Universal Decoding for the Typical Random Code and for the Expurgated Code

Ran Tamir (Averbuch) and Neri Merhav

***Abstract*— We provide two results concerning the optimality of the stochastic-mutual information (SMI) decoder, which chooses the estimated message according to a posterior probability mass function, which is proportional to the exponentiated empirical mutual information induced by the channel output sequence and the different codewords. First, we prove that the error exponents of the typical random codes under the optimal maximum likelihood (ML) decoder and the SMI decoder are equal. As a corollary to this result, we also show that the error exponents of the expurgated codes under the ML and the SMI decoders are equal. These results strengthen the well-known result due to Csiszár and Körner, according to which, the ML and the maximum-mutual information (MMI) decoders achieve equal random-coding error exponents, since the error exponents of the typical random code and the expurgated code are strictly higher than the random-coding error exponents, at least at low coding rates. The universal optimality of the SMI decoder, in the random-coding error exponent sense, is easily proven by commuting the expectation over the channel noise and the expectation over the ensemble. This commutation can no longer be carried out, when it comes to typical and expurgated exponents. Therefore, the proof of the universal optimality of the SMI decoder must be completely different and it turns out to be highly non-trivial.**

Index Terms: Error exponent, expurgated code, mutual information, stochastic decoding, typical random code, universal decoding.

I. Introduction

The error exponent of the typical random code (TRC) [19] is defined as[1]

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

where R is the coding rate, $P_{\text{e}}(\mathcal{C}_n)$ 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.

In [1], 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. In [23] Nazari *et al.* provided bounds on the error exponents of TRCs for both discrete memoryless channels (DMCs) (which coincide with one another for the optimal input distribution) and multiple–access channels. In a recent article by Merhav [19], 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). Recently, Merhav has studied error exponents of TRCs for the colored Gaussian channel [20], typical random trellis codes [21], and has derived a Lagrange–dual lower bound to the TRC exponent [22]. More recently, Tamir *et al.* have studied the large deviations behavior around the TRC exponent [28], and finally, Tamir and Merhav have studied error exponents of typical random Slepian–Wolf codes in [27].

Concerning universal decoding for unknown channels, Goppa [9] was the first to propose the maximum-mutual information (MMI) decoder, which decodes the message as the one whose codeword has the largest empirical mutual information with the channel output sequence. Goppa proved that for DMCs, MMI decoding attains capacity. Csiszár and Körner [2, Theorem 10.2] have further showed that the random-coding error exponent of the MMI decoder, pertaining to the ensemble of the uniform random-coding distribution over a certain type class, is equal to the random-coding error exponent of the optimum ML decoder. Since the seminal work of [2], much work has been done in the area of universal decoding; the interested reader is referred to [5], [6], [10], [11], [13], [14], [16], [29], and the references therein.

In this work, we refer to the stochastic-mutual information (SMI) decoder, which is a GLD with the empirical mutual information metric. We prove that the error exponents of the TRC under ML and SMI decoding are the same[2]. This result improves upon the universal optimality of the MMI decoder proved in [2], since the error exponent of the TRC is strictly larger than the ordinary random-coding error exponent, at least at low coding rates [19]. The fact that the SMI decoder is optimal also w.r.t. the TRC is non-trivial, at least to the authors of this paper. The proof of optimality of the SMI decoder w.r.t. the random-coding error exponent [17, Section 3] relies heavily on the possibility to commute the expectations over the channel noise and the randomness of the ensemble of codes. Here, in case of TRCs, this can no longer be done, because, by

R. Tamir is with the Department of Information Technology and Electrical Engineering, ETH Zurich, 8092 Zürich, Switzerland (e–mail: tamir@isi.ee.ethz.ch).

N. Merhav is with the Andrew & Erna Viterbi Faculty of Electrical and Computer Engineering, Technion – Israel Institute of Technology, Technion City, Haifa 32000, Israel (e–mail: merhav@ee.technion.ac.il).

The research of R. Tamir and N. Merhav was supported by Israel Science Foundation (ISF) grant no. 137/18.

[1]Note that this definition is different from the ordinary random-coding exponent, which is given by $E_{\text{r}}(R) = \lim_{n\to\infty}\left\{-\frac{1}{n}\log\mathbb{E}\left[P_{\text{e}}(\mathcal{C}_n)\right]\right\}$, where the notations are similar to those in (1) above.

[2]Since the proof in this work does not contain any assumptions on the DMC, it actually follows that the ML and SMI decoders attain equal TRC exponents, even if the DMC has a positive zero-error capacity.

definition of the TRC exponent, we first apply the logarithmic function on the error probability and only then average over the randomness of the codebook. Therefore, the proof of our new result is much more involved than in ordinary random coding.

It may seem surprising, at least at first glance, that a stochastic decoder turns out to be universally optimal. To see why this phenomenon should not be very surprising, after all (even for the traditional random coding error exponents), let us recall the competitive minimax approach, which has been developed in [6]. Consider a parametric family of channels $\{P_\theta(\boldsymbol{y}|\boldsymbol{x})\}$, indexed by a parameter θ taking on values in $\Theta \subseteq \mathbb{R}^k$, k being a positive integer (for example, the class of all DMCs with given finite input and output alphabets). A decision rule is a (possibly randomized) map $\Omega : \mathcal{Y} \to \{0, 1, \ldots, M-1\}$, characterized by a conditional probability vector function

$$\Omega = \{(\omega(0|\boldsymbol{y}), \ldots, \omega(M-1|\boldsymbol{y})), \boldsymbol{y} \in \mathcal{Y}^n\}, \tag{2}$$

with $\omega(i|\boldsymbol{y})$ being the conditional probability of deciding in favor of message i given the channel output $\boldsymbol{y}$, $i = 0, 1, \ldots, M-1$. For a given decision rule Ω, the overall probability of error, for a uniform prior on the message set, is given by

$$P_{\mathrm{e}}(\Omega|\theta) = \frac{1}{M} \sum_{i=0}^{M-1} \sum_{\boldsymbol{y} \in \mathcal{Y}^n} [1 - \omega(i|\boldsymbol{y})] P_\theta(\boldsymbol{y}|\boldsymbol{x}_i). \tag{3}$$

Let $\Omega^*(\theta)$ denote the optimum ML decision rule and denote $P_{\mathrm{e}}^*(\theta) = P_{\mathrm{e}}(\Omega^*(\theta)|\theta)$. Finally, the competitive minimax is defined as [6]

$$\inf_{\Omega} \sup_{\theta \in \Theta} \frac{P_{\mathrm{e}}(\Omega|\theta)}{P_{\mathrm{e}}^*(\theta)}. \tag{4}$$

A universal decoder asymptotically minimizes (4) for the average code (i.e., when both $P_{\mathrm{e}}^*(\theta)$ and $P_{\mathrm{e}}(\Omega|\theta)$ are replaced by their ensemble averages). While in the simple case where the channel statistics are perfectly known (i.e., the set Θ being a singleton), the optimal decision rule is clearly deterministic (since the objective in (4) is an affine function of the conditional probabilities composing Ω), it turns out that under channel uncertainty, the inner supremum in (4) convexifies the objective as a functional of Ω (since a supremum over a family of affine functions yields a convex function), thus, the competitive minimax criterion considered in (4), may yield a randomized decision rule as an optimum solution. A more comprehensive discussion on this issue, along with a simple example, can be found in [6, Section 2], where it was also discussed that for ordinary random coding, the optimal randomized decoder can often be well approximated by an asymptotically optimal deterministic decoder, that achieves the same error exponent. Nonetheless, when it comes to TRC error exponents, it is not apparent that this is guaranteed, in general.

Universal decoding w.r.t. TRCs has already been considered in [27]. It was proved in [27] that for Slepian–Wolf source coding, the error exponent of the TRC under the optimal maximum a-posteriori decoder is equal to the TRC exponent under two different universal decoders: the minimum-entropy decoder and its stochastic counterpart. While the universality result of [27] was obtained for some (semi-deterministic) modification (for which the TRC error exponent is given by a relatively simple expression) of the classic random binning scheme, here, the SMI decoder is proved to be optimal w.r.t. the ordinary (fixed-composition) random-coding scheme. In light of this fact, we conjecture that the SMI decoder is optimal w.r.t. the error exponent of the TRC also for more sophisticated random-coding schemes, like the generalized random Gilbert-Varshamov (RGV) code ensemble [26].

Our second result concerns the optimality of SMI decoding w.r.t. expurgated codes. Error exponents of expurgated codes were first developed for the ML decoder [4], [7], later on for a more general family of deterministic decoders [3], and recently for the GLD [17]. In [3, Section V], the question of finding the channels for which the expurgated exponent can be achieved by the minimum-entropy decoder (which is equivalent to the MMI decoder under the fixed-composition code ensemble) was left open. Here, under the assumption that only the decoder is unaware of the channel statistics, we conclude that the SMI decoder is asymptotically optimal also for the expurgated code. Thanks to the intimate relation between the expressions of the TRC exponent [19] and the expurgated bound [17], this result immediately follows.

The remaining part of the paper is organized as follows. In Section II, we establish notation conventions. In Section III, we formalize the model and review some background. In Section IV, we provide and discuss the main results of this work, and in Section V, we prove them.

II. Notation Conventions

Throughout the paper, random variables will be denoted by capital letters, specific values they may take 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 capital letters and the corresponding lower case letters, both in the bold face font. Their alphabets will be superscripted by their dimensions. For example, the random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$, ($n$ – positive integer) may take a specific vector value $\boldsymbol{x} = (x_1, \ldots, x_n)$ in $\mathcal{X}^n$, the n-th order Cartesian power of $\mathcal{X}$, which is the alphabet of each component of this vector. 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_X, $Q_{Y|X}$, and so on. When there is no room for ambiguity, these subscripts will be omitted. For a generic joint distribution $Q_{XY} = \{Q_{XY}(x,y), x \in \mathcal{X}, y \in \mathcal{Y}\}$, which will often be abbreviated by Q, information measures will be denoted in the conventional manner, but with a subscript Q, that is, $H_Q(X)$ is the marginal entropy of X, $H_Q(X|Y)$ is the conditional entropy of X given Y, $I_Q(X;Y) = H_Q(X) - H_Q(X|Y)$ is the mutual information between X and Y, 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, $\lim_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0$. 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 probability simplex for $\mathcal{X}$ will be denoted by $\mathcal{P}(\mathcal{X})$.

The empirical distribution of a sequence $\boldsymbol{x} \in \mathcal{X}^n$, which will be denoted by $\hat{P}_{\boldsymbol{x}}$, is the vector of relative frequencies, $\hat{P}_{\boldsymbol{x}}(x)$, of each symbol $x \in \mathcal{X}$ in $\boldsymbol{x}$. The type class of $\boldsymbol{x} \in \mathcal{X}^n$, denoted $\mathcal{T}^n(\boldsymbol{x})$, is the set of all vectors $\boldsymbol{x}' \in \mathcal{X}^n$ with $\hat{P}_{\boldsymbol{x}'} = \hat{P}_{\boldsymbol{x}}$. When we wish to emphasize the dependence of the type class on the empirical distribution $\hat{P}$, we will denote it by $\mathcal{T}^n(\hat{P})$. 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 $\hat{P}_{\boldsymbol{x}}$, which is the empirical entropy of $\boldsymbol{x}$, will be denoted by $\hat{H}_{\boldsymbol{x}}(X)$. 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, $\hat{P}_{\boldsymbol{xy}}$ would be the joint empirical distribution of $(\boldsymbol{x}, \boldsymbol{y}) = \{(x_i, y_i)\}_{i=1}^n$, $\mathcal{T}^n(Q_{X|Y}|\boldsymbol{y})$ will stand for the conditional type class induced by a sequence $\boldsymbol{y}$ and a relevant empirical conditional distribution $Q_{X|Y}$, which is the set of all vectors $\boldsymbol{x} \in \mathcal{X}^n$ with $\hat{P}_{\boldsymbol{xy}} = Q_{X|Y} \times \hat{P}_{\boldsymbol{y}}$, $\hat{I}_{\boldsymbol{xy}}(X;Y)$ will denote the empirical mutual information induced by $\boldsymbol{x}$ and $\boldsymbol{y}$, and so on. Likewise, when we wish to emphasize the dependence of empirical information measures upon a given empirical distribution given by Q, we denote them using the subscript Q, as described above.

