entropy

MDPI

Article

Trade-offs between Error Exponents and Excess-Rate Exponents of Typical Slepian–Wolf Codes

Ran Tamir (Averbuch) * and Neri Merhav

The Andrew and Erna Viterbi Faculty of Electrical Engineering, Technion—Israel Institute of Technology, Technion City, Haifa 3200003, Israel; merhav@ee.technion.ac.il

* Correspondence: rans@campus.technion.ac.il; Tel.: +972-52-8360585

Abstract: Typical random codes (TRCs) in a communication scenario of source coding with side information in the decoder is the main subject of this work. We study the semi-deterministic code ensemble, which is a certain variant of the ordinary random binning code ensemble. In this code ensemble, the relatively small type classes of the source are deterministically partitioned into the available bins in a one-to-one manner. As a consequence, the error probability decreases dramatically. The random binning error exponent and the error exponent of the TRCs are derived and proved to be equal to one another in a few important special cases. We show that the performance under optimal decoding can be attained also by certain universal decoders, e.g., the stochastic likelihood decoder with an empirical entropy metric. Moreover, we discuss the trade-offs between the error exponent and the excess-rate exponent for the typical random semi-deterministic code and characterize its optimal rate function. We show that for any pair of correlated information sources, both error and excess-rate probabilities exponential vanish when the blocklength tends to infinity.

Keywords: Slepian–Wolf coding; variable-rate coding; error exponent; excess-rate exponent; typical random code

Citation: Tamir, R.; Merhav, N. Trade-offs between Error Exponents and Excess-Rate Exponents of Typical Slepian–Wolf Codes. *Entropy* **2021**, *23*, 265. http://doi.org/10.3390/e23030265

Received: 3 February 2021
Accepted: 21 February 2021
Published: 24 February 2021

1. Introduction

As is well known, the random coding error exponent is defined by

$$\mathsf{E}_{\mathrm{r}}(R) = \lim_{n\to\infty}\Big\{-\tfrac{1}{n}\log\mathbb{E}[P_{\mathrm{e}}(\mathcal{C}_n)]\Big\}, \tag{1}$$

where R is the coding rate, $P_{\mathrm{e}}(\mathcal{C}_n)$ is the error probability of a codebook $\mathcal{C}_n$, and the expectation is with respect to (w.r.t.) the randomness of $\mathcal{C}_n$ across the ensemble of codes. The error exponent of the typical random code (TRC) is defined as [1]

$$\mathsf{E}_{\mathrm{trc}}(R) = \lim_{n\to\infty}\Big\{-\tfrac{1}{n}\mathbb{E}[\log P_{\mathrm{e}}(\mathcal{C}_n)]\Big\}. \tag{2}$$

We believe that the error exponent of the TRC is the more relevant performance metric, as it captures the most likely error exponent of a randomly selected code, as opposed to the random coding error exponent, which is dominated by the relatively poor codes of the ensemble, rather than the channel noise, at relatively low coding rates. In addition, since in random coding analysis, the code is selected at random and remains fixed, it seems reasonable to study the performance of the chosen code itself instead of directly considering the ensemble performance.

To the best of our knowledge, not much is known about TRCs. In [2], Barg and Forney considered TRCs with independently and identically distributed codewords, along with typical linear codes, for the special case of the binary symmetric channel with maximum likelihood (ML) decoding. It was also shown that at a certain range of low rates, $E_{\mathrm{trc}}(R)$

lies between $E_{\mathrm{r}}(R)$ and the expurgated exponent, $E_{\mathrm{ex}}(R)$. In [3] Nazari *et al.* provided bounds on the error exponents of TRCs for both discrete memoryless channels (DMC) and multiple-access channels. In a recent article by Merhav [1], an exact single-letter expression has been derived for the error exponent of typical, random, fixed composition codes, over DMCs, and a wide class of (stochastic) decoders, collectively referred to as the generalized likelihood decoder (GLD). Later, Merhav studied error exponents of TRCs for the colored Gaussian channel [4], typical random trellis codes [5], and a Lagrange dual lower bound for the TRC exponent [6]. Large deviations around the TRC exponent were studied in [7].

While originally defined for pure channel coding [1–3], the notion of TRCs has natural analogues in other settings as well, such as source coding with side information in the decoder [8]. Typical random Slepian–Wolf (SW) code of a certain variant of the ordinary variable-rate random binning code ensemble is the main theme of this work. The random coding error exponent of SW coding, based on fixed-rate (FR) random binning, was first addressed by Gallager in [9], and improved later on by the expurgated bound in [10] and [11]. Variable-rate (VR) SW coding received less attention in the literature; VR codes under average rate constraint have been studied in [12] and proved to outperform FR codes in terms of error exponents. Optimum trade-offs between the error exponent and the excess-rate exponent in VR coding were analyzed in [13]. Sphere-packing upper bounds for source coding with side information in the FR and VR regimes were studied in [9] and [12], respectively. More works where exponential error bounds in source coding have been studied are [14–18].

It turns out that both the FR and VR ensembles suffer from an intrinsic deficiency, caused by statistical fluctuations in the sizes of the bins that are populated by the relatively small type classes of the source. This fundamental problem of the ordinary ensembles is alleviated in some variant of the ordinary VR ensemble-the semi-deterministic (SD) code ensemble, which has already been proposed and studied in its FR version in [18]. In the SD code ensemble, for source type classes which are exponentially larger than the space of the available bins, we just randomly assign each source sequence into one of the bins, as is done in ordinary random binning. Otherwise, for relatively small type classes, we deterministically order each source sequence into a different bin, which provides a one-to-one mapping. This way, all these relatively small source type classes do not contribute to the probability of error. The main results concerning the SD code are the following:

1. The random binning error exponent and the error exponent of the TRC are derived in Theorems 1 and 2, respectively, and proved in Theorem 3 to be equal to one another in a few important special cases, which include the matched likelihood decoder, the MAP decoder, and the universal minimum entropy decoder. To the best of our knowledge, this phenomenon has not been seen elsewhere before, since the TRC exponent usually improves upon the random coding exponent. As a byproduct, we are able to provide a relatively simple expression for the TRC exponent.
2. We prove in Theorem 4 that the error exponent of the TRC under MAP decoding is also attained by two universal decoders: the minimum entropy decoder and the stochastic entropy decoder, which is a GLD with an empirical conditional entropy metric. As far as we know, this result is first of its kind in source coding; in other scenarios, the random coding bound is attained also by universal decoders, but here, we find that the TRC exponent is also universally achievable. Moreover, while the likelihood decoder and the MAP decoder have similar error exponents [19], here we prove a similar result, but for two universal decoders (one stochastic and one deterministic) that share the same metric.
3. We discuss the trade-offs between the error exponent and the excess-rate exponent for a typical random SD code, similarly to [13], but with a different notion of the excess-rate event, which takes into account the available side information. In Theorem 5, we provide an expression for the optimal rate function that guarantees a required level for the error exponent of the typical random SD code. Analogously, Theorem 6 proposes an expression for the optimal rate function that guarantees a required level

for the excess-rate exponent. Furthermore, we show that for any pair of correlated information sources, the typical random SD code attains both exponentially vanishing error and excess-rate probabilities.

The remaining part of the paper is organized as follows. In Section 2, we establish notation conventions. In Section 3, we formalize the model, the coding technique, the main objectives of this work, and we review some background. In Section 4, we provide the main results concerning error exponents and universal decoding in the SD ensemble, and in Section 5, we discuss the trade-offs between the error exponent and the excess-rate exponent.

2. Notation Conventions

Throughout the paper, random variables will be denoted by capital letters, realizations will be denoted by the corresponding lower case letters, and their alphabets will be denoted by calligraphic letters. Random vectors and their realizations will be denoted, respectively, by boldface capital and lower case letters. Their alphabets will be superscripted by their dimensions. Sources and channels will be subscripted by the names of the relevant random variables/vectors and their conditionings, whenever applicable, following the standard notation conventions, e.g., Q_U, $Q_{V|U}$, and so on. When there is no room for ambiguity, these subscripts will be omitted. For a generic joint distribution $Q_{UV} = \{Q_{UV}(u,v), u \in \mathcal{U}, v \in \mathcal{V}\}$, which will often be abbreviated by Q, information measures will be denoted in the conventional manner, but with a subscript Q; that is, $H_Q(U)$ is the marginal entropy of U, $H_Q(U|V)$ is the conditional entropy of U given V, and $I_Q(U;V) = H_Q(U) - H_Q(U|V)$ is the mutual information between U and V. The Kullback–Leibler divergence between two probability distributions, Q_{UV} and P_{UV}, is defined as

$$D(Q_{UV}\|P_{UV}) = \sum_{(u,v)\in\mathcal{U}\times\mathcal{V}} Q_{UV}(u,v)\log\frac{Q_{UV}(u,v)}{P_{UV}(u,v)}, \tag{3}$$

where logarithms, here and throughout the sequel, are understood to be taken to the natural base. The probability of an event $\mathcal{E}$ will be denoted by $\mathbb{P}\{\mathcal{E}\}$, and the expectation operator w.r.t. a probability distribution Q will be denoted by $\mathbb{E}_Q[\cdot]$, where the subscript will often be omitted. For two positive sequences, $\{a_n\}$ and $\{b_n\}$, the notation $a_n \doteq b_n$ will stand for equality in the exponential scale, that is, $\lim_{n\to\infty}(1/n)\log(a_n/b_n) = 0$. Similarly, $a_n \dot{\leq} b_n$ means that $\limsup_{n\to\infty}(1/n)\log(a_n/b_n) \leq 0$, and so on. The indicator function of an event $\mathcal{A}$ will be denoted by $\mathbb{1}\{\mathcal{A}\}$. The notation $[t]_+$ will stand for $\max\{0,t\}$.

The empirical distribution of a sequence $\boldsymbol{u} \in \mathcal{U}^n$, which will be denoted by $\hat{P}_{\boldsymbol{u}}$, is the vector of relative frequencies, $\hat{P}_{\boldsymbol{u}}(u)$, of each symbol $u \in \mathcal{U}$ in $\boldsymbol{u}$. The type class of $\boldsymbol{u} \in \mathcal{U}^n$, denoted $\mathcal{T}(\boldsymbol{u})$, is the set of all vectors $\boldsymbol{u}'$ with $\hat{P}_{\boldsymbol{u}'} = \hat{P}_{\boldsymbol{u}}$. When we wish to emphasize the dependence of the type class on the empirical distribution $\hat{P}$, we will denote it by $\mathcal{T}(\hat{P})$. The set of all types of vectors of length n over $\mathcal{U}$ will be denoted by $\mathcal{P}_n(\mathcal{U})$, and the set of all possible types over $\mathcal{U}$ will be denoted by $\mathcal{P}(\mathcal{U}) = \bigcup_{n=1}^{\infty}\mathcal{P}_n(\mathcal{U})$. Information measures associated with empirical distributions will be denoted with 'hats' and will be subscripted by the sequences from which they are induced. For example, the entropy associated with $\hat{P}_{\boldsymbol{u}}$, which is the empirical entropy of $\boldsymbol{u}$, will be denoted by $\hat{H}_{\boldsymbol{u}}(U)$. Similar conventions will apply to the joint empirical distribution, the joint type class, the conditional empirical distributions and the conditional type classes associated with pairs (and multiples) of sequences of length n. Accordingly, $\hat{P}_{\boldsymbol{u}\boldsymbol{v}}$ would be the joint empirical distribution of $(\boldsymbol{u},\boldsymbol{v}) = \{(u_i,v_i)\}_{i=1}^n$, $\mathcal{T}(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})$ will denote the joint type class of $(\boldsymbol{u},\boldsymbol{v})$, $\mathcal{T}(\hat{P}_{\boldsymbol{u}|\boldsymbol{v}}|\boldsymbol{v})$ will stand for the conditional type class of $\boldsymbol{u}$ given $\boldsymbol{v}$, $\hat{H}_{\boldsymbol{u}\boldsymbol{v}}(U|V)$ will be the empirical conditional entropy, and so on. Likewise, when we wish to emphasize the dependence of empirical information measures upon a given empirical distribution Q, we denote them using the subscript Q, as described above.

3. Problem Formulation and Background

3.1. Problem Formulation

Let $(\boldsymbol{U}, \boldsymbol{V}) = \{(U_t, V_t)\}_{t=1}^{n}$ be n independent copies of a pair of random variables, $(U, V) \sim P_{UV}$, taking on values in finite alphabets, $\mathcal{U}$ and $\mathcal{V}$, respectively. The vector $\boldsymbol{U}$ will designate the source vector to be encoded and the vector $\boldsymbol{V}$ will serve as correlated side information available to the decoder. In ordinary VR binning, the coding rate is not fixed for every $\boldsymbol{u} \in \mathcal{U}^n$, but depends on its empirical distribution. Let us denote a rate function by $R(\cdot)$, which is a given continuous function from the probability simplex of $\mathcal{U}$ to the set of nonnegative reals. In that manner, for every type $Q_U \in \mathcal{P}_n(\mathcal{U})$, all source sequences in $\mathcal{T}(Q_U)$ are randomly partitioned into $e^{nR(Q_U)}$ bins. Every source sequence is encoded by its bin index, denoted by $\mathcal{B}(\boldsymbol{u})$, along with a header that indicates its type index, which requires only a negligible extra rate when n is large enough.

The SD code ensemble is a refinement of the ordinary VR code: for types with $H_Q(U) \geq R(Q_U)$, i.e., type classes which are exponentially larger than the space of available bins, we just randomly assign each source sequence into one out of the $e^{nR(Q_U)}$ bins. For the other types, we deterministically order each member of $\mathcal{T}(Q_U)$ into a different bin. This way, all type classes with $H_Q(U) < R(Q_U)$ do not contribute to the probability of error. The entire binning code of source sequences of blocklength n, i.e., the set $\{\mathcal{B}(\boldsymbol{u})\}_{\boldsymbol{u} \in \mathcal{U}^n}$, is denoted by $\mathcal{B}_n$. A sequence of SW codes, $\{\mathcal{B}_n\}_{n \geq 1}$, indexed by the block length n, will be denoted by $\mathcal{B}$.

The decoder estimates $\boldsymbol{u}$ based on the bin index $\mathcal{B}(\boldsymbol{u})$, the type index $\mathcal{T}(\boldsymbol{u})$, and the side information sequence $\boldsymbol{v}$, which is a realization of $\boldsymbol{V}$. The optimal (MAP) decoder estimates $\boldsymbol{u}$ according to

$$\hat{\boldsymbol{u}} = \underset{\boldsymbol{u}' \in \mathcal{B}(\boldsymbol{u}) \cap \mathcal{T}(\boldsymbol{u})}{\arg\max} \; P(\boldsymbol{u}', \boldsymbol{v}). \tag{4}$$

As in [1,20], we consider here the GLD. The GLD estimates $\boldsymbol{u}$ stochastically, using the bin index $\mathcal{B}(\boldsymbol{u})$, the type index $\mathcal{T}(\boldsymbol{u})$, and the side information sequence $\boldsymbol{v}$, according to the following posterior distribution

$$\mathbb{P}\big\{\hat{\boldsymbol{U}} = \boldsymbol{u}' \big| \boldsymbol{v}, \mathcal{B}(\boldsymbol{u}), \mathcal{T}(\boldsymbol{u})\big\} = \frac{\exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}{\sum_{\tilde{\boldsymbol{u}} \in \mathcal{B}(\boldsymbol{u}) \cap \mathcal{T}(\boldsymbol{u})} \exp\{nf(\hat{P}_{\tilde{\boldsymbol{u}}\boldsymbol{v}})\}}, \tag{5}$$

where $\hat{P}_{\boldsymbol{u}\boldsymbol{v}}$ is the empirical distribution of $(\boldsymbol{u}, \boldsymbol{v})$ and $f(\cdot)$ is a given continuous, real valued functional of this empirical distribution. The GLD provides a unified framework which covers several important special cases, e.g., matched decoding, mismatched decoding, MAP decoding, and universal decoding (similarly to the α-decoders described in [11]). A more detailed discussion is given in [20].

The probability of error is the probability of the event $\{\hat{\boldsymbol{U}} \neq \boldsymbol{U}\}$. For a given binning code $\mathcal{B}_n$, the probability of error is given by

$$P_{\mathrm{e}}(\mathcal{B}_n) = \sum_{\boldsymbol{u}, \boldsymbol{v}} P(\boldsymbol{u}, \boldsymbol{v}) \cdot \mathbb{1}\big\{\hat{H}_{\boldsymbol{u}}(U) \geq R(\hat{P}_{\boldsymbol{u}})\big\} \cdot \frac{\sum_{\boldsymbol{u}' \in \mathcal{B}(\boldsymbol{u}) \cap \mathcal{T}(\boldsymbol{u}), \boldsymbol{u}' \neq \boldsymbol{u}} \exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}{\sum_{\tilde{\boldsymbol{u}} \in \mathcal{B}(\boldsymbol{u}) \cap \mathcal{T}(\boldsymbol{u})} \exp\{nf(\hat{P}_{\tilde{\boldsymbol{u}}\boldsymbol{v}})\}}. \tag{6}$$

For a given rate function, we derive the random binning exponent of this ensemble, which is defined by

$$\mathsf{E}_{\mathrm{r}}(R(\cdot)) = \lim_{n \to \infty} \left\{ -\frac{\log \mathbb{E}[P_{\mathrm{e}}(\mathcal{B}_n)]}{n} \right\}, \tag{7}$$

and compare it to the TRC exponent, which is

$$\mathsf{E}_{\mathrm{trc}}(R(\cdot)) = \lim_{n\to\infty}\left\{-\frac{\mathbb{E}[\log P_{\mathrm{e}}(\mathcal{B}_n)]}{n}\right\}. \tag{8}$$

Although it is unclear that the limits in (7) and (8) exist a priori, it will be evident from the analyses in Appendixes A and B, respectively.

