Dual-Domain Error Exponent Analysis for Type-by-Type Source Coding with Side Information

Mehdi Dabirnia
Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)
mehdi.dabirnia@cttc.cat

Hamdi Joudeh
Eindhoven University of Technology
h.joudeh@tue.nl

Albert Guillén i Fàbregas
University of Cambridge
Universitat Politècnica de Catalunya
guillen@ieee.org

***Abstract*—This paper studies expurgated random coding bounds and exponents for source coding with side information with a given (possibly mismatched) decoding rule. We propose an expurgation technique that is an iterative version of Gallager's expurgation method for channel coding and enables a direct dual domain derivation of non-asymptotic bounds for discrete sources with arbitrary side information alphabets and decoding metrics. Specializing the bounds to memoryless models a dual domain achievable error exponent for type-by-type random coding is derived and shown to coincide with the Csiszár-Körner exponent obtained via graph decomposition.**

I. Introduction

Slepian and Wolf [1] pioneered the study of source coding with side information. They showed that the lack of the side information at the encoder does not affect the minimum compression rate of the source if the side information is available at the decoder. Gallager later studied the Slepian-Wolf coding problem in [2] and via random coding derived the following achievable random coding error exponent

$$E_{\mathrm{r}}(R) = \max_{\rho\in[0,1]} \rho R - E_0(\rho), \tag{1}$$

where the function $E_0(\rho)$ is defined as

$$E_0(\rho) \triangleq \log \sum_{y\in\mathcal{Y}} P_Y(y) \bigg(\sum_{x\in\mathcal{X}} P_{X|Y}(x|y)^{\frac{1}{1+\rho}} \bigg)^{1+\rho}. \tag{2}$$

The form in (1) is known as a dual domain expression.

Using the method of types and considering universal minimum entropy decoding, Csiszár and Körner in [3] derived an equivalent primal domain expression for $E_{\mathrm{r}}(R)$ given by

$$E_{\mathrm{r}}(R) = \min_{P_{\tilde{X}\tilde{Y}}} D(P_{\tilde{X}\tilde{Y}}\|P_{XY}) + \big|R - H(\tilde{X}|\tilde{Y})\big|^{+}. \tag{3}$$

Gallager also derived a corresponding upper bound $E_{\mathrm{sp}}(R)$ by providing the side information sequence to the encoder, which takes a similar form as that in (1) with optimization over $\rho \geq 0$ [2].

Csiszár and Körner [4] showed existence of a code that achieves the best of (the source coding counterparts of) both random coding and expurgated exponents in a unified manner. The achievable exponent of Csiszár and Körner [4] for an arbitrary memoryless decoding metric q is given by

$$E_q^{\mathrm{ck}}(R) = \max\big\{E_{q,\mathrm{r}}^{\mathrm{ck}}(R), E_{q,\mathrm{ex}}^{\mathrm{ck}}(R)\big\}. \tag{4}$$

where

$$E_{q,\mathrm{r}}^{\mathrm{ck}}(R) = \min_{P_{\hat{X}\tilde{X}\tilde{Y}}\in\mathcal{T}} D(P_{\tilde{X}\tilde{Y}}\|P_{XY}) + |R - H(\hat{X}|\tilde{Y})|^{+}, \tag{5}$$

and

$$E_{q,\mathrm{ex}}^{\mathrm{ck}}(R) = \min_{\substack{P_{\hat{X}\tilde{X}\tilde{Y}}\in\mathcal{T} \\ H(\hat{X}|\tilde{X})\geq R}} D(P_{\tilde{X}\tilde{Y}}\|P_{XY}) + R - H(\hat{X}|\tilde{X},\tilde{Y}), \tag{6}$$

respectively, where the set of distributions $\mathcal{T}$ is defined as

$$\mathcal{T} = \Big\{P_{\hat{X}\tilde{X}\tilde{Y}} \in \mathcal{P}(\mathcal{X}\times\mathcal{X}\times\mathcal{Y}) : P_{\hat{X}} = P_{\tilde{X}}, \mathbb{E}\big[\log q(\hat{X},\tilde{Y})\big] \geq \mathbb{E}\big[\log q(\tilde{X},\tilde{Y})\big]\Big\}. \tag{7}$$

Th code proposed in [4] encodes each source type independently, and in this work is termed type-by-type coding. When the decoding metric is chosen to be the ML decoder, i.e., $q(x,y) = P_{Y|X}(y|x)$, which is optimal in this case due to the type-by-type coding, the exponents in (5) and (6), respectively, reduce to their matched counterparts in (3) and

$$E_{\mathrm{ex}}(R) = \min_{P_{\tilde{X}}} \Big\{ D(P_{\tilde{X}}\|P_X) + \min_{P_{\hat{X}\tilde{X}}: P_{\hat{X}}=P_{\tilde{X}}, H(\hat{X}|\tilde{X})\geq R} \Big\{ \mathbb{E}\big[d(\hat{X},\tilde{X})\big] + R - H(\hat{X}|\tilde{X}) \Big\} \Big\} \tag{8}$$

where $d(\hat{x},\tilde{x})$ is the Bhattacharyya distance, defined as

$$d(\hat{x},\tilde{x}) \triangleq -\log \sum_{y\in\mathcal{Y}} \sqrt{P_{Y|X}(y|\hat{x}) P_{Y|X}(y|\tilde{x})}. \tag{9}$$

Exploiting the connection to the counterpart channel coding problem and using permutation codes, the exponent in (8) was also derived later by Ahlswede and Dueck in [5]. The proofs in [4], [5] rely on either graph-theoretic or combinatorial arguments and do not use random coding or expurgation. Furthermore, dual-domain derivations and expressions for (5) and (6), and hence (4), are currently not known. This is a gap in the literature that we aim to address in the present paper.