III. Problem Setting and Background

A. Problem Setting

Consider a DMC $W = \{W(y|x),\ x \in \mathcal{X},\ y \in \mathcal{Y}\}$, where $\mathcal{X}$ and $\mathcal{Y}$ are the finite input and output alphabets, respectively. When the channel is fed with a sequence $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$, it produces $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathcal{Y}^n$ according to

$$W^n(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^n W(y_i|x_i). \tag{5}$$

Let $\mathcal{C}_n$ be a codebook, i.e., a collection $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{M-1}\}$ of $M = e^{nR}$ codewords, n being the block–length and R being the coding rate in nats per channel use. When the transmitter wishes to convey a message, $m \in \{0, 1, \ldots, M-1\}$, it feeds the channel with $\boldsymbol{x}_m$. We assume that messages are chosen with equal probabilities. We consider the ensemble of fixed-composition codes: for a given distribution Q_X over $\mathcal{X}$, all vectors in $\mathcal{C}_n$ are uniformly and independently drawn from the type class $\mathcal{T}^n(Q_X)$.

The optimal (ML) decoder estimates $\hat{m}$, using the channel output $\boldsymbol{y}$, according to

$$\hat{m}(\boldsymbol{y}) = \operatorname*{arg\,max}_{m \in \{0,1,\ldots,M-1\}} W^n(\boldsymbol{y}|\boldsymbol{x}_m). \tag{6}$$

We also consider here the GLD. The GLD chooses the estimated message $\hat{m}$ according to the following posterior distribution, induced by the channel output $\boldsymbol{y}$:

$$\mathbb{P}\left\{\hat{M} = m \middle| \boldsymbol{y}\right\} = \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}}{\sum_{m'=0}^{M-1} \exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}, \tag{7}$$

where $\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}}$ is the empirical distribution of $(\boldsymbol{x}_m, \boldsymbol{y})$ (whose X-marginal, $\hat{P}_{\boldsymbol{x}}$, coincides with Q_X) and $g(\cdot)$ is a given continuous, real–valued functional of this empirical distribution.

This GLD covers several important special cases. Obviously, the choice

$$g(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}}) = \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \hat{P}_{\boldsymbol{x}_m\boldsymbol{y}}(x,y) \log W(y|x) \tag{8}$$

corresponds to the stochastic likelihood (SL) decoder [25]. Slightly more generally, one may introduce a parameter $\beta \ge 0$ and define

$$g(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}}) = \beta \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \hat{P}_{\boldsymbol{x}_m\boldsymbol{y}}(x,y) \log W(y|x). \tag{9}$$

Here, β controls the degree of skewedness of the distribution (7), in the spirit of the notion of finite–temperature decoding [24]: while $\beta = 1$ corresponds to the SL decoder, $\beta \to \infty$ leads to the traditional (deterministic) ML decoder. The decoding metrics in (8) and (9) are well defined even if the channel transition matrix $\{W(y|x), x \in \mathcal{X}, y \in \mathcal{Y}\}$ contains zeros (e.g., the z-channel or the binary erasure channel), since for every $(x,y) \in \mathcal{X} \times \mathcal{Y}$ with $W(y|x) = 0$, obviously also $\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}}(x,y) = 0$, and we use the convention that $0 \log 0 = 0$. Likewise,

$$g(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}}) = \beta \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \hat{P}_{\boldsymbol{x}_m\boldsymbol{y}}(x,y) \log W'(y|x) \tag{10}$$

defines a family of mismatched likelihood decoders, bridging between the mismatched likelihood decoder of [25] and the ordinary, deterministic mismatched decoder (although the parameter β might as well be absorbed in W' in the form of a power of W'). Yet another important case is

$$g(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}}) = \beta \hat{I}_{\boldsymbol{x}_m\boldsymbol{y}}(X;Y), \tag{11}$$

which is a parametric family of mutual information decoders, where $\beta \to \infty$ yields the ordinary MMI universal decoder [2]. The special case of $\beta = 1$ corresponds to the SMI decoder.

Let $\boldsymbol{Y} \in \mathcal{Y}^n$ be the random channel output resulting from the transmission of $\boldsymbol{x}_m$. For a given code $\mathcal{C}_n$, define the error probability as

$$P_{\mathrm{e}}(\mathcal{C}_n) = \frac{1}{M} \sum_{m=0}^{M-1} \mathbb{P}\{\hat{m}(\boldsymbol{Y}) \ne m | m \text{ sent}\}, \tag{12}$$

where $\mathbb{P}\{\cdot\}$ designates the probability measure associated with the randomness of the channel output given its input and the possibly stochastic decoding.

B. Background

Merhav [19] has derived a single-letter expression for the error exponent of the typical random fixed-composition code,

$$E_{\text{trc}}(R, Q_X) = \lim_{n\to\infty} \left\{-\tfrac{1}{n}\mathbb{E}\left[\log P_{\text{e}}(\mathcal{C}_n)\right]\right\}, \quad (13)$$

under the GLD. In order to present the main result of [19], we define first a few quantities. Let

$$\alpha(R, Q_Y) = \max_{\{Q_{\tilde{X}|Y}:\ I_Q(\tilde{X};Y)\leq R,\ Q_{\tilde{X}}=Q_X\}} \left\{g(Q_{\tilde{X}Y}) - I_Q(\tilde{X};Y) + R\right\}, \quad (14)$$

and[3]

$$\begin{aligned}\Gamma(Q_{XX'}, R) &= \min_{Q_{Y|XX'}} \{-\mathbb{E}_Q[\log W(Y|X)] - H_Q(Y|X,X') \\ &\quad + [\max\{g(Q_{XY}), \alpha(R, Q_Y)\} - g(Q_{X'Y})]_+\}. \quad (15)\end{aligned}$$

Then, the error exponent of the TRC is given by [19]

$$\begin{aligned}E_{\text{trc}}(R, Q_X) = &\min_{\{Q_{X'|X}:\ I_Q(X;X')\leq 2R,\ Q_{X'}=Q_X\}} \\ &\{\Gamma(Q_{XX'}, R) + I_Q(X;X') - R\}. \quad (16)\end{aligned}$$

This exponent function is well defined when $g(Q) = \mathbb{E}_Q[\log W(Y|X)]$ is considered (i.e., SL decoding), even if the channel transition matrix $\{W(y|x), x\in\mathcal{X}, y\in\mathcal{Y}\}$ contains zeros, since the minimizing distribution $Q_{XX'Y}$ can always be chosen to be equal to zero whenever W is zero, such that both $g(Q_{XY})$ and $g(Q_{X'Y})$ are finite.

IV. Main Results

A. Typical Random Codes

Our main result is the following, which is proved in Section V.

Theorem 1: For any DMC, the SMI decoder is optimal with respect to the TRC error exponent.

As mentioned before, Csiszár and Körner [2, Theorem 10.2] have proved that the random-coding error exponent of the MMI decoder, pertaining to the ensemble of fixed-composition codes, is as high as the random-coding error exponent of the optimum ML decoder[4]. The fact that the SMI decoder is optimal w.r.t. the TRC is non-trivial. The proof of optimality of the SMI decoder w.r.t. the random-coding error exponent [17, Section 3] relies heavily on the possibility to average directly the error probability, which is defined as

$$P_{\text{e}}(\mathcal{C}_n) = \frac{1}{M}\sum_{m=0}^{M-1}\sum_{\boldsymbol{y}\in\mathcal{Y}^n} W^n(\boldsymbol{y}|\boldsymbol{x}_m)\mathbb{1}\{\hat{m}(\boldsymbol{y})\neq m\}, \quad (17)$$

by first calculating the expectation over the randomness of the ensemble of codes and only then, calculating the expectation over the channel noise. Here, when it comes to TRCs, this can no longer be done, because we first apply the logarithmic function on the probability of error of a given code and only then average over the randomness of the codebook. Therefore, the proof of Theorem 1 is more involved than in ordinary random coding.

We conjecture that the deterministic MMI decoder is also optimal in the TRC sense, from the following considerations. Let $E_{\text{r}}^{\text{ML}}(R, Q_X)$ and $E_{\text{r}}^{\text{MMI}}(R, Q_X)$ be the random-coding error exponents under the ML and MMI decoders, respectively. Also, denote by $E_{\text{trc}}^{\text{ML}}(R, Q_X)$ and $E_{\text{trc}}^{\text{MMI}}(R, Q_X)$ the error exponents of the TRCs under ML and MMI decoding, respectively. Let $R^*(Q_X)$ be the lowest rate for which $E_{\text{trc}}^{\text{ML}}(R, Q_X) = E_{\text{r}}^{\text{ML}}(R, Q_X)$. Then, for all $R\geq R^*(Q_X)$, the following holds:

$$\begin{aligned}E_{\text{trc}}^{\text{MMI}}(R, Q_X) &\geq E_{\text{r}}^{\text{MMI}}(R, Q_X) \quad (18)\\ &= E_{\text{r}}^{\text{ML}}(R, Q_X) \quad (19)\\ &= E_{\text{trc}}^{\text{ML}}(R, Q_X), \quad (20)\end{aligned}$$

where (18) is due to Jensen's inequality, (19) follows from the universal optimality of the MMI decoder in the ordinary random-coding sense, and (20) holds for every $R\geq R^*(Q_X)$. Thus, the desired result holds true almost trivially for the entire range of high coding rates. On the other extreme, one can prove, using similar arguments to those in Appendixes C and D, that

$$E_{\text{trc}}^{\text{MMI}}(0, Q_X) \geq -\sum_{(x,x')\in\mathcal{X}^2} Q_X(x)Q_X(x')\log\left(\sum_{y\in\mathcal{Y}}\sqrt{W(y|x)W(y|x')}\right), \quad (21)$$

which is exactly $E_{\text{ex}}(0, Q_X)$, i.e., the error exponent of the expurgated code at rate zero, hence the optimality of MMI decoding follows at rate zero as well. Unfortunately, the challenge of proving the optimality of the MMI decoder across the *entire* range of rates, has defied our best efforts so far.

B. Expurgated Codes

The result in [17, p. 5045, Theorem 2], which is on the existence of sequences of fixed-composition codes with relatively high error exponents, was stated and proved for the GLD. The proof of [17, Theorem 2] was corrected a short time after in [18], concluding that the general expression of [17, Eq. (36)] is still correct, at least when $g(Q_{XY})$ is an affine functional of Q_{XY}, which is the case of the matched/mismatched stochastic likelihood decoder. Since we need the expurgated exponent to

[3]Here and through the sequel, when $H_Q(Y|X,X')$ is involved in a minimization over $Q_{Y|XX'}$, for a given $Q_{XX'}$, then the subscript Q in H_Q should always be inferred as $Q_{XX'Y} = Q_{Y|XX'}\times Q_{XX'}$. Similarly, the subscript Q in $\mathbb{E}_Q[\log W(Y|X)]$ or in $H_Q(Y|X)$ should be inferred as the XY-marginal of $Q_{XX'Y}$. In other expectations of the form $\mathbb{E}_Q[\log G(X,X',Y)]$, e.g. (B.12), the subscript Q should always be inferred as $Q_{XX'Y}$.

[4]To be more precise, Csiszár and Körner [2, Theorem 10.2] have proved that under MMI decoding, some fixed-composition codes have error probabilities which decay exponentially fast with a rate not smaller than the random-coding error exponent of the optimum ML decoder. Since this result is based on the packing lemma [2, Lemma 10.1], which is proved by random-coding arguments, the above statement easily follows. For the fact that the random-coding error exponent achieved in [2, Theorem 10.2] equals the random-coding error exponent of the random fixed-composition code ensemble under ML decoding, the interested reader is referred to [2, Exercise 10.34] and the references therein.

hold for nonlinear decoding metrics as well (e.g., for SMI decoding), we first prove that [17, Theorem 2] holds for every continuous, real–valued functional $g(Q_{XY})$.