One way to define the excess-rate probability is as $\mathbb{P}\{R(\hat{P}_{\boldsymbol{U}}) \geq R\}$, where R is some target rate [13]. Due to the availability of side information in the decoder, it makes sense to require a target rate which depends on the pair $(\boldsymbol{u}, \boldsymbol{v})$. Since the lowest possible compression rate in this setting is given by $H_P(U|V)$ [8], given $\boldsymbol{U} = \boldsymbol{u}$ and $\boldsymbol{V} = \boldsymbol{v}$, it is reasonable to adopt $\hat{H}_{\boldsymbol{uv}}(U|V)$ as a reference rate. Hence, an alternative definition of the excess-rate probability of a code $\mathcal{B}_n$, is as $p_{\mathrm{er}}(\mathcal{B}_n, R(\cdot), \Delta) = \mathbb{P}\{R(\hat{P}_{\boldsymbol{U}}) \geq \hat{H}_{\boldsymbol{UV}}(U|V) + \Delta\}$, where $\Delta > 0$ is a redundancy threshold. (Note that the entire analysis remains intact if we allow a more general redundancy threshold as $\Delta = \Delta(\hat{P}_{\boldsymbol{uv}})$. This covers other alternatives for the excess-rate probability, e.g., $\mathbb{P}\{R(\hat{P}_{\boldsymbol{U}}) \geq R\}$ or $\mathbb{P}\{R(\hat{P}_{\boldsymbol{U}}) \geq \alpha\hat{H}_{\boldsymbol{U}}(U)\}$, $\alpha \in (0,1)$.) Accordingly, the excess-rate exponent function, achieved by a sequence of codes $\mathcal{B}$, is defined as

$$\mathsf{E}_{\mathrm{er}}(\mathcal{B}, R(\cdot), \Delta) = \liminf_{n\to\infty} -\frac{1}{n}\log p_{\mathrm{er}}(\mathcal{B}_n, R(\cdot), \Delta). \tag{9}$$

The main mission is to characterize the optimal trade-off between the error exponent and the excess-rate exponent for the typical random SD code, and the optimal rate function that attains a prescribed value for the error exponent of the typical random SD code.

3.2. Background

In pure channel coding, Merhav [1] has derived a single-letter expression for the error exponent of the typical random fixed composition code:

$$\mathsf{E}_{\mathrm{trc}}(R, Q_X) = \lim_{n\to\infty}\left\{-\tfrac{1}{n}\mathbb{E}[\log P_{\mathrm{e}}(\mathcal{C}_n)]\right\}. \tag{10}$$

In order to present the main result of [1], we define first a few quantities. Consider a DMC, $W = \{W(y|x), x \in \mathcal{X}, y \in \mathcal{Y}\}$, where $\mathcal{X}$ and $\mathcal{Y}$ are the finite input/output alphabets. Define

$$\alpha(R, Q_Y) = \max_{\{Q_{\tilde{X}|Y}:\ I_Q(\tilde{X};Y)\leq R,\ Q_{\tilde{X}}=Q_X\}} \{g(Q_{\tilde{X}Y}) - I_Q(\tilde{X};Y)\} + R, \tag{11}$$

where the function $g(\cdot)$, which is the decoding metric, is a continuous function that maps joint probability distributions over $\mathcal{X}\times\mathcal{Y}$ to real numbers. Additionally define

$$\begin{aligned}\Gamma(Q_{XX'}, R) = \min_{Q_{Y|XX'}} \{&D(Q_{Y|X}\|W|Q_X) + I_Q(X';Y|X)\\ &+ [\max\{g(Q_{XY}), \alpha(R, Q_Y)\} - g(Q_{X'Y})]_+\},\end{aligned} \tag{12}$$

where $D(Q_{Y|X}\|W|Q_X)$ is the conditional divergence between $Q_{Y|X}$ and W, averaged by Q_X. A brief intuitive explanation on the term $\Gamma(Q_{XX'}, R)$ can be found in [7](Section 4.1). Having defined the above quantities, the error exponent of the TRC is given by [1]

$$\mathsf{E}_{\mathrm{trc}}(R, Q_X) = \min_{\{Q_{X'|X}:\ I_Q(X;X')\leq 2R,\ Q_{X'}=Q_X\}} \{\Gamma(Q_{XX'}, R) + I_Q(X;X') - R\}. \tag{13}$$

Returning to the SW model, several articles have been written on error exponents for the FR and the VR codes. Here, we mention only those results that are directly relevant to the current work. The random binning and expurgated bounds of the FR ensemble in the SW model are given, respectively, by [11] (Section VI, Theorem 2), [10] (Appendix I, Theorem 1)

$$E_{\mathrm{r}}^{\mathrm{fr}}(R) = \min_{Q_U}\Big\{D(Q_U\|P_U) + E_{\mathrm{r}}(Q_U, P_{V|U}, H_Q(U) - R)\Big\}, \tag{14}$$

$$E_{\mathrm{ex}}^{\mathrm{fr}}(R) = \min_{Q_U}\Big\{D(Q_U\|P_U) + E_{\mathrm{ex}}(Q_U, P_{V|U}, H_Q(U) - R)\Big\}, \tag{15}$$

where $E_{\mathrm{r}}(Q_U, P_{V|U}, S)$ and $E_{\mathrm{ex}}(Q_U, P_{V|U}, S)$ are, respectively, the random coding and expurgated bounds associated with the channel $P_{V|U}$ w.r.t. the ensemble of fixed composition code of rate S, whose composition is Q_U. The exponent function $E_{\mathrm{r}}(Q_U, P_{V|U}, S)$ is given by

$$E_{\mathrm{r}}(Q_U, P_{V|U}, S) = \min_{Q_{V|U}}\{D(Q_{V|U}\|P_{V|U}|Q_U) + [I_Q(U;V) - S]_+\}, \tag{16}$$

and $E_{\mathrm{ex}}(Q_U, P_{V|U}, S)$ is given by

$$E_{\mathrm{ex}}(Q_U, P_{V|U}, S) = \min_{\{Q_{U'|U}:\ I_Q(U;U')\le S,\ Q_{U'}=Q_U\}}\{\mathbb{E}_{Q_{UU'}}[d_{P_{V|U}}(U,U')] + I_Q(U;U') - S\}, \tag{17}$$

where

$$d_{P_{V|U}}(u,u') = -\log\left[\sum_{v\in\mathcal{V}}\sqrt{P_{V|U}(v|u)P_{V|U}(v|u')}\right]. \tag{18}$$

The exact error exponent of VR random binning is given by [13] (Equation (34)):

$$E_{\mathrm{r}}^{\mathrm{vr}}(R(\cdot)) = \min_{Q_{UV}}\big\{D(Q_{UV}\|P_{UV}) + [R(Q_U) - H_Q(U|V)]_+\big\}. \tag{19}$$

4. Error Exponents and Universal Decoding

To present some of the results, we need a few more definitions. The minimum conditional entropy (MCE) decoder estimates $\boldsymbol{u}$, using the bin index $\mathcal{B}(\boldsymbol{u})$, the type index $\mathcal{T}(\boldsymbol{u})$, and the side information vector $\boldsymbol{v}$, according to

$$\hat{\boldsymbol{u}} = \operatorname*{arg\,min}_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u})\cap\mathcal{T}(\boldsymbol{u})} \hat{H}_{\boldsymbol{u}'\boldsymbol{v}}(U|V). \tag{20}$$

The stochastic conditional entropy (SCE) decoder is a special case of the GLD with the decoding metric $f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}}) = -\hat{H}_{\boldsymbol{u}\boldsymbol{v}}(U|V)$; i.e., it estimates $\boldsymbol{u}$ according to the following posterior distribution

$$\mathbb{P}\{\hat{\boldsymbol{U}} = \boldsymbol{u}'|\boldsymbol{v}, \mathcal{B}(\boldsymbol{u}), \mathcal{T}(\boldsymbol{u})\} = \frac{\exp\{-n\hat{H}_{\boldsymbol{u}'\boldsymbol{v}}(U|V)\}}{\sum_{\tilde{\boldsymbol{u}}\in\mathcal{B}(\boldsymbol{u})\cap\mathcal{T}(\boldsymbol{u})}\exp\{-n\hat{H}_{\tilde{\boldsymbol{u}}\boldsymbol{v}}(U|V)\}}. \tag{21}$$

First, we present random binning error exponents, which are modifications of (19) to this ensemble. Define the expression

$$E(Q_{UV}, R(\cdot)) = \min_{Q_{U'|V}} \left[R(Q_U) - H_Q(U'|V) + [f(Q_{UV}) - f(Q_{U'V})]_+\right]_+ \tag{22}$$

and the exponent functions:

$$E_{\mathrm{r,GLD}}(R(\cdot)) = \min_{\{Q_{UV}:\, H_Q(U)\geq R(Q_U)\}} \{D(Q_{UV}\|P_{UV}) + E(Q_{UV}, R(\cdot))\}, \tag{23}$$

and

$$E_{\mathrm{r,MAP}}(R(\cdot)) = \min_{\{Q_{UV}:\, H_Q(U)\geq R(Q_U)\}} \big\{D(Q_{UV}\|P_{UV}) + [R(Q_U) - H_Q(U|V)]_+\big\}. \tag{24}$$

The following result is proved in Appendix A.

Theorem 1. *Let $R(\cdot)$ be a given rate function. Then, for the SD ensemble,*

1. $\mathsf{E}_r(R(\cdot)) = E_{r,GLD}(R(\cdot))$ *for the GLD;*
2. $\mathsf{E}_r(R(\cdot)) = E_{r,MAP}(R(\cdot))$ *for the MAP and MCE decoders.*

As a matter of fact, a special case of the second part of Theorem 1 has already been proved in [18] for the FR regime, while here, we prove a stronger result, according to which, the MCE decoder attains the same random binning error exponent as the MAP decoder, in the VR coding regime too. The first part of Theorem 1 is completely new; it proposes a single letter expression for the random binning error exponent, for a wide family of stochastic and deterministic decoders. Additionally, note that an analogous result to the first part of Theorem 1 has been proved in [20]. Comparing the expressions in (19) and (24), namely, the random binning error exponents of the ordinary VR and the SD VR ensembles, respectively, we find that they differ at relatively high coding rates, since these minimization problems share the same objective but (24) also has the constraint $H_Q(U) \geq R(Q_U)$. The origin of this constraint is the deterministic coding of the relatively small type classes.

Next, we provide a single-letter expression for the error exponent of the TRCs in this ensemble. We define

$$\gamma(R(\cdot), Q_U, Q_V) = \max_{\left\{\substack{Q_{\tilde{U}|V}:\, Q_{\tilde{U}}=Q_U,\\ H_Q(\tilde{U}|V)\geq R(Q_{\tilde{U}})}\right\}} \{f(Q_{\tilde{U}V}) + H_Q(\tilde{U}|V)\} - R(Q_{\tilde{U}}) \tag{25}$$

and

$$\Psi(R(\cdot), Q_{UU'V}) = [\max\{f(Q_{UV}), \gamma(R(\cdot), Q_U, Q_V)\} - f(Q_{U'V})]_+. \tag{26}$$

Furthermore, define

$$\Lambda(Q_{UU'}, R(Q_U)) = \min_{Q_{V|UU'}} \{\Psi(R(Q_U), Q_{UU'V}) - H_Q(V|U, U') - \mathbb{E}_Q[\log P(V|U)]\}, \tag{27}$$

and the following exponent function:

$$E_{\text{trc,GLD}}(R(\cdot)) = \min_{\substack{\{Q_{UU'}:\, Q_{U'}=Q_U,\\ H_Q(U)\geq R(Q_U)\}}} \big\{\Lambda(Q_{UU'}, R(Q_U)) - \mathbb{E}_Q[\log P(U)] - H_Q(U,U') + R(Q_U)\big\}. \tag{28}$$

Then, the following theorem is proved in Appendix B.

Theorem 2. *Let $R(\cdot)$ be a given rate function. Then, for the SD ensemble and the GLD,*

$$\mathsf{E}_{trc}(R(\cdot)) = E_{trc,GLD}(R(\cdot)). \tag{29}$$

As explained before, an analogous result has already been proved in pure channel coding [1], and one can find a high degree of similarity between the expressions in (25)–(28) and the expressions in Section 3.2. While in channel coding, the coding rate is fixed, here, on the other hand, we allow the rate to depend on the type class of the source. In order to optimize the rate function, we constrain the problem by introducing the excess-rate exponent (9), which is the exponential rate of decay of the probability that the compression rate will be higher than some predefined level. A detailed discussion on optimal rate functions and optimal trade-offs between these two exponents can be found in Section 5.

The definition of the error exponent of the TRC as in (8) should not be taken for granted. The reason for that is the following. It turns out that the definition in (8) and the value of $-\frac{1}{n}\log P_{\text{e}}(\mathcal{B}_n)$ for the highly probable codes in the ensemble may not be the same, and they coincide if and only if the ensemble does not contain both zero error probability codes and positive error probability codes. For example, the FR ensemble in SW coding contains the one-to-one code (which obviously attains $P_{\text{e}}(\mathcal{B}_n) = 0$) as long as $R \geq \log|\mathcal{U}|$, but it is definitely not a typical code, at least when ordinary random binning is considered. Hence, in this case, we conclude that $-\frac{1}{n}\mathbb{E}[\log P_{\text{e}}(\mathcal{B}_n)] = \infty$, while the value of $-\frac{1}{n}\log P_{\text{e}}(\mathcal{B}_n)$ for the highly probable codes is still finite. As for the SD code ensemble, the definition in (8) indeed provides the error exponent of the highly probable codes in the ensemble, which is explained by the following reasoning. For any given rate function such that $R(Q_U) < H_Q(U)$ for at least one type class, all the type classes with $R(Q_U) < H_Q(U)$ are encoded by random binning; thus, all the codes in the ensemble have a strictly positive error probability, which implies that the value of $-\frac{1}{n}\log P_{\text{e}}(\mathcal{B}_n)$ concentrates around the error exponent of the TRC, as defined in (8).

The proof of Theorem 2 follows exactly the same lines as the proof of ([1] (Theorem 1)), except for one main modification: when we introduce the type class enumerator $N(Q_{UU'})$ (see below) and sum over joint types, the summation set becomes $\{Q_{UU'} : Q_{U'} = Q_U,\ H_Q(U) \geq R(Q_U)\}$, where the constraint $H_Q(U) \geq R(Q_U)$ is due to the indicator function in (6). Afterwards, the analysis of the type class enumerator yields the constraint $H_Q(U,U') \geq R(Q_U)$, which becomes redundant and thus omitted. This constraint is analogous to the constraint $I_Q(X;X') \leq 2R$ in the minimization of (13). The origin of $H_Q(U,U') \geq R(Q_U)$ is the following. Define

$$N(Q_{UU'}) = \sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\big\{\mathcal{B}(\boldsymbol{u}') = \mathcal{B}(\boldsymbol{u})\big\}, \tag{30}$$

which enumerate pairs of source sequences. Then, one of the main steps in the proof of Theorem 2 is deriving the high probability value of $N(Q_{UU'})$, which is 0 if $H_Q(U,U') < R(Q_U)$ (a relatively small set of source pair and relatively large number of bins) and $e^{n[H_Q(U,U')-R(Q_U)]}$ for $H_Q(U,U') \geq R(Q_U)$ (a large set of source sequence pair and a small number of bins). One should note that the analysis of $N(Q_{UU'})$ is not trivial, since it is not a binomial random variable; i.e., the enumerator $N(Q_{UU'})$ is given by the sum of dependent

binary random variables. For a sum N of independent binary random variables, ordinary tools from large deviation theory (e.g., the Chernoff bound) can be invoked for assessing the exponential moments $\mathbb{E}[N^s]$, $s \geq 0$, or the large deviation rate function of $\mathbb{P}\{N \geq e^{n\sigma}\}$, $\sigma \in \mathbb{R}$. For sums of dependent binary random variables, such as $N(Q_{UU'})$ in the current problem, this can no longer be done by the same techniques, and it requires more advanced tools (see, e.g., [1,4–6]).

It is possible to compare (23) and (28) analytically in the special cases of the matched or the mismatched likelihood decoders and the MCE decoder. In the following theorem, the choice $f(Q_{UV}) = \beta \mathbb{E}_Q[\log \tilde{P}(U,V)]$, where $\tilde{P}(U,V)$ is a possibly different source distribution than $P(U,V)$, corresponds to a family of stochastic mismatched decoders. We have the following result, the proof of which is given in Appendix D.

Theorem 3. *Consider the SD ensemble and a given rate function $R(\cdot)$. Then,*

1. *For a GLD with the decoding metric $f(Q) = \beta \mathbb{E}_Q[\log \tilde{P}(U,V)]$, for a given $\beta > 0$,*

$$E_{trc,GLD}(R(\cdot)) = E_{r,GLD}(R(\cdot)). \tag{31}$$

2. *For the MCE decoder,*

$$E_{trc,MCE}(R(\cdot)) = E_{r,MCE}(R(\cdot)). \tag{32}$$

This result is quite surprising at first glance, since one expects the error exponent of the TRC to be strictly better than the random binning error exponent, as in ordinary channel coding at relatively low coding rates [1,2]. This phenomenon is due to the fact that part of the source type classes are deterministically partitioned into bins in a one-to-one fashion, and hence do not affect the probability of error (notice that the constraint $H_Q(U) \geq R(Q_U)$ appears in both the random binning and the TRC exponents, while in the latter, it makes the original constraint $H_Q(U,U') \geq R(Q_U)$ redundant). In the cases of FR or ordinary VR binning, these relatively "thin" type classes dominated the error probability at relatively high binning rates, but now, by encoding them deterministically into the bins; other mechanisms dominate the error event, such as the channel noise (between $\boldsymbol{U}$ and $\boldsymbol{V}$) or the random binning of the type classes with $H_Q(U) \geq R(Q_U)$. The result of the second part of Theorem 3 is also nontrivial, since it establishes an equality between the error exponent of the TRC and the random binning error exponent, but now for a universal decoder.

Concerning universal decoding, it is already known [21] (Exercise 3.1.6), [13] that the random binning error exponents under optimal MAP decoding in both the FR and VR codes, given by (14) and (19), respectively, are also attained by the MCE decoder. Furthermore, a similar result for the SD ensemble has been proved here in Theorem 1. The natural question that arises is whether the error exponent of the TRC is also universally attainable. The following result, which is proved in Appendix E, provides a positive answer to this question.