This work was supported in part by the European Research Council under Grants 101116550, 101142747 and 101158232, and in part by the Spanish Government under Grants PID2020-116683GB-C22 and PID2021-128373OB-I00 and partially by the WAVE-XL project (PID2024-160457OB-I00) funded by MCIN/AEI/10.13039/501100011033 and by FEDER, UE.

The main objective of this paper is to develop an expurgation technique for source coding with side information that allows to derive directly in dual domain achievable non-asymptotic bounds and error exponents for source coding with side information under generic, and possibly mismatched, decoding metrics.

A. Type-by-Type Random Coding

Consider a pair of discrete memoryless correlated sources with finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, and joint distribution P_{XY}. Consider partitioning the codeword set $\mathcal{M} = \{1, \cdots, M\}$ into $|\mathcal{P}_n(\mathcal{X})|$ subsets as: $\mathcal{M} = \bigcup_{i=1}^{|\mathcal{P}_n(\mathcal{X})|} \mathcal{M}_i$ where $\mathcal{M}_i$ is the codeword set for source type $\hat{P}_i$, $i \in \{1, \cdots, |\mathcal{P}_n(\mathcal{X})|\}$, $|\mathcal{M}_i| = \frac{M}{|\mathcal{P}_n(\mathcal{X})|}$ and $\mathcal{M}_i \cap \mathcal{M}_j = \emptyset$ for $i \neq j$.

A type-by-type block source code $\boldsymbol{\mathcal{C}} = \{\mathcal{C}_1, \cdots, \mathcal{C}_{|\mathcal{P}_n(\mathcal{X})|}\}$ is the union of $|\mathcal{P}_n(\mathcal{X})|$ codes where each $\mathcal{C}_i$ is an (n, R_i) block code for source type $\hat{P}_i$ with mapping function $\phi_i : \mathcal{T}_n(\hat{P}_i) \rightarrow \mathcal{M}_i$ and rate $R_i = \frac{\log M}{n} - \delta_n$ for $i \in \{1, \cdots, |\mathcal{P}_n(\mathcal{X})|\}$, where $\delta_n = \frac{\log |\mathcal{P}_n(\mathcal{X})|}{n}$. In other words, code $\mathcal{C}_i$ is a code for the source sequences in source type class $\mathcal{T}_n(\hat{P}_i)$ with codeword set $\mathcal{M}_i$. Observe that by construction, every code $\mathcal{C}_i$ has the same rate, and that an error can only occur between source sequences of the same type. We also note that the effect of partitioning the codeword set on the coding rate vanishes asymptotically since $\delta_n \to 0$ as $n \to \infty$.

Definition 1. *The type-by-type random coding ensemble is the set of all (n, R) block codes for the source alphabet $\mathcal{X}$ with a probability measure over the codes having the following property: For every source type $\hat{P}_i$, $i \in \{1, \cdots, |\mathcal{P}_n(\mathcal{X})|\}$, each source sequence $\boldsymbol{x} \in \mathcal{T}_n(\hat{P}_i)$ is independently assigned with equal probability $\frac{1}{|\mathcal{M}_i|}$ to each of the codewords in $\mathcal{M}_i$.*

The decoder for a type-by-type code $\boldsymbol{\mathcal{C}}$ is a set of mappings $\psi_i : \mathcal{M}_i \times \mathcal{Y}^n \rightarrow \mathcal{T}_n(\hat{P}_i)$ for every $i \in \{1, \cdots, |\mathcal{P}_n(\mathcal{X})|\}$. We consider using a maximum metric decoder as follows. For every $m \in \mathcal{M}_i$

$$\psi_i(m, \boldsymbol{y}) = \operatorname*{arg\,max}_{\boldsymbol{x} \in \mathcal{T}_n(\hat{P}_i) : \phi_i(\boldsymbol{x}) = m} q(\boldsymbol{x}, \boldsymbol{y}), \tag{10}$$

where $q(\boldsymbol{x}, \boldsymbol{y})$ is an arbitrary non-negative decoding metric which is possibly mismatched with respect to the source and side information joint distribution. Given a code and a decoding metric, the probability that source sequence $\boldsymbol{x}$ is decoded incorrectly is denoted by $p_e(\boldsymbol{x}, \boldsymbol{\mathcal{C}})$.