For a given code $\mathcal{C}_n$, the probability of error given that message m was transmitted is given by

$$P_{\mathrm{e}|m}(\mathcal{C}_n) = \sum_{m' \neq m} \sum_{\boldsymbol{y} \in \mathcal{Y}^n} W^n(\boldsymbol{y}|\boldsymbol{x}_m) \cdot \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\sum_{\tilde{m}=0}^{M-1} \exp\{ng(\hat{P}_{\boldsymbol{x}_{\tilde{m}}\boldsymbol{y}})\}}. \quad (22)$$

Then, the following proposition is proved in Appendix A.

Proposition 1: There exists a sequence of fixed-composition codes, $\{\mathcal{C}_n,\ n = 1, 2, \ldots\}$, with composition Q_X, such that

$$\liminf_{n \to \infty} \left[-\frac{\log \max_m P_{\mathrm{e}|m}(\mathcal{C}_n)}{n} \right] \geq E_{\mathrm{ex}}(R, Q_X), \quad (23)$$

where,

$$E_{\mathrm{ex}}(R, Q_X) = \min_{\{Q_{X'|X}:\ I_Q(X;X') \leq R,\ Q_{X'} = Q_X\}} \{\Gamma(Q_{XX'}, R) + I_Q(X;X') - R\}. \quad (24)$$

Although the proof in [18] and the new proof in Appendix A are quite similar, there is still at least one major difference between the two. Both proofs use the inequality $(\sum_i a_i)^s \leq \sum_i a_i^s$, which holds whenever $0 \leq s \leq 1$ and $a_i \geq 0$ for all i, but in a slightly different manner. While here, in Appendix A, this inequality is used in a sum which is only polynomially large, in [18], it is used in an exponentially large sum. As a consequence, in the proof of Proposition 1, the supremum over $\rho \geq 1$ can be switched by the limit of $\rho \to \infty$ to yield the desired result, while in the proof in [18], the supremum over $\rho \geq 1$ and the minimum over $Q_{X'|X}$ must be commuted in order to yield the desired result, but this commutation is much more complicated to be justified for a general decoding metric.

Before stating our main result here, one comment is now in order. One must note that the expurgation process of the codebook relies on the knowledge of the channel statistics, as is evident from the proof in Appendix A. Hence, we assume that only the decoder is ignorant of the channel statistics, while the encoder (or some third party that expurgates the codebook) knows them perfectly. Yet, this assumption can apparently be relaxed by considering more sophisticated code ensembles, like the generalized RGV codes [26]. The RGV code ensemble is, in fact, inherently expurgated, and it is proved in [26] that its random-coding error exponent is at least as high as the expurgated exponent derived by Csiszár and Körner [3]. We conjecture that by relying on the RGV code ensemble and the SMI decoder, one may attain universality (w.r.t. the channel statistics) in both the codebook generation process and the channel decoding, while achieving an error exponent as given in (24). We will not elaborate more on this issue.

Then, our main result is the following.

Theorem 2: For any DMC, the SMI decoder is optimal with respect to the expurgated code.

Since the TRC exponent (16) and the expurgated bound (24) are very similar, and differ only in the constraint of the outer minimization, the proof of this theorem is almost identical to the proof of Theorem 1, and hence omitted.

V. Proof of Theorem 1

Let $E_{\mathrm{trc}}^{\mathrm{ML}}(R, Q_X)$, $E_{\mathrm{trc}}^{\mathrm{SL}}(R, Q_X)$, and $E_{\mathrm{trc}}^{\mathrm{SMI}}(R, Q_X)$ denote the TRC exponents under ML, SL, and SMI decoding, respectively. Since ML decoding is optimal, it immediately follows that $E_{\mathrm{trc}}^{\mathrm{ML}}(R, Q_X) \geq E_{\mathrm{trc}}^{\mathrm{SMI}}(R, Q_X)$. Concerning stochastic decoders, let us recall the result of [12, Theorem 7], which asserts that the probability of error for SL decoding is at most twice the error probability of ML decoding, which guarantees that the error exponents of the TRC under the ML and the SL decoders are equal, i.e., that $E_{\mathrm{trc}}^{\mathrm{ML}}(R, Q_X) = E_{\mathrm{trc}}^{\mathrm{SL}}(R, Q_X)$. Hence, it remains to prove that $E_{\mathrm{trc}}^{\mathrm{SL}}(R, Q_X) \leq E_{\mathrm{trc}}^{\mathrm{SMI}}(R, Q_X)$ also holds. This will be done in a few steps. First, we provide an explicit upper bound on the TRC exponent under SL decoding. To this end, we need two definitions:

$$G(y) \triangleq \sum_{x \in \mathcal{X}} W(y|x) Q_X(x), \quad y \in \mathcal{Y}, \quad (25)$$

and,

$$\Theta(Q_{XX'}, R) \triangleq \max_{\rho \in [0,1]} \max_{\lambda \in [0,1]} \left\{ \rho \bar{\lambda} R - \sum_{(x,x') \in \mathcal{X}^2} Q_{XX'}(x, x') \times \log \left[\sum_{y \in \mathcal{Y}} W(y|x)^{1-\rho\lambda} G(y)^{-\rho\bar{\lambda}} W(y|x')^{\rho} \right] \right\}, \quad (26)$$

where $\bar{\lambda} = 1 - \lambda$.

Then, the following lemma provides such an upper bound. Its proof is in Appendix B.

Lemma 1: The TRC error exponent under SL decoding is upper-bounded by

$$E_{\mathrm{trc}}^{\mathrm{SL}}(R, Q_X) \leq \min_{\{Q_{X'|X}:\ I_Q(X;X') \leq 2R,\ Q_{X'} = Q_X\}} \{\Theta(Q_{XX'}, R) + I_Q(X;X') - R\}. \quad (27)$$

As a second step, we now provide a lower bound on the TRC exponent under SMI decoding. Denoting,

$$\Phi(Q_{XX'}, R) \triangleq \max_{\mu \in [0,1]} \max_{\lambda \in [0,1]} \min_{Q_{Y|XX'}} \{\mu \bar{\lambda} R - \mathbb{E}_Q[\log W(Y|X)] - H_Q(Y|X, X') + \mu\left(\lambda I_Q(X;Y) - I_Q(X';Y)\right)\}, \quad (28)$$

the following lemma is proved in Appendix C.

Lemma 2: The TRC error exponent under SMI decoding is lower-bounded by

$$E_{\mathrm{trc}}^{\mathrm{SMI}}(R, Q_X) \geq \min_{\{Q_{X'|X}:\ I_Q(X;X') \leq 2R,\ Q_{X'} = Q_X\}} \{\Phi(Q_{XX'}, R) + I_Q(X;X') - R\}. \quad (29)$$

Notice that the expressions of (27) and (29) have similar forms, and so, we only have to show that $\Phi(Q_{XX'},R) \geq \Theta(Q_{XX'},R)$ in order to complete the proof. The following lemma, which is proved in Appendix D, establishes this fact.

Lemma 3: For every $Q_{XX'} \in \mathcal{P}(\mathcal{X}^2)$ and $R \geq 0$,

$$\Theta(Q_{XX'},R) \leq \Phi(Q_{XX'},R). \tag{30}$$

Now, we are in a position to compare $E_{\text{trc}}^{\text{SL}}(R,Q_X)$ and $E_{\text{trc}}^{\text{SMI}}(R,Q_X)$:

$$\begin{aligned} E_{\text{trc}}^{\text{SL}}(R,Q_X) &\leq \min_{\{Q_{X'|X}:\ I_Q(X;X')\leq 2R,\ Q_{X'}=Q_X\}} \\ &\qquad \{\Theta(Q_{XX'},R)+I_Q(X;X')-R\} && (31) \\ &\leq \min_{\{Q_{X'|X}:\ I_Q(X;X')\leq 2R,\ Q_{X'}=Q_X\}} \\ &\qquad \{\Phi(Q_{XX'},R)+I_Q(X;X')-R\} && (32) \\ &\leq E_{\text{trc}}^{\text{SMI}}(R,Q_X). && (33) \end{aligned}$$

Hence the optimality of SMI decoding follows and Theorem 1 is proved.

Appendix A

Proof of Proposition 1

Assuming that message m was transmitted, the probability of error, for a given code $\mathcal{C}_n$, is given by

$$P_{\text{e}|m}(\mathcal{C}_n) = \sum_{m'\neq m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n} W^n(\boldsymbol{y}|\boldsymbol{x}_m) \times \frac{\exp\{ng(\hat{P}_{\boldsymbol{x}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}+\sum_{\tilde{m}\neq m}\exp\{ng(\hat{P}_{\boldsymbol{x}_{\tilde{m}}\boldsymbol{y}})\}}. \tag{A.1}$$

Let

$$Z_m(\boldsymbol{y}) = \sum_{\tilde{m}\neq m}\exp\{ng(\hat{P}_{\boldsymbol{x}_{\tilde{m}}\boldsymbol{y}})\}, \tag{A.2}$$

fix $\epsilon > 0$ arbitrarily small, and for every $\boldsymbol{y} \in \mathcal{Y}^n$, define the set

$$\mathcal{B}_\epsilon(m,\boldsymbol{y}) = \left\{\mathcal{C}_n:\ Z_m(\boldsymbol{y}) \leq \exp\{n\alpha(R-\epsilon,\hat{P}_{\boldsymbol{y}})\}\right\}. \tag{A.3}$$

Following the result of [17, Appendix B], we know that, considering the ensemble of randomly selected fixed-composition codes of type Q_X,

$$\mathbb{P}\{\mathcal{B}_\epsilon(m,\boldsymbol{y})\} \leq \exp\{-e^{n\epsilon}+n\epsilon+1\}, \tag{A.4}$$

for every $m \in \{0,1,\ldots,M-1\}$ and $\boldsymbol{y} \in \mathcal{Y}^n$, and so, by the union bound,

$$\begin{aligned} \mathbb{P}\{\mathcal{B}_\epsilon(m)\} &\triangleq \mathbb{P}\left\{\bigcup_{\boldsymbol{y}\in\mathcal{Y}^n}\mathcal{B}_\epsilon(m,\boldsymbol{y})\right\} && (A.5) \\ &\leq \sum_{\boldsymbol{y}\in\mathcal{Y}^n}\mathbb{P}\{\mathcal{B}_\epsilon(m,\boldsymbol{y})\} && (A.6) \\ &\leq \sum_{\boldsymbol{y}\in\mathcal{Y}^n}\exp\{-e^{n\epsilon}+n\epsilon+1\} && (A.7) \\ &= |\mathcal{Y}|^n\cdot\exp\{-e^{n\epsilon}+n\epsilon+1\}, && (A.8) \end{aligned}$$

which still decays double–exponentially fast. Define the set $\mathcal{Q}(Q_X) = \{Q_{X'|X}:\ Q_{X'} = Q_X\}$ and the enumerator

$$N_m(Q_{X'|X}) = \sum_{m'\neq m}\mathbb{1}\left\{\boldsymbol{x}_{m'}\in\mathcal{T}^n(Q_{X'|X}|\boldsymbol{x}_m)\right\}. \tag{A.9}$$

