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

Ran Tamir (Averbuch) and Neri Merhav

May 17, 2020

The Andrew & Erna Viterbi Faculty of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 3200003, ISRAEL
{rans@campus, merhav@ee}.technion.ac.il

Abstract

Typical random codes (TRC) in a communication scenario of source coding with side information at the decoder is the main subject of this work. We derive the TRC error exponent for fixed-rate random binning and show that at relatively high rates, the TRC deviates significantly from the optimal code. We discuss the trade-offs between the error exponent and the excess–rate exponent for the typical random variable-rate code and characterize its optimal rate function. We show that the error exponent of the typical random variable-rate code may be strictly higher than in fixed-rate coding. We propose a new code, the semi-deterministic ensemble, which is a certain variant of the variable-rate code, and show that it dramatically improves upon the later: it is proved that the trade-off function between the error exponent and the excess–rate exponent for the typical random semi-deterministic code may be strictly higher than the same trade-off for the variable-rate code. Moreover, we show that the performance under optimal decoding can be attained also by different universal decoders: the minimum empirical entropy decoder and the generalized (stochastic) likelihood decoder with an empirical entropy metric.

Index Terms: Slepian–Wolf coding, fixed–rate coding, variable–rate coding, error exponent, excess–rate exponent, typical random code.

*This research was supported by the Israel Science Foundation (ISF) grant no. 137/18. This paper was presented in part at the 2019 IEEE Information Theory Workshop, Visby, Gotland, Sweden, 25-28 August, 2019.

1 Introduction

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

$$E_{\mathrm{r}}(R) = \lim_{n\to\infty} \left\{-\frac{1}{n}\log \mathbb{E}\left[P_{\mathrm{e}}(\mathcal{C}_n)\right]\right\}, \tag{1}$$

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 [12]

$$E_{\mathrm{trc}}(R) = \lim_{n\to\infty} \left\{-\frac{1}{n}\mathbb{E}\left[\log P_{\mathrm{e}}(\mathcal{C}_n)\right]\right\}. \tag{2}$$

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 very chosen code instead of directly considering the ensemble performance.

To the best of our knowledge, not much is known on TRCs. In [3], Barg and Forney considered TRCs with independently and identically distributed codewords as well as 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 [16] 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 [12], 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). Lately, Merhav has studied error exponents of TRCs for the colored Gaussian channel [13], typical random trellis codes [14], and a Lagrange–dual lower bound to the TRC exponent [15]. Large deviations around the TRC exponent was studied in [19].

While originally defined for pure channel coding [3], [12], [16], the notion of TRCs has natural analogues in other settings as well, like source coding with side information at the decoder [17]. Typical random Slepian–Wolf (SW) codes for both fixed–rate and variable–rate binning are 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 [8], and improved later on by the expurgated bound in [1] and [5]. Variable–rate (VR) SW coding received less attention in the literature; VR codes under average rate constraint have been studied in [4] 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 [20].

We begin our study by providing a single–letter expression for the error exponent of the TRC for FR random binning. In fact, since SW coding and ordinary channel coding are two sides of the same coin, the same techniques and proof ideas developed in [12] have been useful here too. While the optimal FR code at $R = \log|\mathcal{U}|$ (denoting by $|\mathcal{U}|$ the cardinality of the source alphabet) is a one–to–one mapping between the source sequence set $\mathcal{U}^n$ and the e^{nR} bins, the TRC deviates from this optimal code significantly, and has a finite error exponent as long as $R < 2\log|\mathcal{U}|$. This phenomenon has an intimate relation to the birthday paradox in probability theory. Nevertheless, this analysis of the FR code provides us a much better understanding of the random binning mechanism, and it paves the way to find other classes of codes which performs better than the FR code. This is one of the objectives of this work.

Moving forward, we discuss the trade-offs between the error exponent and the excess–rate exponent for a typical random VR code, similarly to [20], but with a different notion of the excess–rate event, which takes into account the available side information. We provide an expression for the optimal rate function that guarantees a required level for the error exponent of the typical random VR code. We show that upon relaxing the required excess–rate exponent, the resulted error exponent is strictly higher than in FR random binning. Furthermore, we find that for a class of information sources with a certain condition, the typical random VR code attains both exponentially vanishing error and excess–rate probabilities.

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 a new proposed VR ensemble – the semi–deterministic (SD) code ensemble. In this code ensemble, these source type classes are deterministically partitioned into the available bins in a one-to-one manner. As a consequence of this action, the error probability decreases dramatically. 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 and proved to be equal to one another in a few important special cases, that includes 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, in contrast to the two analogous results related to the FR and VR codes.

2. We show 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 many 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 [9], here we prove a similar result, but for two universal decoders (one stochastic and one deterministic) that share the same decoding metric.

3. We derive the trade-off functions between the error exponent and the excess–rate exponent for a typical random SD code and show that they may be strictly better than the trade-off functions for the ordinary VR code. In some cases, the excess–rate exponent for the SD code may reach a strictly positive plateau, while the excess–rate exponent for the ordinary VR code eventually reaches zero. Furthermore, under a strict requirement on the excess–rate exponent, which is equivalent to a FR code, the error exponent of the SD code reaches infinity at $R = \log |\mathcal{U}|$, while the TRC exponent of ordinary VR coding reaches infinity at $R = 2 \log |\mathcal{U}|$.

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 three coding techniques, the main objectives of this work, and we review some background. In Sections 4, 5, and 6, we provide and discuss the main results concerning the fixed–rate, the variable–rate, and the semi–deterministic ensembles, respectively.

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, $I_Q(U;V) = H_Q(U) - H_Q(U|V)$ is the mutual information between U and V, and so on. Logarithms are 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, $a_n \stackrel{\cdot}{\le} b_n$ means that $\limsup_{n\to\infty}(1/n)\log(a_n/b_n) \le 0$, and so on. The indicator function of an event $\mathcal{E}$ will be denoted by $\mathbb{1}\{\mathcal{E}\}$. The notation $[x]_+$ will stand for $\max\{0, x\}$.

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 $\mathcal{P}(\mathcal{U}) \stackrel{\triangle}{=} \bigcup_{n=1}^{\infty} \mathcal{P}_n(\mathcal{U})$. 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, $\hat{P}_{\boldsymbol{uv}}$

would be the joint empirical distribution of $(\boldsymbol{u}, \boldsymbol{v}) = \{(u_i, v_i)\}_{i=1}^n$, $\mathcal{T}(\hat{P}_{\boldsymbol{uv}})$ 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 $\boldsymbol{v}$, $\hat{H}_{\boldsymbol{uv}}(U|V)$ 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. We now distinguish between three different classes of codes:

1. Fixed–rate (FR) binning: When a given realization $\boldsymbol{u} = (u_1, \ldots, u_n) \in \mathcal{U}^n$, of the finite alphabet source vector $\boldsymbol{U}$, is fed into the system, it is encoded into one out of $M = e^{nR}$ bins, denoted by $\mathcal{B}(\boldsymbol{u})$, selected independently at random for every member of $\mathcal{U}^n$. Furthermore, the type index of $\boldsymbol{u}$ is also transmitted to the encoder, but it requires only a negligible extra rate when n is large enough. The entire binning code of source sequences of block–length n, i.e., the set $\{\mathcal{B}(\boldsymbol{u})\}_{\boldsymbol{u}\in\mathcal{U}^n}$, is denoted by $\mathcal{B}_n$. Here, $R > 0$ is referred to as the *binning rate*. 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 $\boldsymbol{u}$, using the bin index $\mathcal{B}(\boldsymbol{u})$, the type index $\mathcal{T}(\boldsymbol{u})$, and the SI vector $\boldsymbol{v} = (v_1, \ldots, v_n)$, according to

$$\hat{\boldsymbol{u}} = \arg \max_{\boldsymbol{u}' \in \mathcal{B}(\boldsymbol{u}) \cap \mathcal{T}(\boldsymbol{u})} P(\boldsymbol{u}', \boldsymbol{v}). \tag{3}$$

As in [11], [12], 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 SI sequence $\boldsymbol{v}$, according to the following posterior distribution

$$\mathbb{P}\left\{\hat{\boldsymbol{U}} = \boldsymbol{u}' \middle| \boldsymbol{v}, \mathcal{B}(\boldsymbol{u}), \mathcal{T}(\boldsymbol{u})\right\} = \frac{\exp\{n f(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}{\sum_{\tilde{\boldsymbol{u}} \in \mathcal{B}(\boldsymbol{u}) \cap \mathcal{T}(\boldsymbol{u})} \exp\{n f(\hat{P}_{\tilde{\boldsymbol{u}}\boldsymbol{v}})\}}, \tag{4}$$

where $\hat{P}_{\boldsymbol{uv}}$ 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 [5]). A more detailed discussion is given in [11].

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

$$P_{\text{e,FR}}(\mathcal{B}_n) = \sum_{\boldsymbol{u},\boldsymbol{v}} P(\boldsymbol{u},\boldsymbol{v}) \frac{\sum_{\boldsymbol{u}' \in \mathcal{B}(\boldsymbol{u}) \cap \mathcal{T}(\boldsymbol{u}), \boldsymbol{u}' \neq \boldsymbol{u}} \exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}{\sum_{\tilde{\boldsymbol{u}} \in \mathcal{B}(\boldsymbol{u}) \cap \mathcal{T}(\boldsymbol{u})} \exp\{nf(\hat{P}_{\tilde{\boldsymbol{u}}\boldsymbol{v}})\}}. \tag{5}$$

The random binning error exponent is defined by

$$E_{\text{r,FR}}(R) = \lim_{n\to\infty} \left\{ -\frac{\log \mathbb{E}[P_{\text{e,FR}}(\mathcal{B}_n)]}{n} \right\}. \tag{6}$$

We wish to derive a single–letter expression for the error exponent of the TRC, which is

$$E_{\text{trc,FR}}(R) = \lim_{n\to\infty} \left\{ -\frac{\mathbb{E}[\log P_{\text{e,FR}}(\mathcal{B}_n)]}{n} \right\}, \tag{7}$$

and then to study some of its basic properties.

2. Variable–rate (VR) binning: We assume that the coding rate is no longer fixed for every $\boldsymbol{u} \in \mathcal{U}^n$, but depends on its empirical distribution. Let us denote the rate function by $R(\hat{P}_{\boldsymbol{u}})$. In that manner, for every type $Q_U \in \mathcal{P}_n(\mathcal{U})$, all the source sequences in the type class $\mathcal{T}(Q_U)$ are randomly partitioned into $e^{nR(Q_U)}$ bins. Every source sequence is encoded by its bin index, prefixed by its type index.

The probability of error is defined similarly to (5) and will be denoted by $P_{\text{e,VR}}(\mathcal{B}_n)$. For a given rate function, the random binning error exponent and the error exponent of the typical random VR code are defined similarly as (6) and (7), respectively, but with $P_{\text{e,FR}}(\mathcal{B}_n)$ being replaced by $P_{\text{e,VR}}(\mathcal{B}_n)$. These will be denoted by $E_{\text{r,VR}}(R(\cdot))$ and $E_{\text{trc,VR}}(R(\cdot))$. One possibility to define the excess–rate probability, which has already extensively studied in [20], is given by $\mathbb{P}\{R(\hat{P}_{\boldsymbol{U}}) \geq R\}$, where R is some target rate. Due to the existence of the available side information at the decoder, it makes sense to require a target rate which depends on the pair $(\boldsymbol{u},\boldsymbol{v})$. Hence, we define the alternative excess–rate probability as $\mathbb{P}\{R(\hat{P}_{\boldsymbol{U}}) \geq \hat{H}_{\boldsymbol{U}\boldsymbol{V}}(U|V)+\Delta\}$, where $\Delta > 0$ is a redundancy threshold. Respectively, the excess–rate exponent is defined as

$$E_{\text{er,VR}}(R(\cdot),\Delta) = \lim_{n\to\infty} \left\{ -\frac{1}{n} \log \mathbb{P}\{R(\hat{P}_{\boldsymbol{U}}) \geq \hat{H}_{\boldsymbol{U}\boldsymbol{V}}(U|V) + \Delta\} \right\}. \tag{8}$$

Now, the main mission is to derive the trade–off between the error exponent and the excess–rate exponent for the typical random VR code, and furthermore, to characterize the optimal rate function that attains a prescribed value for the error exponent of the typical random VR code.

3. Semi–deterministic (SD) binning: This code ensemble is a refinement of the ordinary VR code, which is sensitive to the order between $H_Q(U)$ and $R(Q_U)$. For types with $H_Q(U) \geq R(Q_U)$, i.e., type classes which are exponentially larger than the amount of available bins, we just randomly assign each source sequence into one out of the $e^{nR(Q_U)}$ bins. Otherwise, for types with $H_Q(U) < R(Q_U)$, we deterministically order each member of $\mathcal{T}(Q_U)$ into a different bin, which provides a one–to–one mapping. This way, all type classes with $H_Q(U) < R(Q_U)$ will not affect the probability of error, which is now given by

$$P_{\text{e,SD}}(\mathcal{B}_n) = \sum_{\boldsymbol{u},\boldsymbol{v}} P(\boldsymbol{u},\boldsymbol{v}) \cdot \mathbb{1}\left\{\hat{H}_{\boldsymbol{u}}(U) \geq R(\hat{P}_{\boldsymbol{u}})\right\} \cdot \frac{\sum_{\boldsymbol{u}' \in \mathcal{B}(\boldsymbol{u}), \boldsymbol{u}' \neq \boldsymbol{u}} \exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}{\sum_{\tilde{\boldsymbol{u}} \in \mathcal{B}(\boldsymbol{u})} \exp\{nf(\hat{P}_{\tilde{\boldsymbol{u}}\boldsymbol{v}})\}}. \tag{9}$$

We derive the random binning exponent of this ensemble, which is denoted by $E_{\text{r,SD}}(R(\cdot))$, and compare it to $E_{\text{trc,SD}}(R(\cdot))$, the TRC exponent of the same ensemble. They are defined similarly as (6) and (7), respectively, but with $P_{\text{e,FR}}(\mathcal{B}_n)$ replaced by $P_{\text{e,SD}}(\mathcal{B}_n)$. As for the VR code ensemble defined above, we analyze the trade–off between the error and excess–rate exponents and compare it to the same trade–off in the VR code.

3.2 Background

In pure channel coding, Merhav [12] has derived a single–letter expression for the error exponent of the typical random fixed composition code,

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

In order to present the main result of [12], we define first a few quantities. Define

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

and

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

where $D(Q_{Y|X}\|W|Q_X)$ is the conditional divergence between $Q_{Y|X}$ and W, averaged by Q_X. Under the above defined quantities, the error exponent of the TRC is given by [12]

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

Returning to the SW model, several works have been written on error exponents for the FR and the VR codes. Here, we summarize only those results that are relevant to the current work. The random binning and expurgated bounds of the FR ensemble in the SW model are given respectively by [5, Sec. VI, Th. 2], [1, Appendix I, Th. 1]

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

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

where $E_{\text{r}}(Q_U, P_{V|U}, S)$ and $E_{\text{ex}}(Q_U, P_{V|U}, S)$ are respectively the random coding and expurgated bounds associated with the channel $P_{V|U}$ and a fixed composition code of rate S, all codewords of which belong to $\mathcal{T}(Q_U)$. The exponent function $E_{\text{r}}(Q_U, P_{V|U}, S)$ is given by

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

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

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

where

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

4 Fixed–Rate Binning

In order to characterize the error exponent of the TRC, we define the set $\mathcal{Q} = \{Q_{UU'}:\ Q_U = Q_{U'}\}$ and the quantities

$$\alpha(R, Q_U, Q_V) = \max_{\{Q_{\tilde{U}|V}:\ H_Q(\tilde{U}|V)\geq R,\ Q_{\tilde{U}}=Q_U\}} \{f(Q_{\tilde{U}V}) + H_Q(\tilde{U}|V)\} - R \tag{19}$$

and

$$\Gamma(Q_{UU'}, R) = \min_{Q_{V|UU'}} \left\{[\max\{f(Q_{UV}), \alpha(R, Q_U, Q_V)\} - f(Q_{U'V})]_+ - H_Q(V|U,U') - \mathbb{E}_Q[\log P(V|U)]\right\}. \tag{20}$$

Furthermore, define the following exponent function

$$E_{\rm trc}^{\rm fr}(R) = \min_{\{Q_{UU'}\in\mathcal{Q}:\ H_Q(U,U')\geq R\}} \left\{\Gamma(Q_{UU'},R) - H_Q(U,U') - \mathbb{E}_Q[\log P(U)] + R\right\}. \tag{21}$$

Then, the following theorem is proved in Appendix A.

Theorem 1. *The error exponent of the TRC in the FR ensemble is given by*

$$E_{trc,FR}(R) = E_{trc}^{fr}(R). \tag{22}$$

Discussion

Note that the expression of $E_{\rm trc}^{\rm fr}(R)$ strongly resembles the error exponent of the TRC in channel coding (13). The constraint $H_Q(U,U')\geq R$ in (21) is analogous to the constraint $I_Q(X;X')\leq 2R$ in the minimization of (13). The origin of $H_Q(U,U')\geq R$ is the following. Define

$$N(Q_{UU'}) = \sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\left\{\mathcal{B}(\boldsymbol{u}')=\mathcal{B}(\boldsymbol{u})\right\}, \tag{23}$$

which enumerate pairs of source sequences. Then, one of the main steps in the proof of Theorem 1 is deriving the high probability value of $N(Q_{UU'})$, which is 0 if $H_Q(U,U')<R$ (a relatively small set of source pair and relatively large number of bins) and $\exp\{n[H_Q(U,U')-R]\}$ for $H_Q(U,U')\geq R$ (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, like $N(Q_{UU'})$ in the current problem, this can no longer be done by the same techniques, and it requires more advanced tools (see, e.g., [12]–[15]).

As expected, one can easily prove that $E_{\rm trc}^{\rm fr}(R)$ is at least as large as $E_{\rm r}^{\rm fr}(R)$ (14). In [2, Proposition 3], we present a somewhat stronger result, asserting that at relatively high binning rates, $E_{\rm trc}^{\rm fr}(R)$ is *strictly* larger than $E_{\rm r}^{\rm fr}(R)$. Furthermore, we provide an explicit expression for the rate above which the functions differ.

We may also obtain some intuitive meaning of the term $\alpha(R,Q_U,Q_V)$ by considering the special case of MAP decoding, which corresponds to $f(Q)=\beta\mathbb{E}_Q[\log P(U,V)]$ for $\beta\to\infty$. An analogous explanation was given in [12] for the $\alpha(R, Q_Y)$ term of channel coding, by considering the special case of ML decoding. For very large β, $\alpha(R, Q_U, Q_V) \approx \beta a(R, Q_U, Q_V)$, where

$$a(R, Q_U, Q_V) = \max_{\{Q_{\tilde{U}|V}:\ H_Q(\tilde{U}|V) \geq R,\ Q_{\tilde{U}} = Q_U\}} \mathbb{E}_Q[\log P(\tilde{U}, V)]. \tag{24}$$

Following some basic manipulations, (21) now becomes

$$E_{\text{trc}}^{\text{MAP}}(R) = \min_{\mathcal{Q}(R)} \left\{ D(Q_{UV} \| P_{UV}) - H_Q(U'|U,V) + R \right\}, \tag{25}$$

with the set $\mathcal{Q}(R)$ given by

$$\begin{aligned}\mathcal{Q}(R) = \{Q_{UU'V}:\ & H_Q(U, U') \geq R,\ Q_{U'} = Q_U, \\ & \mathbb{E}_Q[\log P(U', V)] \geq \max\{\mathbb{E}_Q[\log P(U, V)], a(R, Q_U, Q_V)\}\}.\end{aligned} \tag{26}$$

The third constraint in $\mathcal{Q}(R)$ designates the event that an incorrect source sequence (represented by U') is assigned with a log–likelihood score larger than that of the correct source–sequence (represented by U) as well as those of all other source–sequences (represented by the term $a(R, Q_U, Q_V)$). The term $a(R, Q_U, Q_V)$ designates the typical value (with an extremely high probability) of the highest log–likelihood score among all the remaining incorrect source–sequences. A more comprehensive discussion on this point can be found in [12, Sec. 4].