II. Main results

Our first result is an expurgation method for source coding that is valid for general source and side information models and arbitrary decoding metrics. The expurgation method is applied independently for each type and for an arbitrary given type $\hat{P}_i$ it can be summarized as follows. We first use Gallager's expurgation technique, developed for channel coding [6], to show that there exists a code in the ensemble such that at least half of the source sequences of type $\hat{P}_i$ satisfy a desired error bound. We expurgate the "bad" half of source sequences, encode them separately into a new set of codewords, and apply expurgation again. The error bound derived in the previous iteration remains valid here, since we now have fewer source sequences. The procedure stops after (at most) $k = n \log_2 |\mathcal{T}_n(\hat{P}_i)|$ iterations, once all source sequences in the type class are exhausted. Combining the expurgated codes from all iterations and all types, we obtain a code in which all the source sequences satisfy the desired error bound given by the following lemma; the proof can be found in an extended version of this paper [7].

Lemma 1. *There exists a code $\boldsymbol{\mathcal{C}}_{\mathrm{ex}} = \{\mathcal{C}_{\mathrm{ex},1}, \cdots, \mathcal{C}_{\mathrm{ex},|\mathcal{P}_n(\mathcal{X})|}\}$ in the type-by-type random coding ensemble such that for every $i \in \{1, \cdots, |\mathcal{P}_n(\mathcal{X})|\}$ and for every source sequence $\boldsymbol{x} \in \mathcal{T}_n(\hat{P}_i)$ and $\rho \geq 0$*

$$p_e(\boldsymbol{x}, \boldsymbol{\mathcal{C}}_{\mathrm{ex}}) \leq \left(2\mathbb{E}\left[p_e(\boldsymbol{x}, \mathsf{C}_i)^{\frac{1}{\rho}}\right]\right)^{\rho}, \tag{11}$$

where the expectation is over the random coding ensemble for the corresponding type with $\frac{M}{k_i |\mathcal{P}_n(\mathcal{X})|}$ codewords and $k_i = \log_2 |\mathcal{T}_n(\hat{P}_i)| \leq n \log_2 |\mathcal{X}|$.

Applying Lemma 1 we obtain the following achievable error exponent for memoryless sources employing a memoryless decoding metric $q(\boldsymbol{x}, \boldsymbol{y}) = \prod_{i=1}^{n} q(x_i, y_i)$.

Theorem 1. *For every $R > 0$ and every distribution $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ there exists a type-by-type block source code with maximum metric decoder* (10) *employing decoding metric $q(x, y)$ that achieves the exponent*

$$E_q^{\mathrm{tt}}(R) = \max\{E_{q,\mathrm{r}}^{\mathrm{tt}}(R), E_{q,\mathrm{ex}}^{\mathrm{tt}}(R)\} \tag{12}$$

where

$$E_{q,\mathrm{r}}^{\mathrm{tt}}(R) = \sup_{\substack{\rho \in [0,1] \\ s \geq 0, a(\cdot)}} \rho R - E_{\mathrm{s}}(\rho, s, a(\cdot)), \tag{13}$$

$$E_{q,\mathrm{ex}}^{\mathrm{tt}}(R) = \sup_{\substack{\rho \geq 1 \\ s \geq 0, a(\cdot)}} \rho R - E_{\mathrm{x}}(\rho, s, a(\cdot)), \tag{14}$$

with

$$E_{\mathrm{s}}(\rho, s, a(\cdot)) = \log \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} P_{XY}(x, y) \left(\sum_{\bar{x} \in \mathcal{X}} \frac{e^{a(\bar{x})}}{e^{a(x)}} \left(\frac{q(\bar{x}, y)}{q(x, y)} \right)^s \right)^{\rho} \tag{15}$$

$$E_{\mathrm{x}}(\rho, s, a(\cdot)) = \log \sum_{x \in \mathcal{X}} \left(\sum_{\bar{x} \in \mathcal{X}} \left(\sum_{y \in \mathcal{Y}} P_{XY}(x, y) \frac{e^{a(\bar{x})}}{e^{a(x)}} \left(\frac{q(\bar{x}, y)}{q(x, y)} \right)^s \right)^{\frac{1}{\rho}} \right)^{\rho} \tag{16}$$

and the optimization is over real parameters ρ, s and real-valued functions $a : \mathcal{X} \to \mathbb{R}$.

The proof of Theorem 1 can be found in section IV. In deriving the error exponent of Theorem 1 we derive n-letter bounds in (31) and (44) that are valid for any discrete source with arbitrary side information alphabets and arbitrary decoding metrics without the memoryless assumption. These bounds can be used to derive error exponents for general source models.

For the memoryless case, the following result, whose proof can be found in an extended version of this paper [7], shows that the error exponent introduced in Theorem 1 coincides with Csiszár and Körner's [4] for each term in the maximization, namely, $E^{\rm tt}_{q,\rm r}(R) = E^{\rm ck}_{q,\rm r}(R)$ and $E^{\rm tt}_{q,\rm ex}(R) = E^{\rm ck}_{q,\rm ex}(R)$.