Theorem 4. *Consider the SD ensemble and a given rate function $R(\cdot)$. Then, the error exponents of the TRC under the MAP, the MCE, and the SCE decoders are all equal; i.e.,*

$$E_{trc,MAP}(R(\cdot)) = E_{trc,MCE}(R(\cdot)) = E_{trc,SCE}(R(\cdot)). \tag{33}$$

Theorem 4 asserts that the error exponent of the typical random SD code is not affected if the optimal MAP decoder is replaced by a certain universal decoder, which must not even be deterministic. While the left hand equality in (33) follows immediately from the results of Theorems 1 and 3, the right hand equality in (33) is far less trivial, since the SCE

decoder is both universal and stochastic, and hence, its TRC exponent is expected to be inferior w.r.t. the TRC exponent under MAP decoding, but nevertheless, they turn out to be equal. Comparing to channel coding, it has been recently proved in [22] that the error exponent of the typical random fixed composition code (given in (13)) is the same for the ML and the maximum mutual information decoder, but on the other hand, numerical evidence shows that a GLD which is based on an empirical mutual information metric attains a strictly lower exponent.

5. Optimal Trade-off Functions

In this section, we study the optimal trade-off between the threshold Δ, the error exponent of the TRC, and the excess-rate exponent. Since both exponents depend on the rate function, we wish to characterize rate functions that are optimal w.r.t. this trade-off. Since a single-letter characterization of the error exponent of the TRC has already been given in (28), we next provide a single-letter expression for the excess-rate exponent. Define the following exponent function:

$$E_{\text{er}}(R(\cdot),\Delta) = \min_{\{Q_{UV}:\ R(Q_U)\geq H_Q(U|V)+\Delta\}} D(Q_{UV}\|P_{UV}). \tag{34}$$

Then, we have the following.

Proposition 1. *Fix $\Delta > 0$ and let $R(\cdot)$ be any rate function. Then,*

$$\mathsf{E}_{er}(\mathcal{B},R(\cdot),\Delta) = E_{er}(R(\cdot),\Delta). \tag{35}$$

Proof. The excess-rate probability is given by:

$$\begin{aligned}
&\mathbb{P}\{R(\hat{P}_{\boldsymbol{U}}) \geq \hat{H}_{\boldsymbol{UV}}(U|V)+\Delta\} \\
&= \sum_{Q_{UV}} \mathbb{1}\{R(Q_U) \geq H_Q(U|V)+\Delta\}\cdot\mathbb{P}\{(\boldsymbol{U},\boldsymbol{V})\in\mathcal{T}(Q_{UV})\} &(36)\\
&\doteq \sum_{\{Q_{UV}:\ R(Q_U)\geq H_Q(U|V)+\Delta\}} \exp\{-nD(Q_{UV}\|P_{UV})\} &(37)\\
&\doteq \exp\left\{-n\cdot\min_{\{Q_{UV}:\ R(Q_U)\geq H_Q(U|V)+\Delta\}} D(Q_{UV}\|P_{UV})\right\}, &(38)
\end{aligned}$$

which proves the desired result. □

Since Proposition 1 is proved by the method of types [21], we conclude that the excess-rate event is dominated by one specific type class $\mathcal{T}(Q_{UV})$, whose respective rate $R(Q_U)$ has been chosen too large w.r.t. the value of $H_Q(U|V)+\Delta$. One extreme case is when the rate function is given by $H_Q(U)$, which obviously provides a one-to-one mapping, since the size of each $\mathcal{T}(Q_U)$ is upper-bounded by $e^{nH_Q(U)}$. In this case, the probability of error is zero, while the excess-rate probability is one, at least when Δ is not too large. In Section 5.2, we prove that the optimal rate function is indeed upper-bounded by $H_Q(U)$, but can also be strictly smaller, especially when the requirement on the error exponent is not too stringent.

One way to explore the trade-off between the error exponent of the TRC and the excess-rate exponent, which will be presented in Section 5.1, is to require the excess-rate exponent to exceed some value $\mathsf{E}_r > 0$, then solve $E_{\text{er}}(R(\cdot),\Delta) \geq \mathsf{E}_r$ for an optimal rate function $R^*(Q_U)$, and then to substitute this optimal rate function back into the error exponents in (24) and (28) to give expressions for the optimal trade-off function $E_{\text{e}}(\mathsf{E}_r,\Delta)$. In Section 5.2, we present an alternative option to characterize this trade-off, which is to require the error exponent of the TRC to exceed some value $\mathsf{E}_e > 0$, to solve $E_{\text{e}}(R(\cdot)) \geq \mathsf{E}_e$

in order to extract an optimal rate function, and then to substitute it back into the excess-rate exponent in (34) to provide an expression for the optimal trade-off function $E_{\text{er}}(\mathsf{E}_e, \Delta)$.

5.1. Constrained Excess-Rate Exponent

Relying on the exponent function in (34), the following theorem proposes a rate function, whose optimality is proved in Appendix F.

Theorem 5. *Let $\mathsf{E}_r > 0$ be fixed. Then, the constraint $E_{\text{er}}(R(\cdot), \Delta) \geq \mathsf{E}_r$ implies that*

$$R(Q_U) \leq J(Q_U, \mathsf{E}_r, \Delta) \stackrel{\triangle}{=} \min_{\{Q_{V|U}\colon\, D(Q_{UV}\|P_{UV}) \leq \mathsf{E}_r\}} \big\{H_Q(U|V) + \Delta\big\}. \tag{39}$$

This means that we have a dichotomy between two kinds of source types. Each type class that is associated with an empirical distribution that is relatively close to the source distribution, i.e., when $D(Q_{UV}\|P_{UV}) \leq \mathsf{E}_r$ for some $Q_{V|U}$, is partitioned into $e^{nJ(Q_U, \mathsf{E}_r, \Delta)}$ bins, and the rest of the type classes, those that are relatively distant from P_U, are encoded by a one-to-one mapping. Two extreme cases should be considered here. First, when E_r is relatively small, then only the types closest to P_U are encoded with a rate approximately $H_P(U|V) + \Delta$, which can be made arbitrarily close to the SW limit [8], and each a–typical source sequence is allocated with $n \cdot \log_2 |\mathcal{U}|$ bits. This coding scheme is the one related to VR coding with an average rate constraint, like the one discussed in [12]. Second, when E_r is extremely large, then each type class is encoded to $\exp\{n\Delta\}$ bins, which is equivalent to FR coding.

Following the first part of Theorem 3, let us denote the error exponent of the TRC under MAP decoding by $E_{\text{e}}(\cdot)$. Upon substituting the optimal rate function of Theorem 5 back into (24) and (28) and using the fact that $E_{\text{e}}(\cdot)$ is monotonically increasing, we find that the optimal trade-off function for the typical random SD code is given by

$$E_{\text{e}}(\mathsf{E}_r, \Delta) = \min_{\{Q_{UV}\colon\, H_Q(U) \geq J(Q_U)\}} \big\{D(Q_{UV}\|P_{UV}) + [J(Q_U) - H_Q(U|V)]_+\big\}, \tag{40}$$

or, alternatively,

$$E_{\text{e}}(\mathsf{E}_r, \Delta) = \min_{\substack{\{Q_{UU'}\colon\, Q_{U'} = Q_U,\\ H_Q(U) \geq J(Q_U)\}}} \big\{\Lambda(Q_{UU'}, J(Q_U)) - \mathbb{E}_Q[\log P(U)] - H_Q(U, U') + J(Q_U)\big\}, \tag{41}$$

where $J(Q_U) = J(Q_U, \mathsf{E}_r, \Delta)$ is given in (39). The dependence of $E_{\text{e}}(\mathsf{E}_r, \Delta)$ on E_r is as follows. Let $Q^*_{UU'}(\Delta)$ and $Q^*_{V|U}$ be the respective minimizers of the problems which are similar to (39) and (41), except that the constraint $D(Q_{UV}\|P_{UV}) \leq \mathsf{E}_r$ is removed from (39). Furthermore, let $Q^*_U(\Delta)$ be the marginal distribution of $Q^*_{UU'}(\Delta)$. Now, when E_r is sufficiently large, i.e., when $\mathsf{E}_r \geq D(Q^*_U(\Delta) \times Q^*_{V|U}\|P_{UV})$, $E_{\text{e}}(\mathsf{E}_r, \Delta)$ reaches a plateau and is the lowest possible. It follows from the fact that the stringent requirement on the excess-rate forces the encoder to encode each type class Q_U to its target rate Δ, thus all of them affect the error event. Otherwise, when $\mathsf{E}_r < D(Q^*_U(\Delta) \times Q^*_{V|U}\|P_{UV})$, the constraint $D(Q_{UV}\|P_{UV}) \leq \mathsf{E}_r$ is active and $E_{\text{e}}(\mathsf{E}_r, \Delta)$ is a monotonically nonincreasing function of E_r. The reason for that is the fact that as E_r decreases, more and more type classes are encoded with $n \cdot \log_2 |\mathcal{U}|$ bits, and hence do not contribute to the error event. When $\mathsf{E}_r = 0$, necessarily $Q_U = P_U$, only the typical set is encoded, and $E_{\text{e}}(0, \Delta)$ is the highest possible. In this case, $J(Q_U) = H_P(U|V) + \Delta$ and the constraint set in (41) becomes empty when $\Delta > I_P(U;V)$, and then $E_{\text{e}}(0, \Delta) = \infty$.

5.2. Constrained Error Exponent

Based on (24), the following theorem proposes a rate function, whose optimality is proved in Appendix G.

Theorem 6. *Let* $\mathsf{E}_e > 0$ *be fixed. Then, the constraint* $E_e(R(\cdot)) \geq \mathsf{E}_e$ *implies that*

$$R(Q_U) \geq \Omega(Q_U, \mathsf{E}_e) \stackrel{\triangle}{=} \min\{H_Q(U), G(Q_U, \mathsf{E}_e)\}, \tag{42}$$

where,

$$G(Q_U, \mathsf{E}_e) = \max_{\{Q_{V|U}:\, D(Q_{UV}\|P_{UV}) \leq \mathsf{E}_e\}} \{H_Q(U|V) + \mathsf{E}_e - D(Q_{UV}\|P_{UV})\}. \tag{43}$$

The dependence of $G(Q_U, \mathsf{E}_e)$ on E_e is as follows. For any given Q_U, let $\tilde{Q}_{V|U}$ be the minimizer of $D(Q_{UV}\|P_{UV})$. Then, as long as $\mathsf{E}_e < D(Q_U \times \tilde{Q}_{V|U}\|P_{UV})$, the constraint set in (43) is empty, and $R(Q_U)$ can vanish, which practically means that in this range, the entire type class $\mathcal{T}(Q_U)$ can be totally ignored, while still achieving $P_e \approx e^{-n\mathsf{E}_e}$. Only for the unique type $Q_U = P_U$, $G(P_U, \mathsf{E}_e) > 0$ for all $\mathsf{E}_e \geq 0$, and specifically, we find that $G(P_U, 0) = H_P(U|V)$. Furthermore, let $Q^*_{V|U}$ be the maximizer in the unconstrained problem

$$\max_{Q_{V|U}} \{H_Q(U|V) - D(Q_{UV}\|P_{UV})\}. \tag{44}$$

Then, as long as $\mathsf{E}_e \in [D(Q_U \times \tilde{Q}_{V|U}\|P_{UV}), D(Q_U \times Q^*_{V|U}\|P_{UV}))$, $G(Q_U, \mathsf{E}_e)$ is a monotonically nondecreasing function of E_e. When $\mathsf{E}_e \geq D(Q_U \times Q^*_{V|U}\|P_{UV})$, the maximization in (43) reaches its unconstrained optimum, and $G(Q_U, \mathsf{E}_e)$ increases without bound in an affine fashion as $\mathsf{E}_e + H_{Q^*}(U|V) - D(Q_U \times Q^*_{V|U}\|P_{UV})$. As can be seen in (42), $\Omega(Q_U, \mathsf{E}_e)$ finally reaches a plateau at the level of $H_Q(U)$.

Upon substituting $\Omega(Q_U, \mathsf{E}_e)$ back into (34) and using the fact that $E_{\text{er}}(\cdot, \Delta)$ is monotonically nonincreasing, we find that the trade-off function is given by

$$E_{\text{er}}(\mathsf{E}_e, \Delta) = \min_{\{Q_{UV}:\, \Omega(Q_U, \mathsf{E}_e) \geq H_Q(U|V) + \Delta\}} D(Q_{UV}\|P_{UV}). \tag{45}$$

Since $\Omega(Q_U, \mathsf{E}_e)$ is monotonically nondecreasing in E_e for every Q_U, $E_{\text{er}}(\mathsf{E}_e, \Delta)$ is monotonically nonincreasing in E_e, which is not very surprising. The dependence of $E_{\text{er}}(\mathsf{E}_e, \Delta)$ on E_e and Δ is as follows. At $\mathsf{E}_e = 0$, notice that $\Omega(Q_U, 0) = -\infty$ for any $Q_U \neq P_U$ while $\Omega(P_U, 0) = H_P(U|V)$. Thus, $E_{\text{er}}(0, \Delta) = 0$ as long as $\Delta = 0$, and it follows from the monotonicity that $E_{\text{er}}(\mathsf{E}_e, 0) = 0$ everywhere. Otherwise, if $\Delta > 0$, $\{Q_{UV} : \Omega(Q_U, \mathsf{E}_e) \geq H_Q(U|V) + \Delta\}$ is empty as long as $\mathsf{E}_e < \mathsf{E}^*_e(\Delta)$, where an expression for $\mathsf{E}^*_e(\Delta)$ can be found by solving

$$\max_{Q_{UV}} \{\Omega(Q_U, \mathsf{E}_e) - H_Q(U|V)\} \leq \Delta, \tag{46}$$

and then $E_{\text{er}}(\mathsf{E}_e, \Delta) = \infty$ in this range. In the other extreme case of a very large E_e, $\Omega(Q_U, \mathsf{E}_e)$ reaches a plateau at a level of $H_Q(U)$. Then, if $\Delta \leq H_P(U) - H_P(U|V) = I_P(U;V)$, $E_{\text{er}}(\mathsf{E}_e, \Delta)$ reaches zero for a sufficiently large E_e. Else, if $\Delta > I_P(U;V)$, $E_{\text{er}}(\mathsf{E}_e, \Delta)$ reaches a strictly positive plateau, given by

$$\min_{\{Q_{UV}:\ I_Q(U;V)\geq\Delta\}} D(Q_{UV}\|P_{UV}), \tag{47}$$

which is a monotonically nondecreasing function of Δ. Particularly, it means that in this range, the typical random SD code attains both an exponentially vanishing excess-rate probability and $P_{\mathrm{e}} \approx 0$.

It is interesting to relate this to the expurgated bound of the FR code in the SW model, which is given by (15). Comparing $E_{\mathrm{ex}}^{\mathrm{fr}}(R)$ and $E_{\mathrm{e}}(\infty,\Delta)$ analytically is rather difficult. Thus, we examined these two exponent functions numerically. Consider the case of a double binary source with alphabets $\mathcal{U} = \mathcal{V} = \{0,1\}$, and joint probabilities given by $P_{UV}(0,0) = 0.75$, $P_{UV}(0,1) = 0.1$, $P_{UV}(1,0) = 0$, and $P_{UV}(1,1) = 0.15$. We already mentioned before, that in the special case of $\mathsf{E}_r = \infty$, the rate function is given by the threshold Δ, hence we choose $\Delta = R$ in order to have a fair comparison. Graphs of the functions $E_{\mathrm{ex}}^{\mathrm{fr}}(R)$ and $E_{\mathrm{e}}(\infty,R)$ are presented in Figure 1.

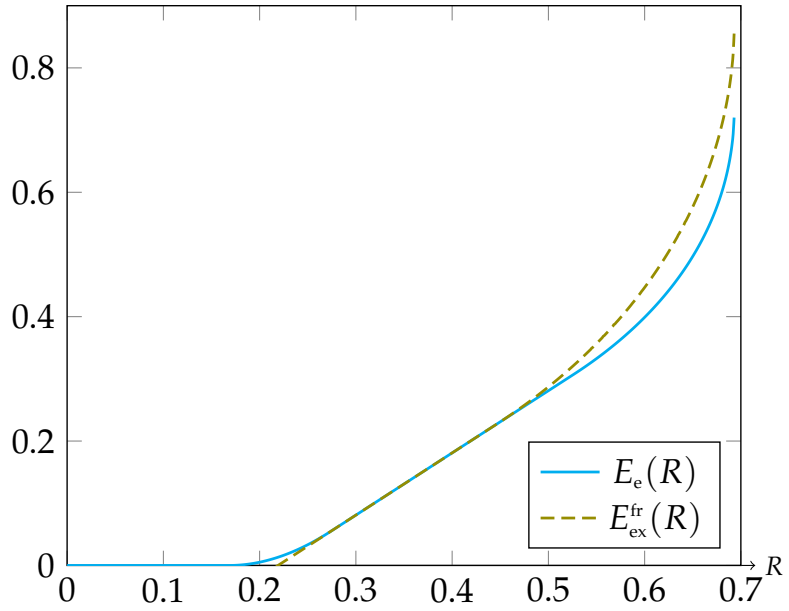

Figure 1. Graphs of the functions $E_{\mathrm{ex}}^{\mathrm{fr}}(R)$ and $E_{\mathrm{e}}(\infty,R)$.