Now, for $\rho \geq 1$, taking the expectation w.r.t. the set $\{\boldsymbol{X}_{m'},\ m'\neq m\}$, while conditioning on $\boldsymbol{X}_m = \boldsymbol{x}_m$, yields

$$\begin{aligned} &\mathbb{E}\left[P_{\text{e}|m}(\mathcal{C}_n)^{1/\rho}\middle|\boldsymbol{x}_m\right] \\ &= \mathbb{E}\left[P_{\text{e}|m}(\mathcal{C}_n)^{1/\rho}\cdot\mathbb{1}\{\mathcal{B}_\epsilon(m)^{\text{c}}\}\middle|\boldsymbol{x}_m\right] \\ &\quad + \mathbb{E}\left[P_{\text{e}|m}(\mathcal{C}_n)^{1/\rho}\cdot\mathbb{1}\{\mathcal{B}_\epsilon(m)\}\middle|\boldsymbol{x}_m\right] && (A.10) \\ &\leq \mathbb{E}\left[\left(\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n}W^n(\boldsymbol{y}|\boldsymbol{x}_m)\right.\right. \\ &\quad \left.\left.\times\frac{\exp\{ng(\hat{P}_{\boldsymbol{X}_{m'}\boldsymbol{y}})\}}{\exp\{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})\}+Z_m(\boldsymbol{y})}\right)^{1/\rho}\cdot\mathbb{1}\{\mathcal{B}_\epsilon(m)^{\text{c}}\}\middle|\boldsymbol{x}_m\right] \\ &\quad + \mathbb{P}\{\mathcal{B}_\epsilon(m)|\boldsymbol{x}_m\} && (A.11) \\ &\leq \mathbb{E}\left[\left(\sum_{m'\neq m}\sum_{\boldsymbol{y}\in\mathcal{Y}^n}W^n(\boldsymbol{y}|\boldsymbol{x}_m)\right.\right. \\ &\quad \left.\left.\times\min\left\{1,\frac{e^{ng(\hat{P}_{\boldsymbol{X}_{m'}\boldsymbol{y}})}}{e^{ng(\hat{P}_{\boldsymbol{x}_m\boldsymbol{y}})}+e^{n\alpha(R-\epsilon,\hat{P}_{\boldsymbol{y}})}}\right\}\right)^{1/\rho}\middle|\boldsymbol{x}_m\right] \\ &\quad + \mathbb{P}\{\mathcal{B}_\epsilon(m)\} && (A.12) \\ &\doteq \mathbb{E}\left[\left(\sum_{m'\neq m}\exp\{-n\Gamma(\hat{P}_{\boldsymbol{x}_m\boldsymbol{X}_{m'}},R-\epsilon)\}\right)^{1/\rho}\middle|\boldsymbol{x}_m\right] \\ &\quad + \mathbb{P}\{\mathcal{B}_\epsilon(m)\} && (A.13) \\ &\leq \mathbb{E}\left[\left(\sum_{\mathcal{Q}(Q_X)}N_m(Q_{X'|X})\cdot e^{-n\Gamma(Q_{XX'},R-\epsilon)}\right)^{1/\rho}\middle|\boldsymbol{x}_m\right] \\ &\quad + |\mathcal{Y}|^n\cdot\exp\{-e^{n\epsilon}+n\epsilon+1\} && (A.14) \\ &\stackrel{\cdot}{\leq} \sum_{\mathcal{Q}(Q_X)}\mathbb{E}\left[N_m(Q_{X'|X})^{1/\rho}\middle|\boldsymbol{x}_m\right]\cdot e^{-n\Gamma(Q_{XX'},R-\epsilon)/\rho}, && (A.15) \end{aligned}$$

where (A.11) follows from the definitions in (A.1) and (A.2) and the fact that $P_{\text{e}|m}(\mathcal{C}_n) \leq 1$. In (A.12), we used the definition of the set $\mathcal{B}_\epsilon(m)$ in order to lower-bound $Z_m(\boldsymbol{y})$ by $\exp\{n\alpha(R-\epsilon,\hat{P}_{\boldsymbol{y}})\}$ for every $\boldsymbol{y} \in \mathcal{Y}^n$. Furthermore, we used the fact that $\mathcal{B}_\epsilon(m)$ and $\boldsymbol{X}_m$ are independent. The passage to (A.13) is due to the method of types and the definition of $\Gamma(Q_{XX'},R)$ in (15). In (A.14) we used the definition of the enumerators $N_m(Q_{X'|X})$ in (A.9) and the upper bound in (A.8). Using the techniques of [15, Section 6.3], the conditional expectation in (A.15) is given by

$$\mathbb{E}\left[N_m(Q_{X'|X})^{1/\rho}\middle|\boldsymbol{x}_m\right] \doteq \begin{cases} \exp\{n(R-I_Q(X;X'))/\rho\} & I_Q(X;X')\leq R \\ \exp\{n(R-I_Q(X;X'))\} & I_Q(X;X')>R \end{cases} \tag{A.16}$$

$$\stackrel{\triangle}{=} \exp\{nE(R,Q,\rho)\}. \tag{A.17}$$

Note that the expression of $E(R,Q,\rho)$ is independent of $\boldsymbol{x}_m$. Substituting it back into (A.15) provides an upper bound on $\mathbb{E}\left[P_{\mathrm{e}|m}(\mathcal{C}_n)^{1/\rho}\middle|\boldsymbol{x}_m\right]$, which is independent of $\boldsymbol{x}_m$, hence, it also holds for the unconditional expectation, i.e.,

$$\mathbb{E}\left[P_{\mathrm{e}|m}(\mathcal{C}_n)^{1/\rho}\right] \stackrel{\cdot}{\le} \sum_{\mathcal{Q}(Q_X)} e^{nE(R,Q,\rho)} \cdot e^{-n\Gamma(Q_{XX'},R-\epsilon)/\rho} \stackrel{\triangle}{=} \boldsymbol{\Delta}. \tag{A.18}$$

According to Markov's inequality, we get

$$\mathbb{P}\left\{\frac{1}{M}\sum_{m=0}^{M-1} P_{\mathrm{e}|m}(\mathcal{C}_n)^{1/\rho} > 2\boldsymbol{\Delta}\right\} \le \frac{1}{2}, \tag{A.19}$$

which means that there exists a code with

$$\frac{1}{M}\sum_{m=0}^{M-1} P_{\mathrm{e}|m}(\mathcal{C}_n)^{1/\rho} \le 2\boldsymbol{\Delta}. \tag{A.20}$$

We conclude that there exists a code $\mathcal{C}_n'$ with $M/2$ codewords for which

$$\max_m P_{\mathrm{e}|m}(\mathcal{C}_n')^{1/\rho} \le 4\boldsymbol{\Delta}, \tag{A.21}$$

and so

$$\max_m P_{\mathrm{e}|m}(\mathcal{C}_n') \le \left(\sum_{\mathcal{Q}(Q_X)} e^{nE(R,Q,\rho)} \cdot e^{-n\Gamma(Q_{XX'},R-\epsilon)/\rho}\right)^{\rho} \tag{A.22}$$

$$\doteq \sum_{\mathcal{Q}(Q_X)} \exp\{n\rho E(R,Q,\rho)\} \cdot \exp\{-n\Gamma(Q_{XX'},R-\epsilon)\} \tag{A.23}$$

$$\doteq \exp\left\{-n \cdot \min_{\mathcal{Q}(Q_X)} \left[\Gamma(Q_{XX'},R-\epsilon) - \rho E(R,Q,\rho)\right]\right\}, \tag{A.24}$$

thus,

$$\liminf_{n\to\infty} -\frac{1}{n}\log\max_m P_{\mathrm{e}|m}(\mathcal{C}_n') \ge \min_{Q_{X'|X}\in\mathcal{Q}(Q_X)} \left[\Gamma(Q_{XX'},R-\epsilon) - \rho E(R,Q,\rho)\right]. \tag{A.25}$$

Since the inequality in (A.25) holds for every $\rho \ge 1$, the negative exponential rate of the maximal probability of error can be bounded as

$$\liminf_{n\to\infty} -\frac{1}{n}\log\max_m P_{\mathrm{e}|m}(\mathcal{C}_n') \ge \sup_{\rho\ge1}\min_{Q_{X'|X}\in\mathcal{Q}(Q_X)} \left[\Gamma(Q_{XX'},R-\epsilon) - \rho E(R,Q,\rho)\right] \tag{A.26}$$

$$\ge \lim_{\rho\to\infty}\min_{Q_{X'|X}\in\mathcal{Q}(Q_X)} \left[\Gamma(Q_{XX'},R-\epsilon) - \rho E(R,Q,\rho)\right] \tag{A.27}$$

$$= \lim_{\rho\to\infty}\min\left\{\min_{\{Q_{X'|X}\in\mathcal{Q}(Q_X):\ I_Q(X;X')\ge R\}} \{\Gamma(Q_{XX'},R-\epsilon) + \rho\cdot(I_Q(X;X')-R)\}, \min_{\{Q_{X'|X}\in\mathcal{Q}(Q_X):\ I_Q(X;X')\le R\}} \{\Gamma(Q_{XX'},R-\epsilon) + I_Q(X;X') - R\}\right\} \tag{A.28}$$

$$= \min\left\{\lim_{\rho\to\infty}\min_{\{Q_{X'|X}\in\mathcal{Q}(Q_X):\ I_Q(X;X')\ge R\}} \{\Gamma(Q_{XX'},R-\epsilon) + \rho\cdot(I_Q(X;X')-R)\}, \min_{\{Q_{X'|X}\in\mathcal{Q}(Q_X):\ I_Q(X;X')\le R\}} \{\Gamma(Q_{XX'},R-\epsilon) + I_Q(X;X') - R\}\right\}, \tag{A.29}$$

where (A.28) follows from (A.16) and (A.29) is due to the fact that if $a_n \to a$ and $b_n \to b$, then also $\min\{a_n,b_n\} \to \min\{a,b\}$.

Concerning the left-hand-term inside the minimum in (A.29), let $\{\rho_n\}$ be any sequence with $\rho_n \to \infty$ and let us denote by $Q(\rho_n)$ the minimizer for ρ_n. Note that the sequence $\{Q(\rho_n)\}_{n=1}^{\infty}$ must have at least one accumulation point since the simplex of finite-alphabet probability distributions is a compact set. Let us define Q^* as an accumulation point (and if there is more than one, choose an arbitrary one), and note that it must satisfy $I_{Q^*}(X;X') = R$, since otherwise, the minimum yields infinity. Let $\{Q(\rho_{n_k})\}_{k=1}^{\infty}$ be a subsequence that tends to Q^*, which exists by the very definition of an accumulation point. Then, we have the following

$$\lim_{\rho\to\infty}\min_{\{Q_{X'|X}\in\mathcal{Q}(Q_X):\ I_Q(X;X')\ge R\}} \{\Gamma(Q_{XX'},R-\epsilon) + \rho\cdot(I_Q(X;X')-R)\} \tag{A.30}$$

$$= \lim_{n\to\infty}\min_{\{Q_{X'|X}\in\mathcal{Q}(Q_X):\ I_Q(X;X')\ge R\}} \{\Gamma(Q_{XX'},R-\epsilon) + \rho_n\cdot(I_Q(X;X')-R)\} \tag{A.31}$$

$$= \lim_{n\to\infty}\{\Gamma(Q_{XX'}(\rho_n),R-\epsilon) + \rho_n\cdot(I_{Q(\rho_n)}(X;X')-R)\} \tag{A.32}$$

$$\ge \limsup_{n\to\infty}\Gamma(Q_{XX'}(\rho_n),R-\epsilon) \tag{A.33}$$

$$\ge \lim_{k\to\infty}\Gamma(Q_{XX'}(\rho_{n_k}),R-\epsilon) \tag{A.34}$$

$$= \Gamma(Q^*_{XX'},R-\epsilon), \tag{A.35}$$

where (A.32) follows from the definition of $Q(\rho_n)$ as the minimizer in the minimization problem in (A.31). The passage to (A.33) is due to the fact that $\rho_n\cdot(I_{Q(\rho_n)}(X;X')-R) \ge 0$ and (A.35) follows from the definition of the accumulation point Q^*. Finally,

$$\Gamma(Q^*_{XX'},R-\epsilon) \ge \min_{\{Q_{X'|X}\in\mathcal{Q}(Q_X):\ I_Q(X;X')=R\}} \Gamma(Q_{XX'},R-\epsilon) \tag{A.36}$$

$$= \min_{\{Q_{X'|X}\in\mathcal{Q}(Q_X):\ I_Q(X;X')=R\}} \{\Gamma(Q_{XX'},R-\epsilon) + I_Q(X;X') - R\}, \tag{A.37}$$

where (A.36) follows from the fact that $I_{Q^*}(X;X') = R$ holds for the accumulation point Q^*, but we are still able to minimize over all distributions with $I_Q(X;X') = R$. Substituting (A.37) back into (A.29), we conclude that

$$\liminf_{n\to\infty} -\frac{1}{n}\log\max_m P_{\mathrm{e}|m}(\mathcal{C}'_n) \geq \min_{\{Q_{X'|X}\in\mathcal{Q}(Q_X):\ I_Q(X;X')\leq R\}} \{\Gamma(Q_{XX'}, R-\epsilon) + I_Q(X;X') - R\}. \quad \text{(A.38)}$$

The proof of Proposition 1 is now complete, thanks to the arbitrariness of $\epsilon > 0$.

