levels (see (24)). Our proof proceeds as follows. First, we derive a non-asymptotic converse (finite blocklength upper) bound on the probability of correct decoding of any code for the k-user causal successive refinement problem using the information spectrum method. Subsequently, by using Cramér's inequality and the variational formulation of the rate-distortion region, we show that the probability of correct decoding decays exponentially fast to zero as the blocklength tends to infinity if the rate-distortion tuple falls outside the rate-distortion region of the causal successive refinement problem.

As far as we are aware, this paper is the first to establish a strong converse theorem for any lossy source coding problem with causal decoder side information. Furthermore, our methods can be used to derive exponential strong converse theorems for other lossy source coding problems with causal decoder side information discussed in Section 1.1. In particular, since the lossy source coding problems with causal decoder side information in [1,14] are special cases of the k-user causal successive refinement problem, the exponential strong converse theorems for the problems in [1,14] follow as a corollary of our result.

To establish the strong converse in this paper, we must overcome several major technical challenges. The main difficulty lies in the fact that for the causal successive refinement problem, the side information is available to the decoder *causally* instead of non-causally. This causal nature of the side information makes the design of the decoder much more complicated and involved, which complicates the analysis of the joint excess-distortion probability. We find that classical strong converse techniques like the image size characterization [34] and the perturbation approach [35] cannot lead to a strong converse theorem due to the above-mentioned difficulty. However, it is possible that other approaches different from ours can be used to obtain a strong converse theorem for the current problem. For example, it is interesting to explore whether two recently proposed strong converse techniques in [36,37] can be used for this purpose considering the fact that the methods in [36,37] have been successfully applied to problems including the Wyner-Ziv problem [29], the Wyner-Ahlswede-Körner (WAK) problem [26,27] and hypothesis testing problems with communication constraints [38–40].

2. Problem Formulation and Existing Results

2.1. Notation

Random variables and their realizations are in upper (e.g., X) and lower case (e.g., x) respectively. Sets are denoted in calligraphic font (e.g., $\mathcal{X}$). We use $\mathcal{X}^{\mathrm{c}}$ to denote the complement of $\mathcal{X}$ and use

$X^n := (X_1, \ldots, X_n)$ to denote a random vector of length n. Furthermore, given any $j \in [n]$, we use $X^{n\backslash j}$ to denote $(X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X^n)$. We use $\mathbb{R}_+$ and $\mathbb{N}$ to denote the set of positive real numbers and integers respectively. Given two integers a and b, we use $[a : b]$ to denote the set of all integers between a and b and use $[a]$ to denote $[1 : a]$. The set of all probability distributions on $\mathcal{X}$ is denoted as $\mathcal{P}(\mathcal{X})$ and the set of all conditional probability distributions from $\mathcal{X}$ to $\mathcal{Y}$ is denoted as $\mathcal{P}(\mathcal{Y}|\mathcal{X})$. For information-theoretic quantities such as entropy and mutual information, we follow the notation in [34]. In particular, when the joint distribution of (X, Y) is $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, we use $I(P_X, P_{Y|X})$ and $I(X; Y)$ interchangeably.

2.2. Problem Formulation

Let $k \in \mathbb{N}$ be a fixed finite integer and let P_{XY^k} be a joint probability mass function (pmf) on the finite alphabet $\mathcal{X} \times (\prod_{j\in[k]} \mathcal{Y}_j)$ with its marginals denoted in the customary way, e.g., P_X, P_{XY_1}. Throughout the paper, we consider memoryless sources $(X^n, Y_1^n, \ldots, Y_k^n)$, which are generated i.i.d. according to P_{XY^k}. Let a finite alphabet $\hat{\mathcal{X}}_j$ be the alphabet of the reproduced source symbol for user $j \in [k]$. Recall the encoder-decoder system model for the k-user causal successive refinement problem in Figure 1.

A formal definition of a code for the causal successive refinement problem is as follows.

Definition 1. *An $(n, M_1, \ldots, M_k)$-code for the causal successive refinement problem consists of*

- *k encoding functions*

$$f_j : \mathcal{X}^n \to \mathcal{M}_j := \{1, \ldots, M_j\},\ j \in [k], \tag{1}$$

- *and kn decoding functions: for each $i \in [n]$*

$$\phi_{j,i} : (\prod_{l\in[j]} \mathcal{M}_l) \times (\mathcal{Y}_j)^i \to \hat{\mathcal{X}}_j,\ j \in [k]. \tag{2}$$

For $j \in [k]$, let $d_j : \mathcal{X} \times \hat{\mathcal{X}}_j \to [0, \infty)$ be a distortion measure. Given the source sequence x^n and a reproduced version $\hat{x}_j^n$, we measure the distortion between them using the additive distortion measure $d_j(x^n, \hat{x}_j^n) := \frac{1}{n}\sum_{i\in[n]} d_j(x_i, \hat{x}_{j,i})$. To evaluate the performance of an $(n, M_1, \ldots, M_k)$-code for the causal successive refinement problem, given distortion specified levels $(D_1, \ldots, D_k)$, we consider the following joint excess-distortion probability

$$\mathrm{P}_{\mathrm{e}}^{(n)}(D_1, \ldots, D_k) := \Pr\left\{\exists\, j \in [k] \text{ s.t. } d_j(X^n, \hat{X}_j^n) > D_j\right\}. \tag{3}$$

For ease of notation, throughout the paper, we use D^k to denote $(D_1, \ldots, D_k)$, M^k to denote $(M_1, \ldots, M_k)$ and R^k to denote $(R_1, \ldots, R_k)$.

Given $\varepsilon \in (0, 1)$, the ε-rate-distortion region for the k-user causal successive refinement problem is defined as follows.

Definition 2. *Given any $\varepsilon \in (0, 1)$, a rate-distortion tuple (R^k, D^k) is said to be ε-achievable if there exists a sequence of (n, M^k)-codes such that*

$$\limsup_{n\to\infty} \frac{1}{n} \log M_1 \le R_1, \tag{4}$$

$$\limsup_{n\to\infty} \frac{1}{n} \log M_j \le R_j - \sum_{l\in[j-1]} R_l,\ \forall\, j \in [2 : k], \tag{5}$$

$$\limsup_{n\to\infty} \mathrm{P}_{\mathrm{e}}^{(n)}(D^k) \le \varepsilon. \tag{6}$$

The closure of the set of all ε-achievable rate-distortion tuples is called the ε-rate-distortion region and is denoted as $\mathcal{R}(\varepsilon)$.

Please note that in Definition 2, R_j is the sum rate of the first j decoders. Using Definition 2, the rate-distortion region for the problem is defined as

$$\mathcal{R} := \bigcap_{\varepsilon \in (0,1)} \mathcal{R}(\varepsilon). \tag{7}$$

2.3. Existing Results

For the two-user causal successive refinement problem, the rate-distortion region was fully characterized by Maor and Merhav (Theorem 1 in [1]). With slight generalization, the result can be extended to the k-user case.

For $j \in [k]$, let W_j be a random variable taking values in a finite alphabet $\mathcal{W}_j$. For simplicity, throughout the paper, we let

$$T := (X, Y^k, W^k, \hat{X}^k), \tag{8}$$

and let $(t, \mathcal{T})$ be a particular realization of T and its alphabet set, respectively.

Define the following set of joint distributions:

$$\begin{aligned} \mathcal{P}^* := \Big\{ & Q_T \in \mathcal{P}(\mathcal{T}) : Q_{XY^k} = P_{XY^k},\ W^k - X - Y^k,\ |\mathcal{W}_1| \le |\mathcal{X}| + 3, \text{ and } \forall\, j \in [k] : \\ & |\mathcal{W}_j| \le |\mathcal{X}| \Big(\prod_{l \in [j-1]} |\mathcal{W}_l| \Big) + 1,\ \hat{X}_j = \phi_j(W^j, Y_j) \text{ for some } \phi_j : \Big(\prod_{l \in [j]} \mathcal{W}_l \Big) \times \mathcal{Y}_j \to \hat{\mathcal{X}}_j \Big\}. \end{aligned} \tag{9}$$

Given any joint distribution $Q_T \in \mathcal{P}(\mathcal{T})$, define the following set of rate-distortion tuples

$$\begin{aligned} \mathcal{R}(Q_T) := \Big\{ & (R^k, D^k) :\ R_1 \ge I(Q_X, Q_{W_1|X}), D_1 \ge \mathbb{E}[d_1(X, \phi_1(W_1, Y_1))], \text{ and } \forall\, j \in [2:k] : \\ & R_j - \sum_{l \in [j-1]} R_l \ge I(Q_{X|W^{j-1}}, Q_{W_j|XW^{j-1}} | Q_{W_{j-1}}),\ D_j \ge \mathbb{E}[d_j(X, \phi_j(W^j, Y_j))] \Big\}. \end{aligned} \tag{10}$$

For $k = 2$, Maor and Merhav [1] defined the following information theoretical sets of rate-distortion tuples

$$\mathcal{R}^* := \bigcup_{Q_T \in \mathcal{P}^*} \mathcal{R}(Q_T). \tag{11}$$

Theorem 1. *The rate-distortion region for the causal successive refinement problem satisfies*

$$\mathcal{R} = \mathcal{R}^*. \tag{12}$$

We remark that in [1], Maor and Merhav considered the average distortion criterion for $k = 2$, i.e.,

$$\limsup_{n \to \infty} \mathbb{E}[d_j(X^n, \hat{X}_j^n)] \le D_k,\ \forall\, j \in [k], \tag{13}$$

instead of the vanishing joint excess-distortion probability criterion (see (6)) in Definition 2. However, with slight modification to the proof of [1], it can be verified (see Appendix A) that the rate-distortion region $\mathcal{R}$ under the vanishing joint excess-distortion probability criterion, is identical to the rate-distortion region $\mathcal{R}^*$ derived by Maor and Merhav under the average distortion criterion.

Theorem 1 implies that if a rate-distortion tuple falls outside the rate-distortion region, i.e., $(R^k, D^k) \notin \mathcal{R}$, then the joint excess-distortion probability $\mathrm{P}_{\mathrm{e}}^{(n)}(D^k)$ is bounded away from zero.

We strengthen the converse proof of Theorem 1 by showing that if $(R^k, D^k) \notin \mathcal{R}$, then the joint excess-distortion probability $\mathrm{P}_{\mathrm{e}}^{(n)}(D^k)$ approaches one exponentially fast as the blocklength n tends to infinity.

3. Main Results

3.1. Preliminaries

In this subsection, we present necessary definitions and a key lemma before stating our main result.

Define the following set of distributions

$$\mathcal{Q} := \{Q_T \in \mathcal{P}(\mathcal{T}) : |\mathcal{W}_j| \leq \big(|\mathcal{X}||\mathcal{Y}||\mathcal{Z}||\hat{\mathcal{X}}_1||\hat{\mathcal{X}}_2|\big)^j,\ \forall\, j \in [k]\}. \tag{14}$$

Throughout the paper, we use α^k to denote $(\alpha_1, \ldots, \alpha_k)$ and use β^k similarly. Given any $(\mu, \alpha^k, \beta^k) \in \mathbb{R}_+ \times [0,1]^{2k}$ such that

$$\sum_{i\in[k]} (\alpha_i + \beta_i) = 1, \tag{15}$$

for any $Q_T \in \mathcal{Q}$, define the following linear combination of log likelihoods

$$\begin{aligned}\omega_{Q_T}^{(\mu,\alpha^k,\beta^k)}(t) :=\ & \log \frac{Q_X(x)}{P_X(x)} + \log \frac{Q_{Y^k|XW^k}(y^k|x,w^k)}{P_{Y^k|X}(y^k|x)} + \log \frac{Q_{XY^{k\setminus 1}W^{k\setminus 1}|Y_1W_1\hat{X}_1}(x,y^{k\setminus 1},w^{k\setminus 1}|y_1,w_1,\hat{x}_1)}{Q_{XY^{k\setminus 1}W^{k\setminus 1}|Y_1W_1}(x,y^{k\setminus 1},w^{k\setminus 1}|y_1,w_1)} \\ &+ \sum_{j\in[2:k]} \log \frac{Q_{\hat{X}_j|XY^kW^k\hat{X}^{j-1}}(\hat{x}_j|x,y^k,w^k,\hat{x}^{j-1})}{Q_{\hat{X}_j|Y_jW^j}(\hat{x}_j|y_j,w^j)} + \mu\alpha_1 \log \frac{Q_{X|W_1}(x|w_1)}{P_X(x)} \\ &+ \sum_{j\in[2:k]} \mu\alpha_j \log \frac{Q_{X|W^j}(x|w^j)}{Q_{X|W^{j-1}}(x|w^{j-1})} + \sum_{j\in[k]} \mu\beta_j d_j(x,\hat{x}_j).\end{aligned} \tag{16}$$

Given any $\theta \in \mathbb{R}_+$ and any $Q_T \in \mathcal{Q}$, define the negative cumulant generating function of $\omega_{Q_T}^{(\mu,\alpha^k,\beta^k)}(\cdot)$ as

$$\Omega^{(\theta,\mu,\alpha^k,\beta^k)}(Q_T) := -\log \mathbb{E}_{Q_T}\big[\exp\big(-\theta\omega_{Q_T}^{(\mu,\alpha^k,\beta^k)}(T)\big)\big]. \tag{17}$$

Furthermore, define the minimal negative cumulant generating function over distributions in $\mathcal{Q}$ as

$$\Omega^{(\theta,\mu,\alpha^k,\beta^k)} := \min_{Q_T\in\mathcal{Q}} \Omega^{(\theta,\mu,\alpha^k,\beta^k)}(Q_T). \tag{18}$$

Finally, given any rate-distortion tuple (R^k, D^k), define

$$\kappa^{(\alpha^k,\beta^k)}(R^k,D^k) := \alpha_1R_1 + \beta_1D_1 + \sum_{j\in[2:k]} \Big(\alpha_j\big(R_j - \sum_{l\in[j-1]} R_l\big) + \beta_jD_j\Big), \tag{19}$$

$$F^{(\theta,\mu,\alpha^k,\beta^k)}(R^k,D^k) := \frac{\Omega^{(\theta,\mu,\alpha^k,\beta^k)} - \theta\mu\kappa^{(\alpha^k,\beta^k)}(R^k,D^k)}{1 + (2k+2)\theta + \sum_{j\in[k]} 2\theta\mu\alpha_j}, \tag{20}$$

$$F(R^k,D^k) := \sup_{(\theta,\mu,\alpha^k,\beta^k)\in\mathbb{R}_+^2\times[0,1]^{2k}:\ \sum_{i\in[k]}(\alpha_i+\beta_i)=1} F^{(\theta,\mu,\alpha^k,\beta^k)}(R^k,D^k). \tag{21}$$

With the above definitions, we have the following lemma establishing the properties of the exponent function $F(R^k, D^k)$.