As can be seen in Figure 1, both $E_{\mathrm{ex}}^{\mathrm{fr}}(R)$ and $E_{\mathrm{e}}(\infty,R)$ tend to infinity as R tends to $\log 2 \approx 0.693$. For relatively high binning rates, $E_{\mathrm{ex}}^{\mathrm{fr}}(R)$ is strictly higher than $E_{\mathrm{e}}(\infty,R)$, which can be explained in the following way: Referring to the analogy between SW coding and channel coding, one can think of each bin as containing a channel code. In general, a channel code behaves well if it does not contain pairs of relatively "close" codewords. Since we randomly assign the source vectors into the bins (even if the populations of the bins are totally equal, which can be attained by randomly partitioning each type class into $\exp\{nR\}$ subsets), it is reasonable to assume that some bins will contain relatively bad codebooks. On the other hand, in the expurgated SW code [11], each type class $\mathcal{T}(Q_U)$ is partitioned into $\exp\{nR\}$ "balanced" subsets in some sense (referring to the enumerators $N(Q_{UU'})$ in (30), they are equally populated in all of the bins), such that the codebooks contained in the bins have approximately equal error probabilities. Moreover, we conclude from (15) that each bin contains a codebook with a quality of an expurgated channel code. This code is certainly better than the TRCs in the SD ensemble.

In channel coding, it is known [23] that the random Gilbert–Varshamov ensemble has an exact random coding error exponent which is as high as the maximum between (16) and (17). In SW source coding, on the other hand, it seems to be a more challenging problem to define an ensemble, such that the error exponent of its TRCs is as high as $E_{\mathrm{ex}}^{\mathrm{fr}}(R)$

of (15). Since the gap between $E_{\mathrm{ex}}^{\mathrm{fr}}(R)$ and $E_{\mathrm{e}}(\infty, R)$ is not necessarily very significant, as can be seen in Figure 1, we conclude that the SD ensemble may be more attractive because the amount of computations needed for drawing a code from it are much lower than the amount of computations required for having an expurgated SW code. In addition, it is important to note that the probability of drawing a SD code with an exponent much lower than $E_{\mathrm{e}}(\infty, R)$ decays exponentially fast, in analogy to the result in pure channel coding [7].

Author Contributions: Both authors contributed equally to this research. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Israel Science Foundation (ISF) grant number 137/18.

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

Appendix A. Proof of Theorem 1

By definition, we have

$$\mathbb{E}[P_{\mathrm{e}}(\mathcal{B}_n)] = \mathbb{E}\left[\frac{\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{U}),\boldsymbol{u}'\neq\boldsymbol{U}} \exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{V}})\}}{\sum_{\tilde{\boldsymbol{u}}\in\mathcal{B}(\boldsymbol{U})} \exp\{nf(\hat{P}_{\tilde{\boldsymbol{u}}\boldsymbol{V}})\}}\right]. \tag{A1}$$

Step 1: Averaging Over the Random Code

We first condition on the true source sequences $(\boldsymbol{U}=\boldsymbol{u}, \boldsymbol{V}=\boldsymbol{v})$ and take the expectation only w.r.t. the random binning. We get

$$\begin{aligned}
&\mathbb{E}[P_{\mathrm{e}}(\mathcal{B}_n)|\boldsymbol{u},\boldsymbol{v}] \\
&= \mathbb{E}\left[\frac{\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}} \exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}{\exp\{n\cdot f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})\} + \sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}} \exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}\right] && \text{(A2)} \\
&= \int_0^1 \mathbb{P}\left\{\frac{\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}} \exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}{\exp\{n\cdot f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})\} + \sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}} \exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}} \geq s\right\} ds && \text{(A3)} \\
&= \int_0^\infty ne^{-n\xi}\cdot\mathbb{P}\left\{\frac{\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}} \exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}{\exp\{n\cdot f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})\} + \sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}} \exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}} \geq e^{-n\xi}\right\} d\xi && \text{(A4)} \\
&= \int_0^\infty ne^{-n\xi}\cdot\mathbb{P}\left\{(1-e^{-n\xi}) \sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}} \exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\} \geq e^{-n\xi}\exp\{nf(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})\}\right\} d\xi && \text{(A5)} \\
&\doteq \int_0^\infty ne^{-n\xi}\cdot\mathbb{P}\left\{\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}} \exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\} \geq \exp\{n[f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})-\xi]\}\right\} d\xi, && \text{(A6)}
\end{aligned}$$

where (A4) follows by changing the integration variable in (A3) according to $s = e^{-n\xi}$. Define

$$N_{\boldsymbol{u},\boldsymbol{v}}(Q_{U|V}) = \sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}} \mathbb{1}\{\boldsymbol{u}'\in\mathcal{T}(Q_{U|V}|\boldsymbol{v})\}, \tag{A7}$$

such that the probability in (A6) is given by

$$\mathbb{P}\left\{\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}}\exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}\geq\exp\{n[f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})-\xi]\}\right\}$$

$$=\mathbb{P}\left\{\sum_{Q_{U'|V}}N_{\boldsymbol{u},\boldsymbol{v}}(Q_{U'|V})\exp\{nf(Q_{U'V})\}\geq\exp\{n[f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})-\xi]\}\right\} \tag{A8}$$

$$\doteq\mathbb{P}\left\{\max_{Q_{U'|V}}N_{\boldsymbol{u},\boldsymbol{v}}(Q_{U'|V})\exp\{nf(Q_{U'V})\}\geq\exp\{n[f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})-\xi]\}\right\} \tag{A9}$$

$$=\mathbb{P}\bigcup_{Q_{U'|V}}\left\{N_{\boldsymbol{u},\boldsymbol{v}}(Q_{U'|V})\exp\{nf(Q_{U'V})\}\geq\exp\{n[f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})-\xi]\}\right\} \tag{A10}$$

$$\doteq\sum_{Q_{U'|V}}\mathbb{P}\left\{N_{\boldsymbol{u},\boldsymbol{v}}(Q_{U'|V})\geq\exp\{n[f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})-f(Q_{U'V})-\xi]\}\right\}, \tag{A11}$$

where $Q_{U'V}=Q_{U'|V}\times\hat{P}_{\boldsymbol{v}}$. Let us denote $B_0=f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})-f(Q_{U'V})$. Now, given $\boldsymbol{u}$ and $\boldsymbol{v}$, $N_{\boldsymbol{u},\boldsymbol{v}}(Q_{U'|V})$ is a binomial sum of $|\mathcal{T}(Q_{U'|V}|\boldsymbol{v})|\doteq e^{nH_Q(U'|V)}$ trials and success rate of the exponential order of $e^{-nR(Q_U)}$. Therefore, using the techniques of [24] (Section 6.3),

$$-\frac{1}{n}\log\mathbb{P}\left\{N_{\boldsymbol{u},\boldsymbol{v}}(Q_{U'|V})\geq\exp\{n[B_0-\xi]\}\right\}$$

$$=\begin{cases}\big[R(Q_U)-H_Q(U'|V)\big]_+ & \big[H_Q(U'|V)-R(Q_U)\big]_+\geq B_0-\xi\\ \infty & \big[H_Q(U'|V)-R(Q_U)\big]_+<B_0-\xi\end{cases} \tag{A12}$$

$$=\begin{cases}\big[R(Q_U)-H_Q(U'|V)\big]_+ & \xi\geq B_0-\big[H_Q(U'|V)-R(Q_U)\big]_+\\ \infty & \xi<B_0-\big[H_Q(U'|V)-R(Q_U)\big]_+\end{cases}, \tag{A13}$$

and so,

$$\int_0^\infty e^{-n\xi}\cdot\mathbb{P}\left\{N_{\boldsymbol{u},\boldsymbol{v}}(Q_{U|V})\geq\exp\{n[B_0-\xi]\}\right\}d\xi$$

$$\doteq\int_{\left[B_0-[H_Q(U'|V)-R(Q_U)]_+\right]_+}^\infty e^{-n\xi}\cdot e^{-n[R(Q_U)-H_Q(U'|V)]_+}d\xi \tag{A14}$$

$$\doteq\exp\left\{-n\Big([R(Q_U)-H_Q(U'|V)]_+ + \big[B_0-[H_Q(U'|V)-R(Q_U)]_+\big]_+\Big)\right\} \tag{A15}$$

$$=\exp\left\{-n\begin{pmatrix}R(Q_U)-H_Q(U'|V)+[B_0]_+ & R(Q_U)\geq H_Q(U'|V)\\ \big[R(Q_U)-H_Q(U'|V)+B_0\big]_+ & R(Q_U)<H_Q(U'|V)\end{pmatrix}\right\} \tag{A16}$$

$$=\exp\left\{-n\begin{pmatrix}\big[R(Q_U)-H_Q(U'|V)+[B_0]_+\big]_+ & R(Q_U)\geq H_Q(U'|V)\\ \big[R(Q_U)-H_Q(U'|V)+[B_0]_+\big]_+ & R(Q_U)<H_Q(U'|V)\end{pmatrix}\right\} \tag{A17}$$

$$=\exp\Big\{-n\cdot\big[R(Q_U)-H_Q(U'|V)+[B_0]_+\big]_+\Big\}. \tag{A18}$$

Finally, we have that

$$\sum_{Q_{U'|V}} \int_0^\infty e^{-n\xi} \cdot \mathbb{P}\Big\{N_{\boldsymbol{u},\boldsymbol{v}}(Q_{U|V}) \geq \exp\{n[B_0 - \xi]\}\Big\} d\xi \tag{A19}$$

$$\doteq \max_{Q_{U'|V}} \exp\Big\{-n \cdot \big[R(Q_U) - H_Q(U'|V) + [B_0]_+\big]_+\Big\} \tag{A20}$$

$$= \exp\left\{-n \cdot \min_{Q_{U'|V}} \big[R(Q_U) - H_Q(U'|V) + [B_0]_+\big]_+\right\}, \tag{A21}$$

thus,

$$E(\boldsymbol{u},\boldsymbol{v}) = \min_{Q_{U'|V}} \big[R(Q_U) - H_Q(U'|V) + [f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}}) - f(Q_{U'V})]_+\big]_+. \tag{A22}$$

Step 2: Averaging Over $\boldsymbol{U}$ and $\boldsymbol{V}$

Notice that the exponent function $E(\boldsymbol{u},\boldsymbol{v})$ depends on $(\boldsymbol{u},\boldsymbol{v})$ only via the empirical distribution $\hat{P}_{\boldsymbol{u}\boldsymbol{v}}$. Averaging over the source and the side information sequences, now yields

$$\mathbb{E}\{P_e(\mathcal{B}_n)\} = \sum_{\boldsymbol{u},\boldsymbol{v}} P(\boldsymbol{u},\boldsymbol{v}) \cdot \mathbb{1}\big\{\hat{H}_{\boldsymbol{u}}(U) \geq R(\hat{P}_{\boldsymbol{u}})\big\} \cdot \exp\big\{-n \cdot E(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})\big\} \tag{A23}$$

$$\doteq \sum_{\{Q_{UV}:\ H_Q(U) \geq R(Q_U)\}} e^{-n \cdot D(Q_{UV}\|P_{UV})} \cdot \exp\big\{-n \cdot E(Q_{UV})\big\} \tag{A24}$$

$$\doteq \exp\left\{-n \cdot \min_{\{Q_{UV}:\ H_Q(U) \geq R(Q_U)\}} [D(Q_{UV}\|P_{UV}) + E(Q_{UV})]\right\}, \tag{A25}$$

which proves the first point of Theorem 1.

Step 3: Moving from Stochastic to Deterministic Decoding

In order to transform the GLD into the general deterministic decoder of

$$\hat{\boldsymbol{u}} = \operatorname*{arg\,max}_{\boldsymbol{u}' \in \mathcal{B}(\boldsymbol{u}) \cap \mathcal{T}(\boldsymbol{u})} f(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}}), \tag{A26}$$

we just have to multiply $f(\cdot)$, in

$$E(Q_{UV}) = \min_{Q_{U'|V}} \big[R(Q_U) - H_Q(U'|V) + [f(Q_{UV}) - f(Q_{U'V})]_+\big]_+, \tag{A27}$$

by $\beta \geq 0$, and then let $\beta \to \infty$. We find that the overall error exponent of the SD ensemble with the general deterministic decoder of (A26) is given by

$$E(P) = \min_{\{Q_{UV}:\ H_Q(U) \geq R(Q_U)\}} \big[D(Q_{UV}\|P_{UV}) + \tilde{E}(Q_{UV})\big], \tag{A28}$$

where,

$$\tilde{E}(Q_{UV}) = \min_{\{Q_{U'|V}:\ f(Q_{U'V}) \geq f(Q_{UV})\}} \big[R(Q_U) - H_Q(U'|V)\big]_+. \tag{A29}$$

Step 4: A Fundamental Limitation on the Error Exponent

Note that the minimum in (A29) can be upper–bounded by choosing a specific distribution in the feasible set. In (A29), we take $Q_{U'|V} = Q_{U|V}$ and then

$$\tilde{E}(Q_{UV}) \leq \big[R(Q_U) - H_Q(U|V)\big]_+. \tag{A30}$$

Hence, the overall error exponent is upper–bounded as

$$E(P) \leq \min_{\{Q_{UV}\colon\, H_Q(U) \geq R(Q_U)\}} \Big[D(Q_{UV}\|P_{UV}) + \big[R(Q_U) - H_Q(U|V)\big]_+\Big]. \tag{A31}$$

Step 5: An Optimal Universal Decoder

We prove that the upper bound of (A31) is attainable by choosing the universal decoding metric $f(Q_{UV}) = -H_Q(U|V)$. Now, we get for (A29)

$$\begin{aligned}
\tilde{E}(Q_{UV}) &= \min_{\{Q_{U'|V}\colon\, f(Q_{U'V}) \geq f(Q_{UV})\}} \big[R(Q_U) - H_Q(U'|V)\big]_+ && \text{(A32)}\\
&= \min_{\{Q_{U'|V}\colon\, H_Q(U|V) \geq H_Q(U'|V)\}} \big[R(Q_U) - H_Q(U'|V)\big]_+ && \text{(A33)}\\
&= \big[R(Q_U) - H_Q(U|V)\big]_+, && \text{(A34)}
\end{aligned}$$

which completes the proof of Theorem 1.

Appendix B. Proof of Theorem 2

Appendix B.1. Lower Bound on the Error Exponent

Our starting point is the following inequality, for any $\rho > 0$,

$$\mathbb{E}[\log P_{\mathrm{e}}(\mathcal{B}_n)] \leq \log\Big(\mathbb{E}[P_{\mathrm{e}}(\mathcal{B}_n)]^{1/\rho}\Big)^{\rho}, \tag{A35}$$

which is due to the following considerations. First, for a positive random variable X, the function

$$f(\rho) = \log\Big(\mathbb{E}\Big[X^{1/\rho}\Big]\Big)^{\rho} \tag{A36}$$

is monotonically decreasing, and second, by L'Hospital's rule,

$$\lim_{\rho\to\infty} \log\Big(\mathbb{E}\Big[X^{1/\rho}\Big]\Big)^{\rho} = \mathbb{E}[\log X]. \tag{A37}$$

Recall that the error probability is given by

$$P_{\mathrm{e}}(\mathcal{B}_n) = \sum_{\boldsymbol{u},\boldsymbol{v}} P(\boldsymbol{u},\boldsymbol{v}) \cdot \mathbb{1}\big\{\hat{H}_{\boldsymbol{u}}(U) \geq R(\hat{P}_{\boldsymbol{u}})\big\} \cdot \frac{\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u})\cap\mathcal{T}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}} \exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}{\sum_{\tilde{\boldsymbol{u}}\in\mathcal{B}(\boldsymbol{u})\cap\mathcal{T}(\boldsymbol{u})} \exp\{nf(\hat{P}_{\tilde{\boldsymbol{u}}\boldsymbol{v}})\}}. \tag{A38}$$

Let

$$Z_{\boldsymbol{u}}(\boldsymbol{v}) = \sum_{\tilde{\boldsymbol{u}}\in\mathcal{B}(\boldsymbol{u})\cap\mathcal{T}(\boldsymbol{u}),\tilde{\boldsymbol{u}}\neq\boldsymbol{u}} \exp\{nf(\hat{P}_{\tilde{\boldsymbol{u}}\boldsymbol{v}})\}, \tag{A39}$$

fix $\epsilon > 0$ arbitrarily small, and for every $\boldsymbol{u} \in \mathcal{U}^n$ and $\boldsymbol{v} \in \mathcal{V}^n$, define the set

$$B_\epsilon(\boldsymbol{u}, \boldsymbol{v}) = \{\mathcal{B}_n : \ Z_{\boldsymbol{u}}(\boldsymbol{v}) \le \exp\{n\alpha(R+\epsilon, \hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}})\}\}. \tag{A40}$$

Following the result of [20] (Appendix B), we prove the following modification in Appendix C.