Although the optimal binning code at $R = \log |\mathcal{U}|$ is a one–to–one mapping between the source sequence set $\mathcal{U}^n$ and the e^{nR} bins, the typical random binning deviates from this optimal code with a relatively high probability. We ask ourselves the following question: at different binning rates, what is the probability to draw a one–to–one code, such that each bin contains no more than one source sequence? Obviously, if $R < \log |\mathcal{U}|$, then it follow from the pigeonhole principle that this probability equals zero. Hence, in what follows, assume that $R \geq \log |\mathcal{U}|$. Consider the FR random binning mechanism of Subsection 3.1 and let $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_M$, $M = e^{nR}$, be the set of bins. Then, we would like to derive $\mathbb{P}\left\{\bigcap_{i=1}^{M} \{|\mathcal{B}_i| \leq 1\}\right\}$, for any $R \geq \log |\mathcal{U}|$. Consider the following

$$\mathbb{P}\left\{\bigcap_{i=1}^{M} \{|\mathcal{B}_i| \leq 1\}\right\} = 1 \cdot \left(1 - \frac{1}{M}\right) \cdot \left(1 - \frac{2}{M}\right) \cdot \ldots \cdot \left(1 - \frac{|\mathcal{U}|^n - 1}{M}\right) \tag{27}$$

$$\cong \exp\left\{-\frac{1}{M}\right\} \cdot \exp\left\{-\frac{2}{M}\right\} \cdot \ldots \cdot \exp\left\{-\frac{|\mathcal{U}|^n - 1}{M}\right\} \tag{28}$$

$$= \exp\left\{-\left(\frac{1}{M} + \frac{2}{M} + \ldots + \frac{|\mathcal{U}|^n - 1}{M}\right)\right\} \tag{29}$$

$$= \exp\left\{-\frac{(|\mathcal{U}|^n - 1)\cdot|\mathcal{U}|^n}{2M}\right\} \tag{30}$$

$$\doteq \exp\left\{-\frac{1}{2}\exp\{n\cdot(2\log|\mathcal{U}| - R)\}\right\}. \tag{31}$$

Hence, we find that a very sharp phase transition occurs at $R = 2\log|\mathcal{U}|$. At rates below $2\log|\mathcal{U}|$, the probability of drawing a one–to–one code converges to zero double–exponentially fast, while at rates above $2\log|\mathcal{U}|$, this probability tends to one. This is in agreement with the simple fact that for any $R > 2\log|\mathcal{U}|$, $E_{\rm trc}^{\rm fr}(R) = \infty$, which immediately follows from the constraint $H_Q(U,U') \geq R$ in (21). This phenomenon is intimately related to the birthday paradox (or the birthday problem) in probability theory. The birthday paradox concerns the probability that among n randomly chosen people, at least two will have the same birth date. More generally, it tells that if one chooses n times from a set of N equiprobable possibilities, then the first repetitions will typically occur when n is at the order of $\sqrt{N}$.

Recall that $[H(U|V), \log|\mathcal{U}|]$ is the relevant range of binning rates in SW coding. We expect $E_{\rm trc}^{\rm fr}(R)$ to improve upon $E_{\rm r}^{\rm fr}(R)$ at relatively high rates, just below $\log|\mathcal{U}|$, in analogy to channel coding, where there is a strict inequality between $E_{\rm trc}(R, Q_X)$ in (13) and the random coding bound in some range of low coding rates. Unfortunately, this is not the case in FR coding, since even for rates near $\log|\mathcal{U}|$, the event that exponentially many bins contain a few source vectors of the same type class is not negligible. While there exists a strong analogy between channel coding and SW coding (see, e.g., eqs. (14)–(15)), it seems to break down for TRCs. E.g., in the expurgated SW code, each bin contains exactly the same number of source vectors, while for the typical random FR code, the sizes of the bins fluctuates. In the extreme case of type classes with $H_Q(U) < R$, it follows by the discussion above that many bins may be populated by few source vectors, while many of them may still be empty. These type classes dominate the error event, and hence, this deficiency found in the FR code may be circumvented by treating them differently. This will be done in two steps. First, we pass from FR codes to VR codes. This improves upon FR codes, but VR codes still suffer from a similar deficiency. Second, we improve further by passing to SD codes.

5 Variable–Rate Binning

In order to characterize the error exponent of the TRCs of the VR ensemble defined in Subsection 3.1, we define first a few quantities. We define

$$\gamma(R(\cdot), Q_U, Q_V) = \max_{\left\{\substack{Q_{\tilde{U}|V}:\ Q_{\tilde{U}}=Q_U,\\ H_Q(\tilde{U}|V)\geq R(Q_{\tilde{U}})}\right\}} \{f(Q_{\tilde{U}V}) + H_Q(\tilde{U}|V)\} - R(Q_{\tilde{U}}) \tag{32}$$

and

$$\Psi(R(\cdot), Q_{UU'V}) = [\max\{f(Q_{UV}), \gamma(R(\cdot), Q_U, Q_V)\} - f(Q_{U'V})]_+ . \tag{33}$$

Furthermore, define

$$\Lambda(Q_{UU'}, R(Q_U)) = \min_{Q_{V|UU'}} \left\{\Psi(R(Q_U), Q_{UU'V}) - H_Q(V|U,U') - \mathbb{E}_Q[\log P(V|U)]\right\}, \tag{34}$$

and the following exponent function

$$E_{\text{e}}(R(\cdot)) = \min_{\left\{\substack{Q_{UU'}:\ Q_{U'}=Q_U,\\ H_Q(U,U')\geq R(Q_U)}\right\}} \left\{\Lambda(Q_{UU'}, R(Q_U)) - \mathbb{E}_Q[\log P(U)] - H_Q(U,U') + R(Q_U)\right\}. \tag{35}$$

Then, we have the following.

Theorem 2. *Let $R(\cdot)$ be any rate function. Then, for VR coding:*

$$E_{trc,VR}(R(\cdot)) = E_{\text{e}}(R(\cdot)). \tag{36}$$

Proof. The proof follows exactly the same lines as in the proof of Theorem 1 (Appendix A), except for one main modification: for a given source vector $\boldsymbol{u} \in \mathcal{T}(Q_U)$, the code has a fixed composition with a rate of $R(Q_U)$. ■

In order to characterize the excess–rate exponent, define the following exponent function:

$$E_{\text{er}}(R(\cdot), \Delta) = \min_{\{Q_{UV}:\ R(Q_U)\geq H_Q(U|V)+\Delta\}} D(Q_{UV}\|P_{UV}). \tag{37}$$

Then, we have the following.

Theorem 3. *Fix $\Delta > 0$ and let $R(\cdot)$ be any rate function. Then, for VR coding:*

$$E_{er,VR}(R(\cdot), \Delta) = E_{er}(R(\cdot), \Delta). \tag{38}$$

Proof. The excess–rate probability is given by:

$$\begin{aligned}
&\mathbb{P}\{R(\hat{P}_{\boldsymbol{U}}) \geq \hat{H}_{\boldsymbol{UV}}(U|V) + \Delta\} \\
&= \sum_{Q_{UV}} \mathbb{1}\{R(Q_U) \geq H_Q(U|V) + \Delta\} \cdot \mathbb{P}\{(\boldsymbol{U}, \boldsymbol{V}) \in \mathcal{T}(Q_{UV})\} && (39)\\
&\doteq \sum_{\{Q_{UV}:\ R(Q_U) \geq H_Q(U|V)+\Delta\}} \exp\{-nD(Q_{UV}\|P_{UV})\} && (40)\\
&\doteq \exp\left\{-n \cdot \min_{\{Q_{UV}:\ R(Q_U) \geq H_Q(U|V)+\Delta\}} D(Q_{UV}\|P_{UV})\right\}, && (41)
\end{aligned}$$

which proves the desired result. ■

5.1 Constrained Excess–Rate Exponent

Next, we would like to study the trade–off between the threshold Δ, the error exponent, and the excess–rate exponent. One possible way to explore this trade–off is to require the excess–rate exponent to exceed some value $\mathsf{E}_r > 0$, to solve $E_{\text{er}}(R(\cdot), \Delta) \geq \mathsf{E}_r$ for an upper bound on the rate function $R(Q_U)$, and then to substitute this upper bound back into the error exponent in (35) to give an expression for $E_{\text{e}}(\mathsf{E}_r, \Delta)$. As for the first step of this procedure, we have the following result, which is proved in Appendix C.

Theorem 4. *Let $\mathsf{E}_r > 0$ be fixed. Then, the requirement $E_{er}(R(\cdot), \Delta) \geq \mathsf{E}_r$ implies that*

$$R(Q_U) \leq J(Q_U, \mathsf{E}_r, \Delta) \stackrel{\triangle}{=} \min_{\{Q_{V|U}:\ D(Q_{UV}\|P_{UV}) \leq \mathsf{E}_r\}} \{H_Q(U|V) + \Delta\}. \tag{42}$$

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 [17], 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 [4]. Second, when E_r is extremely large, then each type class is encoded to $\exp\{n\Delta\}$ bins, which is equivalent to FR coding.

Upon substituting the optimal rate function of Theorem 4 back into (35) and using the fact that $E_{\text{e}}(\cdot)$ is monotonically increasing, we find that the trade–off function for the typical random VR code is given by

$$E_{\mathrm{e}}(\mathsf{E}_r,\Delta)=\min_{\left\{\substack{Q_{UU'}:\ Q_{U'}=Q_U,\\ H_Q(U,U')\geq J(Q_U)}\right\}}\left\{\Lambda(Q_{UU'},J(Q_U))-\mathbb{E}_Q[\log P(U)]-H_Q(U,U')+J(Q_U)\right\}, \quad (43)$$

where $J(Q_U)=J(Q_U,\mathsf{E}_r,\Delta)$. The dependence of $E_{\mathrm{e}}(\mathsf{E}_r,\Delta)$ 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 (42) and (43), but with the only difference that the constraint $D(Q_{UV}\|P_{UV})\leq\mathsf{E}_r$ is removed from (42). Furthermore, let $Q^*_U(\Delta)$ be the marginal distribution of $Q^*_{UU'}(\Delta)$. Now, when E_r is sufficiently large, i.e., when $\mathsf{E}_r\geq D(Q^*_U(\Delta)\times Q^*_{V|U}\|P_{UV})$, $E_{\mathrm{e}}(\mathsf{E}_r,\Delta)$ 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 $E_{\mathrm{e}}(\mathsf{E}_r,\Delta)$ is a monotonically non–increasing 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 $E_{\mathrm{e}}(0,\Delta)$ is the highest possible. In this case, $J(Q_U)=H_P(U|V)+\Delta$ and the constraint set in (43) becomes empty when $\Delta>2H_P(U)-H_P(U|V)$, and then $E_{\mathrm{e}}(0,\Delta)=\infty$.

5.2 Constrained Error Exponent

An alternative option to study the trade–off between the threshold Δ, the error exponent, and the excess–rate exponent is to require the error exponent to exceed some value $\mathsf{E}_e>0$, to solve $E_{\mathrm{e}}(R(\cdot))\geq\mathsf{E}_e$ in order to extract a lower bound on the rate function $R(Q_U)$, and then to substitute this lower bound back into the excess–rate exponent in (37) to provide an expression for $E_{\mathrm{er}}(\mathsf{E}_e,\Delta)$. The main drawback of this option stems from the relatively cumbersome expression of $E_{\mathrm{e}}(R(\cdot))$ in (35), which is given by a nested optimization problem, such that the extraction of the optimal rate function yields an unmanageable expression (like the lower bound in Eq. (E.16) in the proof of Theorem 6), and we have to compromise in some places in order to provide an analytically reasonable result. Thus, the lower bound on the rate function given in Theorem 6 below may not be the lowest possible for all E_e values.

As a matter of fact, any lower bound to the reliability function of VR coding can serve as a basis for exploiting a lower bound on the rate function. Recall that the exact error exponent of VR random binning is given by [20, eq. (34)]

$$E_{\rm r}^{\rm vr}(R(\cdot))=\min_{Q_{UV}}\left\{D(Q_{UV}\|P_{UV})+[R(Q_U)-H_Q(U|V)]_+\right\}. \tag{44}$$

Then, relying on this exponential error bound, the following bound on $R(Q_U)$ is given, which is proved in Appendix D.

Theorem 5. *Let* $\mathsf{E}_e>0$ *be fixed. Then, the requirement* $E_r^{vr}(R(\cdot))\geq\mathsf{E}_e$ *implies that*

$$R(Q_U)\geq G(Q_U,\mathsf{E}_e)\stackrel{\triangle}{=}\max_{\{Q_{V|U}:\ D(Q_{UV}\|P_{UV})\leq\mathsf{E}_e\}}\{\mathsf{E}_e+H_Q(U|V)-D(Q_{UV}\|P_{UV})\}. \tag{45}$$

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 (45) is empty, and $R(Q_U)$ can be as low as $-\infty$, which practically means that in this range, the entire type class $\mathcal{T}(Q_U)$ can be totally ignored, while still achieving $P_{\rm e}\approx e^{-n\mathsf{E}_e}$. 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

$$\max_{Q_{V|U}}\left\{H_Q(U|V)-D(Q_{UV}\|P_{UV})\right\}. \tag{46}$$

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 non–decreasing function of E_e. When $\mathsf{E}_e\geq D(Q_U\times Q_{V|U}^*\|P_{UV})$, the maximization in (45) reaches its unconstrained optimum, and $G(Q_U,\mathsf{E}_e)$ increases without bound in an affine fashion as $\mathsf{E}_e+H_{Q^*}(U|V)-D(Q_U\times Q_{V|U}^*\|P_{UV})$. Of course, the fact that the proposed lower bound of Theorem 5 is unbounded suggests that it cannot be optimal at high E_e values, since by allocating $\exp\{n\cdot 2\log_2|\mathcal{U}|\}$ bins to each source type class, one can attain an infinite error exponent, as we already saw before. Thus, in the sequel, we propose a lower bound on the rate function which is not optimal at relatively low E_e values (because of the compromise in our derivation), but improves upon the bound provided in Theorem 5 at relatively high E_e's. The following result is proved in Appendix E.

Theorem 6. *Let* $\mathsf{E}_e>0$ *be fixed. Then, for the matched likelihood decoder, the requirement* $E_{\rm e}(R(\cdot))\geq\mathsf{E}_e$ *implies that*

$$R(Q_U)\geq F(Q_U,\mathsf{E}_e)\stackrel{\triangle}{=}\max_{\{Q_{U'|U}:\ Q_{U'}=Q_U\}}\left\{H_Q(U,U')-[B_Q(U,U')-\mathsf{E}_e]_+\right\}, \tag{47}$$

where,

$$B_Q(U,U') = -\sum_{(u,u')\in\mathcal{U}^2} Q_{UU'}(u,u')\log\left(\sum_{v\in\mathcal{V}}\sqrt{P(u,v)\cdot P(u',v)}\right). \tag{48}$$

The dependence of $F(Q_U,\mathsf{E}_e)$ on E_e is as follows. Let $Q^*_{U'|U}$ be the maximizer in

$$\max_{\{Q_{U'|U}:\ Q_{U'}=Q_U\}}\left\{H_Q(U,U') - B_Q(U,U')\right\}. \tag{49}$$

Then, as long as $\mathsf{E}_e < B_{Q^*}(U,U')$, the clipping operator in (47) is inactive and $F(Q_U,\mathsf{E}_e)$ increases in an affine manner as $\mathsf{E}_e + H_{Q^*}(U,U') - B_{Q^*}(U,U')$. Whenever $\mathsf{E}_e \geq B_{Q^*}(U,U')$, the clipping operator in (47) becomes active, and $F(Q_U,\mathsf{E}_e)$ is given by

$$F(Q_U,\mathsf{E}_e) = \max_{\{Q_{U'|U}:\ Q_{U'}=Q_U,\ B_Q(U,U')\leq\mathsf{E}_e\}} H_Q(U,U'), \tag{50}$$

which is a concave monotonically non–decreasing function of E_e. When E_e is as large as $\max_{\{Q_{U'|U}:\ Q_{U'}=Q_U\}} B_Q(U,U')$, the constraint set in (50) no longer depends on E_e and $F(Q_U,\mathsf{E}_e)$ reaches a plateau at a level of $2H_Q(U)$. At this range of relatively high error exponents, each type class is encoded at rate which equals twice its empirical entropy. This result agrees with previous findings in this work, i.e., if the binning rate of each type class is double its exponential size, then the entire code will be a one–to–one mapping with a very high probability.

Comparing analytically the lower bounds of Theorems 5 and 6 seems to be complicated. Hence, we demonstrate some of the above discussed characteristics of $F(Q_U,\mathsf{E}_e)$ and $G(Q_U,\mathsf{E}_e)$ by referring to a numerical example. Consider the case of a double binary source with alphabets $\mathcal{U}=\mathcal{V}=\{0,1\}$, and joint probabilities given by $P_{UV}(0,0)=0.8$, $P_{UV}(0,1)=0.05$, $P_{UV}(1,0)=0$, and $P_{UV}(1,1)=0.15$. Fig. 1 displays the two lower bounds on $R(Q_U)$, $F(Q_U,\mathsf{E}_e)$ and $G(Q_U,\mathsf{E}_e)$, for this source and the specific type $Q_U(0)=Q_U(1)=1/2$.

As can be seen, the gap between these two bounds is rather considerable at relatively high E_e values, which means that the typical VR codes require significantly lower rates in order to achieve the same target error exponent. Furthermore, $F(Q_U,\mathsf{E}_e)$ reaches a plateau while $G(Q_U,\mathsf{E}_e)$ grows without bound. In the range where both of the bounds are affine, it seems that they are equal, a fact we conjecture to be true in general, although we were not able to assure an equality between (46) and (49). We also conjecture that the gap in the range of low E_e values is due to the compromise we did in the proof of Theorem 6. This example is quite representative in the sense that other examples yielded qualitatively similar results.

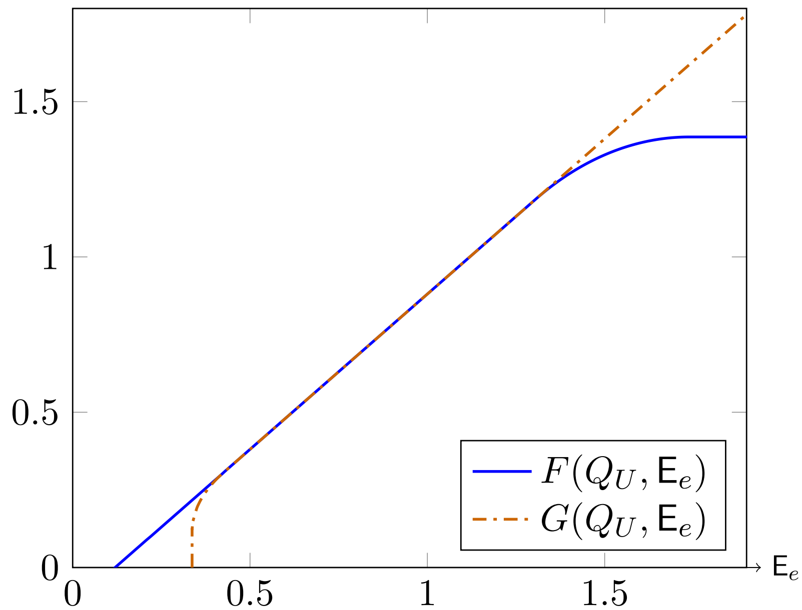

Figure 1: Lower bounds on $R(Q_U)$ for $Q_U = (\frac{1}{2}, \frac{1}{2})$.