Appendix B

Proof of Lemma 1

Recall that our main objective is to upper-bound $E_{\text{trc}}^{\text{SL}}(R, Q_X)$, which is the TRC exponent under SL decoding (i.e., $g(Q) = \mathbb{E}_Q[\log W(Y|X)]$). First, note that

$$\begin{aligned}
&\Gamma(Q_{XX'}, R) \\
&= \min_{Q_{Y|XX'}} \{-\mathbb{E}_Q[\log W(Y|X)] - H_Q(Y|X,X') \\
&\qquad + [\max\{g(Q_{XY}), \alpha(R,Q_Y)\} - g(Q_{X'Y})]_+\} &\text{(B.1)}\\
&= \min_{Q_{Y|XX'}} \max_{\rho\in[0,1]} \{-\mathbb{E}_Q[\log W(Y|X)] - H_Q(Y|X,X') \\
&\qquad + \rho\left(\max\{g(Q_{XY}), \alpha(R,Q_Y)\} - g(Q_{X'Y})\right)\}, &\text{(B.2)}
\end{aligned}$$

where (B.1) is by the definition of $\Gamma(Q_{XX'}, R)$ in (15) and (B.2) is due to the identity $[A]_+ = \max_{\rho\in[0,1]}\{\rho A\}$. Now, we use the following upper bound, which is proved in Appendix E. For every $R \geq 0$ and $Q_Y \in \mathcal{P}(\mathcal{Y})$,

$$\alpha(R, Q_Y) \leq \omega(R, Q_Y) \stackrel{\triangle}{=} R + \sum_{y\in\mathcal{Y}} Q_Y(y)\log G(y). \quad \text{(B.3)}$$

It follows from (B.3) that

$$\begin{aligned}
&\Gamma(Q_{XX'}, R) \\
&\leq \min_{Q_{Y|XX'}} \max_{\rho\in[0,1]} \{-\mathbb{E}_Q[\log W(Y|X)] - H_Q(Y|X,X') \\
&\qquad + \rho\left(\max\{g(Q_{XY}), \omega(R,Q_Y)\} - g(Q_{X'Y})\right)\} &\text{(B.4)}\\
&= \max_{\rho\in[0,1]} \min_{Q_{Y|XX'}} \{-\mathbb{E}_Q[\log W(Y|X)] - H_Q(Y|X,X') \\
&\qquad + \rho\left(\max\{g(Q_{XY}), \omega(R,Q_Y)\} - g(Q_{X'Y})\right)\} &\text{(B.5)}\\
&= \max_{\rho\in[0,1]} \min_{Q_{Y|XX'}} \{-\mathbb{E}_Q[\log W(Y|X)] - H_Q(Y|X,X') \\
&\qquad + \rho\left(\max_{\lambda\in[0,1]}\{\lambda g(Q_{XY}) + \bar{\lambda}\omega(R,Q_Y)\} - g(Q_{X'Y})\right)\} &\text{(B.6)}\\
&= \max_{\rho\in[0,1]} \min_{Q_{Y|XX'}} \max_{\lambda\in[0,1]} \{-\mathbb{E}_Q[\log W(Y|X)] - H_Q(Y|X,X') \\
&\qquad + \rho\left(\lambda g(Q_{XY}) + \bar{\lambda}\omega(R,Q_Y) - g(Q_{X'Y})\right)\} &\text{(B.7)}\\
&= \max_{\rho\in[0,1]} \max_{\lambda\in[0,1]} \min_{Q_{Y|XX'}} \{-\mathbb{E}_Q[\log W(Y|X)] - H_Q(Y|X,X') \\
&\qquad + \rho\left(\lambda g(Q_{XY}) + \bar{\lambda}\omega(R,Q_Y) - g(Q_{X'Y})\right)\}, &\text{(B.8)}
\end{aligned}$$

where (B.5) is because the objective is convex in $Q_{Y|XX'}$ under ML decoding (i.e., when $g(Q) = \mathbb{E}_Q[\log W(Y|X)]$) and concave (affine) in ρ. The convexity is explained as follows: the first term is affine in $Q_{Y|XX'}$ and hence convex. The second term is convex in $Q_{Y|XX'}$ due to the concavity of $H_Q(Y|X,X')$. As for the third term, $\omega(R,Q_Y)$ is affine in $Q_{Y|XX'}$ and the maximum between two affine functions is convex. In (B.6), we used the identity $\max\{A,B\} = \max_{\lambda\in[0,1]}\{\lambda A + \bar{\lambda}B\}$. The passage to (B.8) is due to the fact that the objective function in (B.7) is convex in $Q_{Y|XX'}$ under ML decoding and concave (affine) in λ. Denoting

$$\begin{aligned}
\Xi(Q_{XX'}, R) &= \max_{\rho\in[0,1]} \max_{\lambda\in[0,1]} \min_{Q_{Y|XX'}} \\
&\quad \{-\mathbb{E}_Q[\log W(Y|X)] - H_Q(Y|X,X') \\
&\qquad + \rho\left(\lambda g(Q_{XY}) + \bar{\lambda}\omega(R,Q_Y) - g(Q_{X'Y})\right)\}, \quad \text{(B.9)}
\end{aligned}$$

we have just shown that $\Gamma(Q_{XX'}, R) \leq \Xi(Q_{XX'}, R)$, and so, it follows from (16) that

$$E_{\text{trc}}^{\text{SL}}(R, Q_X) \leq \min_{\{Q_{X'|X}:\ I_Q(X;X')\leq 2R,\ Q_{X'}=Q_X\}} \{\Xi(Q_{XX'}, R) + I_Q(X;X') - R\}. \quad \text{(B.10)}$$

As for the inner minimization in (B.9), we have

$$\begin{aligned}
&\min_{Q_{Y|XX'}} \{-\mathbb{E}_Q[\log W(Y|X)] - H_Q(Y|X,X') \\
&\qquad + \rho\left(\lambda g(Q_{XY}) + \bar{\lambda}\omega(R,Q_Y) - g(Q_{X'Y})\right)\} \\
&= \min_{Q_{Y|XX'}} \{-\mathbb{E}_Q[\log W(Y|X)] - H_Q(Y|X,X') \\
&\qquad + \rho\left(\lambda g(Q_{XY}) + \bar{\lambda}R + \bar{\lambda}\mathbb{E}_Q[\log G(Y)] - g(Q_{X'Y})\right)\} &\text{(B.11)}\\
&= \rho\bar{\lambda}R \\
&\quad + \min_{Q_{Y|XX'}} \mathbb{E}_Q\left[\log\frac{Q_{Y|XX'}(Y|X,X')}{W(Y|X)^{1-\rho\lambda}G(Y)^{-\rho\bar{\lambda}}W(Y|X')^{\rho}}\right] &\text{(B.12)}\\
&= \rho\bar{\lambda}R - \sum_{(x,x')\in\mathcal{X}^2} Q_{XX'}(x,x') \\
&\qquad \times\log\left[\sum_{y\in\mathcal{Y}} W(y|x)^{1-\rho\lambda}G(y)^{-\rho\bar{\lambda}}W(y|x')^{\rho}\right], &\text{(B.13)}
\end{aligned}$$

where (B.11) follows by substituting $\omega(R,Q_Y)$ as defined in (B.3) and in (B.12) we wrote the explicit definition of the conditional entropy $H_Q(Y|X,X')$ and used the fact that $g(Q) = \mathbb{E}_Q[\log W(Y|X)]$. The passage to (B.13) is due to the fact that

$$\min_{Q\in\mathcal{P}(\mathcal{X})} \mathbb{E}_Q\left[\log\frac{Q(X)}{f(X)}\right] = -\log\left(\sum_{x\in\mathcal{X}} f(x)\right) \quad \text{(B.14)}$$

holds for every positive function f. Substituting (B.13) back into (B.9) yields that

$$\begin{aligned}
&\Xi(Q_{XX'}, R) \\
&= \max_{\rho\in[0,1]} \max_{\lambda\in[0,1]} \Bigg\{\rho\bar{\lambda}R - \sum_{(x,x')\in\mathcal{X}^2} Q_{XX'}(x,x') \\
&\qquad \times\log\left[\sum_{y\in\mathcal{Y}} W(y|x)^{1-\rho\lambda}G(y)^{-\rho\bar{\lambda}}W(y|x')^{\rho}\right]\Bigg\} &\text{(B.15)}\\
&= \Theta(Q_{XX'}, R), &\text{(B.16)}
\end{aligned}$$

where (B.16) is by the definition of $\Theta(\cdot,\cdot)$ in (26). Upon substituting (B.16) into (B.10), we complete the proof of Lemma 1.

Appendix C

Proof of Lemma 2

Under SMI decoding, the error exponent of the TRC is given by

$$E_{\text{trc}}^{\text{SMI}}(R,Q_X) = \min_{\{Q_{X'|X}:\ I_Q(X;X')\le 2R,\ Q_{X'}=Q_X\}} \{\Omega(Q_{XX'},R) + I_Q(X;X') - R\}, \quad \text{(C.1)}$$

with

$$\begin{aligned}
&\Omega(Q_{XX'},R)\\
&= \min_{Q_{Y|XX'}} \Big\{-\mathbb{E}_Q[\log W(Y|X)] - H_Q(Y|X,X') \\
&\quad + [\max\{I_Q(X;Y),R\} - I_Q(X';Y)]_+\Big\} && \text{(C.2)}\\
&= \min_{Q_{Y|XX'}} \max_{\mu\in[0,1]} \{-\mathbb{E}_Q[\log W(Y|X)] - H_Q(Y|X,X') \\
&\quad + \mu\left(\max\{I_Q(X;Y),R\} - I_Q(X';Y)\right)\} && \text{(C.3)}\\
&\ge \max_{\mu\in[0,1]} \min_{Q_{Y|XX'}} \{-\mathbb{E}_Q[\log W(Y|X)] - H_Q(Y|X,X') \\
&\quad + \mu\left(\max\{I_Q(X;Y),R\} - I_Q(X';Y)\right)\} && \text{(C.4)}\\
&= \max_{\mu\in[0,1]} \min_{Q_{Y|XX'}} \Big\{-\mathbb{E}_Q[\log W(Y|X)] - H_Q(Y|X,X') \\
&\quad + \mu\left(\max_{\lambda\in[0,1]}\{\lambda I_Q(X;Y) + \bar{\lambda}R\} - I_Q(X';Y)\right)\Big\} && \text{(C.5)}\\
&= \max_{\mu\in[0,1]} \min_{Q_{Y|XX'}} \max_{\lambda\in[0,1]} \{-\mathbb{E}_Q[\log W(Y|X)] \\
&\quad - H_Q(Y|X,X') + \mu\left(\lambda I_Q(X;Y) + \bar{\lambda}R - I_Q(X';Y)\right)\} && \text{(C.6)}\\
&\ge \max_{\mu\in[0,1]} \max_{\lambda\in[0,1]} \min_{Q_{Y|XX'}} \{-\mathbb{E}_Q[\log W(Y|X)] \\
&\quad - H_Q(Y|X,X') + \mu\left(\lambda I_Q(X;Y) + \bar{\lambda}R - I_Q(X';Y)\right)\}, && \text{(C.7)}
\end{aligned}$$

where (C.2) follows from the definition of $\Gamma(Q_{XX'},R)$ in (15) and the fact that $\alpha(R,Q_Y) = R$ under the decoding metric $g(Q) = I_Q(X;Y)$ and (C.3) is due to the identity $[A]_+ = \max_{\rho\in[0,1]}\{\rho A\}$. The passage to (C.4) is due to the commutation between the minimization over $Q_{Y|XX'}$ and the maximization over μ and in (C.5), we used the identity $\max\{A,B\} = \max_{\lambda\in[0,1]}\{\lambda A + \bar{\lambda}B\}$. The passage to (C.7) is due to the commutation between the minimization over $Q_{Y|XX'}$ and the maximization over λ.

Thus, it follows from the definition of $\Phi(Q_{XX'},R)$ in (28), that

$$\Omega(Q_{XX'},R) \ge \Phi(Q_{XX'},R), \quad \text{(C.8)}$$

and hence,

$$E_{\text{trc}}^{\text{SMI}}(R,Q_X) \ge \min_{\{Q_{X'|X}:\ I_Q(X;X')\le 2R,\ Q_{X'}=Q_X\}} \{\Phi(Q_{XX'},R) + I_Q(X;X') - R\}, \quad \text{(C.9)}$$

which completes the proof of Lemma 2.