Lemma A1. *Let $\epsilon > 0$ be arbitrarily small. Then, for every $\boldsymbol{u} \in \mathcal{U}^n$ and $\boldsymbol{v} \in \mathcal{V}^n$,*

$$\mathbb{P}\{Z_{\boldsymbol{u}}(\boldsymbol{v}) \le \exp\{n\alpha(R+\epsilon, \hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}})\}\} \le \exp\{-e^{n\epsilon} + n\epsilon + 1\}. \tag{A41}$$

Thus, by the union bound,

$$\mathbb{P}\left\{\bigcup_{\boldsymbol{u}\in\mathcal{U}^n}\bigcup_{\boldsymbol{v}\in\mathcal{V}^n} B_\epsilon(\boldsymbol{u}, \boldsymbol{v})\right\} \stackrel{\triangle}{=} \mathbb{P}\{B_\epsilon\} \le \sum_{\boldsymbol{u}\in\mathcal{U}^n}\sum_{\boldsymbol{v}\in\mathcal{V}^n} \mathbb{P}\{B_\epsilon(\boldsymbol{u}, \boldsymbol{v})\} \tag{A42}$$

$$\le \sum_{\boldsymbol{u}\in\mathcal{U}^n}\sum_{\boldsymbol{v}\in\mathcal{V}^n} \exp\{-e^{n\epsilon} + n\epsilon + 1\} \tag{A43}$$

$$= |\mathcal{U}\times\mathcal{V}|^n \cdot \exp\{-e^{n\epsilon} + n\epsilon + 1\}, \tag{A44}$$

which still decays double–exponentially fast. Recall that $\mathcal{Q} = \{Q_{UU'} : \ Q_U = Q_{U'}\}$. Then, for any $\rho \ge 1$

$$\mathbb{E}\left\{[P_{\mathrm{e}}(\mathcal{B}_n)]^{1/\rho}\right\}$$

$$= \mathbb{E}\left\{P_{\mathrm{e}}(\mathcal{B}_n)^{1/\rho} \cdot \mathbb{1}\{B_\epsilon^{\mathrm{c}}\}\right\} + \mathbb{E}\left\{P_{\mathrm{e}}(\mathcal{B}_n)^{1/\rho} \cdot \mathbb{1}\{B_\epsilon\}\right\} \tag{A45}$$

$$\le \mathbb{E}\left\{\left[\sum_{\boldsymbol{u},\boldsymbol{v}} P(\boldsymbol{u},\boldsymbol{v})\mathbb{1}\{\hat{H}_{\boldsymbol{u}}(U) \ge R(\hat{P}_{\boldsymbol{u}})\}\frac{\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u})\cap\mathcal{T}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}}\exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}{\exp\{nf(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})\} + Z_{\boldsymbol{u}}(\boldsymbol{v})}\right]^{1/\rho} \mathbb{1}\{B_\epsilon^{\mathrm{c}}\}\right\}$$
$$+ \mathbb{P}\{B_\epsilon\} \tag{A46}$$

$$\le \mathbb{E}\left\{\left[\sum_{\boldsymbol{u},\boldsymbol{v}}\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u})\cap\mathcal{T}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}} P(\boldsymbol{u},\boldsymbol{v})\mathbb{1}\{\hat{H}_{\boldsymbol{u}}(U) \ge R(\hat{P}_{\boldsymbol{u}})\}\right.\right.$$
$$\left.\left.\times \min\left\{1, \frac{\exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}{\exp\{nf(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})\} + \exp\{n\alpha(R+\epsilon, \hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}})\}}\right\}\right]^{1/\rho}\right\}$$
$$+ |\mathcal{U}\times\mathcal{V}|^n \cdot \exp\{-e^{n\epsilon} + n\epsilon + 1\} \tag{A47}$$

$$\doteq \mathbb{E}\left\{\left[\sum_{\boldsymbol{u}}\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u})\cap\mathcal{T}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}} P(\boldsymbol{u})\mathbb{1}\{\hat{H}_{\boldsymbol{u}}(U) \ge R(\hat{P}_{\boldsymbol{u}})\}\right.\right.$$
$$\left.\left.\times \sum_{\boldsymbol{v}} P(\boldsymbol{v}|\boldsymbol{u}) \exp\left\{-n \cdot \left[\max\{f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}}), \alpha(R+\epsilon, \hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}})\} - f(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\right]_+\right\}\right]^{1/\rho}\right\} \tag{A48}$$

$$\doteq \mathbb{E}\left\{\left[\sum_{\boldsymbol{u}}\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u})\cap\mathcal{T}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}} P(\boldsymbol{u}) \cdot \mathbb{1}\{\hat{H}_{\boldsymbol{u}}(U) \ge R(\hat{P}_{\boldsymbol{u}})\} \cdot \exp\{-n \cdot \Lambda(\hat{P}_{\boldsymbol{u}\boldsymbol{u}'}, R+\epsilon)\}\right]^{1/\rho}\right\} \tag{A49}$$

$$= \mathbb{E}\left\{\left[\sum_{\{Q_{UU'}\in\mathcal{Q}:\ H_Q(U)\ge R(Q_U)\}} N(Q_{UU'}) \cdot e^{n\mathbb{E}_Q[\log P(U)]} \cdot \exp\{-n \cdot \Lambda(Q_{UU'}, R+\epsilon)\}\right]^{1/\rho}\right\} \tag{A50}$$

$$\doteq \sum_{\{Q_{UU'}\in\mathcal{Q}:\ H_Q(U)\ge R(Q_U)\}} \mathbb{E}\left\{[N(Q_{UU'})]^{1/\rho}\right\} \cdot e^{n(\mathbb{E}_Q[\log P(U)])/\rho} \cdot \exp\{-n \cdot \Lambda(Q_{UU'}, R+\epsilon)/\rho\}, \tag{A51}$$

where (A47) is due to Lemma A1, (A49) is by the method of types and the definition of $\Lambda(Q_{UU'}, R)$ in (27), and in (A50) we used the definition of $N(Q_{UU'})$ in (30). Therefore, our next task is to evaluate the $1/\rho$–th moment of $N(Q_{UU'})$. Let us define

$$N_{\boldsymbol{u}}(Q_{U'|U}) = \sum_{\boldsymbol{u}' \in \mathcal{T}(Q_{U'|U}|\boldsymbol{u})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}') = \mathcal{B}(\boldsymbol{u})\}. \tag{A52}$$

For a given $\rho \geq 1$, let $s \in [1, \rho]$. Then,

$$\mathbb{E}\Big\{[N(Q_{UU'})]^{1/\rho}\Big\} = \mathbb{E}\left\{\left[\sum_{\boldsymbol{u} \in \mathcal{T}(Q_U)} N_{\boldsymbol{u}}(Q_{U'|U})\right]^{1/\rho}\right\} \tag{A53}$$

$$= \mathbb{E}\left\{\left(\left[\sum_{\boldsymbol{u} \in \mathcal{T}(Q_U)} N_{\boldsymbol{u}}(Q_{U'|U})\right]^{1/s}\right)^{s/\rho}\right\} \tag{A54}$$

$$\leq \mathbb{E}\left\{\left(\sum_{\boldsymbol{u} \in \mathcal{T}(Q_U)} \Big[N_{\boldsymbol{u}}(Q_{U'|U})\Big]^{1/s}\right)^{s/\rho}\right\} \tag{A55}$$

$$\leq \left(\mathbb{E}\left\{\sum_{\boldsymbol{u} \in \mathcal{T}(Q_U)} \Big[N_{\boldsymbol{u}}(Q_{U'|U})\Big]^{1/s}\right\}\right)^{s/\rho} \tag{A56}$$

$$= \left(\sum_{\boldsymbol{u} \in \mathcal{T}(Q_U)} \mathbb{E}\Big\{\Big[N_{\boldsymbol{u}}(Q_{U'|U})\Big]^{1/s}\Big\}\right)^{s/\rho}, \tag{A57}$$

where (A56) follows from Jensen's inequality. Now, $N_{\boldsymbol{u}}(Q_{U'|U})$ is a binomial random variable with $|\mathcal{T}(Q_{U'|U}|\boldsymbol{u})| \doteq e^{nH_Q(U'|U)}$ trials and success rate which is of the exponential order of e^{-nR}. We have that [24] (Section 6.3)

$$\mathbb{E}\Big\{\Big[N_{\boldsymbol{u}}(Q_{U'|U})\Big]^{1/s}\Big\} \doteq \begin{cases} \exp\{n[H_Q(U'|U) - R]/s\} & H_Q(U'|U) \geq R \\ \exp\{n[H_Q(U'|U) - R]\} & H_Q(U'|U) < R \end{cases}, \tag{A58}$$

and so,

$$\mathbb{E}\Big\{[N(Q_{UU'})]^{1/\rho}\Big\} \leq e^{nH_Q(U)\cdot s/\rho} \cdot \left(\mathbb{E}\Big\{\Big[N_{\boldsymbol{u}}(Q_{U'|U})\Big]^{1/s}\Big\}\right)^{s/\rho} \tag{A59}$$

$$\doteq e^{nH_Q(U)\cdot s/\rho} \cdot \begin{cases} \exp\{n[H_Q(U'|U) - R]/\rho\} & H_Q(U'|U) \geq R \\ \exp\{n[H_Q(U'|U) - R]s/\rho\} & H_Q(U'|U) < R \end{cases} \tag{A60}$$

$$= \begin{cases} \exp\{n[H_Q(U)\cdot s + H_Q(U'|U) - R]/\rho\} & H_Q(U'|U) \geq R \\ \exp\{n[H_Q(U) + H_Q(U'|U) - R]s/\rho\} & H_Q(U'|U) < R \end{cases} \tag{A61}$$

$$= \begin{cases} \exp\{n[H_Q(U)\cdot s + H_Q(U'|U) - R]/\rho\} & H_Q(U'|U) \geq R \\ \exp\{n[H_Q(U,U') - R]s/\rho\} & H_Q(U'|U) < R \end{cases}. \tag{A62}$$

After optimizing over s, we get

$$\frac{1}{n}\log\mathbb{E}\Big\{[N(Q_{UU'})]^{1/\rho}\Big\}$$
$$\leq \min_{1\leq s\leq\rho}\begin{cases}\big[H_Q(U)\cdot s+H_Q(U'|U)-R\big]/\rho & H_Q(U'|U)\geq R\\ \big[H_Q(U,U')-R\big]s/\rho & H_Q(U'|U)<R, H_Q(U,U')\geq R\\ \big[H_Q(U,U')-R\big]s/\rho & H_Q(U'|U)<R, H_Q(U,U')<R\end{cases} \tag{A63}$$
$$=\begin{cases}\big[H_Q(U)+H_Q(U'|U)-R\big]/\rho & H_Q(U'|U)\geq R\\ \big[H_Q(U,U')-R\big]/\rho & H_Q(U'|U)<R, H_Q(U,U')\geq R\\ \big[H_Q(U,U')-R\big]\rho/\rho & H_Q(U'|U)<R, H_Q(U,U')<R\end{cases} \tag{A64}$$
$$=\begin{cases}\big[H_Q(U,U')-R\big]/\rho & H_Q(U,U')\geq R\\ \big[H_Q(U,U')-R\big] & H_Q(U,U')<R\end{cases}, \tag{A65}$$

which gives, after raising to the ρ–th power,

$$\Big(\mathbb{E}\Big\{[N(Q_{UU'})]^{1/\rho}\Big\}\Big)^{\rho}\leq\begin{cases}\exp\{n\big[H_Q(U,U')-R\big]\} & H_Q(U,U')\geq R\\ \exp\{n\big[H_Q(U,U')-R\big]\cdot\rho\} & H_Q(U,U')<R\end{cases} \tag{A66}$$
$$=\exp\{n([H_Q(U,U')-R]_+-\rho[R-H_Q(U,U')]_+)\}. \tag{A67}$$

Let us denote $F(Q,R,\rho)=[H_Q(U,U')-R]_+-\rho[R-H_Q(U,U')]_+$. Continuing now from (A51),

$$\Big(\mathbb{E}\Big\{[P_{\mathrm{e}}(\mathcal{B}_n)]^{1/\rho}\Big\}\Big)^{\rho}$$
$$\dot{\leq}\left(\sum_{\{Q_{UU'}\in\mathcal{Q}:\ H_Q(U)\geq R(Q_U)\}}\mathbb{E}\Big\{[N(Q_{UU'})]^{1/\rho}\Big\}\cdot e^{n[\mathbb{E}_Q\log P(U)]/\rho}\cdot\exp\{-n\cdot\Lambda(Q_{UU'},R+\epsilon)/\rho\}\right)^{\rho} \tag{A68}$$
$$\doteq\sum_{\{Q_{UU'}\in\mathcal{Q}:\ H_Q(U)\geq R(Q_U)\}}\Big(\mathbb{E}\Big\{[N(Q_{UU'})]^{1/\rho}\Big\}\Big)^{\rho}\cdot e^{n\mathbb{E}_Q\log P(U)}\cdot\exp\{-n\cdot\Lambda(Q_{UU'},R+\epsilon)\} \tag{A69}$$
$$\leq\sum_{\{Q_{UU'}\in\mathcal{Q}:\ H_Q(U)\geq R(Q_U)\}}\exp\{n(F(Q,R,\rho)+\mathbb{E}_Q[\log P(U)]-\Lambda(Q_{UU'},R+\epsilon))\} \tag{A70}$$
$$\doteq\exp\left\{-n\cdot\min_{\{Q_{UU'}\in\mathcal{Q}:\ H_Q(U)\geq R(Q_U)\}}(\Lambda(Q_{UU'},R+\epsilon)-F(Q,R,\rho)-\mathbb{E}_Q[\log P(U)])\right\}. \tag{A71}$$

where (A70) follows from (A67). Finally, it follows by (A35) that

$$\liminf_{n\to\infty}-\frac{1}{n}\mathbb{E}[\log P_{\mathrm{e}}(\mathcal{B}_n)]$$
$$\geq\liminf_{n\to\infty}-\frac{1}{n}\log\Big(\mathbb{E}[P_{\mathrm{e}}(\mathcal{B}_n)]^{1/\rho}\Big)^{\rho} \tag{A72}$$
$$\geq\min_{\{Q_{UU'}\in\mathcal{Q}:\ H_Q(U)\geq R(Q_U)\}}(\Lambda(Q_{UU'},R+\epsilon)-F(Q,R,\rho)-\mathbb{E}_Q[\log P(U)]). \tag{A73}$$

Letting ρ grow without bound yields that

$$\begin{aligned}
&\liminf_{n\to\infty} -\frac{1}{n}\mathbb{E}[\log P_{\mathrm{e}}(\mathcal{B}_n)] \\
&\geq \min_{\{Q_{UU'}\in\mathcal{Q}:\ H_Q(U)\geq R(Q_U),\ H_Q(U,U')\geq R(Q_U)\}} (\Lambda(Q_{UU'}, R+\epsilon) - H_Q(U,U') \\
&\qquad\qquad + R(Q_U) - \mathbb{E}_Q[\log P(U)]) \qquad \text{(A74)} \\
&= \min_{\{Q_{UU'}\in\mathcal{Q}:\ H_Q(U)\geq R(Q_U)\}} (\Lambda(Q_{UU'}, R+\epsilon) - H_Q(U,U') + R(Q_U) - \mathbb{E}_Q[\log P(U)]). \qquad \text{(A75)}
\end{aligned}$$

Due to the arbitrariness of $\epsilon > 0$, we have proved that

$$\begin{aligned}
&\liminf_{n\to\infty} -\frac{1}{n}\mathbb{E}[\log P_{\mathrm{e}}(\mathcal{B}_n)] \\
&\geq \min_{\{Q_{UU'}\in\mathcal{Q}:\ H_Q(U)\geq R(Q_U)\}} (\Lambda(Q_{UU'}, R) - H_Q(U,U') + R(Q_U) - \mathbb{E}_Q[\log P(U)]). \qquad \text{(A76)}
\end{aligned}$$

completing half of the proof of Theorem 2.

B.2. Upper Bound on the Error Exponent

Consider a joint distribution $Q_{UU'}$, that satisfies $H_Q(U,U') > R$, and define the event $\mathcal{E}(Q_{UU'}) = \{\mathcal{B}_n :\ N(Q_{UU'}) < \exp\{n[H_Q(U,U') - R - \epsilon]\}\}$. We want to show that $\mathbb{P}\{\mathcal{E}(Q_{UU'})\}$ is small. Consider the following:

$$\begin{aligned}
\mathbb{P}\{\mathcal{E}(Q_{UU'})\} &= \mathbb{P}\{N(Q_{UU'}) < \exp\{n[H_Q(U,U') - R - \epsilon]\}\} && \text{(A77)} \\
&= \mathbb{P}\{N(Q_{UU'}) < e^{-n\epsilon}\cdot\mathbb{E}\{N(Q_{UU'})\}\} && \text{(A78)} \\
&= \mathbb{P}\left\{\frac{N(Q_{UU'})}{\mathbb{E}\{N(Q_{UU'})\}} - 1 < -(1-e^{-n\epsilon})\right\} && \text{(A79)} \\
&\leq \mathbb{P}\left\{\left[\frac{N(Q_{UU'}) - \mathbb{E}\{N(Q_{UU'})\}}{\mathbb{E}\{N(Q_{UU'})\}}\right]^2 > (1-e^{-n\epsilon})^2\right\} && \text{(A80)} \\
&\leq \frac{\mathrm{Var}\{N(Q_{UU'})\}}{(1-e^{-n\epsilon})^2\cdot\mathbb{E}^2\{N(Q_{UU'})\}}. && \text{(A81)}
\end{aligned}$$