Lemma 1. *The following holds.*

(i) For any rate-distortion tuple outside the rate-distortion region, i.e., $(R^k, D^k) \notin \mathcal{R}$, we have

$$F(R^k, D^k) > 0, \tag{22}$$

(ii) For any rate-distortion tuple inside the rate-distortion region, i.e., $(R^k, D^k) \in \mathcal{R}$, we have

$$F(R^k, D^k) = 0. \tag{23}$$

The proof of Lemma 1 is inspired by Property 4 in [25], Lemma 2 in [31] and is given in Section 5. As will be shown in Theorem 2, the exponent function $F(R^k, D^k)$ is a lower bound on the exponent of the probability of correct decoding for the k-user causal successive refinement problem. Thus, Claim (i) in Lemma 1 is crucial to establish the exponential strong converse theorem which states that the joint excess-distortion probability (see (3)) approaches one exponentially fast with respect to the blocklength of the source sequences.

3.2. Main Result

Define the probability of correct decoding as

$$\mathrm{P}_{\mathrm{c}}^{(n)}(D^k) := 1 - \mathrm{P}_{\mathrm{e}}^{(n)}(D^k) = \Pr\left\{\forall\, j \in [k],\ d_j(X^n, \hat{X}_j^n) \leq D_j\right\}. \tag{24}$$

Theorem 2. *Given any (n, M^k)-code for the k-user causal successive refinement problem such that*

$$\log M_1 \leq nR_1, \text{and } \forall\, j \in [2:k],\ \log M_j \leq n\left(R_j - \sum_{l \in [j-1]} R_l\right), \tag{25}$$

we have the following non-asymptotic upper bound on the probability of correct decoding

$$\mathrm{P}_{\mathrm{c}}^{(n)}(D^k) \leq (2k+3)\exp(-nF(R^k, D^k)). \tag{26}$$

The proof of Theorem 2 is given in Section 4. Several remarks are in order.

First, our result is non-asymptotic, i.e., the bound in (26) holds for any $n \in \mathbb{N}$. To prove Theorem 2, we adapt the recently proposed strong converse technique by Oohama [25] to analyze the probability of correct decoding. We first obtain a non-asymptotic upper bound using the information spectrum of log-likelihoods involved in the definition of $\omega_{Q_T}^{(\mu,\alpha^k,\beta^k)}$ (see (16)) and then apply Cramér's bound on large deviations (see e.g., Lemma 13 in [31]) to obtain an exponential type non-asymptotic upper bound. Subsequently, we apply the recursive method [25] and proceed similarly as in [31] to obtain the desired result. Our method can also be used to establish similar results for other source coding problems with causal decoder side information [16,19,21].

Second, we do not believe that classical strong converse techniques including the image size characterization [34] and the perturbation approach [35] can be used to obtain a strong converse theorem for the causal successive refinement problem (e.g., Theorem 3). The main obstacle is that the side information is available *causally* and thus complicates the decoding analysis significantly.

Invoking Lemma 1 and Theorem 2, we conclude that the exponent on the right hand side of (26) is positive if and only if the rate-distortion tuple is outside the rate-distortion region, which implies the following exponential strong converse theorem.

Theorem 3. *For any sequence of (n, M^k)-codes satisfying the rate constraints in (25), given any distortion levels D^k, we have that if $(R^k, D^k) \notin \mathcal{R}$, then the probability of correct decoding $\mathrm{P}_{\mathrm{c}}^{(n)}(D^k)$ decays exponentially fast to zero as the blocklength of the source sequences tends to infinity.*

As a result of Theorem 3, we conclude that for every $\varepsilon \in (0,1)$, the ε-rate distortion region (see Definition 2) satisfies that

$$\mathcal{R}(\varepsilon) = \mathcal{R}, \tag{27}$$

i.e., a strong converse holds for the k-user causal successive refinement problem. Using the strong converse theorem and Marton's change-of-measure technique [41], similarly to Theorem 5 in [31], we can also derive an upper bound on the exponent of the joint excess-distortion probability. Furthermore, applying the one-shot techniques in [42], we can also establish a non-asymptotic achievability bound. Applying the Berry-Esseen theorem to the achievability bound and analyzing the non-asymptotic converse bound in Theorem 2, similarly to [25], we conclude that the backoff from the rate-distortion region at finite blocklength scales on the order of $\Theta(\frac{1}{\sqrt{n}})$. However, nailing down the exact second-order asymptotics [43,44] is challenging and is left for future work.

Our main results in Lemma 1, Theorems 2 and 3 can be specialized to the settings in [1,14] with $k = 1$ and $k = 2$ decoders (users) respectively.

4. Proof of the Non-Asymptotic Converse Bound (Theorem 2)

4.1. Preliminaries

Given any (n, M^k)-code with encoding functions $(f_1, \ldots, f_k)$ and and decoding functions $\{(\phi_{1,i}, \ldots, \phi_{k,i})\}_{i\in[n]}$, we define the following induced conditional distributions on the encoders and decoders: for each $j \in [k]$,

$$P_{S_j|X^n}(s_j|x^n) := 1\{s_j = f_j(x^n)\}, \tag{28}$$

$$P_{\hat{X}_j^n|S^jY_j^n}(\hat{x}_j^n|s^j, y_j^n) := \prod_{i\in[n]} 1\{\hat{x}_{j,i} = \phi_{j,i}(s^j, y_{j,1}, \ldots, y_{j,i})\}. \tag{29}$$

For simplicity, in the following, we define

$$G := (X^n, Y_1^n, \ldots, Y_k^n, S^k, \hat{X}_1^n, \ldots, \hat{X}_k^n), \tag{30}$$

and let $(g, \mathcal{G})$ be a particular realization and the alphabet of G respectively. With above definitions, we have that the distribution P_G satisfies that for any $g \in \mathcal{G}$,

$$P_G(g) := P_{XY^k}^n(x^n, y_1^n, \ldots, y_k^n)\big(\prod_{j\in[k]} P_{S_j|X^n}(s_j|x^n)\big)\big(\prod_{j\in[k]} P_{\hat{X}_j^n|S^jY_j^n}(\hat{x}_j^n|s^j, y_j^n)\big). \tag{31}$$

In the remaining part of this section, all distributions denoted by P are induced by the joint distribution P_G.

To simplify the notation, given any $(i, j) \in [n] \times [k]$, we use $Y_{j,1}^{j,i}$ to denote $(Y_{j,1}, \ldots, Y_{j,i})$ and we use $Y_{1,i}^{k,i}$ to denote $(Y_{1,i}, \ldots, Y_{k,i})$. Similarly, we use $W_{1,i}^{k,i}$ and $\hat{X}_{1,i}^{k,i}$. For each $i \in [n]$, let auxiliary random variables be $W_{1,i} := (X^{i-1}, Y_{1,1}^{1,i-1}, \ldots, Y_{k,1}^{k,i-1}, S_1)$ and $W_{j,i} = S_j$ for all $j \in [2:k]$. Please note that as a function of $i \in [n]$, the Markov chain $(W_{1,i}^{k,i}) \leftrightarrow X_i \leftrightarrow (Y_i, Z_i)$ holds under P_G. Throughout the paper, for each $i \in [n]$, we let

$$T_i := (X_i, Y_{1,i}^{k,i}, W_{1,i}^{k,i}, \hat{X}_{1,i}^{k,i}), \tag{32}$$

and let $(t_i, \mathcal{T}_i)$ be a particular realization of T_i and the alphabet of T_i, respectively.

For each $i \in [n]$, let $Q_{C_i|D_i}$ be arbitrary distributions where $C_i \in \mathcal{T}_i$ and $D_i \in \mathcal{T}_i$. Given any positive real number η and rate-distortion tuple (R^k, D^k), define the following subsets of $\mathcal{G}$:

$$\mathcal{B}_1 := \Big\{g : 0 \geq \frac{1}{n}\sum_{i\in[n]} \log \frac{Q_{X_i}(x_i)}{P_X(x_i)} - \eta\Big\}, \tag{33}$$

$$\mathcal{B}_2 := \Big\{g : 0 \geq \frac{1}{n}\sum_{i\in[n]} \log \frac{Q_{Y_{1,i}^{k,i}|X_i W_{1,i}^{k,i}}(y_{1,i}^{k,i}|x_i, w_{1,i}^{k,i})}{P_{Y^k|X}(y_{1,i}^{k,i}|x_i)} - \eta\Big\}, \tag{34}$$

$$\mathcal{B}_3 := \Big\{g : 0 \geq \frac{1}{n}\sum_{i\in[n]} \log \frac{Q_{X_i Y_{2,i}^{k,i} W_{2,i}^{k,i}|Y_{1,i} W_{1,i} \hat{X}_{1,i}}(x_i, y_{2,i}^{k,i}, w_{2,i}^{k,i}|y_{1,i}, w_{1,i}, \hat{x}_{1,i})}{P_{X_i Y_{2,i}^{k,i} W_{2,i}^{k,i}|Y_{1,i} W_{1,i}}(x_i, y_{2,i}^{k,i}, w_{2,i}^{k,i}|y_{1,i}, w_{1,i})} - \eta\Big\}, \tag{35}$$

$$\mathcal{B}_4 := \Big\{g : 0 \geq \frac{1}{n}\sum_{i\in[n]} \log \frac{Q_{\hat{X}_{j,i}|X_i Y_{1,i}^{k,i} W_{1,i}^{k,i} \hat{X}_{1,i}^{j-1,i}}(\hat{x}_{j,i}|x_i, y_{1,i}^{k,i}, w_{1,i}^{k,i}, \hat{x}_{1,i}^{j-1,i})}{P_{\hat{X}_{j,i}|Y_{j,i} W_{1,i}^{j,i}}(\hat{x}_{j,i}|y_{j,i}, w_{1,i}^{j,i})} - \eta,\ \forall\, j \in [2:k]\Big\}, \tag{36}$$

$$\mathcal{B}_5 := \Big\{g : R_1 \geq \frac{1}{n}\sum_{i\in[n]} \log \frac{P_{X_i|W_{1,i}}(x_i|w_{1,i})}{P_X(x_i)} - \eta\Big\}, \tag{37}$$

$$\mathcal{B}_6 := \Big\{g : R_j - \sum_{l\in[j-1]} R_l \geq \frac{1}{n}\sum_{i\in[n]} \log \frac{P_{X_i|W_{1,i}^{j,i}}(x_i|w_{1,i}^{j,i})}{P_{X_i|W_{1,i}^{j-1,i}}(x_i|w_{1,i}^{j-1,i})} - \eta,\ \forall\, j \in [2:k]\Big\} \tag{38}$$

$$\mathcal{B}_7 := \Big\{g : D_j \geq \frac{1}{n}\sum_{i\in[n]} \log \exp(d_j(x_i, \hat{x}_{j,i})),\ \forall\, j \in [k]\Big\}. \tag{39}$$

4.2. Proof Steps of Theorem 2

We first present the following non-asymptotic upper bound on the probability of correct decoding using the information spectrum method.

Lemma 2. *For any (n, M^k)-code satisfying (25), given any distortion levels D^k, we have*

$$\mathrm{P}_{\mathrm{c}}^{(n)}(D^k) \leq \Pr\Big\{\bigcap_{i\in[7]} \mathcal{B}_i\Big\} + (2k+2)\exp(-n\eta). \tag{40}$$

The proof of Lemma 2 is given in Appendix B and is divided into two steps. First, we derive a n-letter non-asymptotic upper bound which holds for certain arbitrary n-letter auxiliary distributions. Subsequently, we single-letterize the derived bound by proper choice of auxiliary distributions and careful decomposition of induced distributions of P_G.

Subsequently, we will apply Cramér's bound on Lemma 2 to obtain an exponential type non-asymptotic upper bound on the probability of correct decoding. For simplicity, we will use P_i to denote P_{T_i} and use Q_i to denote Q_{T_i}. To present our next result, we need the following definitions. Given any $\mu \in \mathbb{R}_+$ and any $(\alpha^k, \beta^k) \in [0,1]^{2k}$ satisfying (15), let $f_{Q_i,P_i}^{(\alpha^k,\beta^k)}(t_i)$ be the weighted sum of log likelihood terms in the summands to the right of the inequalities in $\{\mathcal{B}_i\}_{i\in[7]}$, i.e.,

$$
\begin{aligned}
f_{Q_i,P_i}^{(\alpha^k,\beta^k)}(t_i) := \log \frac{Q_{X_i}(x_i)}{P_X(x_i)} + \log \frac{Q_{Y_{1,i}^{k,i}|X_i W_{1,i}^{k,i}}(y_{1,i}^{k,i}|x_i, w_{1,i}^{k,i})}{P_{Y^k|X}(y_{1,i}^{k,i}|x_i)} + \log \frac{Q_{X_i Y_{2,i}^{k,i} W_{2,i}^{k,i}|Y_{1,i} W_{1,i} \hat{X}_{1,i}}(x_i, y_{2,i}^{k,i}, w_{2,i}^{k,i}|y_{1,i}, w_{1,i}, \hat{x}_{1,i})}{P_{X_i Y_{2,i}^{k,i} W_{2,i}^{k,i}|Y_{1,i} W_{1,i}}(x_i, y_{2,i}^{k,i}, w_{2,i}^{k,i}|y_{1,i}, w_{1,i})} \\
+ \sum_{j\in[2:k]} \log \frac{Q_{\hat{X}_{j,i}|X_i Y_{1,i}^{k,i} W_{1,i}^{k,i} \hat{X}_i^{j-1}}(\hat{x}_{j,i}|x_i, y_{1,i}^{k,i}, w_{1,i}^{k,i}, \hat{x}_{1,i}^{j-1,i})}{P_{\hat{X}_{j,i}|Y_{j,i} W_{1,i}^{j,i}}(\hat{x}_{j,i}|y_{j,i}, w_{1,i}^{j,i})} + \mu\alpha_1 \log \frac{P_{X_i|W_{1,i}}(x_i|w_{1,i})}{P_X(x_i)} \\
+ \sum_{j\in[2:k]} \mu\alpha_j \log \frac{P_{X_i|W_{1,i}^{j,i}}(x_i|w_{1,i}^{j,i})}{P_{X_i|W_{1,i}^{j-1,i}}(x_i|w_{1,i}^{j-1,i})} + \sum_{j\in[k]} \mu\beta_j d_j(x_i, \hat{x}_{j,i}).
\end{aligned} \tag{41}
$$

Furthermore, given any non-negative real number $\lambda \in \mathbb{R}_+$, define the following negative cumulant generating function

$$
\Omega^{(\lambda,\mu,\alpha^k,\beta^k)}(\{P_i, Q_i\}_{i\in[n]}) := -\log \mathbb{E}\Big[\exp\big(-\lambda \sum_{i\in[n]} f_{Q_i,P_i}^{(\mu,\alpha^k,\beta^k)}(T_i)\big)\Big]. \tag{42}
$$

Recall the definition of $\kappa^{(\alpha^k,\beta^k)}(R^k, D^k)$ in (19). Please note that $\kappa^{(\alpha^k,\beta^k)}(R^k, D^k)$ is a linear combination of the rate-distortion tuple. Using Lemma 2 and Cramér's bound (Lemma 13 in [31]), we obtain the following non-asymptotic exponential type upper bound on the probability of correct decoding, whose proof is given in in Appendix D.