Appendix D

Proof of Lemma 3

Recall that the definition of $\Phi(Q_{XX'},R)$ in (28) is

$$\begin{aligned}
&\Phi(Q_{XX'},R)\\
&= \max_{\mu\in[0,1]} \max_{\lambda\in[0,1]} \min_{Q_{Y|XX'}} \{\mu\bar{\lambda}R - \mathbb{E}_Q[\log W(Y|X)] \\
&\quad - H_Q(Y|X,X') + \mu\left(\lambda I_Q(X;Y) - I_Q(X';Y)\right)\}, && \text{(D.1)}
\end{aligned}$$

and note that

$$\begin{aligned}
&\mu\lambda I_Q(X;Y) - \mu I_Q(X';Y)\\
&= \mu\lambda H_Q(Y) - \mu\lambda H_Q(Y|X) - \mu H_Q(Y) + \mu H_Q(Y|X') && \text{(D.2)}\\
&= \mu H_Q(Y|X') - \mu\lambda H_Q(Y|X) - \mu\bar{\lambda}H_Q(Y). && \text{(D.3)}
\end{aligned}$$

Substituting the identity (refer to the steps between (E.9)-(E.11))

$$\begin{aligned}
&\mu H_Q(Y|X') - \mu\lambda H_Q(Y|X) - \mu\bar{\lambda}H_Q(Y)\\
&= \max_V \max_U \min_T \{-\mu\mathbb{E}_Q[\log T(Y|X')] \\
&\quad + \mu\lambda\mathbb{E}_Q[\log V(Y|X)] + \mu\bar{\lambda}\mathbb{E}_Q[\log U(Y)]\}, && \text{(D.4)}
\end{aligned}$$

into (D.1) yields that

$$\begin{aligned}
&\Phi(Q_{XX'},R)\\
&= \max_{\mu\in[0,1]} \max_{\lambda\in[0,1]} \min_{Q_{Y|XX'}} \{\mu\bar{\lambda}R - \mathbb{E}_Q[\log W(Y|X)] \\
&\quad - H_Q(Y|X,X') + \max_V \max_U \min_T \{-\mu\mathbb{E}_Q[\log T(Y|X')] \\
&\quad + \mu\lambda\mathbb{E}_Q[\log V(Y|X)] + \mu\bar{\lambda}\mathbb{E}_Q[\log U(Y)]\}\} && \text{(D.5)}\\
&= \max_{\mu\in[0,1]} \max_{\lambda\in[0,1]} \min_{Q_{Y|XX'}} \max_V \max_U \min_T \{\mu\bar{\lambda}R \\
&\quad - \mathbb{E}_Q[\log W(Y|X)] - H_Q(Y|X,X') \\
&\quad - \mu\mathbb{E}_Q[\log T(Y|X')] + \mu\lambda\mathbb{E}_Q[\log V(Y|X)] \\
&\quad + \mu\bar{\lambda}\mathbb{E}_Q[\log U(Y)]\} && \text{(D.6)}\\
&\ge \max_{\mu\in[0,1]} \max_{\lambda\in[0,1]} \max_V \max_U \min_T \min_{Q_{Y|XX'}} \{\mu\bar{\lambda}R \\
&\quad - \mathbb{E}_Q[\log W(Y|X)] - H_Q(Y|X,X') \\
&\quad - \mu\mathbb{E}_Q[\log T(Y|X')] + \mu\lambda\mathbb{E}_Q[\log V(Y|X)] \\
&\quad + \mu\bar{\lambda}\mathbb{E}_Q[\log U(Y)]\} && \text{(D.7)}\\
&= \max_{\mu\in[0,1]} \max_{\lambda\in[0,1]} \max_V \max_U \min_T \min_{Q_{Y|XX'}} \Big\{\mu\bar{\lambda}R \\
&\quad + \mathbb{E}_Q\left[\log \frac{Q_{Y|XX'}(Y|X,X')}{W(Y|X)V(Y|X)^{-\mu\lambda}U(Y)^{-\mu\bar{\lambda}}T(Y|X')^{\mu}}\right]\Big\}, && \text{(D.8)}
\end{aligned}$$

where (D.7) follows by the commutation between the minimization over $Q_{Y|XX'}$ and the maximizations over V and

U and in (D.8), we grouped the five expectations in (D.7) together. It follows from (B.14) that

$$\Phi(Q_{XX'},R) \geq \max_{\mu\in[0,1]} \max_{\lambda\in[0,1]} \max_{V} \max_{U} \min_{T} \Bigg\{ \mu\bar{\lambda}R - \sum_{(x,x')\in\mathcal{X}^2} Q_{XX'}(x,x') \times \log\Bigg[\sum_{y\in\mathcal{Y}} W(y|x)V(y|x)^{-\mu\lambda}U(y)^{-\mu\bar{\lambda}}T(y|x')^{\mu}\Bigg]\Bigg\}. \tag{D.9}$$

Instead of maximizing over the distributions V and U, we lower-bound by choosing $V(y|x) = W(y|x)$ and $U(y) = G(y)$, where $\{G(y)\}$ is defined in (25). We arrive at

$$\Phi(Q_{XX'},R) \geq \max_{\mu\in[0,1]} \max_{\lambda\in[0,1]} \min_{T} \Bigg\{ \mu\bar{\lambda}R - \sum_{(x,x')\in\mathcal{X}^2} Q_{XX'}(x,x') \times \log\Bigg[\sum_{y\in\mathcal{Y}} W(y|x)^{1-\mu\lambda}G(y)^{-\mu\bar{\lambda}}T(y|x')^{\mu}\Bigg]\Bigg\}. \tag{D.10}$$

Note that

$$\begin{aligned}
&\sum_{y\in\mathcal{Y}} W(y|x)^{1-\mu\lambda}G(y)^{-\mu\bar{\lambda}}T(y|x')^{\mu} \\
&= \sum_{y\in\mathcal{Y}} W(y|x)^{1-\mu\lambda}G(y)^{-\mu\bar{\lambda}}W(y|x')^{\mu}\left(\frac{T(y|x')}{W(y|x')}\right)^{\mu} &\text{(D.11)}\\
&\leq \left[\sum_{y\in\mathcal{Y}}\left(W(y|x)^{1-\mu\lambda}G(y)^{-\mu\bar{\lambda}}W(y|x')^{\mu}\right)^{r}\right]^{1/r} \times \left[\sum_{y\in\mathcal{Y}}\left(\frac{T(y|x')}{W(y|x')}\right)^{s\mu}\right]^{1/s}, &\text{(D.12)}
\end{aligned}$$

where (D.12) is due to the Hölder inequality with $1/r+1/s = 1$ and $r, s \in (1,\infty)$. Substituting (D.12) back into (D.10) yields that

$$\begin{aligned}
&\Phi(Q_{XX'},R) \\
&\geq \max_{\mu\in[0,1]} \max_{\lambda\in[0,1]} \min_{T} \Bigg\{ \mu\bar{\lambda}R - \sum_{(x,x')\in\mathcal{X}^2} Q_{XX'}(x,x') \\
&\quad\times \log\Bigg(\left[\sum_{y\in\mathcal{Y}}\left(W(y|x)^{1-\mu\lambda}G(y)^{-\mu\bar{\lambda}}W(y|x')^{\mu}\right)^{r}\right]^{1/r} \\
&\quad\times \left[\sum_{y\in\mathcal{Y}}\left(\frac{T(y|x')}{W(y|x')}\right)^{s\mu}\right]^{1/s}\Bigg)\Bigg\} &\text{(D.13)}\\
&= \max_{\mu\in[0,1]} \max_{\lambda\in[0,1]} \min_{T} \Bigg\{ \mu\bar{\lambda}R - \sum_{(x,x')\in\mathcal{X}^2} Q_{XX'}(x,x') \\
&\quad\times \log\left[\sum_{y\in\mathcal{Y}}\left(W(y|x)^{1-\mu\lambda}G(y)^{-\mu\bar{\lambda}}W(y|x')^{\mu}\right)^{r}\right]^{1/r} \\
&\quad- \sum_{x\in\mathcal{X}} Q_X(x)\log\left[\sum_{y\in\mathcal{Y}}\left(\frac{T(y|x)}{W(y|x)}\right)^{s\mu}\right]^{1/s}\Bigg\} &\text{(D.14)}\\
&= \max_{\mu\in[0,1]} \max_{\lambda\in[0,1]} \Bigg\{ \mu\bar{\lambda}R - \sum_{(x,x')\in\mathcal{X}^2} Q_{XX'}(x,x') \\
&\quad\times \log\left[\sum_{y\in\mathcal{Y}}\left(W(y|x)^{1-\mu\lambda}G(y)^{-\mu\bar{\lambda}}W(y|x')^{\mu}\right)^{r}\right]^{1/r} \\
&\quad- \max_{T}\left\{\sum_{x\in\mathcal{X}} Q_X(x)\log\left[\sum_{y\in\mathcal{Y}}\left(\frac{T(y|x)}{W(y|x)}\right)^{s\mu}\right]^{1/s}\right\}\Bigg\}, &\text{(D.15)}
\end{aligned}$$

where (D.14) follows from the fact that $Q_{X'} = Q_X$. For the maximization over T in (D.15), we have

$$\begin{aligned}
&\max_{T}\left\{\frac{1}{s}\sum_{x\in\mathcal{X}} Q_X(x)\log\left[\sum_{y\in\mathcal{Y}}\left(\frac{T(y|x)}{W(y|x)}\right)^{s\mu}\right]\right\} \\
&\leq \frac{1}{s}\sum_{x\in\mathcal{X}} Q_X(x)\max_{T(\cdot|x)}\left\{\log\left[\sum_{y\in\mathcal{Y}}\left(\frac{T(y|x)}{W(y|x)}\right)^{s\mu}\right]\right\} &\text{(D.16)}\\
&= \frac{1}{s}\sum_{x\in\mathcal{X}} Q_X(x)\log\left[\max_{T(\cdot|x)}\left\{\sum_{y\in\mathcal{Y}}\left(\frac{T(y|x)}{W(y|x)}\right)^{s\mu}\right\}\right]. &\text{(D.17)}
\end{aligned}$$

We define the Lagrangian function

$$F = \sum_{y\in\mathcal{Y}}\left(\frac{T(y|x)}{W(y|x)}\right)^{s\mu} - \psi\left(\sum_{y\in\mathcal{Y}} T(y|x) - 1\right). \tag{D.18}$$

Now, differentiating with respect to $T(y|x)$ yields

$$\frac{\partial F}{\partial T(y|x)} = \frac{(s\mu)T(y|x)^{s\mu-1}}{W(y|x)^{s\mu}} - \psi. \tag{D.19}$$

The requirement $\partial F/\partial T(y|x) = 0$ is equivalent to

$$T(y|x)^{s\mu-1} = \psi(1/s\mu)W(y|x)^{s\mu}, \tag{D.20}$$

or

$$T(y|x) = \psi' W(y|x)^{s\mu/(s\mu-1)}, \tag{D.21}$$

and thus

$$T^{*}(y|x) = \frac{W(y|x)^{s\mu/(s\mu-1)}}{\sum_{y'\in\mathcal{Y}} W(y'|x)^{s\mu/(s\mu-1)}}. \tag{D.22}$$

Now,

$$\sum_{y\in\mathcal{Y}}\left(\frac{T^*(y|x)}{W(y|x)}\right)^{s\mu}$$

$$=\sum_{y\in\mathcal{Y}}\left(\frac{W(y|x)^{s\mu/(s\mu-1)}}{W(y|x)\sum_{y'\in\mathcal{Y}}W(y'|x)^{s\mu/(s\mu-1)}}\right)^{s\mu} \tag{D.23}$$

$$=\sum_{y\in\mathcal{Y}}\left(\frac{W(y|x)^{1/(s\mu-1)}}{\sum_{y'\in\mathcal{Y}}W(y'|x)^{s\mu/(s\mu-1)}}\right)^{s\mu} \tag{D.24}$$

$$=\frac{\sum_{y\in\mathcal{Y}}W(y|x)^{s\mu/(s\mu-1)}}{\left(\sum_{y'\in\mathcal{Y}}W(y'|x)^{s\mu/(s\mu-1)}\right)^{s\mu}} \tag{D.25}$$

$$=\left(\sum_{y\in\mathcal{Y}}W(y|x)^{s\mu/(s\mu-1)}\right)^{1-s\mu}. \tag{D.26}$$