In order to attain a target error exponent E_e, both $F(Q_U, \mathsf{E}_e)$ and $G(Q_U, \mathsf{E}_e)$ are legitimate lower bounds on $R(Q_U)$, hence also the minimum between them. Let us denote this minimum by $\Omega(Q_U, \mathsf{E}_e)$. Upon substituting $\Omega(Q_U, \mathsf{E}_e)$ back into (37) and using the fact that $E_{\mathrm{r}}(\cdot, \Delta)$ is monotonically non–increasing, we find that the trade–off function is given by

$$E_{\mathrm{er}}(\mathsf{E}_e, \Delta) = \min_{\{Q_{UV}\colon\ \Omega(Q_U, \mathsf{E}_e) \geq H_Q(U|V) + \Delta\}} D(Q_{UV} \| P_{UV}). \tag{51}$$

Since $\Omega(Q_U, \mathsf{E}_e)$ is monotonically non–decreasing in E_e for every Q_U, $E_{\mathrm{er}}(\mathsf{E}_e, \Delta)$ is monotonically non–increasing in E_e, which is not very surprising. The dependence of $E_{\mathrm{er}}(\mathsf{E}_e, \Delta)$ on E_e and Δ is as follows. At $\mathsf{E}_e = 0$, recall that $\Omega(Q_U, 0) = -\infty$ for any $Q_U \neq P_U$ while $\Omega(P_U, 0) = H_P(U|V)$. Thus, $E_{\mathrm{r}}(0, \Delta) = 0$ as long as $\Delta = 0$, and it follows from the monotonicity that $E_{\mathrm{er}}(\mathsf{E}_e, 0) = 0$ everywhere. Otherwise, if $\Delta > 0$, $\{Q_{UV} : \Omega(Q_U, \mathsf{E}_e) \geq H_Q(U|V) + \Delta\}$ is empty as long as $\mathsf{E}_e < \mathsf{E}_e^*(\Delta)$, and then $E_{\mathrm{er}}(\mathsf{E}_e, \Delta) = \infty$ in this range[1]. In the other extreme of a very large E_e, $\Omega(Q_U, \mathsf{E}_e)$ reaches a plateau at a level of $2H_Q(U)$, due to the behavior of $F(Q_U, \mathsf{E}_e)$. Then, if $\Delta \leq 2H_P(U) - H_P(U|V)$, $E_{\mathrm{er}}(\mathsf{E}_e, \Delta)$ reaches zero for a sufficiently large E_e. Else, if $\Delta > 2H_P(U) - H_P(U|V)$, $E_{\mathrm{er}}(\mathsf{E}_e, \Delta)$ reaches a strictly positive plateau, given by $\min_{\{Q_{UV}\colon\ 2H_Q(U) \geq H_Q(U|V) + \Delta\}} D(Q_{UV} \| P_{UV})$, which is a monotonically non–decreasing function of Δ. Particularly, it means that in this range, the typical random VR code attains both an

[1] An expression for $\mathsf{E}_e^*(\Delta)$ can be found by solving $\max_{Q_{UV}} \{\Omega(Q_U, \mathsf{E}_e) - H_Q(U|V)\} \leq \Delta$.

exponentially vanishing excess–rate probability as well as $P_{\mathrm{e}} \approx 0$.

We demonstrate some of the characteristics of $E_{\mathrm{er}}(\mathsf{E}_e, \Delta)$ by referring to a numerical example. Consider again the case of a double binary source, now with joint probabilities given by $P_{UV}(0,0) = 0.899$, $P_{UV}(0,1) = 0.001$, $P_{UV}(1,0) = 0$, and $P_{UV}(1,1) = 0.1$. Graphs of the trade–off function $E_{\mathrm{er}}(\mathsf{E}_e, \Delta)$ as a function of E_e for different values of Δ are presented in Fig. 2. As long as $\Delta \leq 2H_P(U) - H_P(U|V) \approx 0.644$, $E_{\mathrm{er}}(\mathsf{E}_e, \Delta)$ reaches zero, as can be seen for the red solid curve in Fig. 2, which is calculated for $\Delta = 0.64$. For $\Delta > 2H_P(U) - H_P(U|V)$, the curves reach a strictly positive plateau, which increases as Δ grows. Furthermore, each curve equals infinity for sufficiently small E_e values.

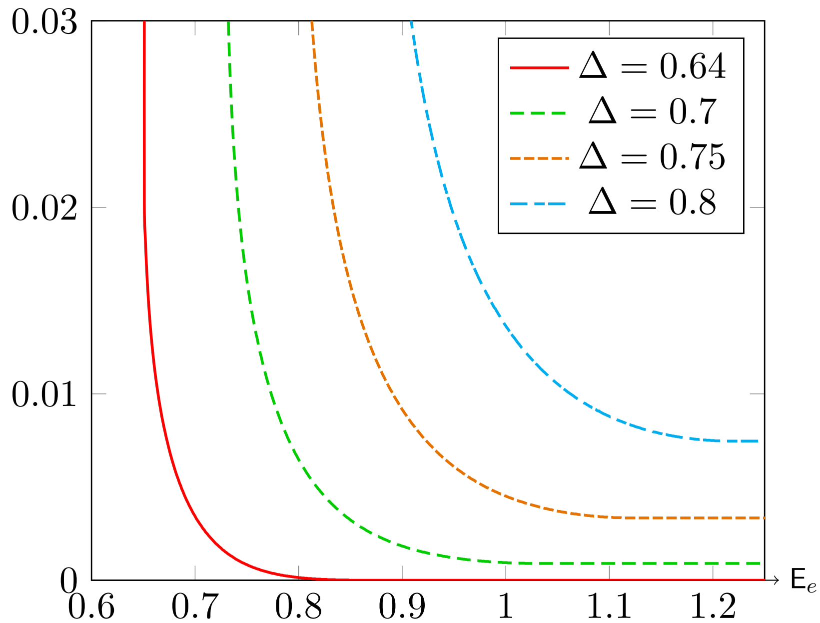

Figure 2: Graphs of $E_{\mathrm{er}}(\mathsf{E}_e, \Delta)$ as a function of E_e for different Δ values.

6 Semi–Deterministic Binning

In order to present some of the results in this section, we make a few more definitions. The minimum conditional entropy (MCE) decoder estimates $\boldsymbol{u}$, using the bin index $\mathcal{B}(\boldsymbol{u})$ and the SI vector $\boldsymbol{v}$, according to

$$\hat{\boldsymbol{u}} = \arg \min_{\boldsymbol{u}' \in \mathcal{B}(\boldsymbol{u})} \hat{H}_{\boldsymbol{u}'\boldsymbol{v}}(U|V). \tag{52}$$

The stochastic conditional entropy (SCE) decoder estimates $\boldsymbol{u}$ according to the following posterior distribution

$$\mathbb{P}\left\{\hat{\boldsymbol{U}}=\boldsymbol{u}'\middle|\boldsymbol{v},\mathcal{B}(\boldsymbol{u})\right\}=\frac{\exp\{-n\hat{H}_{\boldsymbol{u}'\boldsymbol{v}}(U|V)\}}{\sum_{\tilde{\boldsymbol{u}}\in\mathcal{B}(\boldsymbol{u})}\exp\{-n\hat{H}_{\tilde{\boldsymbol{u}}\boldsymbol{v}}(U|V)\}}. \tag{53}$$

6.1 Error Exponents and Universal Decoding

First, we provide random binning error exponents, which generalizes (44) to this new defined ensemble. Define the expression

$$E(Q_{UV})=\min_{Q_{U'|V}}\left[R(Q_U)-H_Q(U'|V)+[f(Q_{UV})-f(Q_{U'V})]_+\right]_+ \tag{54}$$

and the exponent functions:

$$E_{\mathrm{r,GLD}}^{\mathrm{sd}}(R(\cdot))=\min_{\{Q_{UV}:\ H_Q(U)\geq R(Q_U)\}}\left\{D(Q_{UV}\|P_{UV})+E(Q_{UV})\right\}, \tag{55}$$

and

$$E_{\mathrm{r,MAP}}^{\mathrm{sd}}(R(\cdot))=\min_{\{Q_{UV}:\ H_Q(U)\geq R(Q_U)\}}\left\{D(Q_{UV}\|P_{UV})+[R(Q_U)-H_Q(U|V)]_+\right\}. \tag{56}$$

Then, consider the following result, which is proved in Appendix F.

Theorem 7. *Let $R(\cdot)$ be any rate function. Then, for the SD code, $E_{r,SD}(R(\cdot))$ is given by*

1. *$E_{r,GLD}^{sd}(R(\cdot))$ for the GLD,*
2. *$E_{r,MAP}^{sd}(R(\cdot))$ for the MAP and the MCE decoders.*

Next, we provide a single–letter expression for the error exponent of the TRCs in this ensemble. Let $\gamma(R(\cdot),Q_U,Q_V)$, $\Psi(R(\cdot),Q_{UU'V})$, and $\Lambda(Q_{UU'},R(Q_U))$ be defined as in (32)–(34). Define the following exponent function

$$E_{\mathrm{trc,GLD}}^{\mathrm{sd}}(R(\cdot))=\min_{\left\{\substack{Q_{UU'}:\ Q_{U'}=Q_U,\\ H_Q(U)\geq R(Q_U)}\right\}}\left\{\Lambda(Q_{UU'},R(Q_U))-\mathbb{E}_Q[\log P(U)]-H_Q(U,U')+R(Q_U)\right\}. \tag{57}$$

Then, we have the following.

Theorem 8. *Let $R(\cdot)$ be any rate function. Then, for the SD code:*

$$E_{trc,SD}(R(\cdot))=E_{trc,GLD}^{sd}(R(\cdot)). \tag{58}$$

Proof. The proof follows exactly the same lines as the modified proof of Theorem 1 (Appendix A) for Theorem 2, except for one main modification: when we introduce the type class enumerator 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 (9). 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. ∎

It is possible to make an analytic comparison between (55) and (57) in the special cases of the matched/mismatched likelihood decoder and the MCE decoder. We have the following result, the proof of which is given in Appendix G.

Theorem 9. *Consider the SD code 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 $\beta > 0$,*

$$E^{sd}_{trc,GLD}(R(\cdot)) = E^{sd}_{r,GLD}(R(\cdot)). \tag{59}$$

2. *For the MCE decoder,*

$$E^{sd}_{trc,MCE}(R(\cdot)) = E^{sd}_{r,MCE}(R(\cdot)). \tag{60}$$

This result is quite surprising at first glance, since one expect the error exponent of the TRC to be strictly better than the random binning error exponent. We conjecture that this phenomenon is due to the fact that now, 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. In the first place, 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, like 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 9 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 [6, Exercise 3.1.6], [20] that the random binning error exponents under optimal MAP decoding in both the FR and VR codes, given by (14) and (44), respectively, are also attained by the MCE decoder. Furthermore, a similar result for the SD code has been proved here in Theorem 7. A natural question arises whether the error exponent of the TRC is also universally attainable, or only a fraction of it. The following result, which is proved in Appendix H, provide a positive answer to this question.

Theorem 10. *Consider the SD code 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.,*

$$E^{sd}_{trc,MAP}(R(\cdot)) = E^{sd}_{trc,MCE}(R(\cdot)) = E^{sd}_{trc,SCE}(R(\cdot)). \quad (61)$$

The result of Theorem 10 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, that must not even be deterministic. We conjecture that similar results also hold for the FR and the VR ensembles, at least for the MCE decoder, although not being able to provide a proper proof. Comparing to channel coding, numerical evidences shows 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, the GLD which is based on an empirical mutual information metric attains a strictly lower exponent.

6.2 Optimal Trade-off Functions

Following the first point of Theorem 9, let us denote the error exponent of the TRC under MAP decoding by $E^{\mathrm{sd}}_{\mathrm{e}}(\cdot)$. Upon substituting the optimal rate function of Theorem 4 back into (56) and (57) and using the fact that $E^{\mathrm{sd}}_{\mathrm{e}}(\cdot)$ is monotonically increasing, we find that the trade–off function is given by

$$E^{\mathrm{sd}}_{\mathrm{e}}(\mathsf{E}_r, \Delta) = \min_{\{Q_{UV}:\ H_Q(U) \geq J(Q_U)\}} \left\{ D(Q_{UV} \| P_{UV}) + [J(Q_U) - H_Q(U|V)]_+ \right\}, \quad (62)$$

or, alternatively,

$$E^{\mathrm{sd}}_{\mathrm{e}}(\mathsf{E}_r, \Delta) = \min_{\left\{\substack{Q_{UU'}:\ Q_{U'}=Q_U, \\ H_Q(U) \geq J(Q_U)}\right\}} \left\{ \Lambda(Q_{UU'}, J(Q_U)) - \mathbb{E}_Q[\log P(U)] - H_Q(U,U') + J(Q_U) \right\}, \quad (63)$$

where $J(Q_U) = J(Q_U, \mathsf{E}_r, \Delta)$ is given in (42).

The qualitative dependencies of $E^{\mathrm{sd}}_{\mathrm{e}}(\mathsf{E}_r, \Delta)$ and $E_{\mathrm{e}}(\mathsf{E}_r, \Delta)$ (43) on E_r are very similar. Quantitatively, the constraint set in (63) is a subset of the constraint set in (43), which implies that $E^{\mathrm{sd}}_{\mathrm{e}}(\mathsf{E}_r, \Delta) \geq E_{\mathrm{e}}(\mathsf{E}_r, \Delta)$. Referring to the dependence on Δ in the extreme case of a very large

E_r, for which $J(Q_U, \mathsf{E}_r, \Delta) = \Delta$, we find that the difference between $E_{\rm e}^{\rm sd}(\mathsf{E}_r, \Delta)$ and $E_{\rm e}(\mathsf{E}_r, \Delta)$ is quite dramatic; while $E_{\rm e}(\mathsf{E}_r, \Delta)$ is finite as long as $\Delta \le 2\log|\mathcal{U}|$, $E_{\rm e}^{\rm sd}(\mathsf{E}_r, \Delta) = \infty$ for any $\Delta > \log|\mathcal{U}|$. We also demonstrate the difference between $E_{\rm e}^{\rm sd}(\mathsf{E}_r, \Delta)$ and $E_{\rm e}(\mathsf{E}_r, \Delta)$ by referring to a numerical example. Consider once more the case of a double binary source, now with joint probabilities given by $P_{UV}(0,0) = 0.75$, $P_{UV}(0,1) = 0.1$, $P_{UV}(1,0) = 0$, and $P_{UV}(1,1) = 0.15$. Graphs of the trade–off functions $E_{\rm e}^{\rm sd}(\mathsf{E}_r, \Delta)$ and $E_{\rm e}(\mathsf{E}_r, \Delta)$ as a function of Δ in the extreme case of $\mathsf{E}_r = \infty$ are presented in Fig. 3. As can be seen in Fig. 3, $E_{\rm e}^{\rm sd}(\mathsf{E}_r, \Delta)$ and $E_{\rm e}(\mathsf{E}_r, \Delta)$ reach infinity at target thresholds $\log 2 \approx 0.693$ and $2\log 2 \approx 1.386$, respectively.

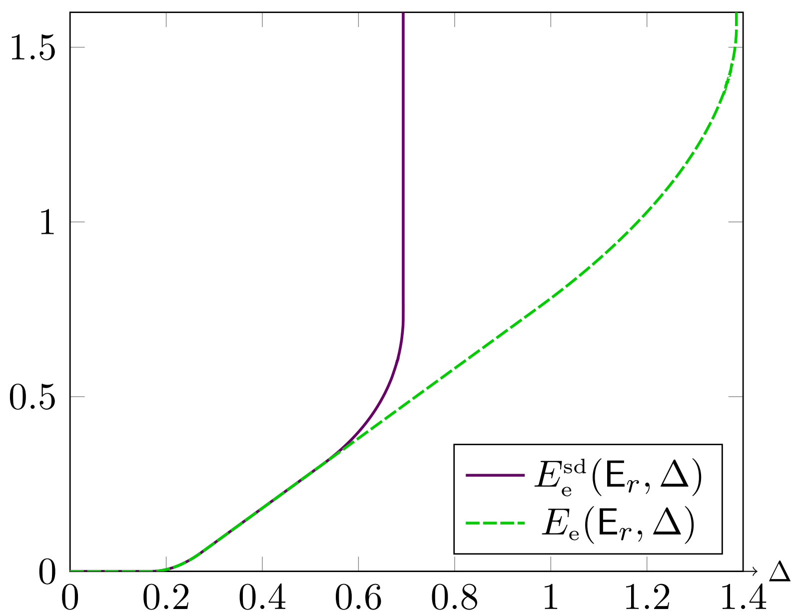

Figure 3: Graphs of different trade–off functions as a function of Δ, when E_r is extremely large.

We also provide an expression for the opposite trade-off function, where the error exponent is constrained to a given threshold. As in Subsection 5.2, the first step will be to solve $E_{\rm e}^{\rm sd}(R(\cdot)) \ge \mathsf{E}_e$ in order to extract a lower bound on the rate function $R(Q_U)$, and then to substitute this lower bound back into the excess–rate exponent in (37) to provide an expression for $E_{\rm er}^{\rm sd}(\mathsf{E}_e, \Delta)$.

Then, relying on the bound of (56), the following bound on $R(Q_U)$ is given, which is proved in Appendix I.

Theorem 11. *Let $\mathsf{E}_e > 0$ be fixed. Then, the requirement $E_e^{sd}(R(\cdot)) \ge \mathsf{E}_e$ implies that*

$$R(Q_U) \ge K(Q_U, \mathsf{E}_e) \stackrel{\triangle}{=} \max_{\{Q_{V|U}:\ D(Q_{UV}\|P_{UV}) \le \mathsf{E}_e\}} \min\{H_Q(U), H_Q(U|V) + \mathsf{E}_e - D(Q_{UV}\|P_{UV})\}. \tag{64}$$

When substituting $K(Q_U, \mathsf{E}_e)$ back into (37) and using the fact that $E_{\rm r}(\cdot, \Delta)$ is monotonically non–increasing, we find that the trade–off function is given by

$$E_{\rm er}^{\rm sd}(\mathsf{E}_e, \Delta) = \min_{\{Q_{UV}:\ K(Q_U,\mathsf{E}_e)\geq H_Q(U|V)+\Delta\}} D(Q_{UV}\|P_{UV}). \tag{65}$$

The qualitative dependencies of $E_{\rm er}^{\rm sd}(\mathsf{E}_e, \Delta)$ and $E_{\rm er}(\mathsf{E}_e, \Delta)$ (51) on E_e and Δ are very similar, hence we only mention some quantitative points where they differ from one another. In the extreme case of a very large E_e, $K(Q_U, \mathsf{E}_e)$ reaches a plateau at a level of $H_Q(U)$. Then, if $\Delta \leq H_P(U) - H_P(U|V) = I_P(U;V)$, $E_{\rm er}^{\rm sd}(\mathsf{E}_e, \Delta)$ reaches zero for a sufficiently large E_e. Else, if $\Delta > I_P(U;V)$, $E_{\rm er}^{\rm sd}(\mathsf{E}_e, \Delta)$ reaches a strictly positive plateau, given by

$$\min_{\{Q_{UV}:\ I_Q(U;V)\geq\Delta\}} D(Q_{UV}\|P_{UV}), \tag{66}$$

which is a monotonically non–decreasing function of Δ. We conclude that the trade-off for the typical random SD code is strictly higher than the trade-off for the ordinary VR code. Moreover, as long as $\Delta \in [I_P(U;V), I_P(U;V) + H_P(U)]$, $E_{\rm er}^{\rm sd}(\mathsf{E}_e, \Delta)$ reaches a strictly positive plateau, while $E_{\rm er}(\mathsf{E}_e, \Delta)$ reaches zero., i.e., in this range, both codes attain $P_{\rm e} \approx 0$, but only the typical random SD code achieves an exponentially vanishing overflow probability.