Proposition 1 (Primal-dual equivalence)**.** *The dual-domain error exponent derived in Theorem 1 coincides with the Csiszár-Körner exponent* (4) *derived in the primal domain via graph decomposition, i.e.,*

$$E^{\rm tt}_q(R) = E^{\rm ck}_q(R). \tag{17}$$

Noticing that exponent is a convex function of the rate and the maximizing ρ is the slope of the exponent curve, similar to [8, Ch. 5] by evaluating the partial derivative of the exponent to find the rate at which $\rho = 0$ maximizes the exponent, we obtain the following rate achieved by type-by-type random coding:

$$H^{\rm tt}_q(X|Y) = \inf_{s\geq 0, a(\cdot)} -\sum_{x,y} P_{XY}(x,y)\log\frac{q(x,y)^s e^{a(x)}}{\sum_{\bar{x}} q(\bar{x},y)^s e^{a(\bar{x})}} \tag{18}$$

$$= H(X|Y) + \inf_{s\geq 0, a(\cdot)} D(P_{X|Y}\|Q^{(s,a(\cdot))}_{X|Y}), \tag{19}$$

where $Q^{(s,a(\cdot))}_{X|Y}(x|y) = \frac{q(x,y)^s e^{a(x)}}{\sum_{\bar{x}\in\mathcal{X}} q(\bar{x},y)^s e^{a(\bar{x})}}$.

III. Example

The joint distribution of the source X with side information Y is defined by the entries of the $|\mathcal{X}|\times|\mathcal{Y}|$ matrix

$$P_{XY} = \begin{bmatrix} 0.588 & 0.006 & 0.006 \\ 0.03 & 0.24 & 0.03 \\ 0.02 & 0.02 & 0.06 \end{bmatrix} \tag{20}$$

with $\mathcal{X} = \mathcal{Y} = \{0,1,2\}$. We consider using a mismatched decoder with a memoryless metric given by the matrix

$$q(x,y) = \begin{bmatrix} 1-2\delta & \delta & \delta \\ \delta & 1-2\delta & \delta \\ \delta & \delta & 1-2\delta \end{bmatrix} \tag{21}$$

with $\delta \in (0,\frac{1}{3})$. For any value of the $\delta \in (0,\frac{1}{3})$, the corresponding decoder is exactly the same and is equivalent to a minimum Hamming distance decoding metric.

Figure 1 illustrates the exponents for type-by-type ensemble with both matched and mismatched decoders as well as those of the standard ensemble which are denoted by $E_q(R)$. The setting considering standard ensemble has been recently studied in [9] and the corresponding exponent $E_q(R)$ can be obtained from the type-by-type exponent $E^{\rm tt}_q(R)$ by setting the cost function $a(\cdot)$ equal to a constant. In standard

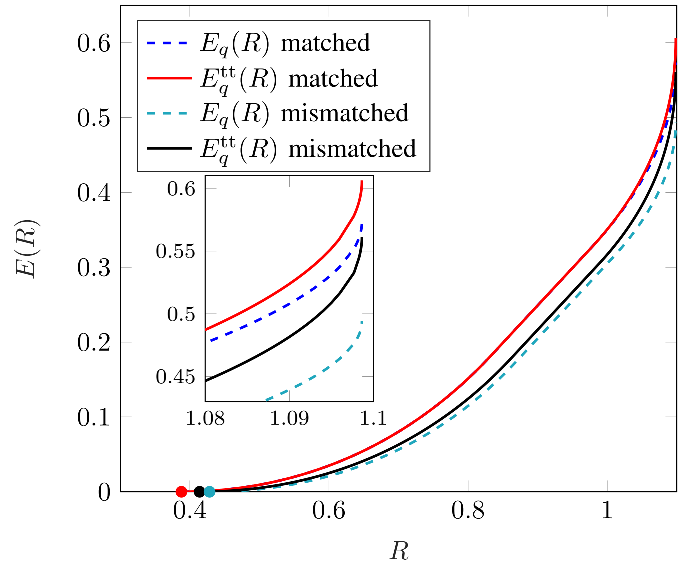

Fig. 1: Error exponents for the source X and side information Y with joint distribution given in (20). The mismatched decoder uses the minimum Hamming distance metric.

ensemble, all source sequences are randomly encoded to the same set of codewords; this is in contrast with type-by-type ensemble where sequences of each type are randomly encoded to a distinct set of codewords. We observe that the type-by-type expurgated exponent is higher for both matched and mismatched decoding. In the matched case, the random coding components of $E_q(R)$ and $E^{\rm tt}_q(R)$ coincide. The corresponding achievable rates are marked with dots in the figure. The conditional entropy $H(X|Y) = 0.3879$ nats is the limit for the matched case while in the mismatched case we have $H_q(X|Y) = 0.4283$ nats for standard ensemble and $H^{\rm tt}_q(X|Y) = 0.4140$ nats for type-by-type ensemble.

