Typical Random Coding Exponent for Finite-State Channels

Giuseppe Cocco
Universitat Pompeu Fabra
giuseppe.cocco@upf.edu

Albert Guillén i Fàbregas
University of Cambridge
Universitat Pompeu Fabra
guillen@ieee.org

Josep Font-Segura
Universitat Pompeu Fabra
josep.font@ieee.org

***Abstract*—We derive a lower bound on the typical random-coding (TRC) exponent of pairwise-independent codeword ensembles used over a finite-state channel (FSC) at rates below capacity. Under some conditions, we also show that the probability of selecting a code from the ensemble with an error exponent larger than our lower bound tends to one as the codeword length tends to infinity. Our result, presented here for the FSC, also applies to compound channels.**

I. Introduction

We consider reliable transmission of information using a code $\mathcal{C}_n = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_{M_n}\}$, a set of M_n equiprobable codewords of length n, over a discrete channel with joint transition probability $W^n(\boldsymbol{y}|\boldsymbol{x})$, where $\boldsymbol{x} \in \mathcal{X}^n$ and $\boldsymbol{y} \in \mathcal{Y}^n$ are respectively the transmitted and received sequences.

The error probability of the code, denoted as $P_{\rm e}(\mathcal{C}_n)$, is a fundamental quantity in information theory. Most of the analysis of the error probability has been centred around discrete memoryless channels (DMC). In a DMC, conditioned on $\boldsymbol{x}$, the channel output sequence $\boldsymbol{y}$ is a sequence of independent random variables, an observation that provides great simplifications in the analysis. A less considered class of channels are channels with memory, such as finite-state channels (FSC), that exhibit great practical importance for example in the modelling of fading [1, Sec. 5.6]. In a FSC, the channel output sequence depends on the current and previous states of the channel [2]. Such dependence due to channel memory poses additional challenges in the analysis of the error probability compared to that in a DMC [3].

We assume that the channel can be in one of a finite set of A possible states $\{1, 2, \dots, A\}$. The statistical behavior of the channel is described by a common conditional probability $p(y_t, s_t|x_t, s_{t-1})$ that links the current channel state s_t and output y_t to the current input x_t and previous state s_{t-1}, recursively. Following [1, Section 4.6], our results are based on the probability of the channel output $\boldsymbol{y}$, given the channel input $\boldsymbol{x}$ and initial state s_0 after summing all the possible state sequences $(s_1, \dots, s_n)$, that is

$$W^n(\boldsymbol{y}|\boldsymbol{x}, s_0) = \sum_{s_n=1}^{A} p_n(\boldsymbol{y}, s_n|\boldsymbol{x}, s_0), \tag{1}$$

where the probability $p_n(\boldsymbol{y}, s_n|\boldsymbol{x}, s_0)$ can be obtained using the recursion

$$p_t(\boldsymbol{y}_t, s_t|\boldsymbol{x}_t, s_0) = \sum_{s_{t-1}=1}^{A} p(y_t, s_t|x_t, s_{t-1}) p_{n-1}(\boldsymbol{y}_{t-1}, s_{t-1}|\boldsymbol{x}_{t-1}, s_0), \tag{2}$$

for $t = 1, \dots, n$. In (2), $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$ are respectively the channel input and output sub-sequences given by $\boldsymbol{x}_t = (x_1, \dots, x_t)$ and $\boldsymbol{y}_t = (y_1, \dots, y_t)$, while for $t = 1$ we have the initial relation $p_1(y_1, s_1|x_1, s_0) = p(y_1, s_1|x_1, s_0)$. Equation (1) represents a family of channel transition probabilities, indexed by the initial state s_0, that also encompasses compound channels since s_0 can be considered as an index that determines the channel from the compound set.

In next section, we use random-coding arguments to study the existence of codes with vanishing error probability $P_{\rm e}(\mathcal{C}_n)$ as $n \to \infty$ when used over a FSC with transition probability (1). In particular, we discuss achievable random coding and typical random coding (TRC) error exponents of pairwise-independent ensembles for rates below the channel capacity.

II. Typical Random Coding

We characterize the error probability of an ensemble of randomly generated codes C_n using the exponent of $P_{\rm e}(\mathcal{C}_n)$, defined as

$$E_n(\mathcal{C}_n) = -\frac{1}{n} \log P_{\rm e}(\mathcal{C}_n). \tag{3}$$

An exponent E is achievable when there exists a sequence of codes $\{\mathcal{C}_n\}_{n=1}^{\infty}$ such that $\liminf_{n\to\infty} E_n(\mathcal{C}_n) \geq E$. Let C_n be the random variable representing a code randomly generated with some probability distribution. In ensembles with pairwise-independent codewords, M_n codewords are generated independently with probability distribution $Q^n(\boldsymbol{x})$. The random-coding error (RCE) exponent is given by the exponent of the limiting expected error probability in the ensemble, that is

$$E_{\rm r}(R, \boldsymbol{Q}) = \lim_{n\to\infty} -\frac{1}{n} \log \mathbb{E}[P_{\rm e}(\mathsf{C}_n)], \tag{4}$$