We demonstrate the difference between $E_{\rm er}^{\rm sd}(\mathsf{E}_e, \Delta)$ and $E_{\rm er}(\mathsf{E}_e, \Delta)$ by referring to the same numerical example as in Fig. 3. Graphs of the trade–off functions $E_{\rm er}^{\rm sd}(\mathsf{E}_e, \Delta)$ and $E_{\rm er}(\mathsf{E}_e, \Delta)$ as a function of E_e in the case of $\Delta = 0.3$ are presented in Fig. 4. This value of Δ is chosen to be between the thresholds $I_P(U;V) \approx 0.254$ and $I_P(U;V) + H_P(U) \approx 0.677$. As can be seen in Fig. 4, $E_{\rm er}^{\rm sd}(\mathsf{E}_e, \Delta)$ reaches a strictly positive plateau, while $E_{\rm er}(\mathsf{E}_e, \Delta)$ reaches zero.

It may also be interesting to make a connection to the expurgated bound of the FR code in the SW model, which is given by (15). Making an analytic comparison between $E_{\rm ex}^{\rm fr}(R)$ and $E_{\rm e}^{\rm sd}(\infty, \Delta)$ is rather difficult. Thus, we examined these two exponent functions via a similar numerical example as in Fig. 3. 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 $E_{\rm ex}^{\rm fr}(R)$ and $E_{\rm e}^{\rm sd}(\infty, R)$ are presented in Fig. 5.

As can be seen in Fig. 5, both $E_{\rm ex}^{\rm fr}(R)$ and $E_{\rm e}^{\rm sd}(\infty, R)$ reach infinity at rates $\log 2 \approx 0.693$. For relatively high binning rates, $E_{\rm ex}^{\rm fr}(R)$ is strictly higher than $E_{\rm e}^{\rm sd}(\infty, R)$, 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

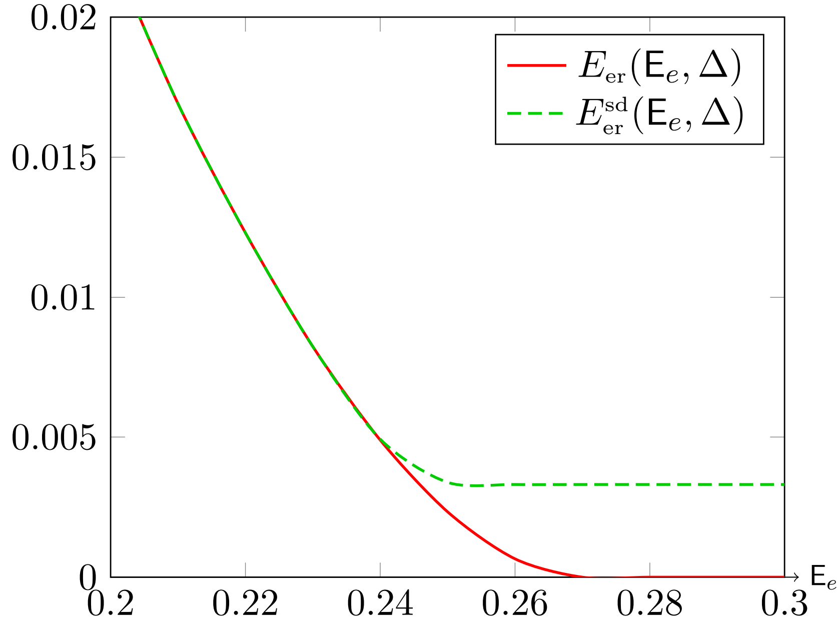

Figure 4: Graphs of $E_{\text{er}}^{\text{sd}}(\mathsf{E}_e, \Delta)$ and $E_{\text{er}}(\mathsf{E}_e, \Delta)$ as a function of E_e for $\Delta = 0.3$.

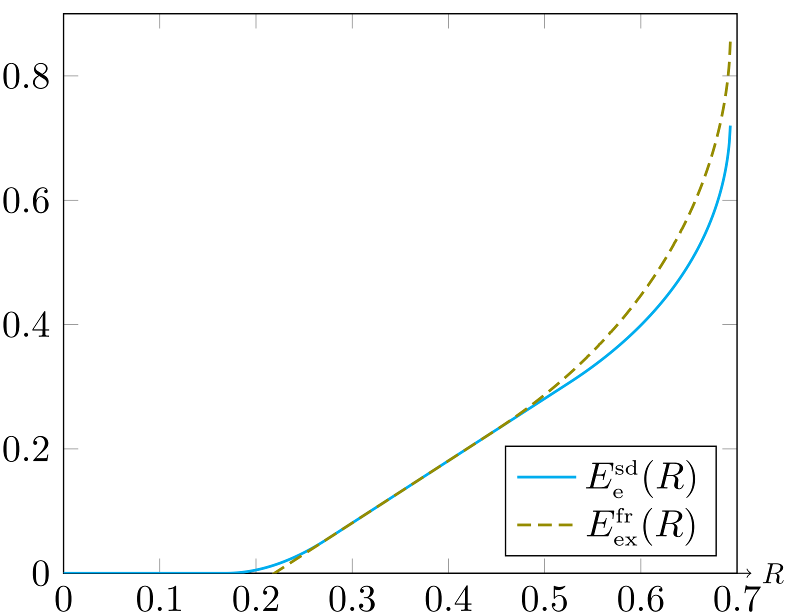

Figure 5: Graphs of the functions $E_{\text{ex}}^{\text{fr}}(R)$ and $E_{\text{e}}^{\text{sd}}(\infty, R)$.

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 opposite side, in the expurgated SW code [5], 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 (23), 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 [18] 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 $E_{\text{ex}}^{\text{fr}}(R)$ of (15). Since the gap between $E_{\text{ex}}^{\text{fr}}(R)$ and $E_{\text{e}}^{\text{sd}}(\infty, R)$ is not necessarily very significant, as can be seen in Fig. 5, 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 $E_{\text{e}}^{\text{sd}}(\infty, R)$ decays exponentially fast to zero, in analogy to the result in pure channel coding [19].

Appendix A

Proof of Theorem 1

Lower Bound on the Error Exponent

Our starting point is the following identity

$$\mathbb{E}[\log P_{\text{e,FR}}(\mathcal{B}_n)] = \lim_{\rho\to\infty} \log\left(\mathbb{E}[P_{\text{e,FR}}(\mathcal{B}_n)]^{1/\rho}\right)^{\rho} = \lim_{\rho\to\infty} \rho \log\left(\mathbb{E}[P_{\text{e,FR}}(\mathcal{B}_n)]^{1/\rho}\right). \tag{A.1}$$

Following the error probability in (5), we have that

$$P_{\text{e,FR}}(\mathcal{B}_n) = \sum_{\boldsymbol{u},\boldsymbol{v}} P(\boldsymbol{u},\boldsymbol{v}) \sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u})\cap\mathcal{T}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}} \frac{\exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}{\sum_{\tilde{\boldsymbol{u}}\in\mathcal{B}(\boldsymbol{u})\cap\mathcal{T}(\boldsymbol{u})} \exp\{nf(\hat{P}_{\tilde{\boldsymbol{u}}\boldsymbol{v}})\}} \tag{A.2}$$

$$= \sum_{\boldsymbol{u},\boldsymbol{v}} P(\boldsymbol{u},\boldsymbol{v}) \sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u})\cap\mathcal{T}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}} \frac{\exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}{\exp\{nf(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})\} + Z_{\boldsymbol{u}}(\boldsymbol{v})}, \tag{A.3}$$

where,

$$Z_{\boldsymbol{u}}(\boldsymbol{v}) = \sum_{\tilde{\boldsymbol{u}}\in\mathcal{B}(\boldsymbol{u})\cap\mathcal{T}(\boldsymbol{u}),\tilde{\boldsymbol{u}}\neq\boldsymbol{u}} \exp\{nf(\hat{P}_{\tilde{\boldsymbol{u}}\boldsymbol{v}})\}. \tag{A.4}$$

We have the following large deviations result concerning $Z_{\boldsymbol{u}}(\boldsymbol{v})$ (proved in Appendix B):

Lemma 1. *Let $\epsilon > 0$ be arbitrarily small. Then,*

$$\mathbb{P}\bigcup_{\boldsymbol{u},\boldsymbol{v}}\left\{Z_{\boldsymbol{u}}(\boldsymbol{v}) \leq \exp\{n\alpha(R+\epsilon, \hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}})\}\right\} \leq |\mathcal{U}\times\mathcal{V}|^n \cdot \exp\{-e^{n\epsilon}+n\epsilon+1\}. \quad \text{(A.5)}$$

Recall that $\mathcal{Q} = \{Q_{UU'}:\ Q_U = Q_{U'}\}$. Then,

$$\begin{aligned}
&\mathbb{E}\left\{[P_{\mathrm{e,FR}}(\mathcal{B}_n)]^{1/\rho}\right\} \\
&= \mathbb{E}\left\{\left[\sum_{\boldsymbol{u},\boldsymbol{v}} P(\boldsymbol{u},\boldsymbol{v}) \sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u})\cap\mathcal{T}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}} \frac{\exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}{\exp\{nf(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})\}+Z_{\boldsymbol{u}}(\boldsymbol{v})}\right]^{1/\rho}\right\} && \text{(A.6)} \\
&\stackrel{\cdot}{\leq} \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}) \right.\right. \\
&\qquad \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\} && \text{(A.7)} \\
&\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}) \right.\right. \\
&\qquad \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\} && \text{(A.8)} \\
&\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\exp\left\{-n\cdot\Gamma(\hat{P}_{\boldsymbol{u}\boldsymbol{u}'},R+\epsilon)\right\}\right]^{1/\rho}\right\} && \text{(A.9)} \\
&= \mathbb{E}\left\{\left[\sum_{Q_{UU'}\in\mathcal{Q}} N(Q_{UU'})\cdot e^{n\mathbb{E}_Q[\log P(U)]}\cdot\exp\left\{-n\cdot\Gamma(Q_{UU'},R+\epsilon)\right\}\right]^{1/\rho}\right\} && \text{(A.10)} \\
&\doteq \sum_{Q_{UU'}\in\mathcal{Q}} \mathbb{E}\left\{[N(Q_{UU'})]^{1/\rho}\right\}\cdot e^{n(\mathbb{E}_Q[\log P(U)])/\rho}\cdot\exp\left\{-n\cdot\Gamma(Q_{UU'},R+\epsilon)/\rho\right\}, && \text{(A.11)}
\end{aligned}$$

where (A.7) is due to Lemma 1, (A.9) is thanks to the method of types and the definition of $\Gamma(Q_{UU'},R)$ in (20), and in (A.10) we used the definition of $N(Q_{UU'})$ in (23). Therefore, our next task is to evaluate the $1/\rho$–th moment of $N(Q_{UU'})$. Let us define

$$N_{\boldsymbol{u}}(Q_{U'|U}) = \sum_{\boldsymbol{u}'\in\mathcal{T}(Q_{U'|U}|\boldsymbol{u})} \mathbb{1}\left\{\mathcal{B}(\boldsymbol{u}')=\mathcal{B}(\boldsymbol{u})\right\}. \quad \text{(A.12)}$$

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

$$\mathbb{E}\left\{[N(Q_{UU'})]^{1/\rho}\right\} = \mathbb{E}\left\{\left[\sum_{\boldsymbol{u}\in\mathcal{T}(Q_U)} N_{\boldsymbol{u}}(Q_{U'|U})\right]^{1/\rho}\right\} \quad \text{(A.13)}$$

$$= \mathbb{E}\left\{\left(\left[\sum_{\boldsymbol{u}\in\mathcal{T}(Q_U)} N_{\boldsymbol{u}}(Q_{U'|U})\right]^{1/s}\right)^{s/\rho}\right\} \tag{A.14}$$

$$\leq \mathbb{E}\left\{\left(\sum_{\boldsymbol{u}\in\mathcal{T}(Q_U)} \left[N_{\boldsymbol{u}}(Q_{U'|U})\right]^{1/s}\right)^{s/\rho}\right\} \tag{A.15}$$

$$\leq \left(\mathbb{E}\left\{\sum_{\boldsymbol{u}\in\mathcal{T}(Q_U)} \left[N_{\boldsymbol{u}}(Q_{U'|U})\right]^{1/s}\right\}\right)^{s/\rho} \tag{A.16}$$

$$= \left(\sum_{\boldsymbol{u}\in\mathcal{T}(Q_U)} \mathbb{E}\left\{\left[N_{\boldsymbol{u}}(Q_{U'|U})\right]^{1/s}\right\}\right)^{s/\rho}, \tag{A.17}$$

where (A.16) 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 [10, Sec. 6.3]

$$\mathbb{E}\left\{\left[N_{\boldsymbol{u}}(Q_{U'|U})\right]^{1/s}\right\} \doteq \begin{cases} \exp\{n[H_Q(U'|U) - R]/s\} & H_Q(U'|U) \geq R \\ \exp\{n[H_Q(U'|U) - R]\} & H_Q(U'|U) < R \end{cases}, \tag{A.18}$$

and so,

$$\mathbb{E}\left\{[N(Q_{UU'})]^{1/\rho}\right\} \leq e^{nH_Q(U)\cdot s/\rho} \cdot \left(\mathbb{E}\left\{\left[N_{\boldsymbol{u}}(Q_{U'|U})\right]^{1/s}\right\}\right)^{s/\rho} \tag{A.19}$$

$$\doteq e^{nH_Q(U)\cdot s/\rho} \cdot \begin{cases} \exp\{n[H_Q(U'|U) - R]/\rho\} & H_Q(U'|U) \geq R \\ \exp\{n[H_Q(U'|U) - R]s/\rho\} & H_Q(U'|U) < R \end{cases} \tag{A.20}$$

$$= \begin{cases} \exp\{n[H_Q(U)\cdot s + H_Q(U'|U) - R]/\rho\} & H_Q(U'|U) \geq R \\ \exp\{n[H_Q(U) + H_Q(U'|U) - R]s/\rho\} & H_Q(U'|U) < R \end{cases} \tag{A.21}$$

$$= \begin{cases} \exp\{n[H_Q(U)\cdot s + H_Q(U'|U) - R]/\rho\} & H_Q(U'|U) \geq R \\ \exp\{n[H_Q(U,U') - R]s/\rho\} & H_Q(U'|U) < R \end{cases}. \tag{A.22}$$

After optimizing over s, we get

$$\frac{1}{n}\log\mathbb{E}\left\{[N(Q_{UU'})]^{1/\rho}\right\}$$

$$\leq \min_{1\leq s\leq\rho} \begin{cases} \left[H_Q(U)\cdot s + H_Q(U'|U) - R\right]/\rho & H_Q(U'|U) \geq R \\ \left[H_Q(U,U') - R\right]s/\rho & H_Q(U'|U) < R, H_Q(U,U') \geq R \\ \left[H_Q(U,U') - R\right]s/\rho & H_Q(U'|U) < R, H_Q(U,U') < R \end{cases} \tag{A.23}$$

$$= \begin{cases} \left[H_Q(U) + H_Q(U'|U) - R\right]/\rho & H_Q(U'|U) \geq R \\ \left[H_Q(U,U') - R\right]/\rho & H_Q(U'|U) < R, H_Q(U,U') \geq R \\ \left[H_Q(U,U') - R\right]\rho/\rho & H_Q(U'|U) < R, H_Q(U,U') < R \end{cases} \tag{A.24}$$

$$= \begin{cases} \left[H_Q(U,U') - R\right]/\rho & H_Q(U,U') \geq R \\ \left[H_Q(U,U') - R\right] & H_Q(U,U') < R \end{cases}, \tag{A.25}$$

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

$$\left(\mathbb{E}\left\{[N(Q_{UU'})]^{1/\rho}\right\}\right)^{\rho} \leq \begin{cases} \exp\{n\,[H_Q(U,U')-R]\} & H_Q(U,U')\geq R \\ \exp\{n\,[H_Q(U,U')-R]\cdot\rho\} & H_Q(U,U')<R \end{cases}, \tag{A.26}$$

and in the limit of ρ tending to infinity,

$$\lim_{\rho\to\infty}\left(\mathbb{E}\left\{[N(Q_{UU'})]^{1/\rho}\right\}\right)^{\rho} \leq \begin{cases} \exp\{n\,[H_Q(U,U')-R]\} & H_Q(U,U')\geq R \\ 0 & H_Q(U,U')<R \end{cases}. \tag{A.27}$$

Continuing now from (A.11),

$$\lim_{\rho\to\infty}\left(\mathbb{E}\left\{[P_{\text{e,FR}}(\mathcal{B}_n)]^{1/\rho}\right\}\right)^{\rho}$$

$$\dot{\leq} \lim_{\rho\to\infty}\left(\sum_{Q_{UU'}\in\mathcal{Q}}\mathbb{E}\left\{[N(Q_{UU'})]^{1/\rho}\right\}\cdot e^{n[\mathbb{E}_Q\log P(U)]/\rho}\cdot\exp\left\{-n\cdot\Gamma(Q_{UU'},R+\epsilon)/\rho\right\}\right)^{\rho} \tag{A.28}$$

$$\doteq \sum_{Q_{UU'}\in\mathcal{Q}}\lim_{\rho\to\infty}\left(\mathbb{E}\left\{[N(Q_{UU'})]^{1/\rho}\right\}\right)^{\rho}\cdot e^{n\mathbb{E}_Q\log P(U)}\cdot\exp\left\{-n\cdot\Gamma(Q_{UU'},R+\epsilon)\right\} \tag{A.29}$$

$$\leq \sum_{\{Q_{UU'}\in\mathcal{Q}:\ H_Q(U,U')\geq R\}}\exp\{n\left[H_Q(U,U')-R\right]\}\cdot e^{n\mathbb{E}_Q[\log P(U)]}\cdot\exp\left\{-n\cdot\Gamma(Q_{UU'},R+\epsilon)\right\} \tag{A.30}$$

$$\doteq \exp\left\{-n\cdot\min_{\{Q_{UU'}\in\mathcal{Q}:\ H_Q(U,U')\geq R\}}\left\{\Gamma(Q_{UU'},R+\epsilon)-H_Q(U,U')-\mathbb{E}_Q[\log P(U)]+R\right\}\right\}, \tag{A.31}$$

where (A.30) follows from (A.27). Due to the arbitrariness of $\epsilon>0$, we have proved that

$$E_{\text{trc,FR}}(R) \geq \min_{\{Q_{UU'}\in\mathcal{Q}:\ H_Q(U,U')\geq R\}}\left\{\Gamma(Q_{UU'},R)-H_Q(U,U')-\mathbb{E}_Q[\log P(U)]+R\right\}, \tag{A.32}$$

completing half of the proof of Theorem 1.