Let us use the shorthand notations $\mathcal{I}(\boldsymbol{u},\boldsymbol{u}') = \mathbb{1}\{\mathcal{B}(\boldsymbol{u}') = \mathcal{B}(\boldsymbol{u})\}$, $K = |\mathcal{T}(Q_{UU'})|$, and $p = e^{-nR}$. Concerning the variance of $N(Q_{UU'})$, we have the following

$$\begin{aligned}
&\mathrm{Var}\{N(Q_{UU'})\} \\
&= \mathbb{E}\{N^2(Q_{UU'})\} - \mathbb{E}^2\{N(Q_{UU'})\} && \text{(A82)} \\
&= \mathbb{E}\left\{\left[\sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})}\mathcal{I}(\boldsymbol{u},\boldsymbol{u}')\right]\times\left[\sum_{(\tilde{\boldsymbol{u}},\hat{\boldsymbol{u}})\in\mathcal{T}(Q_{UU'})}\mathcal{I}(\tilde{\boldsymbol{u}},\hat{\boldsymbol{u}})\right]\right\} - (Kp)^2 && \text{(A83)} \\
&= \sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})}\ \sum_{(\tilde{\boldsymbol{u}},\hat{\boldsymbol{u}})\in\mathcal{T}(Q_{UU'})}\mathbb{E}\{\mathcal{I}(\boldsymbol{u},\boldsymbol{u}')\mathcal{I}(\tilde{\boldsymbol{u}},\hat{\boldsymbol{u}})\} - (Kp)^2 && \text{(A84)} \\
&= \sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})}\mathbb{E}\{\mathcal{I}^2(\boldsymbol{u},\boldsymbol{u}')\} + \sum_{\substack{(\boldsymbol{u},\boldsymbol{u}'),(\tilde{\boldsymbol{u}},\hat{\boldsymbol{u}})\in\mathcal{T}(Q_{UU'})\\(\boldsymbol{u},\boldsymbol{u}')\neq(\tilde{\boldsymbol{u}},\hat{\boldsymbol{u}})}}\mathbb{E}\{\mathcal{I}(\boldsymbol{u},\boldsymbol{u}')\mathcal{I}(\tilde{\boldsymbol{u}},\hat{\boldsymbol{u}})\} - (Kp)^2 && \text{(A85)} \\
&= Kp + K(K-1)p^2 - (Kp)^2 && \text{(A86)} \\
&= Kp(1-p) && \text{(A87)} \\
&\doteq \exp\{n[H_Q(U,U') - R]\}, && \text{(A88)}
\end{aligned}$$

and hence,

$$\mathbb{P}\{\mathcal{E}(Q_{UU'})\} \dot{\leq} \frac{\exp\{n[H_Q(U,U')-R]\}}{\exp\{n[2H_Q(U,U')-2R]\}} \tag{A89}$$

$$= \exp\{-n[H_Q(U,U')-R]\}, \tag{A90}$$

which decays to zero since we have assumed that $H_Q(U,U') > R$. Furthermore, if $H_Q(U,U') \geq R+\epsilon$, then $\mathbb{P}\{\mathcal{E}(Q_{UU'})\}$ tends to zero at least as fast as $e^{-n\epsilon}$. Now, for a given $\epsilon > 0$, and a given joint type $Q_{UU'V}$, such that $H_Q(U,U') \geq R+\epsilon$, let us define

$$Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) = \sum_{\tilde{\boldsymbol{u}}\in\mathcal{B}(\boldsymbol{u})\cap\mathcal{T}(\boldsymbol{u}),\tilde{\boldsymbol{u}}\neq\boldsymbol{u},\boldsymbol{u}'} \exp\{nf(\hat{P}_{\tilde{\boldsymbol{u}}\boldsymbol{v}})\}, \tag{A91}$$

and

$$\mathcal{G}_n(Q_{UU'V}) = \Bigg\{\mathcal{B}_n : \sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}') = \mathcal{B}(\boldsymbol{u})\}\times$$
$$\sum_{\boldsymbol{v}\in\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')} \mathbb{1}\Big\{Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) \leq e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\Big\} \geq$$
$$\exp\{n[H_Q(U,U')-R-3\epsilon/2]\}\cdot|\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')|\Bigg\}, \tag{A92}$$

where $(\boldsymbol{u},\boldsymbol{u}')$ in the expression $|\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')|$ should be understood as any pair of source sequences in $\mathcal{T}(Q_{UU'})$. Next, we define

$$\mathcal{G}_n = \bigcap_{\{Q_{UU'V}:\ H_Q(U,U')\geq R+\epsilon\}} [\mathcal{G}_n(Q_{UU'V}) \cap \mathcal{E}^{\mathrm{c}}(Q_{UU'})]. \tag{A93}$$

We start by proving that $\mathbb{P}\{\mathcal{G}_n\} \to 1$ as $n \to \infty$, or equivalently, that $\mathbb{P}\{\mathcal{G}_n^{\mathrm{c}}\} \to 0$ as $n \to \infty$. Now,

$$\mathbb{P}\{\mathcal{G}_n^{\mathrm{c}}\} = \mathbb{P}\Bigg\{\bigcup_{\{Q_{UU'V}:\ H_Q(U,U')\geq R+\epsilon\}} [\mathcal{G}_n^{\mathrm{c}}(Q_{UU'V}) \cup \mathcal{E}(Q_{UU'})]\Bigg\} \tag{A94}$$

$$\leq \sum_{\{Q_{UU'V}:\ H_Q(U,U')\geq R+\epsilon\}} \mathbb{P}\{\mathcal{G}_n^{\mathrm{c}}(Q_{UU'V}) \cup \mathcal{E}(Q_{UU'})\} \tag{A95}$$

$$= \sum_{\{Q_{UU'V}:\ H_Q(U,U')\geq R+\epsilon\}} [\mathbb{P}\{\mathcal{E}(Q_{UU'})\} + \mathbb{P}\{\mathcal{G}_n^{\mathrm{c}}(Q_{UU'V}) \cap \mathcal{E}^{\mathrm{c}}(Q_{UU'})\}]. \tag{A96}$$

The last summation contains a polynomial number of terms. If we prove that the summand tends to zero exponentially with n, then $\mathbb{P}\{\mathcal{G}_n^{\mathrm{c}}\} \to 0$ as $n \to \infty$. The first term in the summand, $\mathbb{P}\{\mathcal{E}(Q_{UU'})\}$, has already been proved to be upper bounded by $e^{-n\epsilon}$. Concerning the second term, we have the following

$$\mathbb{P}\{\mathcal{G}_n^c(Q_{UU'V}) \cap \mathcal{E}^c(Q_{UU'})\}$$

$$= \mathbb{P}\Bigg[\sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}') = \mathcal{B}(\boldsymbol{u})\} \cdot \sum_{\boldsymbol{v}\in\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')} \mathbb{1}\Big\{Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) \le e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\Big\} < \exp\{n[H_Q(U,U') - R - 3\epsilon/2]\} \cdot |\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')|, \; N(Q_{UU'}) \ge \exp\{n[H_Q(U,U') - R - \epsilon]\}\Bigg] \tag{A97}$$

$$= \mathbb{P}\Bigg[\sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}') = \mathcal{B}(\boldsymbol{u})\} \cdot \sum_{\boldsymbol{v}\in\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')} \mathbb{1}\Big\{Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) > e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\Big\} > [N(Q_{UU'}) - \exp\{n[H_Q(U,U') - R - 3\epsilon/2]\}] \cdot |\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')|, \; N(Q_{UU'}) \ge \exp\{n[H_Q(U,U') - R - \epsilon]\}\Bigg] \tag{A98}$$

$$\le \mathbb{P}\Bigg[\sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}') = \mathcal{B}(\boldsymbol{u})\} \cdot \sum_{\boldsymbol{v}\in\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')} \mathbb{1}\Big\{Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) > e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\Big\} > [\exp\{n[H_Q(U,U') - R - \epsilon]\} - \exp\{n[H_Q(U,U') - R - 3\epsilon/2]\}] \cdot |\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')|, \; N(Q_{UU'}) \ge \exp\{n[H_Q(U,U') - R - \epsilon]\}\Bigg] \tag{A99}$$

$$\le \mathbb{P}\Bigg[\sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}') = \mathcal{B}(\boldsymbol{u})\} \cdot \sum_{\boldsymbol{v}\in\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')} \mathbb{1}\Big\{Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) > e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\Big\} > [\exp\{n[H_Q(U,U') - R - \epsilon]\} - \exp\{n[H_Q(U,U') - R - 3\epsilon/2]\}] \cdot |\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')|\Bigg] \tag{A100}$$

$$\le \frac{\mathbb{E}\Big\{\sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}') = \mathcal{B}(\boldsymbol{u})\} \cdot \sum_{\boldsymbol{v}\in\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')} \mathbb{1}\Big\{Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) > e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\Big\}\Big\}}{[\exp\{n[H_Q(U,U') - R - \epsilon]\} - \exp\{n[H_Q(U,U') - R - 3\epsilon/2]\}] \cdot |\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')|} \tag{A101}$$

$$\doteq \frac{|\mathcal{T}(Q_{UU'})| \cdot |\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')| \cdot \mathbb{P}\Big\{\mathcal{B}(\boldsymbol{u}') = \mathcal{B}(\boldsymbol{u}),\; Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) > e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\Big\}}{\exp\{n[H_Q(U,U') - R - \epsilon]\} \cdot |\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')|} \tag{A102}$$

$$\doteq \frac{\exp\{nH_Q(U,U')\} \cdot \mathbb{P}\{\mathcal{B}(\boldsymbol{u}') = \mathcal{B}(\boldsymbol{u})\} \cdot \mathbb{P}\Big\{Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) > e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\Big\}}{\exp\{n[H_Q(U,U') - R - \epsilon]\}} \tag{A103}$$

$$= e^{n\epsilon} \cdot \mathbb{P}\Big\{Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) > e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\Big\}, \tag{A104}$$

where (A99) follows by using the second event $N(Q_{UU'}) \ge \exp\{n[H_Q(U,U') - R - \epsilon]\}$ to increase the first event inside the probability in (A98), (A100) is true since the second event in (A99) was omitted, (A101) follows from Markov's inequality, and (A103) is due to the independence between the two events inside the probability in (A102). As for the probability in (A104),

$$\begin{aligned}
&\mathbb{P}\Big\{Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) > e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\Big\} \\
&= \mathbb{P}\left\{\sum_{Q_{U|V}} N(Q_{UV})e^{nf(Q_{UV})} > e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\right\} && \text{(A105)}\\
&\doteq \max_{Q_{U|V}} \mathbb{P}\{N(Q_{UV}) > \exp\{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon-f(Q_{UV})]\}\} && \text{(A106)}\\
&\doteq e^{-nE}, && \text{(A107)}
\end{aligned}$$

where $N(Q_{UV})$ is the number of source sequences within $\mathcal{B}(\boldsymbol{u})$, other than $\boldsymbol{u}$ and $\boldsymbol{u}'$, that fall in the conditional type class $\mathcal{T}(Q_{U|V}|\boldsymbol{v})$, which is a binomial random variable with $e^{nH_Q(U|V)} - 2$ trials and success rate of exponential order e^{-nR}, and hence,

$$E = \min_{Q_{U|V}} \begin{cases} [R-H_Q(U|V)]_+ & f(Q_{UV}) + [H_Q(U|V)-R]_+ \geq \alpha(R-2\epsilon,Q_U,Q_V)+\epsilon \\ \infty & f(Q_{UV}) + [H_Q(U|V)-R]_+ < \alpha(R-2\epsilon,Q_U,Q_V)+\epsilon \end{cases} \tag{A108}$$

$$= \min_{\{Q_{U|V}:\, f(Q_{UV})+[H_Q(U|V)-R]_+ \geq \alpha(R-2\epsilon,Q_U,Q_V)+\epsilon\}} [R-H_Q(U|V)]_+. \tag{A109}$$

By definition of the function $\alpha(R,Q_U,Q_V)$, the set $\{Q_{U|V}:\ f(Q_{UV}) + [H_Q(U|V)-R]_+ \geq \alpha(R-2\epsilon,Q_U,Q_V)+\epsilon\}$ is a subset of $\{Q_{U|V}:\ H_Q(U|V) \leq R-2\epsilon\}$. Thus,

$$E \geq \min_{\{Q_{U|V}:\, H_Q(U|V)\leq R-2\epsilon\}} [R-H_Q(U|V)]_+ \geq 2\epsilon, \tag{A110}$$

and hence, $\mathbb{P}\Big\{Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) > e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\Big\} \dot{\leq} e^{-2n\epsilon}$, which provides

$$\mathbb{P}\{\mathcal{G}_n^c(Q_{UU'V}) \cap \mathcal{E}^c(Q_{UU'})\} \dot{\leq} e^{n\epsilon}\cdot e^{-2n\epsilon} = e^{-n\epsilon}, \tag{A111}$$

which proves that $\mathbb{P}\{\mathcal{G}_n\} \to 1$ as $n \to \infty$. Now, for a given $\mathcal{B}_n \in \mathcal{G}_n(Q_{UU'V})$, we define the set

$$\mathcal{K}(\mathcal{B}_n, Q_{UU'V}) = \{(\boldsymbol{u},\boldsymbol{u}',\boldsymbol{v}):\ Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) \leq \exp\{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]\}\}, \tag{A112}$$

and

$$\mathcal{K}(\mathcal{B}_n, Q_{UU'V}|\boldsymbol{u},\boldsymbol{u}') = \{\boldsymbol{v}:\ (\boldsymbol{u},\boldsymbol{u}',\boldsymbol{v}) \in \mathcal{K}(\mathcal{B}_n, Q_{UU'V})\}. \tag{A113}$$

Then, by definition, for any $\mathcal{B}_n \in \mathcal{G}_n(Q_{UU'V})$,

$$\begin{aligned}
&\sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}') = \mathcal{B}(\boldsymbol{u})\} \cdot \frac{|\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}') \cap \mathcal{K}(\mathcal{B}_n, Q_{UU'V}|\boldsymbol{u},\boldsymbol{u}')|}{|\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')|} \\
&\geq \exp\{n[H_Q(U,U') - R - 3\epsilon/2]\}, && \text{(A114)}
\end{aligned}$$

where we have used the fact that $\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')$ has exponentially the same cardinality for all $(\boldsymbol{u},\boldsymbol{u}') \in \mathcal{T}(Q_{UU'})$. Wrapping all up, we get that for any $\mathcal{B}_n \in \mathcal{G}_n$,

$$
\begin{aligned}
&P_{\mathrm{e}}(\mathcal{B}_n)\\
&= \sum_{\boldsymbol{u},\boldsymbol{v}} P(\boldsymbol{u},\boldsymbol{v})\mathbb{1}\{\hat{H}_{\boldsymbol{u}}(U) \geq R(\hat{P}_{\boldsymbol{u}})\}\frac{\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u})\cap\mathcal{T}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}}\exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}{\exp\{nf(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})\}+\exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}+Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v})} && \text{(A115)}\\
&\geq \sum_{\{Q_{UU'}\colon H_Q(U,U')\geq R+\epsilon, H_Q(U)\geq R(Q_U)\}}\ \sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}')=\mathcal{B}(\boldsymbol{u})\}\cdot\exp\{n\mathbb{E}_Q\log P(U)\}\\
&\quad\times\sum_{Q_{V|UU'}}\ \sum_{\boldsymbol{v}\in\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')\cap\mathcal{K}(\mathcal{B}_n,Q_{UU'V}|\boldsymbol{u},\boldsymbol{u}')}\exp\{n\mathbb{E}_Q\log P(V|U)\}\\
&\qquad\times\frac{\exp\{nf(Q_{U'V})\}}{\exp\{nf(Q_{UV})\}+\exp\{nf(Q_{U'V})\}+Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v})} && \text{(A116)}\\
&\geq \sum_{\{Q_{UU'}\colon H_Q(U,U')\geq R+\epsilon, H_Q(U)\geq R(Q_U)\}}\ \sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}')=\mathcal{B}(\boldsymbol{u})\}\cdot\exp\{n\mathbb{E}_Q\log P(U)\}\\
&\quad\times\sum_{Q_{V|UU'}}\ \sum_{\boldsymbol{v}\in\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')\cap\mathcal{K}(\mathcal{B}_n,Q_{UU'V}|\boldsymbol{u},\boldsymbol{u}')}\exp\{n\mathbb{E}_Q\log P(V|U)\}\\
&\qquad\times\frac{\exp\{nf(Q_{U'V})\}}{\exp\{nf(Q_{UV})\}+\exp\{nf(Q_{U'V})\}+\exp\{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]\}} && \text{(A117)}\\
&\doteq \sum_{\{Q_{UU'}\colon H_Q(U,U')\geq R+\epsilon, H_Q(U)\geq R(Q_U)\}}\ \sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}')=\mathcal{B}(\boldsymbol{u})\}\\
&\quad\times\sum_{Q_{V|UU'}}\frac{|\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')\cap\mathcal{K}(\mathcal{B}_n,Q_{UU'V}|\boldsymbol{u},\boldsymbol{u}')|}{|\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')|}\cdot|\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')|\cdot e^{n\mathbb{E}_Q\log P(U,V)}\\
&\quad\times\exp\{-n\cdot[\max\{f(Q_{UV}),\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon\}-f(Q_{U'V})]_+\} && \text{(A118)}\\
&\doteq \sum_{\{Q_{UU'V}\colon H_Q(U,U')\geq R+\epsilon, H_Q(U)\geq R(Q_U)\}}\ \sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}')=\mathcal{B}(\boldsymbol{u})\}\\
&\quad\times\frac{|\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')\cap\mathcal{K}(\mathcal{B}_n,Q_{UU'V}|\boldsymbol{u},\boldsymbol{u}')|}{|\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')|}\cdot e^{nH_Q(V|U,U')}\cdot e^{n\mathbb{E}_Q\log P(U,V)}\\
&\quad\times\exp\{-n\cdot[\max\{f(Q_{UV}),\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon\}-f(Q_{U'V})]_+\} && \text{(A119)}\\
&\geq \sum_{\{Q_{UU'V}\colon H_Q(U,U')\geq R+\epsilon, H_Q(U)\geq R(Q_U)\}}\exp\{n[H_Q(U,U')-R-3\epsilon/2]\}\cdot e^{nH_Q(V|U,U')}\\
&\quad\times e^{n\mathbb{E}_Q\log P(U,V)}\cdot\exp\{-n\cdot[\max\{f(Q_{UV}),\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon\}-f(Q_{U'V})]_+\} && \text{(A120)}\\
&\doteq \exp\Bigg\{-n\cdot\min_{\{Q_{UU'V}\colon H_Q(U,U')\geq R+\epsilon, H_Q(U)\geq R(Q_U)\}}\{-H_Q(U,U')+R+3\epsilon/2-H_Q(V|U,U')\\
&\qquad-\mathbb{E}_Q[\log P(U,V)]+[\max\{f(Q_{UV}),\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon\}-f(Q_{U'V})]_+\}\Bigg\} && \text{(A121)}\\
&\stackrel{\triangle}{=} \exp\{-nE_{\mathrm{trc}}(R,\epsilon)\}, && \text{(A122)}
\end{aligned}
$$

where (A117) follows from the definition of the set $\mathcal{K}(\mathcal{B}_n, Q_{UU'V}|\boldsymbol{u},\boldsymbol{u}')$ in (A113) and (A120) is due to (A114). Consider the following:

$$\begin{aligned}
&\mathbb{E}\left[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{B}_n)\right] \\
&=\sum_{\mathcal{B}_n}\mathbb{P}\{\mathcal{B}_n\}\left(-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{B}_n)\right) && \text{(A123)}\\
&=\sum_{\mathcal{B}_n\in\mathcal{G}_n}\mathbb{P}\{\mathcal{B}_n\}\left(-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{B}_n)\right)+\sum_{\mathcal{B}_n\in\mathcal{G}_n^{\mathrm{c}}}\mathbb{P}\{\mathcal{B}_n\}\left(-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{B}_n)\right) && \text{(A124)}\\
&\le\sum_{\mathcal{B}_n\in\mathcal{G}_n}\mathbb{P}\{\mathcal{B}_n\}\left(-\frac{1}{n}\log e^{-nE_{\mathrm{trc}}(R,\epsilon)}\right)+\sum_{\mathcal{B}_n\in\mathcal{G}_n^{\mathrm{c}}}\mathbb{P}\{\mathcal{B}_n\}\left(-\frac{1}{n}\log e^{-nE_{\mathrm{sp}}(R)}\right) && \text{(A125)}\\
&=\mathbb{P}\{\mathcal{G}_n\}E_{\mathrm{trc}}(R,\epsilon)+\mathbb{P}\{\mathcal{G}_n^{\mathrm{c}}\}E_{\mathrm{sp}}(R), && \text{(A126)}
\end{aligned}$$

which implies that

$$\limsup_{n\to\infty}\mathbb{E}\left[-\frac{1}{n}\log P_{\mathrm{e}}(\mathcal{B}_n)\right]\le E_{\mathrm{trc}}(R,\epsilon). \tag{A127}$$

It follows from the arbitrariness of ϵ that

$$\begin{aligned}
&\limsup_{n\to\infty}\mathbb{E}\left\{-\frac{1}{n}\log[P_{\mathrm{e}}(\mathcal{B}_n)]\right\}\\
&\le\min_{\{Q_{UU'V}\colon\, H_Q(U,U')\ge R, H_Q(U)\ge R(Q_U)\}}\{-H_Q(U,U')+R-H_Q(V|U,U')\\
&\qquad-\mathbb{E}_Q[\log P(U,V)]+[\max\{f(Q_{UV}),\alpha(R,Q_U,Q_V)\}-f(Q_{U'V})]_+\} && \text{(A128)}\\
&=\min_{\{Q_{UU'V}\colon\, H_Q(U)\ge R(Q_U)\}}\{-H_Q(U,U')+R-H_Q(V|U,U')\\
&\qquad-\mathbb{E}_Q[\log P(U,V)]+[\max\{f(Q_{UV}),\alpha(R,Q_U,Q_V)\}-f(Q_{U'V})]_+\} && \text{(A129)}\\
&=\min_{\{Q_{UU'}\in\mathcal{Q}\colon\, H_Q(U)\ge R(Q_U)\}}\{\Lambda(Q_{UU'},R)-H_Q(U,U')+R(Q_U)-\mathbb{E}_Q[\log P(U)]\}, && \text{(A130)}
\end{aligned}$$

which completes the proof of Theorem 2.