Lemma 3. *For any (n, M^k)-code satisfying the conditions in Lemma 2, given any distortion levels D^k, we have*

$$
\mathrm{P}_{\mathrm{c}}^{(n)}(D^k) \le (2k+3) \exp\left(-n \frac{\frac{1}{n}\Omega^{(\lambda,\mu,\alpha^k,\beta^k)}(\{P_i, Q_i\}_{i\in[n]}) - \lambda\mu\kappa^{(\alpha^k,\beta^k)}(R^k, D^k)}{1 + \lambda(k + 2 + \sum_{j\in[k]} \mu\alpha_j)}\right). \tag{43}
$$

For subsequent analyses, let $\underline{\Omega}^{(\lambda,\mu,\alpha^k,\beta^k)}(\{P_i\}_{i\in[n]})$ be the lower bound on the Q-maximal negative cumulant generating function $\Omega^{(\lambda,\mu,\alpha^k,\beta^k)}(\{P_i, Q_i\}_{i\in[n]})$ obtained by optimizing over the choice of auxiliary distributions $\{Q_i\}_{i\in[n]}$, i.e.,

$$
\underline{\Omega}^{(\lambda,\mu,\alpha^k,\beta^k)}(\{P_i\}_{i\in[n]}) := \inf_{n\in\mathbb{N}} \sup_{\{Q_i\}_{i\in[n]}} \Omega^{(\lambda,\mu,\alpha^k,\beta^k)}(\{P_i, Q_i\}_{i\in[n]}). \tag{44}
$$

Here the supremum over $\{Q_i\}_{i\in[n]}$ is taken since we want the bound to hold for favorable auxiliary distributions and the infimum over $n \in \mathbb{N}$ is taken to yield a non-asymptotic bound.

In the following, we derive a relationship between $\underline{\Omega}^{(\lambda,\mu,\alpha^k,\beta^k)}(\{P_i\}_{i\in[n]})$ and $\Omega^{(\theta,\mu,\alpha^k,\beta^k)}$ (cf. (18)), which, as we shall see later, is a crucial step in proving Theorem. For this purpose, given any $(\lambda, \mu, \alpha^k) \in \mathbb{R}_+^2 \times [0,1]^k$ such that

$$
\lambda(k + \sum_{j\in[k]} \mu\alpha_j) \le 1, \tag{45}
$$

let

$$
\theta := \frac{\lambda}{1 - k\lambda - \sum_{j\in[k]} \lambda\mu\alpha_j}. \tag{46}
$$

Then we have the following lemma which shows that $\underline{\Omega}^{(\lambda,\mu,\alpha^k,\beta^k)}(\{P_i\}_{i\in[n]})$ in Equation (44) can be lower bounded by a scaled version of $\Omega^{(\theta,\mu,\alpha^k,\beta^k)}$ in Equation (18).

Lemma 4. *Given any* $(\lambda, \mu, \alpha^k, \beta^k) \in \mathbb{R}_+^2 \times [0,1]^3$ *satisfying* (15) *and* (45), *for* θ *defined in* (46), *we have:*

$$\underline{\Omega}^{(\lambda,\mu,\alpha^k,\beta^k)}(\{P_i\}_{i\in[n]}) \geq \frac{n\Omega^{(\theta,\mu,\alpha^k,\beta^k)}}{1+k\theta+\sum_{j\in[k]}\theta\mu\alpha_j}. \tag{47}$$

The proof of Lemma 4 uses Hölder's inequality and the recursive method in [25] and is given in Appendix E.

Combining Lemmas 3 and 4, we conclude that for any (n, M^k)-code satisfying the conditions in Lemma 2 and for any $(\mu, \alpha^k, \beta^k) \in \mathbb{R}_+ \times [0,1]^3$, given any $\lambda \in \mathbb{R}_+$ satisfying (45), we have

$$\mathrm{P}_{\mathrm{c}}^{(n)}(D^k) \leq (2k+3)\exp\left(-n\frac{\frac{1}{n}\Omega^{(\lambda,\mu,\alpha^k,\beta^k)}(\{P_i,Q_i\}_{i\in[n]})-\lambda\mu\kappa^{(\alpha^k,\beta^k)}(R^k,D^k)}{1+\lambda(k+2+\sum_{j\in[k]}\mu\alpha_j)}\right) \tag{48}$$

$$\leq (2k+3)\exp\left(-n\frac{\Omega^{(\theta,\mu,\alpha^k,\beta^k)}-\theta\mu\kappa^{(\alpha^k,\beta^k)}(R^k,D^k)}{1+(2k+2)\theta+\sum_{j\in[k]}2\theta\mu\alpha_j}\right) \tag{49}$$

$$\leq (2k+3)\exp\left(-nF^{(\theta,\mu,\alpha^k,\beta^k)}(R^k,D^k)\right), \tag{50}$$

where (49) follows from the definitions of $\kappa^{(\alpha^k,\beta^k)}(\cdot)$ in (19) and θ in (46), and (50) is simply due to the definition of $F^{(\theta,\mu,\alpha^k,\beta^k)}(\cdot)$ in (20).

5. Proof of Properties of Strong Converse Exponent: Proof of Lemma 1

5.1. Alternative Expressions for the Rate-Distortion Region

In this section, we present preliminaries for the proof of Lemma 1, including several definitions and two alternative characterizations of the rate-distortion region $\mathcal{R}$ (cf. (7)).

Recall that we use $Y^{k\setminus j}$ to denote $(Y_1,\ldots,Y_{j-1},Y_{j+1},\ldots,Y_k)$. First, paralleling (9), we define the following set of joint distributions

$$\begin{aligned}\mathcal{P} := \{&Q_T \in \mathcal{P}(\mathcal{T}) : Q_{XY^k} = P_{XY^k},\ W^k \leftrightarrow X \leftrightarrow Y^k, \text{ and } \forall\, j\in[k]:\\ &|\mathcal{W}_j| \leq (|\mathcal{X}|+1)^j,\ \hat{X}_j \leftrightarrow (W^j,Y_j) \leftrightarrow (X,Y^{k\setminus j},W_{j+1}^k,\hat{X}^{j-1})\}.\end{aligned} \tag{51}$$

Please note that compared with (9), the deterministic decoding functions ϕ_j are now replaced by stochastic functions, which are characterized by transition matrices and induce Markov chains, and the cardinality bounds on auxiliary random variables are changed accordingly. Using the definitions of $\mathcal{P}$ and $\mathcal{R}(Q_T)$ (cf. (10)), we can define the following rate-distortion region denoted by $\mathcal{R}_{\mathrm{ran}}$ where the subscript "ran" refers to the randomness of the stochastic functions in the definition of $\mathcal{P}$:

$$\mathcal{R}_{\mathrm{ran}} := \bigcup_{Q_T\in\mathcal{P}} \mathcal{R}(Q_T). \tag{52}$$

As we shall see later, $\mathcal{R}_{\mathrm{ran}} = \mathcal{R}^*$.

To present the alternative characterization of the rate-distortion region using supporting hyperplanes, we need the following definitions. First, we let $\mathcal{P}_{\mathrm{sh}}$ be the following set of joint distributions

$$\begin{aligned}\mathcal{P}_{\mathrm{sh}} := \{&Q_T \in \mathcal{P}(\mathcal{T}) : Q_{XY^k} = P_{XY^k},\ W^k - X - Y^k, \text{ and } \forall\, j\in[k],\\ &|\mathcal{W}_j| \leq (|\mathcal{X}|)^j,\ \hat{X}_j - (W^j,Y_j) - (X,Y^{k\setminus j},W_{j+1}^k,\hat{X}^{j-1})\}.\end{aligned} \tag{53}$$

Please note that $\mathcal{P}_{\text{sh}}$ are the same as $\mathcal{P}$ (cf. (51)) except that the cardinality bounds are reduced. Given any $(\alpha^k, \beta^k) \in [0,1]^{2k}$ satisfying (15), define the following linear combination of achievable rate-distortion tuples

$$\mathrm{R}^{(\alpha^k,\beta^k)} := \min_{Q_T \in \mathcal{P}_{\text{sh}}} \{\alpha_1 I(Q_X, Q_{W_1|X}) + \sum_{j\in[2:k]} \alpha_j I(Q_{X|W^{j-1}}, Q_{W_j|XW^{j-1}}|Q_{W^{j-1}}) + \sum_{j\in[k]} \beta_j \mathbb{E}[d_j(X, \hat{X}_j)]\}. \tag{54}$$

Recall the definition of linear combination of rate-distortion tuples $\kappa^{\cdot}(\cdot)$ in (19) and let $\mathcal{R}_{\text{sh}}$ be the following collection of rate-distortion tuples defined using supporting hyperplane $\mathrm{R}^{(\alpha^k,\beta^k)}$:

$$\mathcal{R}_{\text{sh}} := \bigcap_{(\alpha^k,\beta^k)\in[0,1]^{2k}:\ \sum_{i\in[k]}(\alpha_i+\beta_i)=1} \{(R^k, D^k) : \kappa^{(\alpha^k,\beta^k)}(R^k, D^k) \geq \mathrm{R}^{(\alpha^k,\beta^k)}\}. \tag{55}$$

Finally, recall the definitions of the rate-distortion region $\mathcal{R}$ in (7) and the characterization $\mathcal{R}^*$ in (11). Similarly to Properties 2 and 3 in [25], one can establish the following lemma, which states that: (i) the rate-distortion region $\mathcal{R}$ for the k-user causal successive refinement problem remains unchanged even if one uses stochastic decoding functions; and (ii) the rate-distortion region $\mathcal{R}$ has alternative characterization $\mathcal{R}_{\text{sh}}$ in terms of supporting hyperplanes in (54).

Lemma 5. *The rate-distortion region for the causal successive refinement problem satisfies*

$$\mathcal{R} = \mathcal{R}^* = \mathcal{R}_{\text{ran}} = \mathcal{R}_{\text{sh}}. \tag{56}$$

5.2. Proof of Claim (i)

Recall that we use T (cf. (8)) to denote the collection of random variables $(X, Y^k, S^k, \hat{X}^k)$ and use $t, \mathcal{T}$ similarly to denote a realization of T and its alphabet, respectively. For any $P_T \in \mathcal{P}_{\text{sh}}$ (recall (53)), any $(\alpha^k, \beta^k) \in [0,1]^{2k}$ satisfying (15) and any $\lambda \in \mathbb{R}_+$, for any $t \in \mathcal{T}$, paralleling (16) and (17), define the following linear combination of log likelihoods and its negative cumulative generating function:

$$\tilde{\omega}_{P_T}^{(\alpha^k,\beta^k)}(t) := \alpha_1 \log \frac{P_{X|W_1}(x|w_1)}{P_X(x)} + \sum_{j\in[2:k]} \alpha_j \log \frac{P_{X|W^j}(x|w^j)}{P_{X|W_{j-1}}(x|w_{j-1})} + \sum_{j\in[k]} \beta_j d_j(x, \hat{x}_j), \tag{57}$$

$$\tilde{\Omega}^{(\lambda,\alpha^k,\beta^k)}(P_T) := -\log \mathbb{E}_{P_T}\left[\exp(-\lambda \tilde{\omega}_{P_T}^{(\alpha^k,\beta^k)}(T))\right]. \tag{58}$$

For simplicity, we let

$$\alpha^+ := \max_{j\in[k]} \alpha_j. \tag{59}$$

Furthermore, paralleling the steps used to go from (18) to (21) and recalling the definition of $\kappa^{(\alpha^k,\beta^k)}(\cdot)$ in (19), let

$$\tilde{\Omega}^{(\lambda,\alpha^k,\beta^k)} := \min_{P_T\in\mathcal{P}_{\text{sh}}} \tilde{\Omega}^{(\lambda,\alpha^k,\beta^k)}(P_T), \tag{60}$$

$$\tilde{F}^{(\lambda,\alpha^k,\beta^k)}(R^k, D^k) := \frac{\tilde{\Omega}^{(\lambda,\alpha^k,\beta^k)} - \lambda\kappa^{(\alpha^k,\beta^k)}(R^k, D^k)}{2k + 3 + \lambda\alpha^+ + \sum_{j\in[2:k]} \lambda(2k+3)\alpha_j + \sum_{l\in[k]} 2\lambda\alpha_l}, \tag{61}$$

$$\tilde{F}(R^k, D^k) := \sup_{(\lambda,\alpha^k,\beta^k)\in\mathbb{R}_+\times[0,1]^{2k}:\ \sum_{i\in[k]}(\alpha_i+\beta_i)=1} \tilde{F}^{(\lambda,\alpha^k,\beta^k)}(R^k, D^k). \tag{62}$$

To prove Claim (i), we will need the following two definitions of the tilted distribution and the dispersion function:

$$P_T^{(\lambda,\alpha^k,\beta^k)}(t) := \frac{P_T(t)\exp(-\lambda\tilde{\omega}_{P_T}^{(\alpha^k,\beta^k)}(t))}{\mathbb{E}_{P_T}\left[\exp(-\lambda\tilde{\omega}_{P_T}^{(\alpha^k,\beta^k)}(T))\right]}, \tag{63}$$

$$\rho := \sup_{P_T\in\mathcal{P}_{\mathrm{sh}}} \sup_{(\lambda,\alpha^k,\beta^k)\in\mathbb{R}_+\times[0,1]^{2k}:\ \sum_{i\in[k]}(\alpha_i+\beta_i)=1} \mathrm{Var}_{P_T^{(\lambda,\alpha^k,\beta^k)}}\left[\tilde{\omega}_{P_T}^{(\alpha^k,\beta^k)}(T)\right]. \tag{64}$$

Please note that ρ is positive and finite.

The proof of Claim (i) in Lemma 1 is completed by the following lemma which relates $F(R^k,D^k)$ in Equation (21) to $\tilde{F}(R^k,D^k)$ in Equation (62).