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:

$$\mathbb{P}\{\mathcal{E}(Q_{UU'})\} = \mathbb{P}\{N(Q_{UU'})<\exp\{n[H_Q(U,U')-R-\epsilon]\}\} \tag{A.33}$$

$$= \mathbb{P}\{N(Q_{UU'})<e^{-n\epsilon}\cdot\mathbb{E}\{N(Q_{UU'})\}\} \tag{A.34}$$

$$= \mathbb{P}\left\{\frac{N(Q_{UU'})}{\mathbb{E}\{N(Q_{UU'})\}}-1<-(1-e^{-n\epsilon})\right\} \tag{A.35}$$

$$\leq \mathbb{P}\left\{\left[\frac{N(Q_{UU'})-\mathbb{E}\{N(Q_{UU'})\}}{\mathbb{E}\{N(Q_{UU'})\}}\right]^2>(1-e^{-n\epsilon})^2\right\} \tag{A.36}$$

$$\leq \frac{\text{Var}\{N(Q_{UU'})\}}{(1-e^{-n\epsilon})^2 \cdot \mathbb{E}^2\{N(Q_{UU'})\}}. \tag{A.37}$$

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

$$\begin{aligned}
&\text{Var}\{N(Q_{UU'})\} \\
&= \mathbb{E}\{N^2(Q_{UU'})\} - \mathbb{E}^2\{N(Q_{UU'})\} && \text{(A.38)}\\
&= \mathbb{E}\left\{\left[\sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathcal{I}(\boldsymbol{u},\boldsymbol{u}')\right] \times \left[\sum_{(\tilde{\boldsymbol{u}},\hat{\boldsymbol{u}})\in\mathcal{T}(Q_{UU'})} \mathcal{I}(\tilde{\boldsymbol{u}},\hat{\boldsymbol{u}})\right]\right\} - (Kp)^2 && \text{(A.39)}\\
&= \sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \sum_{(\tilde{\boldsymbol{u}},\hat{\boldsymbol{u}})\in\mathcal{T}(Q_{UU'})} \mathbb{E}\left\{\mathcal{I}(\boldsymbol{u},\boldsymbol{u}')\mathcal{I}(\tilde{\boldsymbol{u}},\hat{\boldsymbol{u}})\right\} - (Kp)^2 && \text{(A.40)}\\
&= \sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{E}\left\{\mathcal{I}^2(\boldsymbol{u},\boldsymbol{u}')\right\} + \sum_{\substack{(\boldsymbol{u},\boldsymbol{u}'),(\tilde{\boldsymbol{u}},\hat{\boldsymbol{u}})\in\mathcal{T}(Q_{UU'})\\ (\boldsymbol{u},\boldsymbol{u}')\neq(\tilde{\boldsymbol{u}},\hat{\boldsymbol{u}})}} \mathbb{E}\left\{\mathcal{I}(\boldsymbol{u},\boldsymbol{u}')\mathcal{I}(\tilde{\boldsymbol{u}},\hat{\boldsymbol{u}})\right\} - (Kp)^2 && \text{(A.41)}\\
&= Kp + K(K-1)p^2 - (Kp)^2 && \text{(A.42)}\\
&= Kp(1-p) && \text{(A.43)}\\
&\doteq \exp\{n[H_Q(U,U') - R]\}, && \text{(A.44)}
\end{aligned}$$

and hence,

$$\begin{aligned}
\mathbb{P}\{\mathcal{E}(Q_{UU'})\} &\stackrel{\cdot}{\leq} \frac{\exp\{n[H_Q(U,U') - R]\}}{\exp\{n[2H_Q(U,U') - 2R]\}} && \text{(A.45)}\\
&= \exp\{-n[H_Q(U,U') - R]\}, && \text{(A.46)}
\end{aligned}$$

which decays to zero since we have assumed that $H_Q(U,U') > R$. Furthermore, if $H_Q(U,U') > R+\epsilon$, 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 $H_Q(U,U') > R+\epsilon$, let us define

$$Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) = \sum_{\tilde{\boldsymbol{u}}\in\mathcal{B}(\boldsymbol{u})\cap\mathcal{T}(\boldsymbol{u}),\tilde{\boldsymbol{u}}\neq\boldsymbol{u},\boldsymbol{u}'} \exp\{nf(\hat{P}_{\tilde{\boldsymbol{u}}\boldsymbol{v}})\}, \tag{A.47}$$

and

$$\begin{aligned}
\mathcal{G}_n(Q_{UU'V}) = \Bigg\{\mathcal{B}_n : &\sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}') = \mathcal{B}(\boldsymbol{u})\}\times \\
&\sum_{\boldsymbol{v}\in\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')} \mathbb{1}\left\{Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) \leq e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\right\} \geq \\
&\exp\{n[H_Q(U,U') - R - 3\epsilon/2]\} \cdot |\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')|\Bigg\}, \qquad \text{(A.48)}
\end{aligned}$$

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

$$\mathcal{G}_n = \bigcap_{\{Q_{UU'V}:\ H_Q(U,U')>R+\epsilon\}} [\mathcal{G}_n(Q_{UU'V}) \cap \mathcal{E}^{\rm c}(Q_{UU'})]. \tag{A.49}$$

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

$$\mathbb{P}\{\mathcal{G}_n^{\rm c}\} = \mathbb{P}\left\{\bigcup_{\{Q_{UU'V}:\ H_Q(U,U')>R+\epsilon\}} [\mathcal{G}_n^{\rm c}(Q_{UU'V}) \cup \mathcal{E}(Q_{UU'})]\right\} \tag{A.50}$$

$$\leq \sum_{\{Q_{UU'V}:\ H_Q(U,U')>R+\epsilon\}} \mathbb{P}\left\{\mathcal{G}_n^{\rm c}(Q_{UU'V}) \cup \mathcal{E}(Q_{UU'})\right\} \tag{A.51}$$

$$= \sum_{\{Q_{UU'V}:\ H_Q(U,U')>R+\epsilon\}} [\mathbb{P}\left\{\mathcal{E}(Q_{UU'})\right\} + \mathbb{P}\left\{\mathcal{G}_n^{\rm c}(Q_{UU'V}) \cap \mathcal{E}^{\rm c}(Q_{UU'})\right\}]. \tag{A.52}$$

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

$$\begin{aligned}
&\mathbb{P}\left\{\mathcal{G}_n^{\rm c}(Q_{UU'V}) \cap \mathcal{E}^{\rm c}(Q_{UU'})\right\} \\
&= \mathbb{P}\Bigg[\sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}') = \mathcal{B}(\boldsymbol{u})\} \cdot \sum_{\boldsymbol{v}\in\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')} \mathbb{1}\left\{Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) \leq e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\right\} < \\
&\quad \exp\{n[H_Q(U,U') - R - 3\epsilon/2]\} \cdot |\mathcal{T}(Q_{V|UU'}|\boldsymbol{u}, \boldsymbol{u}')|, \\
&\qquad N(Q_{UU'}) \geq \exp\{n[H_Q(U,U') - R - \epsilon]\}\Bigg] \qquad (A.53) \\
&= \mathbb{P}\Bigg[\sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}') = \mathcal{B}(\boldsymbol{u})\} \cdot \sum_{\boldsymbol{v}\in\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')} \mathbb{1}\left\{Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) > e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\right\} > \\
&\quad [N(Q_{UU'}) - \exp\{n[H_Q(U,U') - R - 3\epsilon/2]\}] \cdot |\mathcal{T}(Q_{V|UU'}|\boldsymbol{u}, \boldsymbol{u}')|, \\
&\qquad N(Q_{UU'}) \geq \exp\{n[H_Q(U,U') - R - \epsilon]\}\Bigg] \qquad (A.54) \\
&\leq \mathbb{P}\Bigg[\sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}') = \mathcal{B}(\boldsymbol{u})\} \cdot \sum_{\boldsymbol{v}\in\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')} \mathbb{1}\left\{Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) > e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\right\} > \\
&\quad [\exp\{n[H_Q(U,U') - R - \epsilon]\} - \exp\{n[H_Q(U,U') - R - 3\epsilon/2]\}] \cdot |\mathcal{T}(Q_{V|UU'}|\boldsymbol{u}, \boldsymbol{u}')|, \\
&\qquad N(Q_{UU'}) \geq \exp\{n[H_Q(U,U') - R - \epsilon]\}\Bigg] \qquad (A.55)
\end{aligned}$$

$$\leq \mathbb{P}\left[\sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}')=\mathcal{B}(\boldsymbol{u})\}\cdot \sum_{\boldsymbol{v}\in\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')} \mathbb{1}\left\{Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) > e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\right\} > \right.$$

$$\left. [\exp\{n[H_Q(U,U')-R-\epsilon]\} - \exp\{n[H_Q(U,U')-R-3\epsilon/2]\}]\cdot|\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')|\right] \quad \text{(A.56)}$$

$$\leq \frac{\mathbb{E}\left\{\sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}')=\mathcal{B}(\boldsymbol{u})\}\cdot\sum_{\boldsymbol{v}\in\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')} \mathbb{1}\left\{Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) > e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\right\}\right\}}{[\exp\{n[H_Q(U,U')-R-\epsilon]\} - \exp\{n[H_Q(U,U')-R-3\epsilon/2]\}]\cdot|\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')|} \quad \text{(A.57)}$$

$$\stackrel{\cdot}{\leq} \frac{|\mathcal{T}(Q_{UU'})|\cdot|\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')|\cdot\mathbb{P}\left\{\mathcal{B}(\boldsymbol{u}')=\mathcal{B}(\boldsymbol{u}),\ Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) > e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\right\}}{\exp\{n[H_Q(U,U')-R-\epsilon]\}\cdot|\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')|} \quad \text{(A.58)}$$

$$\doteq \frac{\exp\{nH_Q(U,U')\}\cdot\mathbb{P}\{\mathcal{B}(\boldsymbol{u}')=\mathcal{B}(\boldsymbol{u})\}\cdot\mathbb{P}\left\{Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) > e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\right\}}{\exp\{n[H_Q(U,U')-R-\epsilon]\}} \quad \text{(A.59)}$$

$$= e^{n\epsilon}\cdot\mathbb{P}\left\{Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) > e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\right\}, \quad \text{(A.60)}$$

where (A.55) follows by using the second event $N(Q_{UU'}) \geq \exp\{n[H_Q(U,U')-R-\epsilon]\}$ to increase the first event inside the probability in (A.54), (A.56) is true since the second event in (A.55) was omitted, (A.57) follows from Markov's inequality, and (A.59) is due to the independence between the two events inside the probability in (A.58). As for the probability in (A.60),

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

$$= \mathbb{P}\left\{\sum_{Q_{U|V}} N(Q_{UV})e^{nf(Q_{UV})} > e^{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon]}\right\} \quad \text{(A.61)}$$

$$\doteq \max_{Q_{U|V}} \mathbb{P}\left\{N(Q_{UV}) > \exp\{n[\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon-f(Q_{UV})]\}\right\} \quad \text{(A.62)}$$

$$\doteq e^{-nE}, \quad \text{(A.63)}$$

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,

$$E = \min_{Q_{U|V}} \begin{cases} [R-H_Q(U|V)]_+ & f(Q_{UV}) + [H_Q(U|V)-R]_+ \geq \alpha(R-2\epsilon,Q_U,Q_V)+\epsilon \\ \infty & f(Q_{UV}) + [H_Q(U|V)-R]_+ < \alpha(R-2\epsilon,Q_U,Q_V)+\epsilon \end{cases} \quad \text{(A.64)}$$

$$= \min_{\{Q_{U|V}:\ f(Q_{UV})+[H_Q(U|V)-R]_+\geq\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon\}} [R-H_Q(U|V)]_+. \quad \text{(A.65)}$$

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,

$$E \geq \min_{\{Q_{U|V}:\ H_Q(U|V)\leq R-2\epsilon\}} [R-H_Q(U|V)]_+ \geq 2\epsilon, \quad \text{(A.66)}$$

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

$$\mathbb{P}\left\{\mathcal{G}_n^{\rm c}(Q_{UU'V}) \cap \mathcal{E}^{\rm c}(Q_{UU'})\right\} \dot{\leq} e^{n\epsilon} \cdot e^{-2n\epsilon} = e^{-n\epsilon}, \tag{A.67}$$

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

$$\mathcal{K}(\mathcal{B}_n, Q_{UU'V}) = \{(\boldsymbol{u}, \boldsymbol{u}', \boldsymbol{v}) : \; Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v}) \leq \exp\{n[\alpha(R-2\epsilon, Q_U, Q_V) + \epsilon]\}\}, \tag{A.68}$$

as well as

$$\mathcal{K}(\mathcal{B}_n, Q_{UU'V}|\boldsymbol{u}, \boldsymbol{u}') = \{\boldsymbol{v} : \; (\boldsymbol{u}, \boldsymbol{u}', \boldsymbol{v}) \in \mathcal{K}(\mathcal{B}_n, Q_{UU'V})\}. \tag{A.69}$$

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

$$\begin{aligned} &\sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}') = \mathcal{B}(\boldsymbol{u})\} \cdot \frac{|\mathcal{T}(Q_{V|UU'}|\boldsymbol{u}, \boldsymbol{u}') \cap \mathcal{K}(\mathcal{B}_n, Q_{UU'V}|\boldsymbol{u}, \boldsymbol{u}')|}{|\mathcal{T}(Q_{V|UU'}|\boldsymbol{u}, \boldsymbol{u}')|} \\ &\geq \exp\{n[H_Q(U, U') - R - 3\epsilon/2]\}, \end{aligned} \tag{A.70}$$

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

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

$$= \mathbb{E}\left\{\left[\sum_{\boldsymbol{u},\boldsymbol{v}} P(\boldsymbol{u}, \boldsymbol{v}) \sum_{\substack{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u})\cap\mathcal{T}(\boldsymbol{u}) \\ \boldsymbol{u}'\neq\boldsymbol{u}}} \frac{\exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}{\exp\{nf(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})\} + \exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\} + Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v})}\right]^{1/\rho}\right\} \tag{A.71}$$

$$= \sum_{\mathcal{B}_n} \mathbb{P}\{\mathcal{B}_n\} \left\{\left[\sum_{\boldsymbol{u},\boldsymbol{v}} P(\boldsymbol{u}, \boldsymbol{v}) \sum_{\substack{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u})\cap\mathcal{T}(\boldsymbol{u}) \\ \boldsymbol{u}'\neq\boldsymbol{u}}} \frac{\exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}{\exp\{nf(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})\} + \exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\} + Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v})}\right]^{1/\rho}\right\} \tag{A.72}$$

$$\begin{aligned} \geq \sum_{\mathcal{B}_n\in\mathcal{G}_n} \mathbb{P}\{\mathcal{B}_n\} &\Bigg[\sum_{\{Q_{UU'}:\; H_Q(U,U')>R+\epsilon\}} \sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}') = \mathcal{B}(\boldsymbol{u})\} \cdot \exp\{n\mathbb{E}_Q \log P(U)\} \\ &\times \sum_{Q_{V|UU'}} \sum_{\boldsymbol{v}\in\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')\cap\mathcal{K}(\mathcal{B}_n,Q_{UU'V}|\boldsymbol{u},\boldsymbol{u}')} \exp\{n\mathbb{E}_Q \log P(V|U)\} \\ &\times \frac{\exp\{nf(Q_{U'V})\}}{\exp\{nf(Q_{UV})\} + \exp\{nf(Q_{U'V})\} + Z_{\boldsymbol{u}\boldsymbol{u}'}(\boldsymbol{v})}\Bigg]^{1/\rho} \end{aligned} \tag{A.73}$$

$$\geq \sum_{\mathcal{B}_n\in\mathcal{G}_n} \mathbb{P}\{\mathcal{B}_n\} \Bigg[\sum_{\{Q_{UU'}:\; H_Q(U,U')>R+\epsilon\}} \sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}') = \mathcal{B}(\boldsymbol{u})\} \cdot \exp\{n\mathbb{E}_Q \log P(U)\}$$

$$\times \sum_{Q_{V|UU'}} \sum_{\boldsymbol{v}\in\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')\cap\mathcal{K}(\mathcal{B}_n,Q_{UU'V}|\boldsymbol{u},\boldsymbol{u}')} \exp\{n\mathbb{E}_Q \log P(V|U)\}$$

$$\times \left. \frac{\exp\{nf(Q_{U'V})\}}{\exp\{nf(Q_{UV})\} + \exp\{nf(Q_{U'V})\} + \exp\{n[\alpha(R-2\epsilon, Q_U, Q_V)+\epsilon]\}} \right]^{1/\rho} \quad \text{(A.74)}$$

$$\doteq \sum_{\mathcal{B}_n\in\mathcal{G}_n} \mathbb{P}\{\mathcal{B}_n\} \left[\sum_{\{Q_{UU'}:\ H_Q(U,U')>R+\epsilon\}} \sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}')=\mathcal{B}(\boldsymbol{u})\} \right.$$

$$\times \sum_{Q_{V|UU'}} \frac{|\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')\cap\mathcal{K}(\mathcal{B}_n,Q_{UU'V}|\boldsymbol{u},\boldsymbol{u}')|}{|\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')|} \cdot |\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')| \cdot e^{n\mathbb{E}_Q\log P(U,V)}$$

$$\times \exp\{-n\cdot[\max\{f(Q_{UV}),\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon\} - f(Q_{U'V})]_+\}]^{1/\rho} \quad \text{(A.75)}$$

$$\doteq \sum_{\mathcal{B}_n\in\mathcal{G}_n} \mathbb{P}\{\mathcal{B}_n\} \left[\sum_{\{Q_{UU'V}:\ H_Q(U,U')>R+\epsilon\}} \sum_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{T}(Q_{UU'})} \mathbb{1}\{\mathcal{B}(\boldsymbol{u}')=\mathcal{B}(\boldsymbol{u})\} \right.$$

$$\times \frac{|\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')\cap\mathcal{K}(\mathcal{B}_n,Q_{UU'V}|\boldsymbol{u},\boldsymbol{u}')|}{|\mathcal{T}(Q_{V|UU'}|\boldsymbol{u},\boldsymbol{u}')|} \cdot e^{nH_Q(V|U,U')} \cdot e^{n\mathbb{E}_Q\log P(U,V)}$$

$$\times \exp\{-n\cdot[\max\{f(Q_{UV}),\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon\} - f(Q_{U'V})]_+\}]^{1/\rho} \quad \text{(A.76)}$$

$$\geq \mathbb{P}\{\mathcal{G}_n\} \left[\sum_{\{Q_{UU'V}:\ H_Q(U,U')>R+\epsilon\}} \exp\{n[H_Q(U,U')-R-3\epsilon/2]\} \cdot e^{nH_Q(V|U,U')} \right.$$

$$\left. \times e^{n\mathbb{E}_Q\log P(U,V)} \cdot \exp\{-n\cdot[\max\{f(Q_{UV}),\alpha(R-2\epsilon,Q_U,Q_V)+\epsilon\} - f(Q_{U'V})]_+\} \right]^{1/\rho}$$

$$\text{(A.77)}$$

$$\doteq \exp\{-n\cdot[E_{\text{trc}}^{\text{fr}}(R)+O(\epsilon)]/\rho\}, \quad \text{(A.78)}$$

where (A.74) follows from the definition of the set $\mathcal{K}(\mathcal{B}_n, Q_{UU'V}|\boldsymbol{u},\boldsymbol{u}')$ in (A.69) and (A.77) is due to (A.70). Finally, raising to the ρ–th power and letting $\rho\to\infty$ yields

$$\lim_{\rho\to\infty} \left(\mathbb{E}\left\{[P_{\text{e,FR}}(\mathcal{B}_n)]^{1/\rho}\right\}\right)^{\rho} \stackrel{\cdot}{\geq} \exp\{-n\cdot[E_{\text{trc}}^{\text{fr}}(R)+O(\epsilon)]\}, \quad \text{(A.79)}$$

which completes the proof of Theorem 1, thanks to the arbitrariness of ϵ.