Appendix C. Proof of Lemma A1

Let $N(\mathcal{T}(Q_{U|V}|\boldsymbol{v}),\mathcal{B}(\boldsymbol{u}))$ be defined as

$$N(\mathcal{T}(Q_{U|V}|\boldsymbol{v}),\mathcal{B}(\boldsymbol{u}))=\sum_{\boldsymbol{u}'\in\mathcal{T}(Q_{U|V}|\boldsymbol{v})}\mathbb{1}\{\mathcal{B}(\boldsymbol{u}')=\mathcal{B}(\boldsymbol{u})\}. \tag{A131}$$

First, note that

$$Z_{\boldsymbol{u}}(\boldsymbol{v})=\sum_{\tilde{\boldsymbol{u}}\in\mathcal{B}(\boldsymbol{u})\cap\mathcal{T}(\boldsymbol{u}),\tilde{\boldsymbol{u}}\neq\boldsymbol{u}}\exp\{nf(\hat{P}_{\tilde{\boldsymbol{u}}\boldsymbol{v}})\}=\sum_{Q_{U|V}\in\mathcal{S}(\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})}N(\mathcal{T}(Q_{U|V}|\boldsymbol{v}),\mathcal{B}(\boldsymbol{u}))e^{nf(Q_{UV})}, \tag{A132}$$

where $\mathcal{S}(\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})=\{Q_{U|V}:\ (\hat{P}_{\boldsymbol{v}}\times Q_{U|V})_U=\hat{P}_{\boldsymbol{u}}\}$. Thus, taking the randomness of $\{\mathcal{B}(\boldsymbol{u})\}_{\boldsymbol{u}\in\mathcal{U}^n}$ into account,

$$\mathbb{P}\big\{Z_{\boldsymbol{v}}(\boldsymbol{u}) \le \exp\{n\alpha(R+\epsilon,\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})\}\big\}$$

$$= \mathbb{P}\left\{\sum_{Q_{U|V}\in\mathcal{S}(\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})} N(\mathcal{T}(Q_{U|V}|\boldsymbol{v}),\mathcal{B}(\boldsymbol{u}))e^{nf(Q_{UV})} \le \exp\{n\alpha(R+\epsilon,\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})\}\right\} \tag{A133}$$

$$\le \mathbb{P}\left\{\max_{Q_{U|V}\in\mathcal{S}(\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})} N(\mathcal{T}(Q_{U|V}|\boldsymbol{v}),\mathcal{B}(\boldsymbol{u}))e^{nf(Q_{UV})} \le \exp\{n\alpha(R+\epsilon,\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})\}\right\} \tag{A134}$$

$$= \mathbb{P}\bigcap_{Q_{U|V}\in\mathcal{S}(\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})}\Big\{N(\mathcal{T}(Q_{U|V}|\boldsymbol{v}),\mathcal{B}(\boldsymbol{u}))e^{nf(Q_{UV})} \le \exp\{n\alpha(R+\epsilon,\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})\}\Big\} \tag{A135}$$

$$= \mathbb{P}\bigcap_{Q_{U|V}\in\mathcal{S}(\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})}\Big\{N(\mathcal{T}(Q_{U|V}|\boldsymbol{v}),\mathcal{B}(\boldsymbol{u})) \le \exp\{n[\alpha(R+\epsilon,\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})-f(Q_{UV})]\}\Big\}. \tag{A136}$$

Now, $N(\mathcal{T}(Q_{U|V}|\boldsymbol{v}),\mathcal{B}(\boldsymbol{u}))$ is a binomial random variable with $|\mathcal{T}(Q_{U|V}|\boldsymbol{v})| \doteq e^{nH_Q(U|V)}$ trials and success rate which is of the exponential order of e^{-nR}. We prove that by the very definition of the function $\alpha(R+\epsilon,\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})$, there must exist some conditional distribution $Q^*_{U|V}\in\mathcal{S}(\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})$ such that for $Q^*_{UV}=\hat{P}_{\boldsymbol{v}}\times Q^*_{U|V}$, the two inequalities $H_{Q^*}(U|V)\ge R+\epsilon$ and $H_{Q^*}(U|V)-R-\epsilon \ge \alpha(R+\epsilon,\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})-f(Q^*_{UV})$ hold. To show that, we assume conversely, i.e., that for every conditional distribution $Q_{U|V}\in\mathcal{S}(\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})$, which defines $Q_{UV}=\hat{P}_{\boldsymbol{v}}\times Q_{U|V}$, either $H_Q(U|V)<R+\epsilon$ or $H_Q(U|V)-R-\epsilon<\alpha(R+\epsilon,\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})-f(Q_{UV})$, which means that for every distribution $Q_{U|V}\in\mathcal{S}(\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})$

$$H_Q(U|V)-\epsilon < \max\{R, R+\alpha(R+\epsilon,\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})-f(Q_{UV})\} \tag{A137}$$

$$= R+[\alpha(R+\epsilon,\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})-f(Q_{UV})]_+. \tag{A138}$$

Writing it slightly differently, for every $Q_{U|V}\in\mathcal{S}(\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})$ there exists some real number $t\in[0,1]$ such that

$$H_Q(U|V)-\epsilon < R+t[\alpha(R+\epsilon,\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})-f(Q_{UV})], \tag{A139}$$

or equivalently,

$$\alpha(R+\epsilon,\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}}) > \max_{Q_{U|V}\in\mathcal{S}(\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})}\min_{t\in[0,1]} f(Q_{UV})+\frac{H_Q(U|V)-R-\epsilon}{t} \tag{A140}$$

$$= \max_{Q_{U|V}\in\mathcal{S}(\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})}\begin{cases} f(Q_{UV})+H_Q(U|V)-R-\epsilon & H_Q(U|V)\ge R+\epsilon \\ -\infty & H_Q(U|V)<R+\epsilon \end{cases} \tag{A141}$$

$$= \max_{\{Q_{U|V}\in\mathcal{S}(\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}}):\ H_Q(U|V)\ge R+\epsilon\}}[f(Q_{UV})+H_Q(U|V)]-R-\epsilon \tag{A142}$$

$$\equiv \alpha(R+\epsilon,\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}}), \tag{A143}$$

which is a contradiction. Let the conditional distribution $Q^*_{U|V}$ be as defined above. Then,

$$\mathbb{P}\bigcap_{Q_{U|V}\in\mathcal{S}(\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})}\Big\{N(\mathcal{T}(Q_{U|V}|\boldsymbol{v}),\mathcal{B}(\boldsymbol{u})) \le \exp\{n[\alpha(R+\epsilon,\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})-f(Q_{UV})]\}\Big\} \tag{A144}$$

$$\le \mathbb{P}\Big\{N(\mathcal{T}(Q^*_{U|V}|\boldsymbol{v}),\mathcal{B}(\boldsymbol{u})) \le \exp\{n[\alpha(R+\epsilon,\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})-f(Q^*_{UV})]\}\Big\}. \tag{A145}$$

Now, we know that both of the inequalities $H_{Q^*}(U|V) \geq R + \epsilon$ and $H_{Q^*}(U|V) - R - \epsilon \geq \alpha(R + \epsilon, \hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}}) - f(Q^*_{UV})$ hold. By the Chernoff bound, the probability of (A145) is upper bounded by

$$\exp\Big\{ - e^{nH_{Q^*}(U|V)} D(e^{-an} \| e^{-bn}) \Big\}, \tag{A146}$$

where $a = H_{Q^*}(U|V) + f(Q^*_{UV}) - \alpha(R + \epsilon, \hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}})$ and $b = R$, and where $D(\alpha \| \beta)$, for $\alpha, \beta \in [0, 1]$, is the binary divergence function, that is

$$D(\alpha \| \beta) = \alpha \log \frac{\alpha}{\beta} + (1 - \alpha) \log \frac{1 - \alpha}{1 - \beta}. \tag{A147}$$

Since $a - b \geq \epsilon$, the binary divergence is lower bounded as follows [24] (Section 6.3):

$$\begin{aligned} D(e^{-an} \| e^{-bn}) &\geq e^{-bn} \Big\{ 1 - e^{-(a-b)n} [1 + n(a - b)] \Big\} && \text{(A148)} \\ &\geq e^{-nR} [1 - e^{-n\epsilon}(1 + n\epsilon)], && \text{(A149)} \end{aligned}$$

where in the second inequality, we invoked the decreasing monotonicity of the function $f(t) = (1 + t)e^{-t}$ for $t \geq 0$. Finally, we get that

$$\begin{aligned} &\mathbb{P}\Big\{ N(\mathcal{T}(Q^*_{U|V}|\boldsymbol{v}), \mathcal{B}(\boldsymbol{u})) \leq \exp\{n[\alpha(R + \epsilon, \hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}}) - f(Q^*_{UV})]\} \Big\} && \text{(A150)} \\ &\leq \exp\Big\{ - e^{nH_{Q^*}(U|V)} \cdot e^{-nR} [1 - e^{-n\epsilon}(1 + n\epsilon)] \Big\} && \text{(A151)} \\ &\leq \exp\Big\{ - e^{n\epsilon} [1 - e^{-n\epsilon}(1 + n\epsilon)] \Big\} && \text{(A152)} \\ &= \exp\Big\{ - e^{n\epsilon} + n\epsilon + 1 \Big\}. && \text{(A153)} \end{aligned}$$

This completes the proof of Lemma A1.

Appendix D. Proof of Theorem 3

By definition of the error exponents, it follows that $E_{\text{trc,GLD}}(R(\cdot)) \geq E_{\text{r,GLD}}(R(\cdot))$. We now prove the other direction. The expression in (28) can also be written as

$$\begin{aligned} &E_{\text{trc,GLD}}(R(\cdot)) \\ &= \min_{\substack{\{Q_{UU'} \colon Q_{U'} = Q_U, \\ H_Q(U) \geq R(Q_U)\}}} \big\{ \Lambda(Q_{UU'}, R(Q_U)) - \mathbb{E}_Q[\log P(U)] - H_Q(U, U') + R(Q_U) \big\} && \text{(A154)} \\ &= \min_{\substack{\{Q_{UU'} \colon Q_{U'} = Q_U, \\ H_Q(U) \geq R(Q_U)\}}} \Bigg\{ \min_{Q_{V|UU'}} \big\{ \Psi(R(Q_U), Q_{UU'V}) - H_Q(V|U, U') - \mathbb{E}_Q[\log P(V|U)] \big\} \\ &\qquad - \mathbb{E}_Q[\log P(U)] - H_Q(U, U') + R(Q_U) \Big\} && \text{(A155)} \\ &= \min_{\substack{\{Q_{UU'V} \colon Q_{U'} = Q_U, \\ H_Q(U) \geq R(Q_U)\}}} \big\{ \Psi(R(Q_U), Q_{UU'V}) - H_Q(U, U', V) - \mathbb{E}_Q[\log P(U, V)] + R(Q_U) \big\} && \text{(A156)} \\ &= \min_{\substack{\{Q_{UU'V} \colon Q_{U'} = Q_U, \\ H_Q(U) \geq R(Q_U)\}}} \big\{ \Psi(R(Q_U), Q_{UU'V}) + D(Q_{UV} \| P_{UV}) - H_Q(U'|U, V) + R(Q_U) \big\} && \text{(A157)} \\ &= \min_{\mathcal{Q}} \big\{ D(Q_{UV} \| P_{UV}) + R(Q_U) - H_Q(U'|U, V) \\ &\qquad + [\max\{f(Q_{UV}), \gamma(R(Q_U), Q_U, Q_V)\} - f(Q_{U'V})]_+ \big\}, && \text{(A158)} \end{aligned}$$

with the set $\mathcal{Q}$ given by $\mathcal{Q} = \{Q_{UU'V} : \ Q_{U'} = Q_U, \ H_Q(U) \geq R(Q_U)\}$, and where,

$$\gamma(R(\cdot), Q_U, Q_V) = \max_{\left\{\substack{Q_{\tilde{U}|V}: \, Q_{\tilde{U}} = Q_U, \\ H_Q(\tilde{U}|V) \geq R(Q_{\tilde{U}})}\right\}} \{f(Q_{\tilde{U}V}) + H_Q(\tilde{U}|V)\} - R(Q_U). \tag{A159}$$

We upper–bound the minimum in (A158) by decreasing the feasible set; we add to $\mathcal{Q}$ the constraint that $U \leftrightarrow V \leftrightarrow U'$ form a Markov chain in that order and denote the new feasible set by $\tilde{\mathcal{Q}}$. We get that

$$\begin{aligned} E_{\text{trc,GLD}}(R(\cdot)) &\leq \min_{\tilde{Q}} \big\{ D(Q_{UV} \| P_{UV}) + R(Q_U) - H_Q(U'|U,V) \\ &\qquad + [\max\{f(Q_{UV}), \gamma(R(Q_U), Q_U, Q_V)\} - f(Q_{U'V})]_+ \big\} \qquad \text{(A160)} \\ &= \min_{\tilde{Q}} \big\{ D(Q_{UV} \| P_{UV}) + R(Q_U) - H_Q(U'|V) \\ &\qquad + [\max\{f(Q_{UV}), \gamma(R(Q_U), Q_U, Q_V)\} - f(Q_{U'V})]_+ \big\} \qquad \text{(A161)} \\ &= \min_{\{Q_{UV}: \, H_Q(U) \geq R(Q_U)\}} \Bigg\{ D(Q_{UV} \| P_{UV}) + \min_{Q_{U'|V} \in \hat{\mathcal{Q}}} \{R(Q_U) - H_Q(U'|V) \\ &\qquad + [\max\{f(Q_{UV}), \gamma(R(Q_U), Q_U, Q_V)\} - f(Q_{U'V})]_+ \} \Bigg\}, \qquad \text{(A162)} \end{aligned}$$

where $\hat{\mathcal{Q}} = \{Q_{U'|V} : \ Q_{U'} = Q_U\}$. In order to upper–bound the inner minimum in (A162), we split into two cases, according to the maximum between $f(Q_{UV})$ and $\gamma(R(Q_U), Q_U, Q_V)$. This is legitimate when the inner minimum and this maximum can be interchanged, which is possible at least in the special cases of the matched/mismatched decoding metrics $f(Q) = \beta \mathbb{E}_Q[\log \tilde{P}(U,V)]$ for some $\beta > 0$, since if $f(Q)$ is linear, then the entire expression inside the inner minimum in (A162) is convex in $Q_{U'|V}$. On the one hand, if the maximum is given by $f(Q_{UV})$, then the inner minimum in (A162) is just

$$\min_{Q_{U'|V} \in \hat{\mathcal{Q}}} \{R(Q_U) - H_Q(U'|V) + [f(Q_{UV}) - f(Q_{U'V})]_+\}. \tag{A163}$$