Lemma 6. *The following holds.*

(i) For any rate-distortion tuple (R^k,D^k),

$$F(R^k,D^k) \geq \tilde{F}(R^k,D^k). \tag{65}$$

(ii) For any rate-distortion tuple (R^k,D^k) outside the rate-distortion region, i.e., $(R^k,D^k)\notin\mathcal{R}$, there exists $\delta\in(0,\rho]$ such that:

$$\tilde{F}(R^k,D^k) \geq \frac{\delta^2}{2(2k+9)\rho} > 0. \tag{66}$$

The proof of Lemma 6 is inspired by [25,31] and given in Appendix F. To prove Lemma 6, we use the alternative characterizations of the rate-distortion region $\mathcal{R}$ in Lemma 5 and analyze the connections between the two exponent functions $F(R^k,D^k)$ and $\tilde{F}(R^k,D^k)$.

5.3. Proof of Claim (ii)

Recall the definition of the linear combination of rate-distortion tuple $\kappa^{(\alpha^k,\beta^k)}(R^k,D^k)$ in Equation (19). If a rate-distortion tuple falls inside the rate-distortion region, i.e., $(R^k,D^k)\in\mathcal{R}$, then there exists a distribution $Q_T^*\in\mathcal{P}_{\mathrm{sh}}$ (see (53)) such that for any $(\alpha^k,\beta^k)\in[0,1]^{2k}$ satisfying (15), we have the following lower bound on $\kappa^{(\alpha^k,\beta^k)}(R^k,D^k)$:

$$\begin{aligned}\kappa^{(\alpha^k,\beta^k)}(R^k,D^k) \geq\ & \alpha_1 I(Q_{X_1}^*,Q_{W_1|X_1}^*) + \beta_1^*\mathbb{E}[d_1(X,\hat{X}_1)] \\ & + \sum_{j\in[2:k]}(\alpha_j^* I(Q_{X_1|W^{j-1}}^*,Q_{W_j|XW^{j-1}}^*|Q_{W^{j-1}}^*) + \beta_j^*\mathbb{E}[d_j(X,\hat{X}_j)]).\end{aligned} \tag{67}$$

Recall the definition of $\Omega^{(\theta,\mu,\alpha^k,\beta^k)}(Q_T)$ in (17). Simple calculation establishes

$$\Omega^{(0,\mu,\alpha^k,\beta^k)}(Q_T) = 0, \tag{68}$$

$$\left.\frac{\partial\Omega^{(\theta,\mu,\alpha^k,\beta^k)}(Q_T)}{\partial\theta}\right|_{\theta=0} = \mathbb{E}_{Q_T}\left[\omega_{Q_T}^{\mu,\alpha^k,\beta^k}(T)\right]. \tag{69}$$

Combining (68) and (69), by concavity of $\Omega^{(\theta,\mu,\alpha^k,\beta^k)}(Q_T)$ in θ, it follows that for any $(\theta,\mu,\alpha^k,\beta^k)\in\mathbb{R}_+^2\times[0,1]^{2k}$,

$$\Omega^{(\theta,\mu,\alpha^k,\beta^k)}(Q_T) \leq \theta\mathbb{E}_{Q_T}\left[\omega_{Q_T}^{\mu,\alpha^k,\beta^k}(T)\right]. \tag{70}$$

Using the definition of $\Omega^{(\theta,\mu,\alpha^k,\beta^k)}$ in (18), it follows that

$$\begin{aligned}
\Omega^{(\theta,\mu,\alpha^k,\beta^k)} &\leq \min_{Q_T \in \mathcal{P}_{\text{sh}}} \Omega^{(\theta,\mu,\alpha^k,\beta^k)}(Q_T) && (71)\\
&\leq \min_{Q_T \in \mathcal{P}_{\text{sh}}} \theta \mathbb{E}_{Q_T}\left[\omega_{Q_T}^{\mu,\alpha^k,\beta^k}(T)\right] && (72)\\
&\leq \alpha_1 I(Q^*_{X_1}, Q^*_{W_1|X_1}) + \beta_1^* \mathbb{E}[d_1(X,\hat{X}_1)] \\
&\quad + \sum_{j\in[2:k]} (\alpha_j^* I(Q^*_{X_1|W^{j-1}}, Q^*_{W_j|XW^{j-1}}|Q^*_{W^{j-1}}) + \beta_j^* \mathbb{E}[d_j(X,\hat{X}_j)]) && (73)\\
&\leq \mu \kappa^{(\alpha^k,\beta^k)}(R^k, D^k), && (74)
\end{aligned}$$

where (71) follows from $\mathcal{P}_{\text{sh}} \subseteq \mathcal{Q}$ (recall (14)), (72) follows from the result in (70), (73) follows from the definitions of $\omega_{Q_T}^{\mu,\alpha^k,\beta^k}(t)$ in (17) and $\mathcal{P}_{\text{sh}}$ in (53), and (74) follows from the result in (67).

Using the definition of $F^{(\theta,\mu,\alpha^k,\beta^k)}(R^k, D^k)$ in (21) and the result in (74), we conclude that for any $(R^k, D^k) \in \mathcal{R}$,

$$F^{(\theta,\mu,\alpha^k,\beta^k)}(R^k, D^k) \leq 0. \tag{75}$$

The proof of Claim (ii) is completed by noting that

$$\lim_{\theta \to 0} F^{(\theta,\mu,\alpha^k,\beta^k)}(R^k, D^k) = 0. \tag{76}$$

6. Conclusions

We considered the k-user causal successive refinement problem [1] and established an exponential strong converse theorem using the strong converse techniques proposed by Oohama [25]. Our work appears to be the first to derive a strong converse theorem for any source coding problem with causal decoder side information. The methods we adopted can also be used to obtain exponential strong converse theorems for other source coding problems with causal decoder side information. This paper further illustrates the usefulness and generality of Oohama's information spectrum method in deriving exponential strong converse theorems. The discovered duality in [45] between source coding with decoder side information [46] and channel coding with encoder state information [47] suggests that Oohama's techniques [25] can also be used to establish the strong converse theorem for channel coding with causal encoder state information, e.g., [48–50].

There are several natural future research directions. In Theorem 2, we presented only a lower bound on the strong converse exponent. It would be worthwhile to obtain an exact expression for the strong converse exponent and thus characterize the speed at which the probability of correct decoding decays exponentially fast with respect to the blocklength of source sequences when the rate-distortion tuple falls outside the rate-distortion region. Furthermore, one can explore whether the methods in this paper can be used to establish strong converse theorems for causal successive refinement under the logarithmic loss [51,52], which corresponds to soft decoding of each source symbol. Finally, one can also explore extensions to continuous alphabet by considering Gaussian memoryless sources under bounded distortion measures and derive second-order asymptotics [44,53–56] for the causal successive refinement problem.

Author Contributions: Formal analysis, L.Z.; Funding acquisition, A.H.; Supervision, A.H.; Writing—original draft, L.Z.; Writing—review & editing, A.H.

Funding: This work was partially supported by ARO grant W911NF-15-1-0479.

Acknowledgments: The authors acknowledge anonymous reviewers for helpful comments.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Theorem 1

Replacing (6) with Definition 2, we can define the ε-rate-distortion region $\mathcal{R}_{\mathrm{ad}}(\varepsilon)$ under the average distortion criterion. Furthermore, let

$$\mathcal{R}_{\mathrm{ad}} := \bigcap_{\varepsilon\in[0,1)} \mathcal{R}_{\mathrm{ad}}(\varepsilon). \tag{A1}$$

Maor and Merhav [1] showed that for $k = 2$,

$$\mathcal{R}_{\mathrm{ad}} = \mathcal{R}^*. \tag{A2}$$

Actually, in Section 7 of [1], in order to prove that $\mathcal{R}^* \subseteq \mathcal{R}_{\mathrm{ad}}$, it was already shown that $\mathcal{R}^* \subseteq \mathcal{R}$. Furthermore, it is straightforward to show that the above results hold for any finite $k \in \mathbb{N}$. Thus, to prove Theorem 1, it suffices to show

$$\mathcal{R} \subseteq \mathcal{R}^* = \mathcal{R}_{\mathrm{ad}}. \tag{A3}$$

For this purpose, given any $j \in [k]$, let

$$\bar{d}_j := \max_{(x,\hat{x}_j)\in\mathcal{X}\times\hat{\mathcal{X}}_j} d_j(x,\hat{x}_j). \tag{A4}$$

From the problem formulation, we know that $\bar{d}_j < \infty$ for all $j \in [k]$. Now consider any rate-distortion tuple $(R^k, D^k) \in \mathcal{R}$, then we have (4) to (6). Therefore, for any $j \in [k]$,

$$\limsup_{n\to\infty} \mathbb{E}[d_j(X^n, \hat{X}_j^n)] \le \limsup_{n\to\infty} \Big(\mathbb{E}[d_j(X^n, \hat{X}_j^n)1\{d_j(X^n, \hat{X}_j^n) \le D_j\}] + \bar{d}_j \Pr\{d_j(X^n, \hat{X}_j^n) > D_j\}\Big) \tag{A5}$$

$$\le D_j. \tag{A6}$$

As a result, we have $(R^k, D^k) \in \mathcal{R}_{\mathrm{ad}}$. Thus establishes that $\mathcal{R} \subseteq \mathcal{R}_{\mathrm{ad}} = \mathcal{R}^*$.

Appendix B. Proof of Lemma 2

Recall the definition of G and $\mathcal{G}$ in (30). Given any $C \in \mathcal{G}$ and $D \in \mathcal{G}$, let $Q_{C|D}$ be arbitrary distributions. For simplicity, given each $j \in [k]$, we use $\mathbf{Y}^j$ to denote $(Y_1^n, \ldots, Y_j^n)$ and use $\mathbf{Y}^{j\backslash l}$ to denote $(Y_1^n, \ldots, Y_{l-1}^n, Y_{l+1}^n, \ldots, Y_l^n)$ where $l \in [j]$. Similarly we use $\hat{\mathbf{X}}^j$ and $\hat{\mathbf{X}}^{j\backslash l}$.

Given any positive real number η, define the following sets:

$$\mathcal{A}_1 := \Big\{g : \frac{1}{n}\log\frac{P_X^n(x^n)}{Q_{X^n}(x^n)} \ge -\eta\Big\}, \tag{A7}$$

$$\mathcal{A}_2 := \Big\{g : \frac{1}{n}\log\frac{P_{Y^k|X}^n(\mathbf{y}^k|x^n)}{Q_{\mathbf{Y}^k|X^nS^k}(\mathbf{y}^k|x^n,s^k)} \ge -\eta\Big\}, \tag{A8}$$

$$\mathcal{A}_3 := \Big\{g : \frac{1}{n}\log\frac{P_{X^n\mathbf{Y}^{k\backslash 1}S^{k\backslash 1}|Y_1^nS_1}(x^n,\mathbf{y}^{k\backslash 1},s^{k\backslash 1}|y_1^n,s_1)}{Q_{X^n\mathbf{Y}^{k\backslash 1}S^{k\backslash 1}|Y_1^nS_1\hat{X}_1^n}(x^n,\mathbf{y}^{k\backslash 1},s^{k\backslash 1}|y_1^n,s_1,\hat{x}_1^n)} \ge -\eta\Big\}, \tag{A9}$$

$$\mathcal{A}_4 := \Big\{g : \frac{1}{n}\log\frac{P_{\hat{X}_j^n|Y_j^nS^j}(\hat{x}_j^n|y_j^n,s^j)}{Q_{\hat{X}_j^n|X^n\mathbf{Y}^kS^k\hat{\mathbf{X}}^{j-1}}(\hat{x}_j^n|x^n,\mathbf{y}^k,s^k,\hat{\mathbf{x}}^{j-1})} \ge -\eta,\ \forall\, j \in [2:k]\Big\}, \tag{A10}$$

$$\mathcal{A}_5 := \Big\{g : R_1 \geq \frac{1}{n}\log\frac{P_{X^n|S_1}(x^n|s_1)}{P_X^n(x^n)} - \eta\Big\}, \tag{A11}$$

$$\mathcal{A}_6 := \Big\{g : R_j - \sum_{l\in[j-1]} R_l \geq \frac{1}{n}\log\frac{P_{X^n|S^j}(x^n|s^j)}{P_{X^n|S^{j-1}}(x^n|s^{j-1})} - \eta,\ \forall\, j \in [2:k]\Big\}, \tag{A12}$$

$$\mathcal{A}_7 := \Big\{g : D_j \geq d_j(x^n, \hat{x}_j^n)\ \forall\, j \in [k]\Big\} = \Big\{g : D_j \geq \frac{1}{n}\sum_{i\in[n]} d_j(x_i, \hat{x}_{j,i})\ \forall\, j \in [k]\Big\}. \tag{A13}$$

Then we have the following non-asymptotic upper bound on the probability of correct decoding.

Lemma A1. *Given any (n, M^k)-code satisfying (25) and any distortion levels D^k, we have*

$$\mathrm{P}_{\mathrm{c}}^{(n)}(D^k) \leq \Pr\Big\{\bigcap_{i\in[7]}\mathcal{A}_i\Big\} + (2k+2)\exp(-n\eta). \tag{A14}$$

The proof of Lemma A1 is given in Appendix C.