Appendix B

Proof of Lemma 1

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

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

First, note that

$$Z_{\boldsymbol{u}}(\boldsymbol{v}) = \sum_{\tilde{\boldsymbol{u}} \in \mathcal{B}(\boldsymbol{u}) \cap \mathcal{T}(\boldsymbol{u}), \tilde{\boldsymbol{u}} \neq \boldsymbol{u}} \exp\{nf(\hat{P}_{\tilde{\boldsymbol{u}}\boldsymbol{v}})\} = \sum_{Q_{U|V} \in \mathcal{S}(\hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}})} N(\mathcal{T}(Q_{U|V}|\boldsymbol{v}), \mathcal{B}(\boldsymbol{u})) e^{nf(Q_{UV})}, \tag{B.2}$$

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,

$$\begin{aligned}
&\mathbb{P}\left\{Z_{\boldsymbol{v}}(\boldsymbol{u}) \leq \exp\{n\alpha(R+\epsilon, \hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}})\}\right\} \\
&= \mathbb{P}\left\{\sum_{Q_{U|V} \in \mathcal{S}(\hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}})} N(\mathcal{T}(Q_{U|V}|\boldsymbol{v}), \mathcal{B}(\boldsymbol{u})) e^{nf(Q_{UV})} \leq \exp\{n\alpha(R+\epsilon, \hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}})\}\right\} && \text{(B.3)} \\
&\leq \mathbb{P}\left\{\max_{Q_{U|V} \in \mathcal{S}(\hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}})} N(\mathcal{T}(Q_{U|V}|\boldsymbol{v}), \mathcal{B}(\boldsymbol{u})) e^{nf(Q_{UV})} \leq \exp\{n\alpha(R+\epsilon, \hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}})\}\right\} && \text{(B.4)} \\
&= \mathbb{P} \bigcap_{Q_{U|V} \in \mathcal{S}(\hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}})} \left\{N(\mathcal{T}(Q_{U|V}|\boldsymbol{v}), \mathcal{B}(\boldsymbol{u})) e^{nf(Q_{UV})} \leq \exp\{n\alpha(R+\epsilon, \hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}})\}\right\} && \text{(B.5)} \\
&= \mathbb{P} \bigcap_{Q_{U|V} \in \mathcal{S}(\hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}})} \left\{N(\mathcal{T}(Q_{U|V}|\boldsymbol{v}), \mathcal{B}(\boldsymbol{u})) \leq \exp\{n[\alpha(R+\epsilon, \hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}}) - f(Q_{UV})]\}\right\}. && \text{(B.6)}
\end{aligned}$$

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 $H_{Q^*}(U|V) \geq R+\epsilon$ and $H_{Q^*}(U|V) - R - \epsilon \geq \alpha(R+\epsilon, \hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}}) - f(Q^*_{UV})$ hold. 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 $Q_{U|V} \in \mathcal{S}(\hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}})$

$$\begin{aligned}
H_Q(U|V) - \epsilon &< \max\{R, R + \alpha(R+\epsilon, \hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}}) - f(Q_{UV})\} && \text{(B.7)} \\
&= R + [\alpha(R+\epsilon, \hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}}) - f(Q_{UV})]_+. && \text{(B.8)}
\end{aligned}$$

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

$$H_Q(U|V) - \epsilon < R + t[\alpha(R+\epsilon, \hat{P}_{\boldsymbol{u}}, \hat{P}_{\boldsymbol{v}}) - f(Q_{UV})], \tag{B.9}$$

or equivalently,

$$\alpha(R+\epsilon,\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}}) > \max_{Q_{U|V}\in\mathcal{S}(\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})} \min_{t\in[0,1]} f(Q_{UV}) + \frac{H_Q(U|V)-R-\epsilon}{t} \tag{B.10}$$

$$= \max_{Q_{U|V}\in\mathcal{S}(\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})} \begin{cases} f(Q_{UV}) + H_Q(U|V) - R - \epsilon & H_Q(U|V) \geq R+\epsilon \\ -\infty & H_Q(U|V) < R+\epsilon \end{cases} \tag{B.11}$$

$$= \max_{\{Q_{U|V}\in\mathcal{S}(\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}}):\ H_Q(U|V)\geq R+\epsilon\}} [f(Q_{UV}) + H_Q(U|V)] - R - \epsilon \tag{B.12}$$

$$\equiv \alpha(R+\epsilon,\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}}), \tag{B.13}$$

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

$$\mathbb{P} \bigcap_{Q_{U|V}\in\mathcal{S}(\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}})} \Big\{ N(\mathcal{T}(Q_{U|V}|\boldsymbol{v}),\mathcal{B}(\boldsymbol{u})) \leq \exp\{n[\alpha(R+\epsilon,\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}}) - f(Q_{UV})]\} \Big\} \tag{B.14}$$

$$\leq \mathbb{P}\Big\{ N(\mathcal{T}(Q^*_{U|V}|\boldsymbol{v}),\mathcal{B}(\boldsymbol{u})) \leq \exp\{n[\alpha(R+\epsilon,\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}}) - f(Q^*_{UV})]\} \Big\}. \tag{B.15}$$

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 (B.15) is upper bounded by

$$\exp\Big\{ - e^{nH_{Q^*}(U|V)} D(e^{-an} \| e^{-bn}) \Big\}, \tag{B.16}$$

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

$$D(\alpha\|\beta) = \alpha\log\frac{\alpha}{\beta} + (1-\alpha)\log\frac{1-\alpha}{1-\beta}. \tag{B.17}$$

Since $a - b \geq \epsilon$, the binary divergence is lower bounded as follows ([10, Sec. 6.3]):

$$D(e^{-an}\|e^{-bn}) \geq e^{-bn}\Big\{1 - e^{-(a-b)n}[1 + n(a-b)]\Big\} \tag{B.18}$$

$$\geq e^{-nR}[1 - e^{-n\epsilon}(1+n\epsilon)], \tag{B.19}$$

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

$$\mathbb{P}\Big\{ N(\mathcal{T}(Q^*_{U|V}|\boldsymbol{v}),\mathcal{B}(\boldsymbol{u})) \leq \exp\{n[\alpha(R+\epsilon,\hat{P}_{\boldsymbol{u}},\hat{P}_{\boldsymbol{v}}) - f(Q^*_{UV})]\} \Big\} \tag{B.20}$$

$$\leq \exp\Big\{ - e^{nH_{Q^*}(U|V)} \cdot e^{-nR}[1 - e^{-n\epsilon}(1+n\epsilon)] \Big\} \tag{B.21}$$

$$\leq \exp\big\{ - e^{n\epsilon}[1 - e^{-n\epsilon}(1+n\epsilon)] \big\} \tag{B.22}$$

$$= \exp\big\{ - e^{n\epsilon} + n\epsilon + 1 \big\}. \tag{B.23}$$

Finally, the factor of $|\mathcal{U}\times\mathcal{V}|^n$ comes from the union bound, taking into account all $|\mathcal{U}\times\mathcal{V}|^n$ possible pairs $\{(\boldsymbol{u},\boldsymbol{v})\}$. This completes the proof of Lemma 1.

Appendix C

Proof of Theorem 4

We start by writing the expression in (37) in a slightly different way using $\min_{\{Q:\ g(Q)\le 0\}} f(Q) = \min_Q \sup_{s\ge 0}\{f(Q)+s\cdot g(Q)\}$:

$$E_{\mathrm{er}}(R(\cdot),\Delta) = \min_{\{Q_{UV}:\ R(Q_U)\ge H_Q(U|V)+\Delta\}} D(Q_{UV}\|P_{UV}) \tag{C.1}$$

$$= \min_{Q_{UV}} \sup_{\sigma\ge 0}\{D(Q_{UV}\|P_{UV}) + \sigma\cdot(H_Q(U|V)+\Delta - R(Q_U))\}. \tag{C.2}$$

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

$$\min_{Q_{UV}} \sup_{\sigma\ge 0}\{D(Q_{UV}\|P_{UV}) + \sigma\cdot(H_Q(U|V)+\Delta - R(Q_U))\} \ge \mathsf{E}_r \tag{C.3}$$

or,

$$\forall Q_{UV},\ \exists\sigma\ge 0,\ D(Q_{UV}\|P_{UV}) + \sigma\cdot(H_Q(U|V)+\Delta - R(Q_U)) \ge \mathsf{E}_r \tag{C.4}$$

or,

$$\forall Q_U,\ \forall Q_{V|U},\ \exists\sigma\ge 0,\ R(Q_U)\le H_Q(U|V)+\Delta+\frac{D(Q_{UV}\|P_{UV})-\mathsf{E}_r}{\sigma} \tag{C.5}$$

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

$$R(Q_U)\le \min_{Q_{V|U}}\sup_{\sigma\ge 0}\left\{H_Q(U|V)+\Delta+\frac{D(Q_{UV}\|P_{UV})-\mathsf{E}_r}{\sigma}\right\} \tag{C.6}$$

$$= \min_{Q_{V|U}}\begin{cases} H_Q(U|V)+\Delta & D(Q_{UV}\|P_{UV})\le \mathsf{E}_r \\ \infty & D(Q_{UV}\|P_{UV}) > \mathsf{E}_r \end{cases} \tag{C.7}$$

$$= \min_{\{Q_{V|U}:\ D(Q_{UV}\|P_{UV})\le \mathsf{E}_r\}}\{H_Q(U|V)+\Delta\}, \tag{C.8}$$

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

Appendix D

Proof of Theorem 5

It follows by the identity $[A]_+ = \max_{\mu\in[0,1]}\mu A$ that (44) can also be written as

$$E_{\mathrm{r}}^{\mathrm{vr}}(R(\cdot)) = \min_{Q_U}\min_{Q_{V|U}}\max_{\mu\in[0,1]}\{D(Q_{UV}\|P_{UV})+\mu\cdot(R(Q_U)-H_Q(U|V))\}, \tag{D.1}$$

such that $E_{\mathrm{r}}^{\mathrm{vr}}(R(\cdot)) \geq \mathsf{E}_e$ is equivalent to

$$\forall Q_U,\ \forall Q_{V|U},\ \exists \mu \in [0,1]:\ D(Q_{UV}\|P_{UV}) + \mu \cdot (R(Q_U) - H_Q(U|V)) \geq \mathsf{E}_e, \tag{D.2}$$

or,

$$\forall Q_U,\ \forall Q_{V|U},\ \exists \mu \in [0,1]:\ R(Q_U) \geq H_Q(U|V) + \frac{\mathsf{E}_e - D(Q_{UV}\|P_{UV})}{\mu}, \tag{D.3}$$

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

$$R(Q_U) \geq \max_{Q_{V|U}} \min_{\mu\in[0,1]} \left\{ H_Q(U|V) + \frac{\mathsf{E}_e - D(Q_{UV}\|P_{UV})}{\mu} \right\} \tag{D.4}$$

$$= \max_{Q_{V|U}} \begin{cases} H_Q(U|V) + \mathsf{E}_e - D(Q_{UV}\|P_{UV}) & \mathsf{E}_e \geq D(Q_{UV}\|P_{UV}) \\ -\infty & \mathsf{E}_e < D(Q_{UV}\|P_{UV}) \end{cases} \tag{D.5}$$

$$= \max_{\{Q_{V|U}:\ D(Q_{UV}\|P_{UV}) \leq \mathsf{E}_e\}} \{H_Q(U|V) + \mathsf{E}_e - D(Q_{UV}\|P_{UV})\}, \tag{D.6}$$

and the proof is complete.

Appendix E

Proof of Theorem 6

We start by writing the expressions in (32), (33), and (35) in a slightly different way. First, (32) can be written as

$$\gamma(R(\cdot), Q_U, Q_V) = \max_{\left\{\substack{Q_{\tilde{U}|V}:\ Q_{\tilde{U}}=Q_U,\\ H_Q(\tilde{U}|V)\geq R(Q_U)}\right\}} \{f(Q_{\tilde{U}V}) + H_Q(\tilde{U}|V)\} - R(Q_U) \tag{E.1}$$

$$= \max_{\{Q_{\tilde{U}|V}:\ Q_{\tilde{U}}=Q_U\}} \inf_{\theta\geq 0} \{\theta \cdot (H_Q(\tilde{U}|V) - R(Q_U)) + f(Q_{\tilde{U}V}) + H_Q(\tilde{U}|V) - R(Q_U)\} \tag{E.2}$$

$$= \max_{\{Q_{\tilde{U}|V}:\ Q_{\tilde{U}}=Q_U\}} \inf_{\theta\geq 0} \{(\theta+1) \cdot (H_Q(\tilde{U}|V) - R(Q_U)) + f(Q_{\tilde{U}V})\} \tag{E.3}$$

$$= \max_{\{Q_{\tilde{U}|V}:\ Q_{\tilde{U}}=Q_U\}} \inf_{\theta\geq 1} \{\theta \cdot (H_Q(\tilde{U}|V) - R(Q_U)) + f(Q_{\tilde{U}V})\}, \tag{E.4}$$

where (E.2) is due to the identity $\max_{\{Q:\ g(Q)\geq 0\}} f(Q) = \max_Q \inf_{\mu\geq 0}\{f(Q)+\mu\cdot g(Q)\}$. Similarly for (35) provides

$$E_{\mathrm{e}}(R(\cdot))$$

$$= \min_{\left\{\substack{Q_{UU'}:\ Q_{U'}=Q_U,\\ H_Q(U,U')\geq R(Q_U)}\right\}} \left\{\Lambda(Q_{UU'}, R(Q_U)) - \mathbb{E}_Q[\log P(U)] - H_Q(U,U') + R(Q_U)\right\} \tag{E.5}$$

$$= \min_{Q_U} \min_{\{Q_{U'|U}:\ Q_{U'}=Q_U\}} \sup_{\sigma\geq 1} \left\{\sigma \cdot (R(Q_U) - H_Q(U,U')) + \Lambda(Q_{UU'}, R(Q_U)) - \mathbb{E}_Q[\log P(U)]\right\}, \tag{E.6}$$

where (E.6) follows by the identity $\min_{\{Q:\ g(Q)\leq 0\}} f(Q) = \min_Q \sup_{s\geq 0}\{f(Q) + s\cdot g(Q)\}$. As for (33), we have

$$\Psi(R(\cdot), Q_{UU'V}) = [\max\{f(Q_{UV}), \gamma(R(\cdot), Q_U, Q_V)\} - f(Q_{U'V})]_+ \tag{E.7}$$

$$= \max_{\mu\in[0,1]} \mu\cdot\left\{\max_{\tau\in[0,1]}\{(1-\tau)\cdot f(Q_{UV}) + \tau\cdot\gamma(R(\cdot), Q_U, Q_V)\} - f(Q_{U'V})\right\} \tag{E.8}$$

$$= \max_{\mu\in[0,1]}\max_{\tau\in[0,1]}\{\mu(1-\tau)\cdot f(Q_{UV}) + \mu\tau\cdot\gamma(R(\cdot), Q_U, Q_V) - \mu\cdot f(Q_{U'V})\}, \tag{E.9}$$

where (E.8) is due to the identities $[A]_+ = \max_{\mu\in[0,1]}\mu A$ and $\max\{B, C\} = \max_{\tau\in[0,1]}\{(1-\tau)\cdot B + \tau\cdot C\}$. Substituting (E.4), (E.9), and (34) into (E.6), yields

$$\begin{aligned}
&E_{\mathrm{e}}(R(\cdot)) \\
&= \min_{Q_U}\min_{\{Q_{U'|U}:\ Q_{U'}=Q_U\}}\sup_{\sigma\geq 1}\min_{Q_{V|UU'}}\max_{\mu\in[0,1]}\max_{\tau\in[0,1]}\max_{\{Q_{\tilde{U}|V}:\ Q_{\tilde{U}}=Q_U\}}\inf_{\theta\geq 1}\{\sigma\cdot(R(Q_U) - H_Q(U,U')) \\
&\quad + \mu(1-\tau)\cdot f(Q_{UV}) + \mu\tau\cdot[\theta\cdot(H_Q(\tilde{U}|V) - R(Q_U)) + f(Q_{\tilde{U}V})] \\
&\quad - \mu\cdot f(Q_{U'V}) - H_Q(V|U,U') - \mathbb{E}_Q[\log P(U,V)]\} &\text{(E.10)}\\
&= \min_{Q_U}\min_{\{Q_{U'|U}:\ Q_{U'}=Q_U\}}\sup_{\sigma\geq 1}\min_{Q_{V|UU'}}\max_{\mu\in[0,1]}\max_{\tau\in[0,1]}\max_{\{Q_{\tilde{U}|V}:\ Q_{\tilde{U}}=Q_U\}}\inf_{\theta\geq 1}\{\sigma\cdot R(Q_U) - \sigma\cdot H_Q(U,U') \\
&\quad + \mu(1-\tau)\cdot f(Q_{UV}) + \mu\tau\theta\cdot H_Q(\tilde{U}|V) - \mu\tau\theta\cdot R(Q_U) + \mu\tau\cdot f(Q_{\tilde{U}V}) \\
&\quad - \mu\cdot f(Q_{U'V}) - H_Q(V|U,U') - \mathbb{E}_Q[\log P(U,V)]\} &\text{(E.11)}\\
&= \min_{Q_U}\min_{\{Q_{U'|U}:\ Q_{U'}=Q_U\}}\sup_{\sigma\geq 1}\min_{Q_{V|UU'}}\max_{\mu\in[0,1]}\max_{\tau\in[0,1]}\max_{\{Q_{\tilde{U}|V}:\ Q_{\tilde{U}}=Q_U\}}\inf_{\theta\geq 1}\{(\sigma-\mu\tau\theta)\cdot R(Q_U) \\
&\quad - \sigma\cdot H_Q(U,U') + \mu(1-\tau)\cdot f(Q_{UV}) + \mu\tau\theta\cdot H_Q(\tilde{U}|V) + \mu\tau\cdot f(Q_{\tilde{U}V}) \\
&\quad - \mu\cdot f(Q_{U'V}) - H_Q(V|U,U') - \mathbb{E}_Q[\log P(U,V)]\}. &\text{(E.12)}
\end{aligned}$$