On the other hand, if the maximum is given by $\gamma(R(Q_U), Q_U, Q_V)$, let $Q^* = Q^*_{\tilde{U}|V}$ be the maximizer in (A159), and then

$$\begin{aligned} &\min_{Q_{U'|V} \in \hat{\mathcal{Q}}} \{R(Q_U) - H_Q(U'|V) + [\gamma(R(Q_U), Q_U, Q_V) - f(Q_{U'V})]_+\} \\ &= \min_{Q_{U'|V} \in \hat{\mathcal{Q}}} \left\{ R(Q_U) - H_Q(U'|V) + \left[f(Q^*_{\tilde{U}V}) + H_{Q^*}(\tilde{U}|V) - R(Q_U) - f(Q_{U'V}) \right]_+ \right\} \qquad \text{(A164)} \\ &\leq R(Q_U) - H_{Q^*}(U'|V) + \left[f(Q^*_{\tilde{U}V}) + H_{Q^*}(\tilde{U}|V) - R(Q_U) - f(Q^*_{U'V}) \right]_+ \qquad \text{(A165)} \\ &= R(Q_U) - H_{Q^*}(U'|V) + \left[H_{Q^*}(\tilde{U}|V) - R(Q_U) \right]_+ \qquad \text{(A166)} \\ &= R(Q_U) - H_{Q^*}(U'|V) + H_{Q^*}(\tilde{U}|V) - R(Q_U) \qquad \text{(A167)} \\ &= 0, \qquad \text{(A168)} \end{aligned}$$

where (A165) is because we choose $Q^*_{U'|V} = Q^*_{\tilde{U}|V}$ instead of minimizing over all $Q_{U'|V} \in \hat{\mathcal{Q}}$ and (A167) is true since $H_{Q^*}(\tilde{U}|V) \geq R(Q_U)$ by the definition of $\gamma(R(Q_U), Q_U, Q_V)$. Combining (A163) and (A168), we find that (A162) is upper–bounded by

$$E_{\text{trc,GLD}}(R(\cdot)) \leq \min_{\{Q_{UV}\colon\, H_Q(U)\geq R(Q_U)\}} \{D(Q_{UV}\|P_{UV}) + \max\left\{\min_{Q_{U'|V}\in\hat{\mathcal{Q}}} \{R(Q_U) - H_Q(U'|V) + [f(Q_{UV}) - f(Q_{U'V})]_+\}, 0\right\}\} \tag{A169}$$

$$= \min_{\{Q_{UV}\colon\, H_Q(U)\geq R(Q_U)\}} \{D(Q_{UV}\|P_{UV}) + \left[\min_{Q_{U'|V}\in\hat{\mathcal{Q}}} \{R(Q_U) - H_Q(U'|V) + [f(Q_{UV}) - f(Q_{U'V})]_+\}\right]_+\} \tag{A170}$$

$$= \min_{\{Q_{UV}\colon\, H_Q(U)\geq R(Q_U)\}} \{D(Q_{UV}\|P_{UV}) + \min_{Q_{U'|V}\in\hat{\mathcal{Q}}} \left\{\left[R(Q_U) - H_Q(U'|V) + [f(Q_{UV}) - f(Q_{U'V})]_+\right]_+\right\}\} \tag{A171}$$

$$= E_{\text{r,GLD}}(R(\cdot)), \tag{A172}$$

which proves the first point of the theorem. Moving forward, consider the following:

$$E_{\text{trc,MAP}}(R(\cdot)) \overset{\text{(a)}}{=} E_{\text{r,MAP}}(R(\cdot)) \overset{\text{(b)}}{=} E_{\text{r,MCE}}(R(\cdot)) \overset{\text{(c)}}{\leq} E_{\text{trc,MCE}}(R(\cdot)) \overset{\text{(d)}}{\leq} E_{\text{trc,MAP}}(R(\cdot)), \tag{A173}$$

where (a) follows from the first point in this theorem by using the matched decoding metric $f(Q) = \beta\mathbb{E}_Q[\log P(U,V)]$ and letting $\beta \to \infty$. Equality (b) is due to the second point of Theorem 1, which ensures that the random binning error exponents of the MAP and the MCE decoders are equal. Passage (c) is thanks to the fact that for any decoder, the error exponent of the typical random code is always at least as high as the random coding error exponent and (d) is due to the fact that the MAP decoder is optimal. Finally, the leftmost and the rightmost sides of (A173) are the same, which implies that passages (c) and (d) must hold with equalities. The equality in passage (c) concludes the second point of the theorem.

Appendix E. Proof of Theorem 4

The left equality in (33) is implied by the proved equality in passage (d) in (A173). In order to prove the right equality in (33), first note that $E_{\text{trc,SCE}}(R(\cdot)) \leq E_{\text{trc,MAP}}(R(\cdot))$ by the optimality of the MAP decoder. For the other direction, consider the universal decoding metric of $f(Q_{UV}) = -H_Q(U|V)$. Then, trivially,

$$\gamma(R(\cdot), Q_U, Q_V) = \max_{\left\{\substack{Q_{\tilde{U}|V}\colon\, Q_{\tilde{U}}=Q_U,\\ H_Q(\tilde{U}|V)\geq R(Q_{\tilde{U}})}\right\}} \{f(Q_{\tilde{U}V}) + H_Q(\tilde{U}|V)\} - R(Q_U) = -R(Q_U), \tag{A174}$$

and

$$\Psi(R(\cdot), Q_{UU'V}) = [\max\{f(Q_{UV}), \gamma(R(\cdot), Q_U, Q_V)\} - f(Q_{U'V})]_+ \tag{A175}$$

$$= \left[\max\{-H_Q(U|V), -R(Q_U)\} + H_Q(U'|V)\right]_+ \tag{A176}$$

$$= \left[H_Q(U'|V) - \min\{H_Q(U|V), R(Q_U)\}\right]_+ \tag{A177}$$

$$\geq \left[H_Q(U'|U,V) - \min\{H_Q(U|V), R(Q_U)\}\right]_+. \tag{A178}$$

We have the following

$$
\begin{aligned}
&E_{\text{trc,SCE}}(R(\cdot)) \\
&= \min_{Q}\big\{D(Q_{UV}\|P_{UV}) + R(Q_U) - H_Q(U'|U,V) \\
&\qquad +[\max\{f(Q_{UV}), \gamma(R(Q_U), Q_U, Q_V)\} - f(Q_{U'V})]_+\big\} && \text{(A179)} \\
&\geq \min_{Q}\big\{D(Q_{UV}\|P_{UV}) + R(Q_U) - H_Q(U'|U,V) \\
&\qquad +\big[H_Q(U'|U,V) - \min\{H_Q(U|V), R(Q_U)\}\big]_+\big\} && \text{(A180)} \\
&= \min_{Q}\big\{D(Q_{UV}\|P_{UV}) - \min\{H_Q(U|V), H_Q(U'|U,V), R(Q_U)\} + R(Q_U)\big\} && \text{(A181)} \\
&\geq \min_{Q}\big\{D(Q_{UV}\|P_{UV}) - \min\{H_Q(U|V), H_Q(U'), R(Q_U)\} + R(Q_U)\big\} && \text{(A182)} \\
&= \min_{\{Q_{UV}:\ H_Q(U)\geq R(Q_U)\}}\big\{D(Q_{UV}\|P_{UV}) - \min\{H_Q(U|V), H_Q(U), R(Q_U)\} + R(Q_U)\big\} && \text{(A183)} \\
&= \min_{\{Q_{UV}:\ H_Q(U)\geq R(Q_U)\}}\big\{D(Q_{UV}\|P_{UV}) - \min\{H_Q(U|V), R(Q_U)\} + R(Q_U)\big\} && \text{(A184)} \\
&= \min_{\{Q_{UV}:\ H_Q(U)\geq R(Q_U)\}}\big\{D(Q_{UV}\|P_{UV}) + \max\{R(Q_U) - H_Q(U|V), 0\}\big\} && \text{(A185)} \\
&= \min_{\{Q_{UV}:\ H_Q(U)\geq R(Q_U)\}}\big\{D(Q_{UV}\|P_{UV}) + [R(Q_U) - H_Q(U|V)]_+\big\} && \text{(A186)} \\
&= E_{\text{trc,MAP}}(R(\cdot)), && \text{(A187)}
\end{aligned}
$$

which completes the proof of the theorem.

Appendix F. Proof of Theorem 5

We start by writing the expression in (34) in a slightly different way using $\min_{\{Q:\ g(Q)\leq 0\}} f(Q) = \min_Q \sup_{s\geq 0}\{f(Q) + s\cdot g(Q)\}$:

$$
\begin{aligned}
E_{\text{er}}(R(\cdot),\Delta) &= \min_{\{Q_{UV}:\ R(Q_U)\geq H_Q(U|V)+\Delta\}} D(Q_{UV}\|P_{UV}) && \text{(A188)} \\
&= \min_{Q_{UV}}\sup_{\sigma\geq 0}\{D(Q_{UV}\|P_{UV}) + \sigma\cdot(H_Q(U|V) + \Delta - R(Q_U))\}. && \text{(A189)}
\end{aligned}
$$

Now, the requirement $E_{\text{er}}(R(\cdot),\Delta)\geq \mathsf{E}_r$ is equivalent to

$$
\min_{Q_{UV}}\sup_{\sigma\geq 0}\{D(Q_{UV}\|P_{UV}) + \sigma\cdot(H_Q(U|V) + \Delta - R(Q_U))\}\geq \mathsf{E}_r \qquad \text{(A190)}
$$

or,

$$
\forall Q_{UV},\ \exists\sigma\geq 0,\ D(Q_{UV}\|P_{UV}) + \sigma\cdot(H_Q(U|V) + \Delta - R(Q_U))\geq \mathsf{E}_r \qquad \text{(A191)}
$$

or,

$$
\forall Q_U,\ \forall Q_{V|U},\ \exists\sigma\geq 0,\ R(Q_U)\leq H_Q(U|V) + \Delta + \frac{D(Q_{UV}\|P_{UV}) - \mathsf{E}_r}{\sigma} \qquad \text{(A192)}
$$

or that for any $Q_U\in\mathcal{P}(\mathcal{U})$,

$$R(Q_U) \leq \min_{Q_{V|U}} \sup_{\sigma \geq 0} \left\{ H_Q(U|V) + \Delta + \frac{D(Q_{UV}\|P_{UV}) - \mathsf{E}_r}{\sigma} \right\} \tag{A193}$$

$$= \min_{Q_{V|U}} \begin{cases} H_Q(U|V) + \Delta & D(Q_{UV}\|P_{UV}) \leq \mathsf{E}_r \\ \infty & D(Q_{UV}\|P_{UV}) > \mathsf{E}_r \end{cases} \tag{A194}$$

$$= \min_{\{Q_{V|U}\colon\, D(Q_{UV}\|P_{UV}) \leq \mathsf{E}_r\}} \{H_Q(U|V) + \Delta\}, \tag{A195}$$

with the understanding that a minimum over an empty set equals infinity.

Appendix G. Proof of Theorem 6

It follows by the identities $\min_{\{Q\colon\, g(Q)\leq 0\}} f(Q) = \min_Q \sup_{s\geq 0}\{f(Q) + s \cdot g(Q)\}$ and $[A]_+ = \max_{\mu\in[0,1]} \mu A$ that (24) can also be written as

$$E_e(R(\cdot)) = \min_{Q_U} \min_{Q_{V|U}} \max_{\mu\in[0,1]} \sup_{\sigma\geq 0} \{D(Q_{UV}\|P_{UV}) + \mu \cdot (R(Q_U) - H_Q(U|V)) + \sigma \cdot (R(Q_U) - H_Q(U))\}, \tag{A196}$$

such that $E_e(R(\cdot)) \geq \mathsf{E}_e$ is equivalent to

$$\begin{aligned} &\forall Q_U,\ \forall Q_{V|U},\ \exists \mu \in [0,1],\ \exists \sigma \geq 0 : \\ &D(Q_{UV}\|P_{UV}) + \mu \cdot (R(Q_U) - H_Q(U|V)) + \sigma \cdot (R(Q_U) - H_Q(U)) \geq \mathsf{E}_e, \end{aligned} \tag{A197}$$

or,

$$\begin{aligned} &\forall Q_U,\ \forall Q_{V|U},\ \exists \mu \in [0,1],\ \exists \sigma \geq 0 : \\ &R(Q_U) \geq \frac{\mu \cdot H_Q(U|V) + \sigma \cdot H_Q(U) + \mathsf{E}_e - D(Q_{UV}\|P_{UV})}{\mu + \sigma}, \end{aligned} \tag{A198}$$

or that for any $Q_U \in \mathcal{P}(\mathcal{U})$,

$$R(Q_U) \geq \max_{Q_{V|U}} \min_{\mu\in[0,1]} \inf_{\sigma\geq 0} \left\{ \frac{\mu \cdot H_Q(U|V) + \sigma \cdot H_Q(U) + \mathsf{E}_e - D(Q_{UV}\|P_{UV})}{\mu + \sigma} \right\} \tag{A199}$$

$$= \max_{Q_{V|U}} \min_{\mu\in[0,1]} \min \left\{ H_Q(U), H_Q(U|V) + \frac{\mathsf{E}_e - D(Q_{UV}\|P_{UV})}{\mu} \right\} \tag{A200}$$

$$= \max_{Q_{V|U}} \min \left\{ H_Q(U), \min_{\mu\in[0,1]} \left\{ H_Q(U|V) + \frac{\mathsf{E}_e - D(Q_{UV}\|P_{UV})}{\mu} \right\} \right\} \tag{A201}$$

$$= \max_{Q_{V|U}} \begin{cases} \min\{H_Q(U), H_Q(U|V) + \mathsf{E}_e - D(Q_{UV}\|P_{UV})\} & \mathsf{E}_e \geq D(Q_{UV}\|P_{UV}) \\ -\infty & \mathsf{E}_e < D(Q_{UV}\|P_{UV}) \end{cases} \tag{A202}$$

$$= \max_{\{Q_{V|U}\colon\, D(Q_{UV}\|P_{UV}) \leq \mathsf{E}_e\}} \min\{H_Q(U), H_Q(U|V) + \mathsf{E}_e - D(Q_{UV}\|P_{UV})\} \tag{A203}$$

$$= \min \left\{ H_Q(U), \max_{\{Q_{V|U}\colon\, D(Q_{UV}\|P_{UV}) \leq \mathsf{E}_e\}} \{H_Q(U|V) + \mathsf{E}_e - D(Q_{UV}\|P_{UV})\} \right\}, \tag{A204}$$

and the proof is complete.

References

1. Merhav, N. Error exponents of typical random codes. *IEEE Trans. Inf. Theory* **2018**, *64*, 6223–6235. [CrossRef]
2. Barg, A.; Forney, G.D., Jr. Random codes: Minimum distances and error exponents. *IEEE Trans. Inf. Theory* **2003**, *48*, 2568–2573. [CrossRef]
3. Nazari, A.; Anastasopoulos, A.; Pradhan, S.S. Error exponent for multiple–access channels: Lower bounds. *IEEE Trans. Inf. Theory* **2014**, *60*, 5095–5115. [CrossRef]
4. Merhav, N. Error exponents of typical random codes for the colored Gaussian channel. *IEEE Trans. Inf. Theory* **2019**, *65*, 8164–8179. [CrossRef]
5. Merhav, N. Error exponents of typical random trellis codes. *IEEE Trans. Inf. Theory* **2020**, *66*, 2067–2077. [CrossRef]
6. Merhav, N. A Lagrange–dual lower bound to the error exponent of the typical random code. *IEEE Trans. Inf. Theory* **2020**, *66*, 3456–3464. [CrossRef]
7. Tamir (Averbuch), R.; Merhav, N.; Weinberger, N.; Guillén i Fàbregas, A. Large deviations behavior of the logarithmic error probability of random codes. *IEEE Trans. Inf. Theory* **2020**, *66*, 6635–6659. [CrossRef]
8. Slepian, D.; Wolf, J. Noiseless coding of correlated information sources. *IEEE Trans. Inf. Theory* **1973**, *19*, 471–480. [CrossRef]
9. Gallager, R.G. Source coding with side information and universal coding. LIDS-P-937. Available online: http://web.mit.edu/gallager/www/papers/paper5.pdf (accessed on 20 February 2021).
10. Ahlswede, R.; Dueck, G. Good codes can be produced by a few permutations. *IEEE Trans. Inf. Theory* **1982**, *28*, 430–443. [CrossRef]
11. Csiszár, I.; Körner, J. Graph decomposition: A new key to coding theorems. *IEEE Trans. Inf. Theory* **1981**, *27*, 5–12. [CrossRef]
12. Chen, J.; He, D.-K.; Jagmohan, A.; Lastras-Montaño, L.A. On the Reliability Function of Variable-Rate Slepian-Wolf Coding. *Entropy* **2017**, *19*, 389. [CrossRef]
13. Weinberger, N.; Merhav, N. Optimum Tradeoffs Between the Error Exponent and the Excess-Rate Exponent of Variable-Rate Slepian–Wolf Coding. *IEEE Trans. Inf. Theory* **2015**, *61*, 2165–2190. [CrossRef]
14. Csiszár, I. Linear codes for sources and source networks: Error exponents, universal coding. *IEEE Trans. Inf. Theory* **1982**, *28*, 585–592. [CrossRef]
15. Csiszár, I.; Körner, J. Towards a general theory of source networks. *IEEE Trans. Inf. Theory* **1980**, *26*, 155–165. [CrossRef]
16. Kelly, B.G.; Wagner, A.B. Improved Source Coding Exponents via Witsenhausen's Rate. *IEEE Trans. Inf. Theory* **2011**, *57*, 5615–5633. [CrossRef]
17. Kelly, B.G.; Wagner, A.B. Reliability in Source Coding With Side Information. *IEEE Trans. Inf. Theory* **2012**, *58*, 5086–5111. [CrossRef]
18. Oohama, Y.; Han, T. Universal coding for the Slepian-Wolf data compression system and the strong converse theorem. *IEEE Trans. Inf. Theory* **1994**, *40*, 1908–1919. [CrossRef]
19. Liu, J.; Cuff, P.; Verdú, S. On α–decodability and α–likelihood decoder. In Proceedings of the 2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 3–6 October 2017.
20. Merhav, N. The generalized stochastic likelihood decoder: random coding and expurgated bounds. *IEEE Trans. Inf. Theory* **2017**, *63*, 5039–5051. See also a correction at *IEEE Trans. Inf. Theory* **2017**, *63*, 6827–6829. [CrossRef]
21. Csiszár, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*; Cambridge University Press: Cambridge, UK, 2011.
22. Tamir (Averbuch), R.; Merhav, N. The MMI decoder is asymptotically optimal for the typical random code and for the expurgated code. *arXiv* **2020**, arXiv:2007.12225.
23. Somekh-Baruch, A.; Scarlett, J.; Guillén i Fàbregas, A. Generalized Random Gilbert-Varshamov Codes. *IEEE Trans. Inf. Theory* **2019**, *65*, 3452–3469. [CrossRef]
24. Merhav, N. Statistical physics and information theory. *Found. Trends Commun. Inf. Theory* **2009**, *6*, pp. 1–212.