In the remainder of this subsection, we single-letterize the bound in Lemma A1. Recall that given any $(i,j) \in [n]\times[k]$, we use $Y_{j,1}^{j,i}$ to denote $(Y_{j,1},\ldots,Y_{j,i})$.Recalling that the distributions starting with P are all induced by the joint distribution P_G in (31) and using the choice of auxiliary random variables $(W_{1,i},\ldots,W_{k,i},V_i)$, we have

$$\begin{aligned}
&P_{X^n\mathbf{Y}^{k\setminus 1}S^{k\setminus 1}|Y_1^nS_1}(x^n,\mathbf{y}^{k\setminus 1},s^{k\setminus 1}|y_1^n,s_1)\\
&= \prod_{i\in[n]} P_{X_iY_{2,i}^{k,i}S^{k\setminus 1}|X^{i-1},Y_{2,1}^{2,i-1},\ldots,Y_{k,1}^{k,i-1},Y_1^n,S_1}(x_i,y_{2,i}^{k,i},s^{k\setminus 1}|x^{i-1},y_{2,1}^{2,i-1},\ldots,y_{k,1}^{k,i-1},y_1^n,s_1) &\text{(A15)}\\
&= \prod_{i\in[n]} P_{X_iY_{2,i}^{k,i}S^{k\setminus 1}|X^{i-1},Y_{1,1}^{1,i-1},\ldots,Y_{k,1}^{i-1},Y_{1,i},S_1}(x_i,y_{2,i}^{k,i},s^{k\setminus 1}|x^{i-1},y_{1,1}^{1,i-1},\ldots,y_{k,1}^{k,i-1},y_{1,i},s_1) &\text{(A16)}\\
&= \prod_{i\in[n]} P_{X_iY_{2,i}^{k,i}W_{2,i}^{k,i}|Y_{1,i}W_{1,i}}(x_i,y_{2,i}^{k,i},w_{2,i}^{k,i}|y_{1,i},w_{1,i}) &\text{(A17)}
\end{aligned}$$

$$\begin{aligned}
P_{\hat{X}_j^n|Y_j^nS^j}(\hat{x}_j^n|y_j^n,s_j) &= \prod_{i\in[n]} P_{\hat{X}_{j,i}|Y_{j,1}^{j,i}S^j}(\hat{x}_{j,i}|y_{j,1}^{j,i},s^j) &\text{(A18)}\\
&= \prod_{i\in[n]} P_{\hat{X}_{j,i}|X^{i-1},Y_{1,1}^{1,i-1},\ldots,Y_{k,1}^{k,i-1},Y_{j,i},S^j}(\hat{x}_{j,i}|x^{i-1},y_{1,1}^{1,i-1},\ldots,y_{k,1}^{k,i-1},y_{j,i},s^j) &\text{(A19)}\\
&= \prod_{i\in[n]} P_{\hat{X}_{j,i}|Y_{j,i}W_{1,i}^{k,i}}(\hat{x}_{j,i}|y_{j,i},w_{1,i}^{k,i}), &\text{(A20)}\\
P_{X^n|S_1}(x^n|s_1) &= \prod_{i\in[n]} P_{X_i|X^{i-1}S_1}(x_i|x^{i-1},S_1) &\text{(A21)}\\
&= \prod_{i\in[n]} P_{X_i|X^{i-1}Y_{1,1}^{1,i-1},\ldots,Y_{k,1}^{k,i-1}S_1}(x_i|x^{i-1}y_{1,1}^{1,i-1},\ldots,y_{k,1}^{k,i-1},S_1) &\text{(A22)}\\
&= \prod_{i\in[n]} P_{X_i|W_{1,i}}(x_i|w_{1,i}) &\text{(A23)}\\
P_{X^n|S^{j-1}}(x^n|s^{j-1}) &= \prod_{i\in[n]} P_{X_i|X^{i-1}S^{j-1}}(x_i|x^{i-1},s^{j-1}) &\text{(A24)}\\
&= \prod_{i\in[n]} P_{X_i|X^{i-1}Y_{1,1}^{1,i-1},\ldots,Y_{k,1}^{k,i-1},S^{j-1}}(x_i|x^{i-1},y_{1,1}^{1,i-1},\ldots,y_{k,1}^{k,i-1},s^{j-1}) &\text{(A25)}\\
&= \prod_{i\in[n]} P_{X_i|W_{1,i}^{j-1,i}}(x_i,w_{1,i}^{j-1,i}), &\text{(A26)}\\
P_{X^n|S^j}(x^n|s^j) &= \prod_{i\in[n]} P_{X_i|W_{1,i}^{j,i}}(x_i,w_{1,i}^{j,i}), &\text{(A27)}
\end{aligned}$$

where (A16) follows from the Markov chain $(X_i, Y_{2,i}^{k,i}, S_2^k) - (X^{i-1}, Y_{1,1}^{1,i-1}, \ldots, Y_{k,1}^{k,i-1}, Y_{1,i}, S_1) - Y_{1,i+1}^{1,n}$, (A19) follows from the Markov chain $\hat{X}_{j,i} - (Y_{j,1}^{j,i}, S^j) - (X^{i-1}, Y_{1,1}^{1,i-1}, \ldots, Y_{j-1,1}^{j-1,i-1}, Y_{j+1,1}^{j+1,i-1}, \ldots, Y_{k,1}^{k,i-1})$, (A22) follows from the Markov chain $X_i - (X^{i-1}, S_1) - (Y_1^{i-1}, \ldots, Y_k^{i-1})$, and (A25) follows from the Markov chain $X_i - (X^{i-1}, S^{j-1}) - (Y_{1,1}^{1,i-1}, \ldots, Y_{k,1}^{k,i-1})$.

Furthermore, recall that for $i \in [n]$, $Q_{C_i|D_i}$ are arbitrary distributions where $C_i \in \mathcal{T}_i$ and $D_i \in \mathcal{T}_i$. Please note that Lemma A1 holds for arbitrary choices of distributions $Q_{C|D}$ where $C \in \mathcal{G}$ and $D \in \mathcal{G}$. The proof of Lemma 2 is completed by using Lemma A1 with the following choices of auxiliary distributions and noting that $\mathcal{B}_7 = \mathcal{A}_7$:

$$Q_{X^n}(x^n) := \prod_{i\in[n]} Q_{X_i}(x_i), \tag{A28}$$

$$Q_{\mathbf{Y}^k|X^nS^k}(\mathbf{y}^k|x^n, s^k) := \prod_{i\in[n]} Q_{Y_{1,i}^{k,i}|X_i,W_{1,i}^{k,i}}(y_{1,i}^{k,i}|x_i, w_{1,i}^{k,i}), \tag{A29}$$

$$Q_{X^n\mathbf{Y}^{k\setminus 1}S_2^k|Y_1^nS_1\hat{X}_1^n}(x^n, \mathbf{y}^{k\setminus 1}, s_2^k|y_1^n, s_1, \hat{x}_1^n) := \prod_{i\in[n]} Q_{X_iY_{2,i}^{k,i}W_{2,i}^{k,i}|Y_{1,i}W_{1,i}\hat{X}_{1,i}}(x_i, y_{2,i}^{k,i}, w_{2,i}^{k,i}|y_{1,i}, w_{1,i}, \hat{x}_{1,i}) \tag{A30}$$

$$Q_{\hat{X}_j^n|X^n\mathbf{Y}^kS^k\hat{\mathbf{X}}^{j-1}}(\hat{x}_j^n|x^n, \mathbf{y}^k, s^k, \hat{\mathbf{x}}^{k\setminus j}) := \prod_{i\in[n]} Q_{\hat{X}_{j,i}|X_i,Y_{1,i}^{k,i},W_{1,i}^{k,i},\hat{X}_{1,i}^{j-1,i}}(\hat{x}_{j,i}|x_i, y_{1,i}^{k,i}, w_{1,i}^{k,i}, \hat{x}_{1,i}^{j-1,i}). \tag{A31}$$

Appendix C. Proof of Lemma A1

Recall the definition of the probability of correct decoding $\mathrm{P}_{\mathrm{c}}^{(n)}(D^k)$ in (24) and the definitions of sets $\{\mathcal{A}_j\}_{j\in[7]}$ in (A7) to (A13). For any (n, M^k)-code, we have that

$$\mathrm{P}_{\mathrm{c}}^{(n)}(D^k) = \Pr\{\mathcal{A}_7\} \tag{A32}$$

$$= \Pr\left\{\mathcal{A}_7 \bigcap \Big(\bigcap_{j\in[6]} \mathcal{A}_j\Big)\right\} + \Pr\left\{\mathcal{A}_7 \bigcap \Big(\bigcup_{j\in[6]} \mathcal{A}_j^{\mathrm{c}}\Big)\right\} \tag{A33}$$

$$\leq \Pr\left\{\bigcap_{j\in[7]} \mathcal{A}_j\right\} + \sum_{j\in[6]} \Pr\{\mathcal{A}_j^{\mathrm{c}}\}, \tag{A34}$$

where (A34) follows from the union bound and the fact that $\Pr\{\mathcal{A} \cap \mathcal{B}\} \leq \Pr\{\mathcal{B}\}$ for any two sets $\mathcal{A}$ and $\mathcal{B}$. The proof of Lemma A1 is completed by showing that

$$\sum_{j\in[6]} \Pr\{\mathcal{A}_j^{\mathrm{c}}\} \leq (2k+2)\exp(-n\eta). \tag{A35}$$

In the remainder of this subsection, we show that (A35) holds. Recall the joint distribution of G in (31). In the following, when we use a (conditional) distribution starting with P, we mean that the (conditional) distribution is induced by the joint distribution P_G in (31).

Using the definition of $\mathcal{A}_1$ in (A7),

$$\Pr\{\mathcal{A}_1^{\mathrm{c}}\} = \sum_{x^n\in\mathcal{X}^n} P_X^n(x^n) 1\{P_X^n(x^n) \leq \exp(-n\eta) Q_{X^n}(x^n)\} \tag{A36}$$

$$\leq \exp(-n\eta). \tag{A37}$$

Similarly to (A37), it follows that

$$\Pr\{\mathcal{A}_2^{\mathrm{c}}\} = \sum_{g\in\mathcal{A}_2^{\mathrm{c}}} P_G(g) \tag{A38}$$

$$= \sum_{x^n,s^k,\mathbf{y}^k} P_{XY^k}(x^n,\mathbf{y}^k)\Big(\prod_{j\in[k]} P_{S_j|X^n}(s_j|x^n)\Big)1\{P^n_{Y^k|X}(\mathbf{y}^k|x^n) \le \exp(-n\eta)Q_{\mathbf{Y}^k|X^nS^k}(\mathbf{y}^k|x^n,s^k)\} \tag{A39}$$

$$\le \exp(-n\eta)\sum_{x^n,s^k,\mathbf{y}^k} P^n_X(x^n)Q_{\mathbf{Y}^k|X^nS^k}(\mathbf{y}^k|x^n,s^k)\Big(\prod_{j\in[k]} P_{S_j|X^n}(s_j|x^n)\Big) \tag{A40}$$

$$\le \exp(-n\eta), \tag{A41}$$

$$\Pr\{\mathcal{A}_3^{\mathrm{c}}\} = \sum_{g\in\mathcal{A}_3^{\mathrm{c}}} P_G(g) \tag{A42}$$

$$\le \exp(-n\eta)\sum_{x^n,\mathbf{y}^k,s^k,\hat{x}_1^n} P_{Y_1^nS_1}(y_1^n,s_1)P_{\hat{X}_1^n|Y_1^nS_1}(\hat{x}_1^n|y_1^n,s_1)Q_{X^n\mathbf{Y}^{k\setminus 1}S_2^k|Y_1^nS_1\hat{X}_1^n}(x^n,\mathbf{y}^{k\setminus 1},s_2^k|y_1^n,s_1,\hat{x}_1^n) \tag{A43}$$

$$\le \exp(-n\eta), \tag{A44}$$

Furthermore, using the definition of $\mathcal{A}_4$ in (A10) and the union bound,

$$\Pr\{\mathcal{A}_4^{\mathrm{c}}\} \le \sum_{j\in[2:k]} \exp(-n\eta)\sum_{x^n,\mathbf{y}^k,s^k,\hat{\mathbf{x}}^j} P^n_{XY^k}(x^n,\mathbf{y}^k)\Big(\prod_{l\in[k]} P_{S_l|X^n}(s_l|x^n)\Big)\Big(\prod_{l\in[j-1]} P_{\hat{X}_l^n|Y_l^nS^l}(\hat{x}_l^n|y_l^n,s_l)\Big) \tag{A45}$$

$$\times\, Q_{\hat{X}_j^n|X^n\mathbf{Y}^kS^k\hat{\mathbf{X}}^{j-1}}(\hat{x}_j^n|x^n,\mathbf{y}^k,s^k,\hat{\mathbf{x}}^{j-1}) \tag{A46}$$

$$\le (k-1)\exp(-n\eta). \tag{A47}$$

Furthermore, using the definition of $\mathcal{A}_5$ in (A11),

$$\Pr\{\mathcal{A}_5^{\mathrm{c}}\} \le \sum_{x^n,s_1} P_{S_1|X^n}(s_1|x^n)\exp(-n(R_1+\eta))P_{X^n|S_1}(x^n|s_1) \tag{A48}$$

$$\le \sum_{x^n,s_1} \exp(-n(R_1+\eta))P_{X^n|S_1}(x^n|s_1) \tag{A49}$$

$$= \sum_{s_1} \exp(-n(\eta+R_1)) \tag{A50}$$

$$\le \exp(-n\eta), \tag{A51}$$

where (A49) follows since $P_{S_1|X^n}(s_1|x^n) \le 1$ for all (x^n,s_1), and (A51) follows since $\sum_{s_1} = |\mathcal{W}_1| = M_1 \le \exp(nR_1)$.

Using the definition of $\mathcal{A}_6$ in (A12) and the union bound similarly to (A47), it follows that

$$\Pr\{\mathcal{A}_6^{\mathrm{c}}\} \le \sum_{j\in[2:k]}\sum_{x^n,s^j} P_{S^{j-1}}(s^{j-1})\exp(-n\eta)P_{X^n|S^j}(x^n|s^j)\exp\Big(-n\Big(R_j - \sum_{l\in[j-1]} R_l\Big)\Big)P_{S_j|X^n}(s_j|x^n) \tag{A52}$$

$$\le \sum_{j\in[2:k]} \exp(-n\eta)\sum_{x^n,s^j} P_{S^{j-1}}(s^{j-1})P_{X^n|S^j}(x^n|s^j)\exp\Big(-n\Big(R_j - \sum_{l\in[j-1]} R_l\Big)\Big) \tag{A53}$$

$$\le \sum_{j\in[2:k]} \exp(-n\eta)\sum_{s_j}\exp\Big(-n\Big(R_j - \sum_{l\in[j-1]} R_l\Big)\Big) \tag{A54}$$

$$\le (k-1)\exp(-n\eta), \tag{A55}$$

where (A53) follows since $P_{S_j|X^n}(s_j|x^n) \le 1$ for all (x^n,s_j) and (A55) follows since $\sum_{s_j} = |\mathcal{M}_j| = M_j \le \exp(n(R_j - \sum_{l\in[j-1]} R_l))$.