IV. Proof of Theorem 1

Lemma 1 showed the existence of a good code in the type-by-type random coding ensemble which satisfies an upper bound on the error probability for every source sequence. Here we show that this code achieves the exponent of Theorem 1. We first show the achievability of $E^{\rm tt}_{q,\rm ex}(R)$ and then that of $E^{\rm tt}_{q,\rm r}(R)$.

We start by bounding $P_e(\boldsymbol{x},\mathcal{C}_i)$ for a given source sequence $\boldsymbol{x}\in\mathcal{T}_n(\hat{P}_i)$ and given type-by-type code $\boldsymbol{\mathcal{C}}$ with subcode $\mathcal{C}_i$ for type $\hat{P}_i$ as follows

$$P_e(\boldsymbol{x},\mathcal{C}_i)$$

$$= \sum_{\boldsymbol{y}\in\mathcal{Y}^n} P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})\mathbb{1}\left[\bigcup_{\substack{\bar{\boldsymbol{x}}\in\mathcal{T}_n(\hat{P}_i)\setminus\boldsymbol{x}\\ \phi_i(\bar{\boldsymbol{x}})=\phi_i(\boldsymbol{x})}}\{q(\bar{\boldsymbol{x}},\boldsymbol{Y})\geq q(\boldsymbol{x},\boldsymbol{Y})\}\right] \tag{22}$$

$$\leq \sum_{\boldsymbol{y}\in\mathcal{Y}^n} P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \sum_{\substack{\bar{\boldsymbol{x}}\in\mathcal{T}_n(\hat{P}_i)\setminus\boldsymbol{x}\\ \phi_i(\bar{\boldsymbol{x}})=\phi_i(\boldsymbol{x})}} \mathbb{1}\left[q(\bar{\boldsymbol{x}},\boldsymbol{y})\geq q(\boldsymbol{x},\boldsymbol{y})\right] \tag{23}$$

$$\leq \sum_{\bar{\boldsymbol{x}} \in \mathcal{T}_n(\hat{P}_i)} \mathbb{1}\left[\phi_i(\bar{\boldsymbol{x}}) = \phi_i(\boldsymbol{x})\right] \sum_{\boldsymbol{y}} P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \left(\frac{q(\bar{\boldsymbol{x}}, \boldsymbol{y})}{q(\boldsymbol{x}, \boldsymbol{y})}\right)^s \tag{24}$$

where we use union bound in (23) and (24) follows from including $\bar{\boldsymbol{x}} = \boldsymbol{x}$ in the summation and holds for any $s \geq 0$.

Now considering the ensemble of random type-by-type codes and denoting the induced random encoding function by $\Phi_i(\cdot)$, we upper bound the $\mathbb{E}\left[P_e(\boldsymbol{x}, \mathsf{C}_i)^{\frac{1}{\rho}}\right]$ using (24) as follows

$$\begin{aligned}
&\mathbb{E}\left[P_e(\boldsymbol{x}, \mathsf{C}_i)^{\frac{1}{\rho}}\right] \\
&\leq \mathbb{E}\left[\left(\sum_{\bar{\boldsymbol{x}} \in \mathcal{T}_n(\hat{P}_i)} \mathbb{1}\left[\Phi_i(\bar{\boldsymbol{x}}) = \Phi_i(\boldsymbol{x})\right] \right.\right. \\
&\qquad \left.\left. \times \sum_{\boldsymbol{y} \in \mathcal{Y}^n} P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \left(\frac{q(\bar{\boldsymbol{x}}, \boldsymbol{y})}{q(\boldsymbol{x}, \boldsymbol{y})}\right)^s\right)^{\frac{1}{\rho}}\right] &(25)\\
&\leq \sum_{\bar{\boldsymbol{x}} \in \mathcal{T}_n(\hat{P}_i)} \mathbb{E}\left[\mathbb{1}\left[\Phi_i(\bar{\boldsymbol{x}}) = \Phi_i(\boldsymbol{x})\right]\right] \\
&\qquad \times \left(\sum_{\boldsymbol{y} \in \mathcal{Y}^n} P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \left(\frac{q(\bar{\boldsymbol{x}}, \boldsymbol{y})}{q(\boldsymbol{x}, \boldsymbol{y})}\right)^s\right)^{\frac{1}{\rho}} &(26)\\
&\leq \frac{k_i |\mathcal{P}_n(\mathcal{X})|}{M} \sum_{\bar{\boldsymbol{x}} \in \mathcal{T}_n(\hat{P}_i)} \left(\sum_{\boldsymbol{y} \in \mathcal{Y}^n} P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \left(\frac{q(\bar{\boldsymbol{x}}, \boldsymbol{y})}{q(\boldsymbol{x}, \boldsymbol{y})}\right)^s\right)^{\frac{1}{\rho}} &(27)
\end{aligned}$$

where (26) follows from inequality $(\sum_i a_i)^{\frac{1}{\rho}} \leq \sum_i a_i^{\frac{1}{\rho}}$ for $\rho \geq 1$, (27) follows from $\mathbb{E}\left[\mathbb{1}\left[\Phi_i(\bar{\boldsymbol{x}}) = \Phi_i(\boldsymbol{x})\right]\right] \leq \frac{k_i |\mathcal{P}_n(\mathcal{X})|}{M}$ where $k_i = \log_2 |\mathcal{T}_n(\hat{P}_i)|$ is an upper bound on the number of iterations in the expurgation method.