Now, we want to solve $E_{\mathrm{e}}(R(\cdot)) \geq \mathsf{E}_e$ and arrive at a lower bound on the rate function $R(Q_U)$. Requiring that (E.12) is greater or equal to E_e is equivalent to

$$\begin{aligned}
&\forall Q_U,\ \forall\{Q_{U'|U}:\ Q_{U'}=Q_U\},\ \exists\sigma\geq 1,\ \forall Q_{V|UU'}, \\
&\exists\mu\in[0,1],\ \exists\tau\in[0,1],\ \exists\{Q_{\tilde{U}|V}:\ Q_{\tilde{U}}=Q_U\},\ \forall\theta\geq 1: \\
&(\sigma-\mu\tau\theta)\cdot R(Q_U) - \sigma\cdot H_Q(U,U') + \mu(1-\tau)\cdot f(Q_{UV}) + \mu\tau\theta\cdot H_Q(\tilde{U}|V) + \mu\tau\cdot f(Q_{\tilde{U}V}) \\
&\quad - \mu\cdot f(Q_{U'V}) - H_Q(V|U,U') - \mathbb{E}_Q[\log P(U,V)] \geq \mathsf{E}_e, &\text{(E.13)}
\end{aligned}$$

which is the same as

$$\begin{aligned}
&\forall Q_U,\ \forall\{Q_{U'|U}:\ Q_{U'}=Q_U\},\ \exists\sigma\geq 1,\ \forall Q_{V|UU'},\\
&\exists\mu\in[0,1],\ \exists\tau\in[0,1],\ \exists\{Q_{\tilde{U}|V}:\ Q_{\tilde{U}}=Q_U\},\ \forall\theta\geq 1:\\
&(\sigma-\mu\tau\theta)\cdot R(Q_U)\geq \mathsf{E}_e+H_Q(V|U,U')+\mathbb{E}_Q[\log P(U,V)]+\sigma\cdot H_Q(U,U')\\
&\quad+\mu\cdot f(Q_{U'V})-\mu(1-\tau)\cdot f(Q_{UV})-\mu\tau\theta\cdot H_Q(\tilde{U}|V)-\mu\tau\cdot f(Q_{\tilde{U}V}),
\end{aligned}\tag{E.14}$$

or

$$\begin{aligned}
&\forall Q_U,\ \forall\{Q_{U'|U}:\ Q_{U'}=Q_U\},\ \exists\sigma\geq 1,\ \forall Q_{V|UU'},\\
&\exists\mu\in[0,1],\ \exists\tau\in[0,1],\ \exists\{Q_{\tilde{U}|V}:\ Q_{\tilde{U}}=Q_U\},\ \forall\theta\geq 1:\\
&(\sigma-\mu\tau\theta)\cdot R(Q_U)\geq \mathsf{E}_e+H_Q(V|U,U')+\mathbb{E}_Q[\log P(U,V)]+\sigma\cdot H_Q(U,U')\\
&\quad+\mu\cdot[f(Q_{U'V})-(1-\tau)\cdot f(Q_{UV})-\tau\cdot(\theta H_Q(\tilde{U}|V)+f(Q_{\tilde{U}V}))],
\end{aligned}\tag{E.15}$$

which is, in turn, equivalent to require that for any $Q_U\in\mathcal{P}(\mathcal{U})$,

$$\begin{aligned}
R(Q_U)\geq &\max_{\{Q_{U'|U}:\ Q_{U'}=Q_U\}}\inf_{\sigma\geq 1}\max_{Q_{V|UU'}}\min_{\mu\in[0,1]}\min_{\tau\in[0,1]}\min_{\{Q_{\tilde{U}|V}:\ Q_{\tilde{U}}=Q_U\}}\sup_{\theta\geq 1}\\
&\frac{1}{(\sigma-\mu\tau\theta)}\cdot\{\mathsf{E}_e+H_Q(V|U,U')+\mathbb{E}_Q[\log P(U,V)]+\sigma\cdot H_Q(U,U')\\
&+\mu\cdot[f(Q_{U'V})-(1-\tau)\cdot f(Q_{UV})-\tau\cdot(\theta H_Q(\tilde{U}|V)+f(Q_{\tilde{U}V}))]\}.
\end{aligned}\tag{E.16}$$

It turns out that the expression in (E.16) is relatively cumbersome and cannot be recast into a more simple expression, hence, at that point, we must compromise on the tightness of the lower bound for the rate function. Note that all the maximizations on the right–hand–side of (E.16) are mandatory, but the minimizations are not, such that we may choose any value we like and still have a valid lower bound on the rate function. Having said that, let us choose $\tau=0$, which saves us the need to maximize over $\theta\geq 1$ and minimize over $\{Q_{\tilde{U}|V}:\ Q_{\tilde{U}}=Q_U\}$. We get the following weakened lower bound on the rate function:

$$\begin{aligned}
R(Q_U)\geq &\max_{\{Q_{U'|U}:\ Q_{U'}=Q_U\}}\inf_{\sigma\geq 1}\max_{Q_{V|UU'}}\min_{\mu\in[0,1]}\frac{1}{\sigma}\cdot\{\mathsf{E}_e+H_Q(V|U,U')+\mathbb{E}_Q[\log P(U,V)]\\
&+\sigma\cdot H_Q(U,U')+\mu\cdot(f(Q_{U'V})-f(Q_{UV}))\}.
\end{aligned}\tag{E.17}$$

At this point, we assume that $f(Q_{UV}) = \mathbb{E}_Q[\log P(U,V)]$, and so

$$R(Q_U) \geq \max_{\{Q_{U'|U}:\ Q_{U'}=Q_U\}} \inf_{\sigma\geq 1} \max_{Q_{V|UU'}} \min_{\mu\in[0,1]} \frac{1}{\sigma}\cdot\{\mathsf{E}_e + H_Q(V|U,U') + \mathbb{E}_Q[\log P(U,V)] + \sigma\cdot H_Q(U,U') + \mu\cdot(\mathbb{E}_Q[\log P(U',V)] - \mathbb{E}_Q[\log P(U,V)])\} \tag{E.18}$$

$$= \max_{\{Q_{U'|U}:\ Q_{U'}=Q_U\}} \inf_{\sigma\geq 1} \min_{\mu\in[0,1]} \max_{Q_{V|UU'}} \frac{1}{\sigma}\cdot\{\mathsf{E}_e + H_Q(V|U,U') + \mathbb{E}_Q[\log P(U,V)] + \sigma\cdot H_Q(U,U') + \mu\cdot(\mathbb{E}_Q[\log P(U',V)] - \mathbb{E}_Q[\log P(U,V)])\} \tag{E.19}$$

$$= \max_{\{Q_{U'|U}:\ Q_{U'}=Q_U\}} \inf_{\sigma\geq 1} \min_{\mu\in[0,1]} \frac{1}{\sigma}\cdot\{\mathsf{E}_e + \sigma\cdot H_Q(U,U') - \min_{Q_{V|UU'}}\{-H_Q(V|U,U') - \mathbb{E}_Q[\log P(U,V)] - \mu\cdot(\mathbb{E}_Q[\log P(U',V)] - \mathbb{E}_Q[\log P(U,V)])\}\}, \tag{E.20}$$

where (E.19) is because the objective in (E.18) is concave in $Q_{V|UU'}$ and affine in μ. Now, the minimization over $\{Q_{V|UU'}\}$ can be carried out as follows:

$$\min_{Q_{V|UU'}}\{-H_Q(V|U,U') - \mathbb{E}_Q[\log P(U,V)] - \mu\cdot(\mathbb{E}_Q[\log P(U',V)] - \mathbb{E}_Q[\log P(U,V)])\}$$

$$= \min_{Q_{V|UU'}} \sum_{u,u'} Q_{UU'}(u,u') \sum_v Q_{V|UU'}(v|u,u') \log\frac{Q_{V|UU'}(v|u,u')}{P(u,v)^{1-\mu}\cdot P(u',v)^{\mu}} \tag{E.21}$$

$$= -\sum_{u,u'} Q_{UU'}(u,u') \log\left(\sum_v P(u,v)^{1-\mu}\cdot P(u',v)^{\mu}\right). \tag{E.22}$$

At this stage, denote the expression in (E.22) as $B_Q^{\mu}(U,U')$. Substituting (E.22) back into (E.20) yields

$$R(Q_U) \geq \max_{\{Q_{U'|U}:\ Q_{U'}=Q_U\}} \inf_{\sigma\geq 1} \min_{\mu\in[0,1]} \frac{1}{\sigma}\cdot\{\mathsf{E}_e + \sigma\cdot H_Q(U,U') - B_Q^{\mu}(U,U')\}. \tag{E.23}$$

It is relatively easy to prove that the function

$$f(\mu) = \sum_{(u,u')\in\mathcal{U}^2} Q_{UU'}(u,u') \log\left(\sum_{v\in\mathcal{V}} P(u,v)^{1-\mu}\cdot P(u',v)^{\mu}\right) \tag{E.24}$$

is minimized at $\mu^* = \frac{1}{2}$. We use the definition of $B_Q(U,U')$ in (48), which finally provides

$$R(Q_U) \geq \max_{\{Q_{U'|U}:\ Q_{U'}=Q_U\}} \inf_{\sigma\geq 1} \left\{H_Q(U,U') + \frac{\mathsf{E}_e - B_Q(U,U')}{\sigma}\right\} \tag{E.25}$$

$$= \max_{\{Q_{U'|U}:\ Q_{U'}=Q_U\}} \begin{cases} H_Q(U,U') & \mathsf{E}_e \geq B_Q(U,U') \\ H_Q(U,U') + \mathsf{E}_e - B_Q(U,U') & \mathsf{E}_e < B_Q(U,U') \end{cases} \tag{E.26}$$

$$= \max_{\{Q_{U'|U}:\ Q_{U'}=Q_U\}} \left\{H_Q(U,U') - [B_Q(U,U') - \mathsf{E}_e]_+\right\}, \tag{E.27}$$

which completes the proof.

Appendix F

Proof of Theorem 7

We have that

$$\mathbb{E}[P_{\mathrm{e,SD}}(\mathcal{B}_n)] = \mathbb{E}\left[\frac{\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{U}),\boldsymbol{u}'\neq\boldsymbol{U}}\exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{V}})\}}{\sum_{\tilde{\boldsymbol{u}}\in\mathcal{B}(\boldsymbol{U})}\exp\{nf(\hat{P}_{\tilde{\boldsymbol{u}}\boldsymbol{V}})\}}\right]. \tag{F.1}$$

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

$$\begin{aligned}
&\mathbb{E}[P_{\mathrm{e,SD}}(\mathcal{B}_n)|\boldsymbol{u},\boldsymbol{v}] \\
&= \mathbb{E}\left[\frac{\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}}\exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}{\exp\{n\cdot f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})\}+\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}}\exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}\right] &\text{(F.2)}\\
&= \int_0^1 \mathbb{P}\left\{\frac{\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}}\exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}{\exp\{n\cdot f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})\}+\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}}\exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}\geq s\right\}ds &\text{(F.3)}\\
&= \int_0^\infty ne^{-n\xi}\cdot\mathbb{P}\left\{\frac{\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}}\exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}{\exp\{n\cdot f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})\}+\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}}\exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}}\geq e^{-n\xi}\right\}d\xi &\text{(F.4)}\\
&= \int_0^\infty ne^{-n\xi}\cdot\mathbb{P}\left\{(1-e^{-n\xi})\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}}\exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}\geq e^{-n\xi}\exp\{nf(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})\}\right\}d\xi &\text{(F.5)}\\
&\doteq \int_0^\infty ne^{-n\xi}\cdot\mathbb{P}\left\{\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}}\exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}\geq \exp\{n[f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})-\xi]\}\right\}d\xi. &\text{(F.6)}
\end{aligned}$$

Define

$$N_{\boldsymbol{u},\boldsymbol{v}}(Q_{U|V}) = \sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}}\mathbb{1}\{\boldsymbol{u}'\in\mathcal{T}(Q_{U|V}|\boldsymbol{v})\}, \tag{F.7}$$

such that the probability in (F.6) is given by

$$\begin{aligned}
&\mathbb{P}\left\{\sum_{\boldsymbol{u}'\in\mathcal{B}(\boldsymbol{u}),\boldsymbol{u}'\neq\boldsymbol{u}}\exp\{nf(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}})\}\geq\exp\{n[f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})-\xi]\}\right\}\\
&= \mathbb{P}\left\{\sum_{Q_{U'|V}} N_{\boldsymbol{u},\boldsymbol{v}}(Q_{U'|V})\exp\{nf(Q_{U'V})\}\geq\exp\{n[f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})-\xi]\}\right\} &\text{(F.8)}\\
&\doteq \mathbb{P}\left\{\max_{Q_{U'|V}} N_{\boldsymbol{u},\boldsymbol{v}}(Q_{U'|V})\exp\{nf(Q_{U'V})\}\geq\exp\{n[f(\hat{P}_{\boldsymbol{u}\boldsymbol{v}})-\xi]\}\right\} &\text{(F.9)}
\end{aligned}$$

$$= \mathbb{P} \bigcup_{Q_{U'|V}} \left\{ N_{\boldsymbol{u},\boldsymbol{v}}(Q_{U'|V}) \exp\{nf(Q_{U'V})\} \geq \exp\{n[f(\hat{P}_{\boldsymbol{uv}}) - \xi]\} \right\} \tag{F.10}$$

$$\doteq \sum_{Q_{U'|V}} \mathbb{P} \left\{ N_{\boldsymbol{u},\boldsymbol{v}}(Q_{U'|V}) \geq \exp\{n[f(\hat{P}_{\boldsymbol{uv}}) - f(Q_{U'V}) - \xi]\} \right\}, \tag{F.11}$$

where $Q_{U'V} = Q_{U'|V} \times \hat{P}_{\boldsymbol{v}}$. Let us denote $B_0 = f(\hat{P}_{\boldsymbol{uv}}) - f(Q_{U'V})$. Since $N_{\boldsymbol{u},\boldsymbol{v}}(Q_{U'|V})$ is a binomial sum of $|\mathcal{T}(Q_{U'|V}|\boldsymbol{v})| \doteq e^{nH_Q(U'|V)}$ trials and probability of success $e^{-nR(Q_U)}$,

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

$$= \begin{cases} [R(Q_U) - H_Q(U'|V)]_+ & [H_Q(U'|V) - R(Q_U)]_+ \geq B_0 - \xi \\ \infty & [H_Q(U'|V) - R(Q_U)]_+ < B_0 - \xi \end{cases} \tag{F.12}$$

$$= \begin{cases} [R(Q_U) - H_Q(U'|V)]_+ & \xi \geq B_0 - [H_Q(U'|V) - R(Q_U)]_+ \\ \infty & \xi < B_0 - [H_Q(U'|V) - R(Q_U)]_+ \end{cases}, \tag{F.13}$$

and so,

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

$$\doteq \int_{\left[B_0 - [H_Q(U'|V) - R(Q_U)]_+\right]_+}^\infty e^{-n\xi} \cdot e^{-n[R(Q_U) - H_Q(U'|V)]_+} d\xi \tag{F.14}$$

$$\doteq \exp\left\{ -n \left([R(Q_U) - H_Q(U'|V)]_+ + \left[B_0 - [H_Q(U'|V) - R(Q_U)]_+\right]_+ \right) \right\} \tag{F.15}$$

$$= \exp\left\{ -n \begin{pmatrix} R(Q_U) - H_Q(U'|V) + [B_0]_+ & R(Q_U) \geq H_Q(U'|V) \\ [R(Q_U) - H_Q(U'|V) + B_0]_+ & R(Q_U) < H_Q(U'|V) \end{pmatrix} \right\} \tag{F.16}$$

$$= \exp\left\{ -n \begin{pmatrix} [R(Q_U) - H_Q(U'|V) + [B_0]_+]_+ & R(Q_U) \geq H_Q(U'|V) \\ [R(Q_U) - H_Q(U'|V) + [B_0]_+]_+ & R(Q_U) < H_Q(U'|V) \end{pmatrix} \right\} \tag{F.17}$$

$$= \exp\left\{ -n \cdot \left[R(Q_U) - H_Q(U'|V) + [B_0]_+\right]_+ \right\}. \tag{F.18}$$

Finally, we have that

$$\sum_{Q_{U'|V}} \int_0^\infty e^{-n\xi} \cdot \mathbb{P} \left\{ N_{\boldsymbol{u},\boldsymbol{v}}(Q_{U|V}) \geq \exp\{n[B_0 - \xi]\} \right\} d\xi \tag{F.19}$$

$$\doteq \max_{Q_{U'|V}} \exp\left\{ -n \cdot \left[R(Q_U) - H_Q(U'|V) + [B_0]_+\right]_+ \right\} \tag{F.20}$$

$$= \exp\left\{ -n \cdot \min_{Q_{U'|V}} \left[R(Q_U) - H_Q(U'|V) + [B_0]_+\right]_+ \right\}, \tag{F.21}$$

thus,

$$E(\boldsymbol{u}, \boldsymbol{v}) = \min_{Q_{U'|V}} \left[R(Q_U) - H_Q(U'|V) + [f(\hat{P}_{\boldsymbol{uv}}) - f(Q_{U'V})]_+ \right]_+. \tag{F.22}$$

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 $\hat{P}_{\boldsymbol{uv}}$. Averaging over the source and the SI sequences, now provides

$$\mathbb{E}\left\{P_{\text{e,SD}}(\mathcal{B}_n)\right\} = \sum_{\boldsymbol{u},\boldsymbol{v}} P(\boldsymbol{u}, \boldsymbol{v}) \cdot \mathbb{1}\left\{\hat{H}_{\boldsymbol{u}}(U) \geq R(\hat{P}_{\boldsymbol{u}})\right\} \cdot \exp\left\{-n \cdot E(\hat{P}_{\boldsymbol{uv}})\right\} \tag{F.23}$$

$$\doteq \sum_{\{Q_{UV}:\ H_Q(U) \geq R(Q_U)\}} e^{-n \cdot D(Q_{UV} \| P_{UV})} \cdot \exp\left\{-n \cdot E(Q_{UV})\right\} \tag{F.24}$$

$$\doteq \exp\left\{-n \cdot \min_{\{Q_{UV}:\ H_Q(U) \geq R(Q_U)\}} \left[D(Q_{UV} \| P_{UV}) + E(Q_{UV})\right]\right\}, \tag{F.25}$$

which proves the first point of Theorem 7.

Step 3: Moving from Stochastic to Deterministic Decoding

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

$$\hat{\boldsymbol{u}} = \arg \max_{\boldsymbol{u}' \in \mathcal{B}(\boldsymbol{u})} f(\hat{P}_{\boldsymbol{u}'\boldsymbol{v}}), \tag{F.26}$$

we just have to multiply $f(\cdot)$ by $\beta \geq 0$, and then let $\beta \to \infty$. We find that the overall error exponent of the semi–deterministic variable–rate code with the general deterministic decoder of (F.26) is given by

$$E(P) = \min_{\{Q_{UV}:\ H_Q(U) \geq R(Q_U)\}} \left[D(Q_{UV} \| P_{UV}) + \tilde{E}(Q_{UV})\right], \tag{F.27}$$

where,

$$\tilde{E}(Q_{UV}) = \min_{\{Q_{U'|V}:\ f(Q_{U'V}) \geq f(Q_{UV})\}} \left[R(Q_U) - H_Q(U'|V)\right]_+. \tag{F.28}$$

Step 4: A Fundamental Limitation on the Error Exponent

Note that the minimum in (F.28) can be upper–bounded by choosing a specific distribution in the feasible set. In (F.28), we take $Q_{U'|V} = Q_{U|V}$ and then

$$\tilde{E}(Q_{UV}) \leq \left[R(Q_U) - H_Q(U|V)\right]_+. \tag{F.29}$$

Hence, the overall error exponent is upper–bounded as

$$E(P) \leq \min_{\{Q_{UV}:\ H_Q(U) \geq R(Q_U)\}} \left[D(Q_{UV} \| P_{UV}) + \left[R(Q_U) - H_Q(U|V)\right]_+\right]. \tag{F.30}$$

Step 5: An Optimal Universal Decoder

We prove that the upper bound of (F.30) is attainable by choosing the universal decoding metric $f(Q_{UV}) = -H_Q(U|V)$. Now, we get for (F.28)

$$\begin{aligned}
\tilde{E}(Q_{UV}) &= \min_{\{Q_{U'|V}:\ f(Q_{U'V}) \geq f(Q_{UV})\}} \left[R(Q_U) - H_Q(U'|V)\right]_+ && \text{(F.31)}\\
&= \min_{\{Q_{U'|V}:\ H_Q(U|V) \geq H_Q(U'|V)\}} \left[R(Q_U) - H_Q(U'|V)\right]_+ && \text{(F.32)}\\
&= \left[R(Q_U) - H_Q(U|V)\right]_+, && \text{(F.33)}
\end{aligned}$$

which completes the proof of Theorem 7.