Appendix D. Proof of Lemma 3

For any $(\mu, \alpha^k, \beta^k) \in \mathbb{R}_+ \times [0,1]^{2k}$ satisfying (15), for $i \in [4]$, define $\mathcal{F}_i = \mathcal{B}_i$ (cf. (33) to (36)) and for $i \in [5:7]$, define

$$\mathcal{F}_5 := \left\{ g : \mu\alpha_1 R_1 \geq \frac{\mu\alpha_1}{n} \sum_{i\in[n]} \log \frac{Q_{X_i|W_{1,i}}(x_i|w_{1,i})}{P_X(x_i)} - \mu\alpha_1\eta \right\}, \tag{A56}$$

$$\mathcal{F}_6 := \left\{ g : \mu\alpha_j (R_j - \sum_{l\in[j-1]} R_l) \geq \sum_{i\in[n]} \frac{\mu\alpha_j}{n} \log \frac{Q_{X_i|W_{1,i}W_{2,i}}(x_i|w_{1,i}, w_{2,i})}{P_{X_i|W_{1,i}}(x_i|w_{1,i})} - \mu\alpha_j\eta, \ \forall\, j \in [2:k] \right\}, \tag{A57}$$

$$\mathcal{F}_7 := \left\{ g : \mu\beta_j D_j \geq \frac{\mu\beta_j}{n} \sum_{i\in[n]} \log\exp(d_1(x_i, \hat{x}_{1,i})), \ \forall j \in [k] \right\}. \tag{A58}$$

Furthermore, let

$$c(\mu, \alpha^k) := k + 2 + \sum_{j\in[k]} \mu\alpha_j. \tag{A59}$$

Using Lemma 2 and definitions in (A56) to (A59), we obtain that

$$\begin{aligned}
&\mathrm{P}_{\mathrm{c}}^{(n)}(D^k) - (2k+2)\exp(-n\eta) \\
&\leq \Pr\Big\{ \bigcap_{i\in[7]} \mathcal{F}_i \Big\} && \text{(A60)} \\
&\leq \Pr\Big\{ n(\mu\kappa^{(\alpha^k,\beta^k)}(R^k, D^k) + c(\mu, \alpha^k)\eta) \geq \sum_{i\in[n]} f_{Q_i,P_i}^{(\mu,\alpha^k,\beta^k)}(T_i) \Big\} && \text{(A61)} \\
&\leq \exp\Big\{ n\lambda(\mu\kappa^{(\alpha^k,\beta^k)}(R^k, D^k) + c(\mu, \alpha^k)\eta) + \log\mathbb{E}\Big[\exp\Big(-\lambda \sum_{i\in[n]} f_{Q_i,P_i}^{(\mu,\alpha^k,\beta^k)}(T_i) \Big) \Big] \Big\} && \text{(A62)} \\
&= \exp\Big\{ n\Big(\lambda\mu\kappa^{(\alpha^k,\beta^k)}(R^k, D^k) + \lambda c(\mu, \alpha^k)\eta - \frac{1}{n}\Omega^{(\lambda,\mu,\alpha^k,\beta^k)}(\{P_i, Q_i\}_{i\in[n]}) \Big) \Big\}, && \text{(A63)}
\end{aligned}$$

where (A62) follows from Cramér's bound in Lemma 13 of [31] and (A63) follows from the definition of $\Omega^{(\lambda,\mu,\alpha^k,\beta^k)}(\{P_i, Q_i\}_{i\in[n]})$ in (42).

Choose η such that

$$-\eta = \lambda\mu\kappa^{(\alpha^k,\beta^k)}(R^k, D^k) + \lambda c(\mu, \alpha^k)\eta - \frac{1}{n}\Omega^{(\lambda,\mu,\alpha^k,\beta^k)}(\{P_i, Q_i\}_{i\in[n]}), \tag{A64}$$

i.e.,

$$\eta = \frac{\frac{1}{n}\Omega^{(\lambda,\mu,\alpha^k,\beta^k)}(\{P_i, Q_i\}_{i\in[n]}) - \lambda\mu\kappa^{(\alpha^k,\beta^k)}(R^k, D^k)}{1 + \lambda c(\mu, \alpha^k)}. \tag{A65}$$

The proof of Lemma 3 is completed by combining (A63) and (A65).

Appendix E. Proof of Lemma 4

Recall that for each $i \in [n]$, we use t_i to denote $(x_i, y_{1,i}^{k,i}, w_{1,i}^{k,i}, \hat{x}_{1,i}^{k,i})$ and use T_i similarly. Recall that the auxiliary random variables are chosen as $w_{1,i} = (x^{i-1}, y_1^{i-1}, \ldots, y_k^{i-1}, s_1)$ and $w_{j,i} = s_j$ for all $j \in [2:k]$. Using the definition of $f_{Q_i,P_i}^{(\mu,\alpha^k,\beta^k)}$ in (41), define

$$h_{Q_i,P_i}^{(\lambda,\mu,\alpha^k,\beta^k)}(t_i) := \exp\Big(-\lambda f_{Q_i,P_i}^{(\mu,\alpha^k,\beta^k)}(t_i) \Big). \tag{A66}$$

Recall the joint distribution of G in (31). For each $j \in [n]$, define

$$\tilde{C}_j := \sum_g P_G(g) \prod_{i\in[j]} h_{Q_i,P_i}^{(\lambda,\mu,\alpha^k,\beta^k)}(t_i), \tag{A67}$$

$$P_G^{(\lambda,\mu,\alpha^k,\beta^k)|j}(g) := \frac{P_G(g)\prod_{i\in[j]} h_{Q_i,P_i}^{(\lambda,\mu,\alpha^k,\beta^k)}(t_i)}{\tilde{C}_i}, \tag{A68}$$

$$\Lambda_j^{(\lambda,\mu,\alpha^k,\beta^k)}(\{Q_i,P_i\}_{i\in[n]}) := \frac{\tilde{C}_j}{\tilde{C}_{j-1}}. \tag{A69}$$

Combining (42) and (A69),

$$\exp\Big(-\Omega_{(\{P_i,Q_i\}_{i\in[n]})}^{(\lambda,\mu,\alpha^k,\beta^k)}\Big) = \mathbb{E}\Big[\prod_{i\in[n]} h_{Q_i,P_i}^{(\lambda,\mu,\alpha^k,\beta^k)}(T_i)\Big] \tag{A70}$$

$$= \sum_{g\in\mathcal{G}} P_G(g) \prod_{i\in[n]} h_{Q_i,P_i}^{(\lambda,\mu,\alpha^k,\beta^k)}(t_i) \tag{A71}$$

$$= \prod_{i\in[n]} \Lambda_i^{(\lambda,\mu,\alpha^k,\beta^k)}(\{Q_i,P_i\}). \tag{A72}$$

Furthermore, similar to Lemma 5 of [25], we obtain the following lemma, which is critical in the proof of Lemma 4.

Lemma A2. *For each $j \in [n]$,*

$$\Lambda_j^{(\lambda,\mu,\alpha^k,\beta^k)}(\{Q_i,P_i\}_{i\in[n]}) = \sum_{g\in\mathcal{G}} P_G^{(\lambda,\mu,\alpha^k,\beta^k)|j-1}(g) h_{Q_j,P_j}^{(\mu,\alpha^k,\beta^k)}(t_i). \tag{A73}$$

Furthermore, for each $j \in [n]$, define

$$P^{(\lambda,\mu,\alpha^k,\beta^k)}(t_j) := \sum_{\substack{x_{j+1}^n, y_{1,j+1}^n,\ldots,y_{k,j+1}^n,\\ \hat{x}_1^{j-1},\ldots,\hat{x}_k^{j-1},\hat{x}_{1,j+1}^n,\ldots,\hat{x}_{k,j+1}^n}} P_G^{(\lambda,\mu,\alpha^k,\beta^k)|j-1}(g). \tag{A74}$$

Using Lemma A2 and (A74), it follows that for each $j \in [n]$,

$$\Lambda_j^{(\lambda,\mu,\alpha^k,\beta^k)}(\{Q_i,P_i\}_{i\in[n]}) = \sum_{t_j} P^{(\lambda,\mu,\alpha^k,\beta^k)}(t_j) h_{Q_j,P_j}^{(\mu,\alpha^k,\beta^k)}(t_j). \tag{A75}$$

Recall that the auxiliary distributions $\{Q_i\}_{i\in[n]}$ can be arbitrary distributions. Following the recursive method in [25], for each $i \in [n]$, we choose Q_i such that

$$Q_i(t_i) = P^{(\lambda,\mu,\alpha^k,\beta^k)}(t_i). \tag{A76}$$

Let $Q_{C_i|D_i}$, where $C_i \in \mathcal{T}_i$ and $D_i \in \mathcal{T}_i$, be induced by Q_i. Using the definition of $h_{Q_i,P_i}^{(\lambda,\mu,\alpha^k,\beta^k)}(t_i)$ in (A66), we define

$$\xi_{Q_i,P_i}^{(\lambda,\mu,\alpha^k,\beta^k)}(t_i) := h_{Q_i,P_i}^{(\lambda,\mu,\alpha^k,\beta^k)}(t_i)\left(\frac{P_{X_iY_{2,i}^{k,i}W_{2,i}^{k,i}|Y_{1,i}W_{1,i}}(x_i,y_{2,i}^{k,i},w_{2,i}^{k,i}|y_{1,i},w_{1,i})}{Q_{X_iY_{2,i}^{k,i}W_{2,i}^{k,i}|Y_{1,i}W_{1,i}}(x_i,y_{2,i}^{k,i},w_{2,i}^{k,i}|y_{1,i},w_{1,i})}\right)^{-\lambda}$$
$$\times\left(\prod_{j\in[2:k]}\frac{P_{\hat{X}_{j,i}|Y_{j,i}W_{1,i}^{j,i}}(\hat{x}_{j,i}|y_{j,i},w_{1,i}^{j,i})}{Q_{\hat{X}_{j,i}|Y_{j,i}W_{1,i}^{j,i}}(\hat{x}_{j,i}|y_{j,i},w_{1,i}^{j,i})}\right)^{-\lambda}\left(\frac{P_{X_i|W_{1,i}}(x_i|w_{1,i})}{Q_{X_i|W_{1,i}}(x_i|w_{1,i})}\right)^{-\lambda\mu\alpha_1} \tag{A77}$$
$$\times\prod_{j\in[2:k]}\left(\frac{P_{X_i|W_{1,i}^{j-1,i}}(x_i,w_{1,i}^{j-1,i})}{Q_{X_i|W_{1,i}^{j-1,i}}(x_i,w_{1,i}^{j-1,i})}\right)^{-\lambda\mu\alpha_j}.$$

In the following, for simplicity, we let $\Psi := 1-k\lambda-\sum_{j\in[k]}\lambda\mu\alpha_j$. Combining (A74) and (A75), we obtain that for each $l\in[n]$,

$$\Lambda_l^{(\lambda,\mu,\alpha^k,\beta^k)}(\{Q_i,P_i\}_{i\in[n]})$$
$$=\mathbb{E}_{Q_l}\left[h_{Q_l,P_l}^{(\mu,\alpha^k,\beta^k)}(T_l)\right] \tag{A78}$$
$$=\mathbb{E}_{Q_l}\left[\xi_{Q_l,P_l}^{(\mu,\alpha^k,\beta^k)}(T_l)\left(\frac{P_{X_lY_{2,l}^{k,l}W_{2,l}^{k,l}|Y_{1,l}W_{1,l}}(x_l,y_{2,l}^{k,l},w_{2,l}^{k,l}|y_{1,l},w_{1,l})}{Q_{X_lY_{2,l}^{k,l}W_{2,l}^{k,l}|Y_{1,l}W_{1,l}}(x_l,y_{2,l}^{k,l},w_{2,l}^{k,l}|y_{1,l},w_{1,l})}\right)^{\lambda}\left(\prod_{j\in[2:k]}\frac{P_{\hat{X}_{j,l}|Y_{j,l}W_{1,l}^{j,l}}(\hat{x}_{j,l}|y_{j,l},w_{1,l}^{j,l})}{Q_{\hat{X}_{j,l}|Y_{j,l}W_{1,l}^{j,l}}(\hat{x}_{j,l}|y_{j,l},w_{1,l}^{j,l})}\right)^{\lambda}\right.$$
$$\left.\times\left(\frac{P_{X_l|W_{1,l}}(x_l|w_{1,l})}{Q_{X_l|W_{1,l}}(x_l|w_{1,l})}\right)^{\lambda\mu\alpha_1}\prod_{j\in[2:k]}\left(\frac{P_{X_l|W_{1,l}^{j-1,l}}(x_l,w_{1,l}^{j-1,l})}{Q_{X_l|W_{1,l}^{j-1,l}}(x_l,w_{1,l}^{j-1,l})}\right)^{\lambda\mu\alpha_j}\right] \tag{A79}$$
$$\leq\left(\mathbb{E}_{Q_l}\left[\left(\xi_{Q_l,P_l}^{(\mu,\alpha^k,\beta^k)}(T_l)\right)^{\frac{1}{\Psi}}\right]\right)^{\Psi}\left(\mathbb{E}\left[\frac{P_{X_lY_{2,l}^{k,l}W_{2,l}^{k,l}|Y_{1,l}W_{1,l}}(x_l,y_{2,l}^{k,l},w_{2,l}^{k,l}|y_{1,l},w_{1,l})}{Q_{X_lY_{2,l}^{k,l}W_{2,l}^{k,l}|Y_{1,l}W_{1,l}}(x_l,y_{2,l}^{k,l},w_{2,l}^{k,l}|y_{1,l},w_{1,l})}\right]\right)^{\lambda}$$
$$\times\prod_{j\in[2:k]}\left(\mathbb{E}\left[\frac{P_{\hat{X}_{j,l}|Y_{j,l}W_{1,l}^{j,l}}(\hat{x}_{j,l}|y_{j,l},w_{1,l}^{j,l})}{Q_{\hat{X}_{j,l}|Y_{j,l}W_{1,l}^{j,l}}(\hat{x}_{j,l}|y_{j,l},w_{1,l}^{j,l})}\right]\right)^{\lambda}\left(\mathbb{E}\left[\frac{P_{X_l|W_{1,l}}(x_l|w_{1,l})}{Q_{X_l|W_{1,l}}(x_l|w_{1,l})}\right]\right)^{\lambda\mu\alpha_1}$$
$$\times\prod_{j\in[2:k]}\left(\mathbb{E}\left[\frac{P_{X_l|W_{1,l}^{j-1,l}}(x_l,w_{1,l}^{j-1,l})}{Q_{X_l|W_{1,l}^{j-1,l}}(x_l,w_{1,l}^{j-1,l})}\right]\right)^{\lambda\mu\alpha_j} \tag{A80}$$
$$\leq\exp\left(-\Psi\Omega^{(\frac{\lambda}{\Psi},\mu,\alpha^k,\beta^k)}(Q_j)\right) \tag{A81}$$
$$=\exp\left(-\frac{\Omega^{(\theta,\mu,\alpha^k,\beta^k)}(Q_j)}{1+k\theta+\sum_{j\in[k]}\theta\mu\alpha_j}\right) \tag{A82}$$
$$\leq\exp\left(-\min_{Q_j\in\mathcal{P}(\mathcal{T}_j)}\frac{\Omega^{(\theta,\mu,\alpha^k,\beta^k)}(Q_j)}{1+k\theta+\sum_{j\in[k]}\theta\mu\alpha_j}\right) \tag{A83}$$
$$=\exp\left(-\frac{\Omega^{(\theta,\mu,\alpha^k,\beta^k)}}{1+k\theta+\sum_{j\in[k]}\theta\mu\alpha_j}\right) \tag{A84}$$

where (A80) results from Hölder's inequality, (A81) follows from the definitions of $\Omega^{(\theta,\mu,\alpha^k,\beta^k)}(\cdot)$ in (17) and $\xi_{Q_j,P_j}^{(\mu,\alpha^k,\beta^k)}(\cdot)$ in (A77), (A82) follows from the result in (46), and (A84) follows from the definition of $\Omega^{(\theta,\mu,\alpha^k,\beta^k)}$ in (18) and the fact it is sufficient to consider distributions Q_j with cardinality bounds $W_{1,j}\leq|\mathcal{X}|$ and $W_{2,j}\leq|\mathcal{X}|^2$ for the optimization problem in (A83) (the proof of this fact is similar to Property 4(a) in [25] and thus omitted).