Now substituting (27) in (11) from Lemma 1 and summing over all source sequences of all types we find an upper bound on the error probability of the codebook $\boldsymbol{\mathcal{C}}_{\text{ex}}$ as

$$p_e(\boldsymbol{\mathcal{C}}_{\text{ex}}) = \sum_{i=1}^{|\mathcal{P}_n(\mathcal{X})|} p_e(\mathcal{C}_{\text{ex},i}) \tag{28}$$

where

$$\begin{aligned}
&p_e(\mathcal{C}_{\text{ex},i}) \leq \left(\frac{2k_i |\mathcal{P}_n(\mathcal{X})|}{M}\right)^{\rho} \times \\
&\sum_{\boldsymbol{x} \in \mathcal{T}_n(\hat{P}_i)} \left(\sum_{\bar{\boldsymbol{x}} \in \mathcal{T}_n(\hat{P}_i)} \left(\sum_{\boldsymbol{y}} P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \left(\frac{q(\bar{\boldsymbol{x}}, \boldsymbol{y})}{q(\boldsymbol{x}, \boldsymbol{y})}\right)^s\right)^{\frac{1}{\rho}}\right)^{\rho} &(29)
\end{aligned}$$

We can further upper bound the average error probability $p_e(\boldsymbol{\mathcal{C}}_{\text{ex}})$ as

$$p_e(\boldsymbol{\mathcal{C}}_{\text{ex}}) = \max_i p_e(\mathcal{C}_{\text{ex},i}) \tag{30}$$

which implies that the corresponding error exponent is dominated by the exponent of the worst type.

In order to find a simpler bound that does not require maximization over the type, we can weaken the bound in (29) by including all $\bar{\boldsymbol{x}}$ in the sum, however to keep the bound tight we introduce ratio of an arbitrary cost function as $\frac{e^{a(\bar{\boldsymbol{x}})}}{e^{a(\boldsymbol{x})}}$ where the cost function $a(\boldsymbol{x})$ depends on the sequence $\boldsymbol{x}$ only through its type. Observe that the ratio is equal to 1 when both $\boldsymbol{x}$ and $\bar{\boldsymbol{x}}$ have the same type, also this cost function can be optimized for to obtain the tightest bound. Simplifying (29) and upper-bounding k_i by $k = n \log_2 |\mathcal{X}|$ we obtain

$$\begin{aligned}
&p_e(\boldsymbol{\mathcal{C}}_{\text{ex}}) \leq \left(\frac{2k |\mathcal{P}_n(\mathcal{X})|}{M}\right)^{\rho} \sum_{\boldsymbol{x} \in \mathcal{X}^n} P_{\boldsymbol{X}}(\boldsymbol{x}) \\
&\times \left(\sum_{\bar{\boldsymbol{x}} \in \mathcal{X}^n} \left(\sum_{\boldsymbol{y} \in \mathcal{Y}^n} P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \frac{e^{a(\bar{\boldsymbol{x}})}}{e^{a(\boldsymbol{x})}} \left(\frac{q(\bar{\boldsymbol{x}}, \boldsymbol{y})}{q(\boldsymbol{x}, \boldsymbol{y})}\right)^s\right)^{\frac{1}{\rho}}\right)^{\rho}. &(31)
\end{aligned}$$

Equation (31) is valid for any discrete source and any decoding metric. We now specialize this to the case of memoryless sources, metrics, and cost functions using $q(\boldsymbol{x}, \boldsymbol{y}) = \prod_{i=1}^n q(x_i, y_i)$ and $a(\boldsymbol{x}) = \sum_{i=1}^n a(x_i)$. Therefore we obtain

$$\begin{aligned}
&p_e(\boldsymbol{\mathcal{C}}_{\text{ex}}) \leq \left(\frac{2k |\mathcal{P}_n(\mathcal{X})|}{M}\right)^{\rho} \left(\sum_{x \in \mathcal{X}} P_X(x) \right. \\
&\left. \times \left(\sum_{\bar{x} \in \mathcal{X}} \left(\sum_{y \in \mathcal{Y}} P_{Y|X}(y|x) \frac{e^{a(\bar{x})}}{e^{a(x)}} \left(\frac{q(\bar{x}, y)}{q(x, y)}\right)^s\right)^{\frac{1}{\rho}}\right)^{\rho}\right)^n &(32)\\
&= e^{-n(\rho(R - \delta_n) - E_{\text{x}}(\rho, s, a(\cdot)))} &(33)
\end{aligned}$$

where $\delta_n = \frac{\log 2k |\mathcal{P}_n(\mathcal{X})|}{n} \to 0$ as $n \to \infty$ and $E_{\text{x}}(\rho, s, a(\cdot))$ is given in (16). Hence, we show the achievability of the exponent $E_{q,\text{ex}}(R)$ by optimizing over $\rho, s, a(\cdot)$ as

$$E_{q,\text{ex}}^{\text{tt}}(R) = \max_{\rho \geq 1, s \geq 0, a(\cdot)} \rho R - E_{\text{x}}(\rho, s, a(\cdot)). \tag{34}$$