Appendix G

Proof of Theorem 9

By definition of the error exponents, it follows that $E^{\text{sd}}_{\text{trc,GLD}}(R(\cdot)) \geq E^{\text{sd}}_{\text{r,GLD}}(R(\cdot))$. We now prove the other direction. The expression in (57) can also be written as

$$\begin{aligned}
&E^{\text{sd}}_{\text{trc,GLD}}(R(\cdot))\\
&= \min_{\left\{\substack{Q_{UU'}:\ Q_{U'}=Q_U,\\ H_Q(U)\geq R(Q_U)}\right\}} \left\{\Lambda(Q_{UU'}, R(Q_U)) - \mathbb{E}_Q[\log P(U)] - H_Q(U,U') + R(Q_U)\right\} && \text{(G.1)}\\
&= \min_{\left\{\substack{Q_{UU'}:\ Q_{U'}=Q_U,\\ H_Q(U)\geq R(Q_U)}\right\}} \left\{\min_{Q_{V|UU'}} \left\{\Psi(R(Q_U), Q_{UU'V}) - H_Q(V|U,U') - \mathbb{E}_Q[\log P(V|U)]\right\}\right.\\
&\qquad\qquad \left. - \mathbb{E}_Q[\log P(U)] - H_Q(U,U') + R(Q_U)\right\} && \text{(G.2)}\\
&= \min_{\left\{\substack{Q_{UU'V}:\ Q_{U'}=Q_U,\\ H_Q(U)\geq R(Q_U)}\right\}} \left\{\Psi(R(Q_U), Q_{UU'V}) - H_Q(U,U',V) - \mathbb{E}_Q[\log P(U,V)] + R(Q_U)\right\} && \text{(G.3)}\\
&= \min_{\left\{\substack{Q_{UU'V}:\ Q_{U'}=Q_U,\\ H_Q(U)\geq R(Q_U)}\right\}} \left\{\Psi(R(Q_U), Q_{UU'V}) + D(Q_{UV}\|P_{UV}) - H_Q(U'|U,V) + R(Q_U)\right\} && \text{(G.4)}\\
&= \min_{\mathcal{Q}} \left\{D(Q_{UV}\|P_{UV}) + R(Q_U) - H_Q(U'|U,V)\right.\\
&\qquad \left. + \left[\max\{f(Q_{UV}), \gamma(R(Q_U), Q_U, Q_V)\} - f(Q_{U'V})\right]_+\right\}, && \text{(G.5)}
\end{aligned}$$

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,

$$\gamma(R(\cdot), Q_U, Q_V) = \max_{\left\{\substack{Q_{\tilde{U}|V}:\ Q_{\tilde{U}}=Q_U,\\ H_Q(\tilde{U}|V)\geq R(Q_{\tilde{U}})}\right\}} \{f(Q_{\tilde{U}V}) + H_Q(\tilde{U}|V)\} - R(Q_U). \tag{G.6}$$

We upper–bound the minimum in (G.5) 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

$$E^{\text{sd}}_{\text{trc,GLD}}(R(\cdot)) \leq \min_{\tilde{\mathcal{Q}}} \big\{ D(Q_{UV} \| P_{UV}) + R(Q_U) - H_Q(U'|U,V) + [\max\{f(Q_{UV}), \gamma(R(Q_U), Q_U, Q_V)\} - f(Q_{U'V})]_+ \big\} \tag{G.7}$$

$$= \min_{\tilde{\mathcal{Q}}} \big\{ D(Q_{UV} \| P_{UV}) + R(Q_U) - H_Q(U'|V) + [\max\{f(Q_{UV}), \gamma(R(Q_U), Q_U, Q_V)\} - f(Q_{U'V})]_+ \big\} \tag{G.8}$$

$$= \min_{\{Q_{UV}:\ H_Q(U) \geq R(Q_U)\}} \Bigg\{ D(Q_{UV} \| P_{UV}) + \min_{Q_{U'|V} \in \hat{\mathcal{Q}}} \{ R(Q_U) - H_Q(U'|V) + [\max\{f(Q_{UV}), \gamma(R(Q_U), Q_U, Q_V)\} - f(Q_{U'V})]_+ \} \big\}, \tag{G.9}$$

where $\hat{\mathcal{Q}} = \{Q_{U'|V} :\ Q_{U'} = Q_U\}$. In order to upper–bound the inner minimum in (G.9), 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 (G.9) is convex in $Q_{U'|V}$. On the one hand, if the maximum is given by $f(Q_{UV})$, then the inner minimum in (G.9) is just

$$\min_{Q_{U'|V} \in \hat{\mathcal{Q}}} \big\{ R(Q_U) - H_Q(U'|V) + [f(Q_{UV}) - f(Q_{U'V})]_+ \big\}. \tag{G.10}$$

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 (G.6), and then

$$\min_{Q_{U'|V} \in \hat{\mathcal{Q}}} \big\{ R(Q_U) - H_Q(U'|V) + [\gamma(R(Q_U), Q_U, Q_V) - f(Q_{U'V})]_+ \big\}$$

$$= \min_{Q_{U'|V} \in \hat{\mathcal{Q}}} \left\{ R(Q_U) - H_Q(U'|V) + \left[f(Q^*_{\tilde{U}V}) + H_{Q^*}(\tilde{U}|V) - R(Q_U) - f(Q_{U'V}) \right]_+ \right\} \tag{G.11}$$

$$\leq R(Q_U) - H_{Q^*}(U'|V) + \left[f(Q^*_{\tilde{U}V}) + H_{Q^*}(\tilde{U}|V) - R(Q_U) - f(Q^*_{U'V}) \right]_+ \tag{G.12}$$

$$= R(Q_U) - H_{Q^*}(U'|V) + \left[H_{Q^*}(\tilde{U}|V) - R(Q_U) \right]_+ \tag{G.13}$$

$$= R(Q_U) - H_{Q^*}(U'|V) + H_{Q^*}(\tilde{U}|V) - R(Q_U) \tag{G.14}$$

$$= 0, \tag{G.15}$$

where (G.12) is because we choose $Q^*_{U'|V} = Q^*_{\tilde{U}|V}$ instead of minimizing over all $Q_{U'|V} \in \hat{\mathcal{Q}}$ and (G.14) 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 (G.10) and (G.15), we find that (G.9) is upper–bounded by

$$\begin{aligned}
E^{\text{sd}}_{\text{trc,GLD}}(R(\cdot)) \leq & \min_{\{Q_{UV}\colon\, H_Q(U) \geq R(Q_U)\}} \Big\{ D(Q_{UV} \| P_{UV}) \\
& + \max \left\{ \min_{Q_{U'|V} \in \hat{\mathcal{Q}}} \left\{ R(Q_U) - H_Q(U'|V) + [f(Q_{UV}) - f(Q_{U'V})]_+ \right\}, 0 \right\} \Big\} \qquad \text{(G.16)} \\
= & \min_{\{Q_{UV}\colon\, H_Q(U) \geq R(Q_U)\}} \Big\{ D(Q_{UV} \| P_{UV}) \\
& + \left[\min_{Q_{U'|V} \in \hat{\mathcal{Q}}} \left\{ R(Q_U) - H_Q(U'|V) + [f(Q_{UV}) - f(Q_{U'V})]_+ \right\} \right]_+ \Big\} \qquad \text{(G.17)} \\
= & \min_{\{Q_{UV}\colon\, H_Q(U) \geq R(Q_U)\}} \Big\{ D(Q_{UV} \| P_{UV}) \\
& + \min_{Q_{U'|V} \in \hat{\mathcal{Q}}} \left\{ \left[R(Q_U) - H_Q(U'|V) + [f(Q_{UV}) - f(Q_{U'V})]_+ \right]_+ \right\} \Big\} \qquad \text{(G.18)} \\
= & E^{\text{sd}}_{\text{r,GLD}}(R(\cdot)), \qquad \text{(G.19)}
\end{aligned}$$

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

$$E^{\text{sd}}_{\text{trc,MAP}}(R(\cdot)) \overset{\text{(a)}}{=} E^{\text{sd}}_{\text{r,MAP}}(R(\cdot)) \overset{\text{(b)}}{=} E^{\text{sd}}_{\text{r,MCE}}(R(\cdot)) \overset{\text{(c)}}{\leq} E^{\text{sd}}_{\text{trc,MCE}}(R(\cdot)) \overset{\text{(d)}}{\leq} E^{\text{sd}}_{\text{trc,MAP}}(R(\cdot)), \qquad \text{(G.20)}$$

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 7, 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 (G.20) 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 H

Proof of Theorem 10

The left equality in (61) is implied by the proved equality in passage (d) in (G.20). In order to prove the right equality in (61), first note that $E^{\text{sd}}_{\text{trc,SCE}}(R(\cdot)) \leq E^{\text{sd}}_{\text{trc,MAP}}(R(\cdot))$ 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,

$$\gamma(R(\cdot), Q_U, Q_V) = \max_{\left\{\substack{Q_{\tilde{U}|V}:\ Q_{\tilde{U}}=Q_U,\\ H_Q(\tilde{U}|V)\geq R(Q_{\tilde{U}})}\right\}} \{f(Q_{\tilde{U}V}) + H_Q(\tilde{U}|V)\} - R(Q_U) = -R(Q_U), \tag{H.1}$$

as well as

$$\begin{aligned}
\Psi(R(\cdot), Q_{UU'V}) &= [\max\{f(Q_{UV}), \gamma(R(\cdot), Q_U, Q_V)\} - f(Q_{U'V})]_+ && \text{(H.2)}\\
&= \left[\max\{-H_Q(U|V), -R(Q_U)\} + H_Q(U'|V)\right]_+ && \text{(H.3)}\\
&= \left[H_Q(U'|V) - \min\{H_Q(U|V), R(Q_U)\}\right]_+ && \text{(H.4)}\\
&\geq \left[H_Q(U'|U,V) - \min\{H_Q(U|V), R(Q_U)\}\right]_+ . && \text{(H.5)}
\end{aligned}$$

We have the following

$$\begin{aligned}
&E^{\mathrm{sd}}_{\mathrm{trc,SCE}}(R(\cdot))\\
&= \min_{\mathcal{Q}} \big\{D(Q_{UV}\|P_{UV}) + R(Q_U) - H_Q(U'|U,V)\\
&\qquad + [\max\{f(Q_{UV}), \gamma(R(Q_U), Q_U, Q_V)\} - f(Q_{U'V})]_+\big\} && \text{(H.6)}\\
&\geq \min_{\mathcal{Q}} \big\{D(Q_{UV}\|P_{UV}) + R(Q_U) - H_Q(U'|U,V)\\
&\qquad + \left[H_Q(U'|U,V) - \min\{H_Q(U|V), R(Q_U)\}\right]_+\Big\} && \text{(H.7)}\\
&= \min_{\mathcal{Q}} \big\{D(Q_{UV}\|P_{UV}) - \min\{H_Q(U|V), H_Q(U'|U,V), R(Q_U)\} + R(Q_U)\big\} && \text{(H.8)}\\
&\geq \min_{\mathcal{Q}} \big\{D(Q_{UV}\|P_{UV}) - \min\{H_Q(U|V), H_Q(U'), R(Q_U)\} + R(Q_U)\big\} && \text{(H.9)}\\
&= \min_{\{Q_{UV}:\ H_Q(U)\geq R(Q_U)\}} \{D(Q_{UV}\|P_{UV}) - \min\{H_Q(U|V), H_Q(U), R(Q_U)\} + R(Q_U)\} && \text{(H.10)}\\
&= \min_{\{Q_{UV}:\ H_Q(U)\geq R(Q_U)\}} \{D(Q_{UV}\|P_{UV}) - \min\{H_Q(U|V), R(Q_U)\} + R(Q_U)\} && \text{(H.11)}\\
&= \min_{\{Q_{UV}:\ H_Q(U)\geq R(Q_U)\}} \{D(Q_{UV}\|P_{UV}) + \max\{R(Q_U) - H_Q(U|V), 0\}\} && \text{(H.12)}\\
&= \min_{\{Q_{UV}:\ H_Q(U)\geq R(Q_U)\}} \{D(Q_{UV}\|P_{UV}) + [R(Q_U) - H_Q(U|V)]_+\} && \text{(H.13)}\\
&= E^{\mathrm{sd}}_{\mathrm{trc,MAP}}(R(\cdot)), && \text{(H.14)}
\end{aligned}$$

which completes the proof of the theorem.

Appendix I

Proof of Theorem 11

It follows by the identities $\min_{\{Q:\ g(Q)\leq 0\}} f(Q) = \min_Q \sup_{s\geq 0}\{f(Q) + s\cdot g(Q)\}$ and $[A]_+ = \max_{\mu\in[0,1]} \mu A$ that (56) can also be written as

$$E_{\mathrm{e}}^{\mathrm{sd}}(R(\cdot)) = \min_{Q_U}\min_{Q_{V|U}}\max_{\mu\in[0,1]}\sup_{\sigma\geq 0}\{D(Q_{UV}\|P_{UV}) + \mu\cdot(R(Q_U) - H_Q(U|V)) \\ + \sigma\cdot(R(Q_U) - H_Q(U))\}, \tag{I.1}$$

such that $E_{\mathrm{e}}^{\mathrm{sd}}(R(\cdot)) \geq \mathsf{E}_e$ is equivalent to

$$\forall Q_U,\ \forall Q_{V|U},\ \exists\mu\in[0,1],\ \exists\sigma\geq 0: \\ D(Q_{UV}\|P_{UV}) + \mu\cdot(R(Q_U) - H_Q(U|V)) + \sigma\cdot(R(Q_U) - H_Q(U)) \geq \mathsf{E}_e, \tag{I.2}$$

or,

$$\forall Q_U,\ \forall Q_{V|U},\ \exists\mu\in[0,1],\ \exists\sigma\geq 0: \\ R(Q_U) \geq \frac{\mu\cdot H_Q(U|V) + \sigma\cdot H_Q(U) + \mathsf{E}_e - D(Q_{UV}\|P_{UV})}{\mu+\sigma}, \tag{I.3}$$

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

$$\begin{aligned} R(Q_U) &\geq \max_{Q_{V|U}}\min_{\mu\in[0,1]}\inf_{\sigma\geq 0}\left\{\frac{\mu\cdot H_Q(U|V) + \sigma\cdot H_Q(U) + \mathsf{E}_e - D(Q_{UV}\|P_{UV})}{\mu+\sigma}\right\} && \text{(I.4)}\\ &= \max_{Q_{V|U}}\min_{\mu\in[0,1]}\min\left\{H_Q(U), H_Q(U|V) + \frac{\mathsf{E}_e - D(Q_{UV}\|P_{UV})}{\mu}\right\} && \text{(I.5)}\\ &= \max_{Q_{V|U}}\min\left\{H_Q(U), \min_{\mu\in[0,1]}\left\{H_Q(U|V) + \frac{\mathsf{E}_e - D(Q_{UV}\|P_{UV})}{\mu}\right\}\right\} && \text{(I.6)}\\ &= \max_{Q_{V|U}}\begin{cases}\min\{H_Q(U), H_Q(U|V) + \mathsf{E}_e - D(Q_{UV}\|P_{UV})\} & \mathsf{E}_e \geq D(Q_{UV}\|P_{UV})\\ -\infty & \mathsf{E}_e < D(Q_{UV}\|P_{UV})\end{cases} && \text{(I.7)}\\ &= \max_{\{Q_{V|U}:\ D(Q_{UV}\|P_{UV})\leq\mathsf{E}_e\}}\min\{H_Q(U), H_Q(U|V) + \mathsf{E}_e - D(Q_{UV}\|P_{UV})\}, && \text{(I.8)}\end{aligned}$$

and the proof is complete.

References

[1] R. Ahlswede and G. Dueck, “Good codes can be produced by a few permutations,” *IEEE Trans. Inf. Theory*, vol. 28, no. 3, pp. 430–443, May 1982.

[2] R. Averbuch and N. Merhav, “Error exponents of typical random codes of source–channel coding,” *Proc. ITW 2019*, pp. 1–5, Visby, Gotland, Sweden, 2019.

[3] A. Barg and G. D. Forney, Jr., “Random codes: minimum distances and error exponents,” *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2568–2573, September 2002.

[4] J. Chen, D.-K. He, A. Jagmohan, and L. A. Lastras-Montãno, “On the reliability function of variable–rate Slepian–Wolf coding,” Entropy, vol. 19, 389, 2017. doi:10.3390/e19080389

[5] I. Csiszár and J. Körner, “Graph decomposition: a new key to coding theorems,” *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 5–12, January 1981.

[6] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Cambridge University Press, 2011.

[7] R. G. Gallager, *Information Theory and Reliable Communication*, New York, Wiley 1968.

[8] R. G. Gallager, “Source coding with side information and universal coding,” LIDS-P-937, M.I.T., 1976.

[9] J. Liu, P. Cuff, and S. Verdú, “On α–decodability and α–likelihood decoder,” in *Proc. 55th Ann. Allerton Conf. Comm. Control Comput.*, Monticello, IL, Oct. 2017.

[10] N. Merhav,“Statistical physics and information theory,” *Foundations and Trends in Communications and Information Theory,* vol. 6, nos. 12, pp. 1–212, 2009.

[11] N. Merhav, “The generalized stochastic likelihood decoder: random coding and expurgated bounds,” *IEEE Trans. on Inform. Theory*, vol. 63, no. 8, pp. 5039–5051, August 2017. See also a correction at *IEEE Trans. on Inform. Theory*, vol. 63, no. 10, pp. 6827–6829, October 2017.

[12] N. Merhav, “Error exponents of typical random codes,” *IEEE Trans. on Inform. Theory*, vol. 64, no. 9, pp. 6223–6235, September 2018.

[13] N. Merhav, “Error exponents of typical random codes for the colored Gaussian channel,” *IEEE Trans. on Inform. Theory*, vol. 65, no. 12, pp. 8164–8179, December 2019.

[14] N. Merhav, “Error exponents of typical random trellis codes,” *IEEE Trans. on Inform. Theory*, vol. 66, no. 4, pp. 2067–2077, April 2020.

[15] N. Merhav, "A Lagrange–dual lower bound to the error exponent of the typical random code," to appear in *IEEE Trans. on Inform. Theory*, 2020.

[16] A. Nazari, A. Anastasopoulos, and S. S. Pradhan, "Error exponent for multiple–access channels: lower bounds," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5095–5115, September 2014.

[17] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Inform. Theory*, vol. 19, no. 4, pp. 471–480, July 1973.

[18] A. Somekh–Baruch, J. Scarlett, and A. Guillén i Fàbregas, "Generalized random Gilbert–Varshamov codes," *IEEE Trans. on Inform. Theory*, vol. 65, no. 5, pp. 3452–3469, May 2019.

[19] R. Tamir (Averbuch), N. Merhav, N. Weinberger, and A. Guillén i Fàbregas, "Large deviations behavior of the logarithmic error probability of random codes," submitted for publication, December 2019.

[20] N. Weinberger and N. Merhav, "Optimum tradeoffs between the error exponent and the excess–rate exponent of variable–rate Slepian–Wolf coding," *IEEE Trans. on Inform. Theory*, vol. 61, no. 4, pp. 2165–2190, April 2015.