Continuing from (D.17), one obtains

$$\frac{1}{s}\sum_{x\in\mathcal{X}}Q_X(x)\log\left[\max_{T(\cdot|x)}\left\{\sum_{y\in\mathcal{Y}}\left(\frac{T(y|x)}{W(y|x)}\right)^{s\mu}\right\}\right] \tag{D.27}$$

$$=\frac{1}{s}\sum_{x\in\mathcal{X}}Q_X(x)\log\left[\left(\sum_{y\in\mathcal{Y}}W(y|x)^{s\mu/(s\mu-1)}\right)^{1-s\mu}\right] \tag{D.28}$$

$$=\frac{1-s\mu}{s}\sum_{x\in\mathcal{X}}Q_X(x)\log\left[\sum_{y\in\mathcal{Y}}W(y|x)^{s\mu/(s\mu-1)}\right]. \tag{D.29}$$

Lower-bounding (D.15) using (D.29) yields

$$\Phi(Q_{XX'},R)$$

$$\geq\max_{\mu\in[0,1]}\max_{\lambda\in[0,1]}\left\{\mu\bar{\lambda}R-\sum_{(x,x')\in\mathcal{X}^2}Q_{XX'}(x,x')\right.$$

$$\times\log\left[\sum_{y\in\mathcal{Y}}\left(W(y|x)^{1-\mu\lambda}G(y)^{-\mu\bar{\lambda}}W(y|x')^{\mu}\right)^r\right]^{1/r}$$

$$\left.-\frac{1-s\mu}{s}\sum_{x\in\mathcal{X}}Q_X(x)\log\left[\sum_{y\in\mathcal{Y}}W(y|x)^{s\mu/(s\mu-1)}\right]\right\} \tag{D.30}$$

$$\triangleq J(\mu,\lambda,r,s). \tag{D.31}$$

Optimizing over (r,s) in their feasible set yields

$$\Phi(Q_{XX'},R)$$

$$\geq\sup_{\left\{\substack{r,s\in(1,\infty),\\ 1/r+1/s=1}\right\}}\max_{\mu\in[0,1]}\max_{\lambda\in[0,1]}J(\mu,\lambda,r,s) \tag{D.32}$$

$$=\max_{\mu\in[0,1]}\max_{\lambda\in[0,1]}\sup_{\left\{\substack{r,s\in(1,\infty),\\ 1/r+1/s=1}\right\}}J(\mu,\lambda,r,s) \tag{D.33}$$

$$\geq\max_{\mu\in[0,1]}\max_{\lambda\in[0,1]}\lim_{(r,s)\to(1,\infty)}J(\mu,\lambda,r,s) \tag{D.34}$$

$$=\max_{\mu\in[0,1]}\max_{\lambda\in[0,1]}\left\{\mu\bar{\lambda}R-\sum_{(x,x')\in\mathcal{X}^2}Q_{XX'}(x,x')\right.$$

$$\times\log\left(\sum_{y\in\mathcal{Y}}W(y|x)^{1-\mu\lambda}G(y)^{-\mu\bar{\lambda}}W(y|x')^{\mu}\right)$$

$$\left.+\mu\sum_{x\in\mathcal{X}}Q_X(x)\log\left(\sum_{y\in\mathcal{Y}}W(y|x)\right)\right\} \tag{D.35}$$

$$=\max_{\mu\in[0,1]}\max_{\lambda\in[0,1]}\left\{\mu\bar{\lambda}R-\sum_{(x,x')\in\mathcal{X}^2}Q_{XX'}(x,x')\right.$$

$$\left.\times\log\left(\sum_{y\in\mathcal{Y}}W(y|x)^{1-\mu\lambda}G(y)^{-\mu\bar{\lambda}}W(y|x')^{\mu}\right)\right\}, \tag{D.36}$$

where (D.33) follows by commuting the supremum over $\{r,s\in(1,\infty),\ 1/r+1/s=1\}$ and the maxima over μ and λ. In passage (D.34), we lower-bounded by switching the supremum over the set $\{r,s\in(1,\infty),\ 1/r+1/s=1\}$ to the value of $J(\mu,\lambda,r,s)$ at the limiting point $(1,\infty)$, which belongs to the constraint set of the supremum.

Comparing $\Theta(Q_{XX'},R)$ in (26), and $\Phi(Q_{XX'},R)$ yields

$$\Phi(Q_{XX'},R)$$

$$\geq\max_{\mu\in[0,1]}\max_{\lambda\in[0,1]}\left\{\mu\bar{\lambda}R-\sum_{(x,x')\in\mathcal{X}^2}Q_{XX'}(x,x')\right.$$

$$\left.\times\log\left(\sum_{y\in\mathcal{Y}}W(y|x)^{1-\mu\lambda}G(y)^{-\mu\bar{\lambda}}W(y|x')^{\mu}\right)\right\} \tag{D.37}$$

$$=\Theta(Q_{XX'},R), \tag{D.38}$$

which completes the proof of Lemma 3.

Appendix E

Proof of Eq. (B.3)

We upper-bound $\alpha(R,Q_Y)$:

$$\alpha(R,Q_Y)$$

$$=\max_{\{Q_{\tilde{X}|Y}:\ I_Q(\tilde{X};Y)\leq R,\ Q_{\tilde{X}}=Q_X\}}\left\{g(Q_{\tilde{X}Y})-I_Q(\tilde{X};Y)+R\right\} \tag{E.1}$$

$$=\max_{\{Q_{\tilde{X}|Y}:\ Q_{\tilde{X}}=Q_X\}}\inf_{\sigma\geq1}\left\{g(Q_{\tilde{X}Y})+\sigma(R-I_Q(\tilde{X};Y))\right\} \tag{E.2}$$

$$=\max_{\{Q_{\tilde{X}|Y}:\ Q_{\tilde{X}}=Q_X\}}\inf_{\sigma\geq1}\left\{g(Q_{\tilde{X}Y})+\sigma(R-H_Q(\tilde{X})+H_Q(\tilde{X}|Y))\right\} \tag{E.3}$$

$$=\max_{\{Q_{\tilde{X}|Y}:\ Q_{\tilde{X}}=Q_X\}}\inf_{\sigma\geq1}\left\{g(Q_{\tilde{X}Y})+\sigma(R-H_Q(X)+H_Q(\tilde{X}|Y))\right\} \tag{E.4}$$

$$=\inf_{\sigma\geq1}\max_{\{Q_{\tilde{X}|Y}:\ Q_{\tilde{X}}=Q_X\}}\left\{g(Q_{\tilde{X}Y})+\sigma(R-H_Q(X)+H_Q(\tilde{X}|Y))\right\} \tag{E.5}$$

$$=\inf_{\sigma\geq1}\{\sigma(R-H_Q(X))$$

$$+ \max_{\{Q_{\tilde{X}|Y}\colon Q_{\tilde{X}}=Q_X\}} \{g(Q_{\tilde{X}Y}) + \sigma H_Q(\tilde{X}|Y)\}\Bigg\} \quad \text{(E.6)}$$

$$= \inf_{\sigma\geq 1} \Bigg\{\sigma(R - H_Q(X)) + \max_{Q_{\tilde{X}|Y}} \inf_{\tau\geq 0} \{g(Q_{\tilde{X}Y}) + \sigma H_Q(\tilde{X}|Y) - \tau D(Q_{\tilde{X}}\|Q_X)\}\Bigg\}, \quad \text{(E.7)}$$

where (E.1) is by the definition of $\alpha(R, Q_Y)$ in (14), (E.2) is due to the identity

$$\max_{\{Q\colon g(Q)\geq 0\}} f(Q) = \max_Q \inf_{\sigma\geq 0} \{f(Q) + \sigma \cdot g(Q)\}, \quad \text{(E.8)}$$

and (E.4) is by the constraint $Q_{\tilde{X}} = Q_X$. The passage to (E.5) follows from the minimax theorem, since the objective function in (E.4) is concave in $Q_{\tilde{X}|Y}$ and convex (affine) in σ. In (E.7), we replaced the constraint $Q_{\tilde{X}} = Q_X$ by the addition of the term $-\sup_{\tau\geq 0} \tau D(Q_{\tilde{X}}\|Q_X)$. Note that

$$\begin{aligned}
&- \tau D(Q_{\tilde{X}}\|Q_X) \\
&= \tau H_Q(\tilde{X}) + \tau\mathbb{E}_Q[\log Q_X(\tilde{X})] && \text{(E.9)} \\
&= \min_V \left\{\tau H_Q(\tilde{X}) + \tau D(Q_{\tilde{X}}\|V)\right\} + \tau\mathbb{E}_Q[\log Q_X(\tilde{X})] && \text{(E.10)} \\
&= \min_V \left\{-\tau\mathbb{E}_Q[\log V(\tilde{X})]\right\} + \tau\mathbb{E}_Q[\log Q_X(\tilde{X})], && \text{(E.11)}
\end{aligned}$$

where (E.9) follows directly from the definition of the relative entropy and (E.10) is since $\min_V D(Q_{\tilde{X}}\|V) = 0$. The passage to (E.11) follows again from the definition of the relative entropy. Substituting (E.11) into the inner optimizations of (E.7), yields:

$$\begin{aligned}
&\max_{Q_{\tilde{X}|Y}} \inf_{\tau\geq 0} \left\{g(Q_{\tilde{X}Y}) + \sigma H_Q(\tilde{X}|Y) - \tau D(Q_{\tilde{X}}\|Q_X)\right\} \\
&= \max_{Q_{\tilde{X}|Y}} \inf_{\tau\geq 0} \Big\{g(Q_{\tilde{X}Y}) + \sigma H_Q(\tilde{X}|Y) \\
&\qquad + \min_V\{-\tau\mathbb{E}_Q[\log V(\tilde{X})]\} + \tau\mathbb{E}_Q[\log Q_X(\tilde{X})]\Big\} && \text{(E.12)} \\
&= \max_{Q_{\tilde{X}|Y}} \inf_{\tau\geq 0} \min_V \Big\{\mathbb{E}_Q[\log W(Y|\tilde{X})] + \sigma H_Q(\tilde{X}|Y) \\
&\qquad - \tau\mathbb{E}_Q[\log V(\tilde{X})] + \tau\mathbb{E}_Q[\log Q_X(\tilde{X})]\Big\} && \text{(E.13)} \\
&\leq \inf_{\tau\geq 0} \min_V \max_{Q_{\tilde{X}|Y}} \Big\{\mathbb{E}_Q[\log W(Y|\tilde{X})] + \sigma H_Q(\tilde{X}|Y) \\
&\qquad - \tau\mathbb{E}_Q[\log V(\tilde{X})] + \tau\mathbb{E}_Q[\log Q_X(\tilde{X})]\Big\} && \text{(E.14)} \\
&= \inf_{\tau\geq 0} \min_V \max_{Q_{\tilde{X}|Y}} \Bigg\{- \sum_{y\in\mathcal{Y}} Q_Y(y) \sum_{x\in\mathcal{X}} Q_{\tilde{X}|Y}(x|y) \\
&\qquad \times \log\left[\frac{Q_{\tilde{X}|Y}(x|y)^\sigma}{W(y|x)Q_X(x)^\tau V(x)^{-\tau}}\right]\Bigg\} && \text{(E.15)} \\
&= \sigma \inf_{\tau\geq 0} \min_V \max_{Q_{\tilde{X}|Y}} \Bigg\{- \sum_{y\in\mathcal{Y}} Q_Y(y) \sum_{x\in\mathcal{X}} Q_{\tilde{X}|Y}(x|y) \\
&\qquad \times \log\left[\frac{Q_{\tilde{X}|Y}(x|y)}{W(y|x)^{1/\sigma}Q_X(x)^{\tau/\sigma}V(x)^{-\tau/\sigma}}\right]\Bigg\} && \text{(E.16)} \\
&= \sigma \inf_{\tau\geq 0} \min_V \Bigg\{\sum_{y\in\mathcal{Y}} Q_Y(y) \\
&\qquad \times \log\left(\sum_{x\in\mathcal{X}} W(y|x)^{1/\sigma}Q_X(x)^{\tau/\sigma}V(x)^{-\tau/\sigma}\right)\Bigg\} && \text{(E.17)} \\
&= \inf_{\tau\geq 0} \min_V \Bigg\{\sum_{y\in\mathcal{Y}} Q_Y(y) \\
&\qquad \times \log\left(\sum_{x\in\mathcal{X}} W(y|x)^{1/\sigma}Q_X(x)^{\tau/\sigma}V(x)^{-\tau/\sigma}\right)^\sigma\Bigg\}, && \text{(E.18)}
\end{aligned}$$