The proof of Lemma 4 is completed by combining (A72) and (A84).

Appendix F. Proof of Lemma 6

Appendix F.1. Proof of Claim (i)

For any $Q_T \in \mathcal{Q}$ (see (14)), let $P_T \in \mathcal{P}_{\rm sh}$ (see (53)) be chosen such that $P_{W^k|X} = Q_{W^k|X}$ and $P_{\hat{X}_j|Y_jW^j} = Q_{\hat{X}_j|Y_jW^j}$ for all $j \in [k]$.

In the following, we drop the subscript of distributions when there is no confusion. For any $(\theta, \mu, \alpha^k, \beta^k) \in \mathbb{R}_+^2 \times [0,1]^{2k}$ satisfying (15) and

$$\sum_{j\in[2:k]} \mu\alpha_j \le 1 \text{ and } \forall\, l \in [k],\ \theta \le \frac{1}{1+\mu\alpha_l}, \tag{A85}$$

using the definition of $\Omega^{(\theta,\mu,\alpha^k,\beta^k)}(Q_T)$ in (17), we obtain

$$\begin{aligned}
&\exp\Big(-\Omega^{(\theta,\mu,\alpha^k,\beta^k)}(Q_T)\Big)\\
&= \mathbb{E}_{Q_T}\Bigg[\Bigg(\frac{P(X,Y^k)Q(X,Y^{k\setminus 1},W^{k\setminus 1}|Y_1,W_1)\big(\prod_{j\in[2:k]}Q(\hat{X}_j|Y,W^j)\big)}{Q(X)Q(Y^k|X,W^k)Q(X,Y^{k\setminus 1},W^{k\setminus 1}|Y_1,W_1,\hat{X}_1)\big(\prod_{j\in[2:k]}Q(\hat{X}_j|X,Y^k,W^k,\hat{X}^{j-1})\big)}\Bigg)^{\theta}\\
&\qquad\times\Big(\frac{P(X)}{Q(X|W_1)}\Big)^{\theta\mu\alpha_1}\Big(\prod_{j\in[2:k]}\Big(\frac{Q(X|W^{j-1})}{Q(X|W^j)}\Big)^{\theta\mu\alpha_j}\Big)\exp\Big(-\theta\mu\big(\sum_{j\in[k]}\beta_j d_j(X,\hat{X}_j)\big)\Big)\Bigg] &\text{(A86)}\\
&= \mathbb{E}_{Q_T}\Bigg[\Big(\frac{P(T)}{Q(T)}\Big)^{\theta}\Big(\frac{P(X)}{Q(X|W_1)}\Big)^{\theta\mu\alpha_1}\Big(\prod_{j\in[2:k]}\Big(\frac{Q(X|W^{j-1})}{Q(X|W^j)}\Big)^{\theta\mu\alpha_j}\Big)\exp\Big(-\theta\mu\big(\sum_{j\in[k]}\beta_j d_j(X,\hat{X}_j)\big)\Big)\Bigg] &\text{(A87)}\\
&= \mathbb{E}_{Q_T}\Bigg[\Big(\frac{P(T)}{Q(T)}\Big)^{\theta}\Big(\frac{P(X)}{P(X|W_1)}\Big)^{\theta\mu\alpha_1}\Big(\prod_{j\in[2:k]}\Big(\frac{Q(X|W^{j-1})}{P(X|W^j)}\Big)^{\theta\mu\alpha_j}\Big)\exp\Big(-\theta\mu\big(\sum_{j\in[k]}\beta_j d_j(X,\hat{X}_j)\big)\Big)\Bigg]\\
&\qquad\times\Big(\prod_{j\in[k]}\Big(\frac{P(X|W^j)}{Q(X|W^j)}\Big)^{\theta\mu\alpha_j}\Big) &\text{(A88)}\\
&\le \Bigg(\mathbb{E}_{Q_T}\Bigg[\Big(\frac{P(T)}{Q(T)}\Big)\Big(\frac{P(X)}{P(X|W_1)}\Big)^{\mu\alpha_1}\Big(\prod_{j\in[2:k]}\Big(\frac{Q(X|W^{j-1})}{P(X|W^j)}\Big)^{\mu\alpha_j}\Big)\exp\Big(-\mu\big(\sum_{j\in[k]}\beta_j d_j(X,\hat{X}_j)\big)\Big)\Bigg]\Bigg)^{\theta}\\
&\qquad\times\prod_{j\in[k]}\Big(\mathbb{E}_{Q_T}\Big[\Big(\frac{P(X|W^j)}{Q(X|W^j)}\Big)^{\frac{\theta\mu\alpha_j}{1-\theta}}\Big]\Big)^{1-\theta} &\text{(A89)}\\
&\le \Bigg(\mathbb{E}_{P_T}\Bigg[\Big(\frac{P(X)}{P(X|W_1)}\Big)^{\mu\alpha_1}\Big(\prod_{j\in[2:k]}\Big(\frac{Q(X|W^{j-1})}{P(X|W^j)}\Big)^{\mu\alpha_j}\Big)\exp\Big(-\mu\big(\sum_{j\in[k]}\beta_j d_j(X,\hat{X}_j)\big)\Big)\Bigg]\Bigg)^{\theta} &\text{(A90)}\\
&= \Bigg(\mathbb{E}_{P_T}\Bigg[\Big(\frac{P(X)}{P(X|W_1)}\Big)^{\mu\alpha_1}\Big(\prod_{j\in[2:k]}\Big(\frac{P(X|W^{j-1})}{P(X|W^j)}\Big)^{\mu\alpha_j}\Big)\exp\Big(-\mu\big(\sum_{j\in[k]}\beta_j d_j(X,\hat{X}_j)\big)\Big)\\
&\qquad\times\prod_{j\in[2:k]}\Big(\frac{Q(X|W^{j-1})}{P(X|W^{j-1})}\Big)^{\mu\alpha_j}\Bigg]\Bigg)^{\theta} &\text{(A91)}\\
&= \Bigg(\mathbb{E}_{P_T}\Bigg[\Big(\Big(\frac{P(X)}{P(X|W_1)}\Big)^{\mu\alpha_1}\Big(\prod_{j\in[2:k]}\Big(\frac{P(X|W^{j-1})}{P(X|W^j)}\Big)^{\mu\alpha_j}\Big)\exp\Big(-\mu\big(\sum_{j\in[k]}\beta_j d_j(X,\hat{X}_j)\big)\Big)\Big)^{\frac{1}{1-\sum_{j\in[2:k]}\mu\alpha_j}}\Bigg]\Bigg)^{\theta(1-\sum_{j\in[2:k]}\mu\alpha_j)}\\
&\qquad\times\prod_{j\in[2:k]}\Big(\mathbb{E}_{P_T}\Big[\Big(\frac{Q(X|W^{j-1})}{P(X|W^{j-1})}\Big)\Big]\Big)^{\theta\mu\alpha_j} &\text{(A92)}\\
&= \exp\Big(-\theta(1-\sum_{j\in[2:k]}\mu\alpha_j)\tilde{\Omega}\Big(\frac{\mu}{1-\sum_{j\in[2:k]}\mu\alpha_j},\alpha^k,\beta^k\Big)\Big), &\text{(A93)}
\end{aligned}$$

where (A87) follows since (i) with our choice of $P_T \in \mathcal{P}_{\rm sh}$, we have

$$P(T) = P(X,Y^k)P(W^k|X)\big(\prod_{j\in[k]}P(\hat{X}_j|Y_j,W^j)\big) \tag{A94}$$

and (ii) the following equality holds

$$\frac{Q(X,Y^{k\setminus 1},W^{k\setminus 1}|Y_1,W_1)}{Q(X,Y^{k\setminus 1},W^{k\setminus 1}|Y_1,W_1,\hat{X}_1)} = \frac{Q(\hat{X}_1|Y_1,W_1)}{Q(\hat{X}_1|X,Y^k,W^k)}, \tag{A95}$$

Equation (A89) follows from Hölder's inequality, (A90) follows from the concavity of X^a for $a \in [0,1]$ and the choice of θ which ensures $\frac{\theta\mu\alpha_j}{1-\theta} \leq 1$ for all $j \in [k]$, (A92) follows by applying Hölder's inequality and recalling that $\sum_{j\in[2:k]} \mu\alpha_j \leq 1$, and (A93) follows from the definition of $\tilde{\Omega}^{(\lambda,\alpha^k,\beta^k)}(P_T)$ in (58).

Therefore, for any $(\theta,\mu,\alpha^k,\beta^k) \in \mathbb{R}_+^2 \times [0,1]^{2k}$ satisfying (15) and (A85), using the definition of $\Omega^{(\theta,\mu,\alpha^k,\beta^k)}$ in (18) and the result in (A93), we have that

$$\Omega^{(\theta,\mu,\alpha^k,\beta^k)} \geq \theta(1-\sum_{j\in[2:k]}\mu\alpha_j)\tilde{\Omega}\Big(\frac{\mu}{1-\sum_{j\in[2:k]}\mu\alpha_j},\alpha^k,\beta^k\Big). \tag{A96}$$

Recalling the definition of $F(R^k,D^k)$ in (21) and using the result in (A96), we have

$$\begin{aligned}
&F(R^k,D^k)\\
&= \sup_{(\theta,\mu,\alpha^k,\beta^k)\in\mathbb{R}_+^2\times[0,1]^{2k}:\ \sum_{i\in[k]}(\alpha_i+\beta_i)=1} \frac{\Omega^{(\theta,\mu,\alpha^k,\beta^k)} - \theta\mu\kappa^{(\alpha^k,\beta^k)}(R^k,D^k)}{1+(2k+2)\theta+\sum_{j\in[k]}2\theta\mu\alpha_j} &\text{(A97)}\\
&\geq \sup_{\substack{(\theta,\mu,\alpha^k,\beta^k)\in\mathbb{R}_+^2\times[0,1]^{2k}:\\ (15)\text{ and }(A85)}} \frac{\theta(1-\sum_{j\in[2:k]}\mu\alpha_j)\tilde{\Omega}\Big(\frac{\mu}{1-\sum_{j\in[2:k]}\mu\alpha_j},\alpha^k,\beta^k\Big) - \theta\mu\kappa^{(\alpha^k,\beta^k)}(R^k,D^k)}{1+(2k+2)\theta+\sum_{j\in[k]}2\theta\mu\alpha_j} &\text{(A98)}\\
&= \sup_{\substack{(\mu,\alpha^k,\beta^k)\in\mathbb{R}_+\times[0,1]^{2k}:\\ (15)\text{ and }\mu\leq\frac{1}{\sum_{j\in[2:k]}\alpha_j}}} \sup_{\theta\in\mathbb{R}_+:\max_{j\in[k]}\theta(1+\mu\alpha_j)\leq 1} \frac{\theta(1-\sum_{j\in[2:k]}\mu\alpha_j)\tilde{\Omega}\Big(\frac{\mu}{1-\sum_{j\in[2:k]}\mu\alpha_j},\alpha^k,\beta^k\Big) - \theta\mu\kappa^{(\alpha^k,\beta^k)}(R^k,D^k)}{1+(2k+2)\theta+\sum_{j\in[k]}2\theta\mu\alpha_j} &\text{(A99)}\\
&= \sup_{\substack{(\mu,\alpha^k,\beta^k)\in\mathbb{R}_+\times[0,1]^{2k}:\\ (15)\text{ and }\mu\leq\frac{1}{\sum_{j\in[2:k]}\alpha_j}}} \frac{(1-\sum_{j\in[2:k]}\mu\alpha_j)\tilde{\Omega}\Big(\frac{\mu}{1-\sum_{j\in[2:k]}\mu\alpha_j},\alpha^k,\beta^k\Big) - \mu\kappa^{(\alpha^k,\beta^k)}(R^k,D^k)}{2k+3+\mu\alpha^+ +\sum_{l\in[k]}2\mu\alpha_l} &\text{(A100)}\\
&= \sup_{(\lambda,\alpha^k,\beta^k)\in\mathbb{R}_+\times[0,1]^{2k}:\ (15)} \frac{\tilde{\Omega}^{(\lambda,\alpha^k,\beta^k)} - \lambda\kappa^{(\alpha^k,\beta^k)}(R^k,D^k)}{2k+3+\lambda\alpha^+ +\sum_{j\in[2:k]}\lambda(2k+3)\alpha_j+\sum_{l\in[k]}2\lambda\alpha_l} &\text{(A101)}\\
&= \tilde{F}(R^k,D^k), &\text{(A102)}
\end{aligned}$$

where (A100) follows since

$$\begin{aligned}
\sup_{\theta\in\mathbb{R}_+:\max_{j\in[k]}\theta(1+\mu\alpha_j)\leq 1} \frac{\theta}{1+(2k+2)\theta+\sum_{j\in[k]}2\theta\mu\alpha_j} &= \min_{j\in[k]} \frac{1}{2k+3+\mu\alpha_j+\sum_{l\in[k]}2\mu\alpha_l} &\text{(A103)}\\
&= \frac{1}{2k+3+\mu\alpha^+ +\sum_{l\in[k]}2\mu\alpha_l}, &\text{(A104)}
\end{aligned}$$

and (A101) follows by choosing $\lambda = \frac{\mu}{1-\sum_{j\in[2:k]}\mu\alpha_j}$ and (A102) follows from the definition of $\tilde{F}$ in (62).

Appendix F.2. Proof of Claim (ii)

Recall the definitions of $\tilde{\Omega}^{(\lambda,\alpha^k,\beta^k)}(P_T)$ in (57) and $P_T^{(\lambda,\alpha^k,\beta^k)}$ in (63). By simple calculation, one can verify that

$$\begin{aligned}
\frac{\partial\tilde{\Omega}^{(\lambda,\alpha^k,\beta^k)}(P_T)}{\partial\lambda} &= \mathbb{E}_{P_T^{(\lambda,\alpha^k,\beta^k)}}\big[\tilde{\omega}_{P_T}^{(\alpha^k,\beta^k)}(T)\big], &\text{(A105)}\\
\frac{\partial^2\tilde{\Omega}^{(\lambda,\alpha^k,\beta^k)}(P_T)}{\partial\lambda^2} &= -\mathrm{Var}_{P_T^{(\lambda,\alpha^k,\beta^k)}}\big[\tilde{\omega}_{P_T}^{(\alpha^k,\beta^k)}(T)\big]. &\text{(A106)}
\end{aligned}$$

Applying Taylor expansion to $\tilde{\Omega}^{(\lambda,\alpha^k,\beta^k)}(P_T)$ at around $\lambda = 0$ and combining (A105), (A106), we have that for any $P_T \in \mathcal{P}_{\text{sh}}$ and any $\lambda \in [1, \frac{1}{\sum_{j\in[k]}\alpha_j}]$, there exists $\tau \in [0, \lambda]$ such that

$$\tilde{\Omega}^{(\lambda,\alpha^k,\beta^k)}(P_T) = \tilde{\Omega}^{(0,\alpha^k,\beta^k)}(P_T) + \lambda \mathbb{E}_{P_T^{(0,\alpha^k,\beta^k)}}\left[\tilde{\omega}_{P_T}^{(\alpha^k,\beta^k)}(T)\right] - \frac{\lambda^2}{2}\text{Var}_{P_T^{(\tau,\alpha^k,\beta^k)}}\left[\tilde{\omega}_{P_T}^{(\alpha^k,\beta^k)}(T)\right] \tag{A107}$$

$$\geq \lambda \mathbb{E}_{P_T}\left[\tilde{\omega}_{P_T}^{(\alpha^k,\beta^k)}(T)\right] - \frac{\lambda^2\rho}{2}, \tag{A108}$$

where (A108) follows from the definitions in (57), (63) and (64).