This work has been funded in part by the postdoctoral fellowship programme Beatriu de Pinós, funded by the Catalan Secretary of Universities and Research, by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 801370, Spanish Ministry of Economy and Competitiveness under Grant PID2020-116683GB-C22, and by the European Research Council under ERC grant agreement 725411.

where $R = \lim_{n\to\infty} \frac{1}{n} \log M_n$ is the code rate and $\boldsymbol{Q}$ is the limiting distribution of Q^n, which is assumed to exist. The random-coding exponent for FSC was studied in [4] and [5] and further developed in [1, Sec. 5.9]. The channel coding theorem for finite-state indecomposable channels was first proved by Blackwell, Breiman and Thomasian [6]. Among recent works on error exponent for FSC, most deal with FSC in the presence of feedback. In [7] it is shown that a universal decoder over finite-state channels can achieve an error exponent equal to the one obtained by maximum likelihood (ML) decoding despite channel statistics being not known. In [8] an algorithm to estimate the information rates in channels with memory by using finite-state approximations is presented.

Instead of the random-coding exponent, the typical error exponent has emerged as the error exponent of the random-coding ensemble [9], [10]. The TRC exponent is defined as the limiting expected error exponent in the ensemble

$$E_{\rm trc}(R, \boldsymbol{Q}) = \lim_{n\to\infty} -\frac{1}{n}\mathbb{E}[\log P_{\rm e}(\mathsf{C}_n)]. \tag{5}$$

The importance of the TRC is argued in [10] based on the fact that the code is selected randomly only once and then kept fixed, hence the interest in the error exponent of a typical random code rather than the exponent of the average error probability across the ensemble. For the independently and identically distributed (i.i.d.) [9] and the constant composition (CC) [10] [11, Lemma 3] ensembles, the typical error exponent satisfies $E_{\rm trc}(R, Q) = E_{\rm ex}(2R, Q) + R \le E_{\rm ex}(R, Q)$, Q being the asymptotic single-letter version of Q^n, with equality for $R = 0$[1]. In [13] (resp. [14]) it is shown that, for the i.i.d. (resp. constant composition) code ensemble over discrete memoryless channel (DMC), the exponent concentrates around $E_{\rm trc}(R, Q)$.

In [15], we studied the probability of selecting a code $\mathcal{C}_n$ from the ensemble with an error exponent (3) larger than a finite-length lower bound version of the TRC exponent (5). We showed that such probability converges to one as the codeword length n tends to infinity for channels with arbitrary alphabets or memory, with the only assumption of pairwise-independent codewords. We next state a refinement of [15, Th. 1], the proof of which is provided in [16], using a tighter TRC lower bound.

Theorem 1. *Consider any channel W^n and pairwise-independent ensemble with codeword distribution Q^n and rate $R = \liminf_{n\to\infty} \frac{1}{n} \log M_n$. For all such channels and code ensembles it holds that*

$$\mathbb{P}\Big[\liminf_{n\to\infty} E_n(\mathsf{C}_n) > \liminf_{n\to\infty} E_{\rm trc,lb}(R, Q^n)\Big] = 1, \tag{6}$$

where $E_{\rm trc,lb}(R, Q^n)$ is a TRC lower bound given by

$$E_{\rm trc,lb}(R, Q^n) = \max\left\{E_{\rm trc,x}(R, Q^n), E_{\rm trc,r}(R, Q^n)\right\}, \tag{7}$$

being $E_{\rm trc,x}(R, Q^n)$ and $E_{\rm trc,r}(R, Q^n)$ two error exponents given in terms of the expurgated and the random-coding exponents, respectively.

[1]In [10] an inequality sign is used, but this is only because the improved expurgated presented in [12, Section 1, point 4.] is used instead of Gallager's.

In the next section, we give closed-form expressions for the TRC lower bounds $E_{\rm trc,x}(R, Q^n)$ and $E_{\rm trc,r}(R, Q^n)$, again valid for channels with arbitrary alphabets and memory, and specialize such bounds to FSCs.

III. TRC Lower Bounds for Finite-State Channels

For FSCs, the relevant channel probability is $W^n(\boldsymbol{y}|\boldsymbol{x}, s_0)$ given in (1). However, since the receiver may not know in advance what the initial state s_0 is, we consider a maximum-metric decoder with decoding metric the average over equiprobable states, that is

$$W^n(\boldsymbol{y}|\boldsymbol{x}) = \sum_{s_0=1}^{A} \frac{1}{A} W^n(\boldsymbol{y}|\boldsymbol{x}, s_0). \tag{8}$$

The results in this section are valid for all rates below capacity. For FSC, we use the notion of channel capacity of indecomposable channels in [1, Eq. (4.6.3)], the value C such that the ensemble-average error probability vanishes as $n \to \infty$ for all rates R such that $R < C$.