To show that the same code of Lemma 1 achieves $E_{q,\text{r}}^{\text{tt}}(R)$ we use (11) of Lemma 1 with $\rho = 1$, and obtain that for every source sequence $\boldsymbol{x} \in \mathcal{T}_n(\hat{P}_i)$

$$p_e(\boldsymbol{x}, \boldsymbol{\mathcal{C}}_{\text{ex}}) \leq 2\mathbb{E}\left[P_e(\boldsymbol{x}, \mathsf{C}_i)\right]. \tag{35}$$

Averaging over all source sequences we obtain

$$\begin{aligned}
p_e(\boldsymbol{\mathcal{C}}_{\text{ex}}) &\leq 2 \sum_{i=1}^{|\mathcal{P}_n(\mathcal{X})|} \sum_{\boldsymbol{x} \in \mathcal{T}_n(\hat{P}_i)} P_{\boldsymbol{X}}(\boldsymbol{x}) \mathbb{E}\left[P_e(\boldsymbol{x}, \mathsf{C}_i)\right] &(36)\\
&= 2\mathbb{E}\left[p_e(\mathbf{C})\right], &(37)
\end{aligned}$$

which shows that the error probability of $\boldsymbol{\mathcal{C}}_{\text{ex}}$ is upper bounded by twice the ensemble average error probability.

Now we derive an upper bound on the type-by-type ensemble average error probability following the derivations in [2].

We start by bounding $p_e(\boldsymbol{y}, \boldsymbol{\mathcal{C}})$ for a given side information sequence $\boldsymbol{y}$ and given type-by-type code $\boldsymbol{\mathcal{C}}$ as follows

$$
\begin{aligned}
&p_e(\boldsymbol{y},\boldsymbol{\mathcal{C}})\\
&=\sum_{i=1}^{|\mathcal{P}_n(\mathcal{X})|}\sum_{\boldsymbol{x}\in\mathcal{T}_n(\hat{P}_i)}P_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})\\
&\qquad\times\mathbb{1}\left[\bigcup_{\substack{\bar{\boldsymbol{x}}\in\mathcal{T}_n(\hat{P}_i)\setminus\boldsymbol{x}\\ \phi_i(\bar{\boldsymbol{x}})=\phi_i(\boldsymbol{x})}}\{q(\bar{\boldsymbol{x}},\boldsymbol{y})\geq q(\boldsymbol{x},\boldsymbol{y})\}\right] \qquad (38)\\
&\leq\sum_{i=1}^{|\mathcal{P}_n(\mathcal{X})|}\sum_{\boldsymbol{x}\in\mathcal{T}_n(\hat{P}_i)}P_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})\left(\sum_{\substack{\bar{\boldsymbol{x}}\in\mathcal{T}_n(\hat{P}_i)\setminus\boldsymbol{x}\\ \phi_i(\bar{\boldsymbol{x}})=\phi_i(\boldsymbol{x})}}\left(\frac{q(\bar{\boldsymbol{x}},\boldsymbol{y})}{q(\boldsymbol{x},\boldsymbol{y})}\right)^s\right)^\rho \qquad (39)
\end{aligned}
$$

where (39) follows from using the inequality $\mathbb{1}\left[\bigcup_i A_i\right]\leq\left(\sum_i\mathbb{1}\left[A_i\right]\right)^\rho$ for any set of events $\{A_i\}$ and $\rho\in[0,1]$ and it holds for any $s\geq 0$.

Now considering the ensemble of random type-by-type block source codes we upper bound the $\mathbb{E}\left[p_e(\boldsymbol{y},\mathbf{C})\right]$ using (39) as follows

$$
\begin{aligned}
&\mathbb{E}\left[p_e(\boldsymbol{y},\mathbf{C})\right] \qquad (40)\\
&\leq\sum_{i=1}^{|\mathcal{P}_n(\mathcal{X})|}\sum_{\boldsymbol{x}\in\mathcal{T}_n(\hat{P}_i)}P_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})\\
&\times\left(\sum_{\bar{\boldsymbol{x}}\in\mathcal{T}_n(\hat{P}_i)\setminus\boldsymbol{x}}\left(\frac{q(\bar{\boldsymbol{x}},\boldsymbol{y})}{q(\boldsymbol{x},\boldsymbol{y})}\right)^s\mathbb{E}\left[\mathbb{1}\left[\Phi_i(\bar{\boldsymbol{x}})=\Phi_i(\boldsymbol{x})\right]\right]\right)^\rho \qquad (41)\\
&\leq\left(\frac{k|\mathcal{P}_n(\mathcal{X})|}{M}\right)^\rho\\
&\qquad\times\sum_{\boldsymbol{x}\in\mathcal{X}^n}P_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})\left(\sum_{\bar{\boldsymbol{x}}\in\mathcal{X}^n}\frac{e^{a(\bar{\boldsymbol{x}})}}{e^{a(\boldsymbol{x})}}\left(\frac{q(\bar{\boldsymbol{x}},\boldsymbol{y})}{q(\boldsymbol{x},\boldsymbol{y})}\right)^s\right)^\rho, \qquad (42)
\end{aligned}
$$