Using the definitions in (54), (57) and (60) and the result in (A108), we have that for any $\lambda \in [0, \frac{1}{\sum_{j\in[k]}\alpha_j}]$,

$$\tilde{\Omega}^{(\lambda,\alpha^k,\beta^k)} = \min_{P_T\in\mathcal{P}_{\text{sh}}} \tilde{\Omega}^{(\lambda,\alpha^k,\beta^k)}(P_T) \tag{A109}$$

$$\geq \lambda \text{R}^{(\alpha^k,\beta^k)} - \frac{\lambda^2\rho}{2}. \tag{A110}$$

For any rate-distortion tuple outside the rate-distortion region, i.e., $(R^k, D^k) \notin \mathcal{R}$, from Lemma 5, we conclude that there exists $(\alpha^{k,*}, \beta^{k,*}) \in [0,1]^{2k}$ satisfying (15) such that for some positive $\delta \in [0, \rho]$

$$\kappa^{(\alpha^{k,^*},\beta^{k,*})}(R^k, D^k) \leq R^{(\alpha^{k,^*},\beta^{k,*})} - \delta. \tag{A111}$$

Using the definition of $\tilde{F}(R^k, D^k)$ in (62), we have

$$\tilde{F}(R^k, D^k) = \sup_{(\lambda,\alpha^k,\beta^k)\in\mathbb{R}_+\times[0,1]^{2k}:\ (15)} \frac{\tilde{\Omega}^{(\lambda,\alpha^k,\beta^k)} - \lambda\kappa^{(\alpha^k,\beta^k)}(R^k, D^k)}{2k+3+\lambda\alpha^+ + \sum_{j\in[2:k]}\lambda(2k+3)\alpha_j + \sum_{l\in[k]}2\lambda\alpha_l} \tag{A112}$$

$$\geq \sup_{\lambda\in[0,1]} \frac{\tilde{\Omega}^{(\lambda,\alpha^{k,*},\beta^{k,*})} - \lambda\kappa^{(\alpha^{k,*},\beta^{k,*})}(R^k, D^k)}{2k+3+\lambda\max_{j\in[k]}\alpha_j^* + \sum_{j\in[2:k]}\lambda(2k+3)\alpha_j^* + \sum_{l\in[k]}2\lambda\alpha_l^*} \tag{A113}$$

$$\geq \sup_{\lambda\in[0,1]} \frac{\lambda\delta - \frac{\lambda^2\rho}{2}}{2k+9} \tag{A114}$$

$$= \frac{\delta^2}{2(2k+9)\rho}, \tag{A115}$$

where (A114) follows from the results in (A110), (A111) and the inequality

$$2k+3+\lambda\max_{j\in[k]}\alpha_j^* + \sum_{j\in[2:k]}\lambda(2k+3)\alpha_j^* + \sum_{l\in[k]}2\lambda\alpha_l^* \leq 2k+9, \tag{A116}$$

resulting from the constraints that $(\alpha^{k,*}, \beta^{k,^*}) \in [0,2]^{2k}$ satisfying (15) and $\lambda \in [0,1]$.

References

1. Maor, A.; Merhav, N. On Successive Refinement with Causal Side Information at the Decoders. *IEEE Trans. Inf. Theory* **2008**, *54*, 332–343. [CrossRef]
2. Tian, C.; Diggavi, S.N. On multistage successive refinement for Wyner–Ziv source coding with degraded side informations. *IEEE Trans. Inf. Theory* **2007**, *53*, 2946–2960. [CrossRef]
3. Steinberg, Y.; Merhav, N. On successive refinement for the Wyner-Ziv problem. *IEEE Trans. Inf. Theory* **2004**, *50*, 1636–1654. [CrossRef]
4. Equitz, W.H.; Cover, T.M. Successive refinement of information. *IEEE Trans. Inf. Theory* **1991**, *37*, 269–275. [CrossRef]

5. Koshelev, V. Estimation of mean error for a discrete successive-approximation scheme. *Probl. Peredachi Informatsii* **1981**, *17*, 20–33.
6. Rimoldi, B. Successive refinement of information: Characterization of the achievable rates. *IEEE Trans. Inf. Theory* **1994**, *40*, 253–259. [CrossRef]
7. Kanlis, A.; Narayan, P. Error exponents for successive refinement by partitioning. *IEEE Trans. Inf. Theory* **1996**, *42*, 275–282. [CrossRef]
8. No, A.; Ingber, A.; Weissman, T. Strong Successive Refinability and Rate-Distortion-Complexity Tradeoff. *IEEE Trans. Inf. Theory* **2016**, *62*, 3618–3635. [CrossRef]
9. Zhou, L.; Tan, V.Y.F.; Motani, M. Second-Order and Moderate Deviation Asymptotics for Successive Refinement. *IEEE Trans. Inf. Theory* **2017**, *63*, 2896–2921. [CrossRef]
10. Tuncel, E.; Rose, K. Additive successive refinement. *IEEE Trans. Inf. Theory* **2003**, *49*, 1983–1991. [CrossRef]
11. Chow, J.; Berger, T. Failure of successive refinement for symmetric Gaussian mixtures. *IEEE Trans. Inf. Theory* **1997**, *43*, 350–352. [CrossRef]
12. Tuncel, E.; Rose, K. Error exponents in scalable source coding. *IEEE Trans. Inf. Theory* **2003**, *49*, 289–296. [CrossRef]
13. Effros, M. Distortion-rate bounds for fixed- and variable-rate multiresolution source codes. *IEEE Trans. Inf. Theory* **1999**, *45*, 1887–1910. [CrossRef]
14. Weissman, T.; Gamal, A.E. Source Coding With Limited-Look-Ahead Side Information at the Decoder. *IEEE Trans. Inf. Theory* **2006**, *52*, 5218–5239. [CrossRef]
15. El Gamal, A.; Kim, Y.H. *Network Information Theory*; Cambridge University Press: Cambridge, UK, 2011.
16. Timo, R.; Vellambi, B.N. Two lossy source coding problems with causal side-information. In Proceedings of the 2009 IEEE International Symposium on Information Theory, Seoul, South Korea, 28 June–3 July 2009; pp. 1040–1044. [CrossRef]
17. Gu, W.H.; Effros, M. Source coding for a simple multi-hop network. In Proceedings of the International Symposium on Information Theory (ISIT 2005), Adelaide, SA, Australia, 4–9 September 2005.
18. Gray, R.; Wyner, A. Source coding for a simple network. *Bell Syst. Tech. J.* **1974**, *53*, 1681–1721. [CrossRef]
19. Maor, A.; Merhav, N. On Successive Refinement for the Kaspi/Heegard–Berger Problem. *IEEE Trans. Inf. Theory* **2010**, *56*, 3930–3945. [CrossRef]
20. Heegard, C.; Berger, T. Rate distortion when side information may be absent. *IEEE Trans. Inf. Theory* **1985**, *31*, 727–734. [CrossRef]
21. Chia, Y.K.; Weissman, T. Cascade and Triangular source coding with causal side information. In Proceedings of the 2011 IEEE International Symposium on Information Theory Proceedings, St. Petersburg, Russia, 31 July–5 August 2011; pp. 1683–1687. [CrossRef]
22. Fong, S.L.; Tan, V.Y.F. A Proof of the Strong Converse Theorem for Gaussian Multiple Access Channels. *IEEE Trans. Inf. Theory* **2016**, *62*, 4376–4394. [CrossRef]
23. Oohama, Y. Exponent Function for One Helper Source Coding Problem at Rates outside the Rate Region. *arXiv* **2015**, arXiv:1504.05891.
24. Oohama, Y. New Strong Converse for Asymmetric Broadcast Channels. *arXiv* **2016**, arXiv:1604.02901.
25. Oohama, Y. Exponential Strong Converse for Source Coding with Side Information at the Decoder. *Entropy* **2018**, *20*, 352. [CrossRef]
26. Ahlswede, R.; Korner, J. Source coding with side information and a converse for degraded broadcast channels. *IEEE Trans. Inf. Theory* **1975**, *21*, 629–637. [CrossRef]
27. Wyner, A.D. On source coding with side information at the decoder. *IEEE Trans. Inf. Theory* **1975**, *21*, 294–300. [CrossRef]
28. Korner, J.; Marton, K. General broadcast channels with degraded message sets. *IEEE Trans. Inf. Theory* **1977**, *23*, 60–64. [CrossRef]
29. Wyner, A.D.; Ziv, J. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Inf. Theory* **1976**, *22*, 1–10. [CrossRef]
30. Tuncel, E.; Gündüz, D. Identification and Lossy Reconstruction in Noisy Databases. *IEEE Trans. Inf. Theory* **2014**, *60*, 822–831. [CrossRef]
31. Zhou, L.; Tan, V.Y.F.; Motani, M. Exponential Strong Converse for Content Identification with Lossy Recovery. *IEEE Trans. Inf. Theory* **2018**, *64*, 5879—5897. [CrossRef]

32. Wyner, A. The common information of two dependent random variables. *IEEE Trans. Inf. Theory* **1975**, *21*, 163–179. [CrossRef]
33. Yu, L.; Tan, V.Y.F. Wyner's Common Information Under Rényi Divergence Measures. *IEEE Trans. Inf. Theory* **2018**, *64*, 3616–3632. [CrossRef]
34. Csiszár, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*; Cambridge University Press: Cambridge, UK, 2011.
35. Gu, W.; Effros, M. A strong converse for a collection of network source coding problems. In Proceedings of the 2009 IEEE International Symposium on Information Theory, Seoul, South Korea, 28 June–3 July 2009; pp. 2316–2320. [CrossRef]
36. Liu, J.; van Handel, R.; Verdú, S. Beyond the Blowing-Up Lemma: Optimal Second-Order Converses via Reverse Hypercontractivity. Preprint. Available online: http://web.mit.edu/jingbo/www/preprints/msl-blup.pdf (accessed on 17 April 2019).
37. Tyagi, H.; Watanabe, S. Strong Converse using Change of Measure. *arXiv* **2018**, arXiv:1805.04625.
38. Ahlswede, R.; Csiszár, I. Hypothesis testing with communication constraints. *IEEE Trans. Inf. Theory* **1986**, *32*, 533–542. [CrossRef]
39. Salehkalaibar, S.; Wigger, M.; Wang, L. Hypothesis testing in multi-hop networks. *arXiv* **2017**, arXiv:1708.05198.
40. Cao, D.; Zhou, L.; Tan, V.Y.F. Strong Converse for Hypothesis Testing Against Independence over a Two-Hop Network. *arXiv* **2018**, arXiv:1808.05366.
41. Marton, K. Error exponent for source coding with a fidelity criterion. *IEEE Trans. Inf. Theory* **1974**, *20*, 197–199. [CrossRef]
42. Yassaee, M.H.; Aref, M.R.; Gohari, A. A technique for deriving one-shot achievability results in network information theory. In Proceedings of the 2013 IEEE International Symposium on Information Theory, Istanbul, Turkey, 7–12 July 2013; pp. 1287–1291.
43. Kostina, V.; Verdú, S. Fixed-length lossy compression in the finite blocklength regime. *IEEE Trans. Inf. Theory* **2012**, *58*, 3309–3338. [CrossRef]
44. Tan, V.Y.F. Asymptotic Estimates in Information Theory with Non-Vanishing Error Probabilities. *Found. Trends Commun. Inf. Theory* **2014**, *11*, 1–184. [CrossRef]
45. Pradhan, S.S.; Chou, J.; Ramchandran, K. Duality between source coding and channel coding and its extension to the side information case. *IEEE Trans. Inf. Theory* **2003**, *49*, 1181–1203. [CrossRef]
46. Slepian, D.; Wolf, J.K. Noiseless coding of correlated information sources. *IEEE Trans. Inf. Theory* **1973**, *19*, 471–480. [CrossRef]
47. Gelfand, S. Coding for channel with random parameters. *Probl. Contr. Inf. Theory* **1980**, *9*, 19–31.
48. Shannon, C.E. Channels with side information at the transmitter. *IBM J. Res. Dev.* **1958**, *2*, 289–293. [CrossRef]
49. Sigurjónsson, S.; Kim, Y.H. On multiple user channels with state information at the transmitters. In Proceedings of the International Symposium on Information Theory (ISIT 2005), Adelaide, SA, Australia, 4–9 September 2005; pp. 72–76.
50. Zaidi, A.; Shitz, S.S. On cooperative multiple access channels with delayed CSI at transmitters. *IEEE Trans. Inf. Theory* **2014**, *60*, 6204–6230. [CrossRef]
51. Shkel, Y.Y.; Verdú, S. A single-shot approach to lossy source coding under logarithmic loss. *IEEE Trans. Inf. Theory* **2018**, *64*, 129–147. [CrossRef]
52. Courtade, T.A.; Weissman, T. Multiterminal source coding under logarithmic loss. *IEEE Trans. Inf. Theory* **2014**, *60*, 740–761. [CrossRef]
53. Strassen, V. Asymptotische abschätzungen in shannons informationstheorie. In *Transactions of the Third Prague Conference on Information Theory etc*; Czechoslovak Academy of Sciences: Prague, Czech Republic, 1962; pp. 689–723.
54. Hayashi, M. Information spectrum approach to second-order coding rate in channel coding. *IEEE Trans. Inf. Theory* **2009**, *55*, 4947–4966. [CrossRef]

55. Polyanskiy, Y.; Poor, H.V.; Verdú, S. Channel coding rate in the finite blocklength regime. *IEEE Trans. Inf. Theory* **2010**, *56*, 2307–2359. [CrossRef]
56. Kostina, V. Lossy Data Compression: Non-Asymptotic Fundamental Limits. Ph.D. Thesis, Department of Electrical Engineering, Princeton University, Princeton, NJ, USA, 2013.