where (E.13) is since $g(Q) = \mathbb{E}_Q[\log W(Y|X)]$, (E.14) is due to the commutation between the maximization over $Q_{\tilde{X}|Y}$ and the minimizations over τ and V, and in (E.15), we wrote explicitly the expectations and grouped the four terms together. The passage to (E.17) is due to (B.14). Substituting (E.18) back into (E.7) and using the notation

$$G(y, \sigma, \tau, V) = \left(\sum_{x\in\mathcal{X}} W(y|x)^{1/\sigma}Q_X(x)^{\tau/\sigma}V(x)^{-\tau/\sigma}\right)^\sigma, \quad \text{(E.19)}$$

we obtain

$$\begin{aligned}
&\alpha(R, Q_Y) \\
&\leq \inf_{\sigma\geq 1} \inf_{\tau\geq 0} \min_V \Bigg\{\sigma(R - H_Q(X)) \\
&\qquad + \sum_{y\in\mathcal{Y}} Q_Y(y)\log G(y, \sigma, \tau, V)\Bigg\} && \text{(E.20)} \\
&= \inf_{\sigma\geq 1} \inf_{\tau\geq 0} \min_V \{\sigma(R - H_Q(X)) \\
&\qquad + \mathbb{E}_Q[\log G(Y, \sigma, \tau, V)]\}. && \text{(E.21)}
\end{aligned}$$

Defining

$$V^*(x) = \frac{Q_X(x)^{1-1/\tau}}{\sum_{x'\in\mathcal{X}} Q_X(x')^{1-1/\tau}} \quad \text{(E.22)}$$

yields

$$\begin{aligned}
&G(y, \sigma, \tau, V^*) \\
&= \left(\sum_{x\in\mathcal{X}} W(y|x)^{1/\sigma}Q_X(x)^{\tau/\sigma}V^*(x)^{-\tau/\sigma}\right)^\sigma && \text{(E.23)} \\
&= \Bigg[\sum_{x\in\mathcal{X}} W(y|x)^{1/\sigma}Q_X(x)^{[1-(1-1/\tau)]\tau/\sigma} \\
&\qquad \times \left(\sum_{x'\in\mathcal{X}} Q_X(x')^{1-1/\tau}\right)^{\tau/\sigma}\Bigg]^\sigma && \text{(E.24)} \\
&= \left(\sum_{x\in\mathcal{X}} W(y|x)^{1/\sigma}Q_X(x)^{1/\sigma}\right)^\sigma \left(\sum_{x\in\mathcal{X}} Q_X(x)^{1-1/\tau}\right)^\tau. && \text{(E.25)}
\end{aligned}$$

Taking the limit $\tau \to \infty$ in the second factor of (E.25), we have

$$\lim_{\tau\to\infty}\left(\sum_{x\in\mathcal{X}} Q_X(x)^{1-1/\tau}\right)^{\tau}$$
$$= \lim_{\tau\to\infty} \exp\left\{\frac{\log\left(\sum_{x\in\mathcal{X}} Q_X(x)^{1-1/\tau}\right)}{1/\tau}\right\} \tag{E.26}$$
$$= \lim_{s\to 0} \exp\left\{\frac{\log\left(\sum_{x\in\mathcal{X}} Q_X(x)^{1-s}\right)}{s}\right\} \tag{E.27}$$
$$= \lim_{s\to 0} \exp\left\{\frac{\sum_{x\in\mathcal{X}} Q_X(x)^{1-s}\log(Q_X(x))(-1)}{\sum_{x\in\mathcal{X}} Q_X(x)^{1-s}}\right\} \tag{E.28}$$
$$= \exp\left\{-\frac{\sum_{x\in\mathcal{X}} Q_X(x)\log(Q_X(x))}{\sum_{x\in\mathcal{X}} Q_X(x)}\right\} \tag{E.29}$$
$$= \exp\left\{H_Q(X)\right\}, \tag{E.30}$$

where (E.28) is due to L'Hospital's rule. Continuing from (E.21), we obtain

$$\alpha(R, Q_Y)$$
$$\leq \inf_{\sigma\geq 1}\inf_{\tau\geq 0}\left\{\sigma(R - H_Q(X)) + \mathbb{E}_Q[\log G(Y,\sigma,\tau,V^*)]\right\} \tag{E.31}$$
$$\leq \inf_{\sigma\geq 1}\lim_{\tau\to\infty}\left\{\sigma(R - H_Q(X)) + \mathbb{E}_Q[\log G(Y,\sigma,\tau,V^*)]\right\} \tag{E.32}$$
$$= \inf_{\sigma\geq 1}\left\{\sigma(R - H_Q(X)) + H_Q(X) + \mathbb{E}_Q\left[\log\left\{\left(\sum_{x\in\mathcal{X}} W(Y|x)^{1/\sigma} Q_X(x)^{1/\sigma}\right)^{\sigma}\right\}\right]\right\} \tag{E.33}$$
$$\leq R + \sum_{y\in\mathcal{Y}} Q_Y(y)\log\left(\sum_{x\in\mathcal{X}} W(y|x)Q_X(x)\right), \tag{E.34}$$

where (E.34) follows from the choice $\sigma = 1$. The proof of Eq. (B.3) is now complete.

Acknowledgment

The authors are very grateful to the anonymous referees and to the Associate Editor, Marco Dalai, for pointing out some gaps in the earlier versions of the proofs of Proposition 1 and Lemma 1.

References

[1] 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.

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

[3] 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.

[4] I. Csiszár, J. Körner, and K. Marton, "A new look at the error exponent of discrete memoryless channels," in *Proc. ISIT '77*, p. 107 (abstract), Cornell University, Ithaca, New York, 1977.

[5] M. Feder and A. Lapidoth, "Universal decoding for channels with memory," *IEEE Trans. Inf. Theory*, vol. 44, no. 5, pp. 1726–1745, September 1998.

[6] M. Feder and N. Merhav, "Universal composite hypothesis testing: a competitive minimax approach," *IEEE Trans. Inf. Theory*, special issue in memory of Aaron D. Wyner, vol. 48, no. 6, pp. 1504–1517, June 2002.

[7] 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.

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

[9] V. D. Goppa, "Nonprobabilistic mutual information without memory," *Probl. Cont. Information Theory*, vol. 4, pp. 97–102, 1975.

[10] W. Huleihel, N. Weinberger and N. Merhav, "Erasure/list random coding error exponents are not universally achievable," *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5403–5421, October 2016.

[11] A. Lapidoth and J. Ziv, "On the universality of the LZ-based decoding algorithm," *IEEE Trans. Inf. Theory*, vol. 44, no. 5, pp. 1746–1755, September 1998.

[12] 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, October 2017.

[13] Y. Lomnitz and M. Feder, "Communication over individual channels – a general framework," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7333–7358, November 2011.

[14] N. Merhav, "Universal decoding for memoryless Gaussian channels with a deterministic interference," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1261–1269, July 1993.

[15] N. Merhav, "Statistical physics and information theory," *Foundations and Trends in Communications and Information Theory,* vol. 6, nos. 1–2, pp. 1–212, 2009.

[16] N. Merhav, "Universal decoding for arbitrary channels relative to a given class of decoding metrics," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5566–5576, September 2013.

[17] N. Merhav, "The generalized stochastic likelihood decoder: random coding and expurgated bounds," *IEEE Trans. Inf. Theory*, vol. 63, no. 8, pp. 5039–5051, August 2017.

[18] N. Merhav, Correction to "The generalized stochastic likelihood decoder: random coding and expurgated bounds," *IEEE Trans. Inf. Theory*, vol. 63, no. 10, pp. 6827–6829, October 2017.

[19] N. Merhav, "Error exponents of typical random codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6223–6235, September 2018.

[20] N. Merhav, "Error exponents of typical random codes for the colored Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 8164–8179, December 2019.

[21] N. Merhav, "Error exponents of typical random trellis codes," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2067–2077, April 2020.

[22] N. Merhav, "A Lagrange–dual lower bound to the error exponent of the typical random code," *IEEE Trans. Inf. Theory*, vol. 66, no. 6, pp. 3456–3464, June 2020.

[23] 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.

[24] P. Ruján, "Finite temperature error–correcting codes," *Phys. Rev. Let.*, vol. 70, no. 19, pp. 2968–2971, May 1993.

[25] J. Scarlett, A. Martinéz and A. Guillén i Fàbregas, "The likelihood decoder: error exponents and mismatch," *Proc. 2015 IEEE International Symposium on Information Theory (ISIT 2015)*, pp. 86–90, Hong Kong, June 2015.

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

[27] R. Tamir (Averbuch) and N. Merhav, "Trade-offs between error exponents and excess–rate exponents of typical Slepian–Wolf codes," *Entropy* 23, 265. `http://doi.org/10.3390/e23030265` February 2021.

[28] R. Tamir (Averbuch), N. Merhav, N. Weinberger, and A. Guillén i Fàbregas, "Large deviations behavior of the logarithmic error probability of random codes," *IEEE Trans. Inf. Theory*, vol. 66, no. 11, pp. 6635–6659, November 2020.

[29] J. Ziv, "Universal decoding for finite-state channels," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 4, pp. 453–460, July 1985.

Ran Tamir (Averbuch) (S'17–M'21) is a post-doctoral fellow at the Swiss Federal Institute of Technology (ETH), Zurich, Switzerland. He received the B.Sc., M.Sc., and Ph.D. degrees from the Technion, Israel Institute of Technology, in 2010, 2017, and 2021, respectively, all in electrical engineering. His research interests are in the areas of information theory, large deviations theory, and coded communications.

Neri Merhav (S'86–M'87–SM'93–F'99) was born in Haifa, Israel, on March 16, 1957. He received the B.Sc., M.Sc., and D.Sc. degrees from the Technion, Israel Institute of Technology, in 1982, 1985, and 1988, respectively, all in electrical engineering.

From 1988 to 1990 he was with AT&T Bell Laboratories, Murray Hill, NJ, USA. Since 1990 he has been with the Viterbi Faculty of Electrical and Computer Engineering of the Technion, where he is the Irving Shepard Professor. During 1994–2000 he was also serving as a consultant to the Hewlett–Packard Laboratories – Israel (HPL-I). His research interests include

information theory, statistical communications, and statistical signal processing. He is especially interested in the areas of lossless/lossy source coding and prediction/filtering, relationships between information theory and statistics, detection, estimation, as well as in the area of Shannon Theory, including topics in joint source–channel coding, source/channel simulation, and coding with side information with applications to information hiding and watermarking systems. Another recent research interest concerns the relationships between Information Theory and statistical physics.

Dr. Merhav was a co-recipient of the 1993 Paper Award of the IEEE Information Theory Society and he is a Fellow of the IEEE since 1999. He also received the 1994 American Technion Society Award for Academic Excellence and the 2002 Technion Henry Taub Prize for Excellence in Research. During 1996-1999 he served as an Associate Editor for Source Coding to the IEEE TRANSACTIONS ON INFORMATION THEORY, and during 2017-2020 – as an Associate Editor for Shannon Theory in the same journal. He also served as a co–chairman of the Program Committee of the 2001 IEEE International Symposium on Information Theory. He is currently on the Editorial Board of FOUNDATIONS AND TRENDS IN COMMUNICATIONS AND INFORMATION THEORY.