where (41) follows from Jensen's inequality and the concavity of x^ρ for $\rho\in[0,1]$ and (42) follows from $\mathbb{E}\left[\mathbb{1}\left[\Phi(\bar{\boldsymbol{x}})=\Phi(\boldsymbol{x})\right]\right]=\frac{k_i}{M}$ and upper bounding k_i by $k=n\log_2|\mathcal{X}|$ and including all $\bar{\boldsymbol{x}}$ in the sum, however to keep the bound tight we introduce ratio of an arbitrary cost function as $\frac{e^{a(\bar{\boldsymbol{x}})}}{e^{a(\boldsymbol{x})}}$ where the cost function $a(\boldsymbol{x})$ depends on the sequence $\boldsymbol{x}$ only through its type. Notice that the ratio is equal to 1 when both $\boldsymbol{x}$ and $\bar{\boldsymbol{x}}$ have the same type, also this cost function can be optimized for to obtain the tightest bound.

Averaging over all side information sequences we obtain an upper bound on the type-by-type ensemble average error probability as

$$
\begin{aligned}
&\mathbb{E}\left[p_e(\mathbf{C})\right]=\sum_{\boldsymbol{y}\in\mathcal{Y}^n}P_{\boldsymbol{Y}}(\boldsymbol{y})\mathbb{E}\left[p_e(\boldsymbol{y},\mathbf{C})\right] \qquad (43)\\
&\leq\left(\frac{k|\mathcal{P}_n(\mathcal{X})|}{M}\right)^\rho\\
&\qquad\times\sum_{\boldsymbol{x},\boldsymbol{y}}P_{\boldsymbol{X}\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y})\left(\sum_{\bar{\boldsymbol{x}}\in\mathcal{X}^n}\frac{e^{a(\bar{\boldsymbol{x}})}}{e^{a(\boldsymbol{x})}}\left(\frac{q(\bar{\boldsymbol{x}},\boldsymbol{y})}{q(\boldsymbol{x},\boldsymbol{y})}\right)^s\right)^\rho \qquad (44)
\end{aligned}
$$

where (44) holds for any $\rho\in[0,1]$ and $s\geq 0$. Introducing (44) in (37) and particularizing it to the case of memoryless sources, metrics and cost functions we obtain

$$
\begin{aligned}
&p_e(\boldsymbol{\mathcal{C}}_{\mathrm{ex}})\leq 2\left(\frac{k|\mathcal{P}_n(\mathcal{X})|}{M}\right)^\rho\\
&\times\left(\sum_{x\in\mathcal{X},y\in\mathcal{Y}}P_{XY}(x,y)\left(\sum_{\bar{x}\in\mathcal{X}}\frac{e^{a(\bar{x})}}{e^{a(x)}}\left(\frac{q(\bar{x},y)}{q(x,y)}\right)^s\right)^\rho\right)^n \qquad (45)\\
&=e^{-n(\rho(R-\delta_n)-E_{\mathrm{s}}(\rho,s,a(\cdot))-\delta'_n)} \qquad (46)
\end{aligned}
$$

where $\delta_n=\frac{\log k|\mathcal{P}_n(\mathcal{X})|}{n}\to 0$ and $\delta'_n=\frac{\log 2}{n}\to 0$ as $n\to\infty$ and and $E_{\mathrm{s}}(\rho,s,a(\cdot))$ is given in (15). Hence, we show the achievability of the exponent $E^{\mathrm{tt}}_{q,\mathrm{r}}(R)$ by optimizing over $\rho,s,a(\cdot)$ as in (13).

REFERENCES

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

[2] R. G. Gallager, "Source coding with side information and universal coding," M.I.T. Technical Report LIDS-P-937, Tech. Rep., 1979.

[3] I. Csiszár and J. Körner, "Towards a general theory of source networks," *IEEE Trans. Inf. Theory*, vol. 26, no. 2, pp. 155–165, 1980.

[4] ——, "Graph decomposition: A new key to coding theorems," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 5–12, Jan. 1981.

[5] 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, 1982.

[6] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. 11, no. 1, pp. 3–18, 1965.

[7] M. Dabirnia, H. Joudeh, and A. Guillén i Fàbregas, "Dual domain expurgated error exponents for source coding with side information," *IEEE Trans. Inf. Theory*, early access, 2026.

[8] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, Inc. New York, NY, USA, 1968.

[9] M. Dabirnia, H. Joudeh, and A. Guillén i Fàbregas, "Source coding with side information: A dual-domain expurgated error exponent," in *2026 Proc. Int. Zürich Seminar Inf. Commun. (IZS)), Zürich, Switzerland*, Feb. 2026.