A. Lower Bound $E_{\rm trc,x}(R, Q^n)$

The first lower bound used in (7), meaningful for low rates and based on our previous work in [15], is given by

$$E_{\rm trc,x}(R, Q^n) = E^n_{\rm ex}(2R, Q^n) + R - \delta_n, \tag{9}$$

where $E^n_{\rm ex}(R, Q^n)$ is the multi-letter version of the expurgated exponent [1, Eq. (5.7.7)], that is

$$E^n_{\rm ex}(R, Q^n) = E^n_{\rm x}(\hat{\lambda}_n, Q^n) - \hat{\lambda}_n R, \tag{10}$$

with $\hat{\lambda}_n$ the parameter yielding to the highest exponent,

$$\hat{\lambda}_n = \arg\max_{\lambda\ge 1}\left\{E^n_{\rm x}(\lambda, Q^n) - \lambda 2R\right\} \tag{11}$$

and δ_n is a backoff given by $\delta_n = \frac{\hat{\lambda}_n}{n}\log\gamma_n$, where γ_n is a positive-defined sequence such that $\gamma_n \to \infty$ and $\frac{\log\gamma_n}{n} \to 0$ as the code length $n \to \infty$.

Following the footsteps in the proof of [15, Th. 1] using (8) as the decoding metric, we obtain that with high probability [16] a code randomly selected from an ensemble has an error probability, averaged over all initial states, upper bounded by

$$P_{\rm e}(\mathcal{C}_n) \le \frac{1}{M_n}\left(\gamma_n M_n(M_n-1)\right)^{\lambda} \cdot \cdot \left(\sum_{\boldsymbol{x}}\sum_{\boldsymbol{x}'} Q^n(\boldsymbol{x})Q^n(\boldsymbol{x}')Z_n(\boldsymbol{x},\boldsymbol{x}')^{\frac{1}{\lambda}}\right)^{\lambda}, \tag{12}$$

where for convenience we defined the Bhattacharyya coefficient $Z^n(\boldsymbol{x}, \boldsymbol{x}')$ for FSC as

$$Z^n(\boldsymbol{x},\boldsymbol{x}') = \sum_{\boldsymbol{y}}\sqrt{\sum_{s_0=1}^{A}\frac{1}{A}W^n(\boldsymbol{y}|\boldsymbol{x}, s_0)\sum_{s_0'=1}^{A}\frac{1}{A}W^n(\boldsymbol{y}|\boldsymbol{x}', s_0')}. \tag{13}$$

From (12) it is possible to find a bound that holds for any given initial state, useful in case a distribution over $\bar{s}_0$ is unknown.

Referring to the error probability for code $\mathcal{C}_n$ given an initial state $\bar{s}_0$ as $P_{\mathrm{e}}(\mathcal{C}_n, \bar{s}_0)$, we have the bounds

$$P_{\mathrm{e}}(\mathcal{C}_n) = \sum_{\bar{s}_0} q(\bar{s}_0) P_{\mathrm{e}}(\mathcal{C}_n, \bar{s}_0) \tag{14}$$
$$\geq \max_{\bar{s}_0} q(\bar{s}_0) P_{\mathrm{e}}(\mathcal{C}_n, \bar{s}_0) \tag{15}$$
$$\geq \frac{1}{A} P_{\mathrm{e}}(\mathcal{C}_n, \bar{s}_0). \tag{16}$$

From (16) we obtain

$$P_{\mathrm{e}}(\mathcal{C}_n, \bar{s}_0) \leq A P_{\mathrm{e}}(\mathcal{C}_n), \tag{17}$$

a bound that holds independently of the initial state distribution or on whether such distribution exists or not. Plugging (12) into (17) and using the same arguments in [15, Th. 1], we have that with high probability and for any initial state $\bar{s}_0$, the error probability of a code in the ensemble satisfies

$$P_{\mathrm{e}}(\mathcal{C}_n, \bar{s}_0) \leq \frac{1}{M_n} \left(\gamma_n M_n (M_n - 1)\right)^{\lambda} \cdot \cdot \left(\sum_{\boldsymbol{x}} \sum_{\boldsymbol{x}'} Q^n(\boldsymbol{x}) Q^n(\boldsymbol{x}') \left(A Z_n(\boldsymbol{x}, \boldsymbol{x}')\right)^{\frac{1}{\lambda}} \right)^{\lambda}, \tag{18}$$

where we note that the outer A simplifies with the $\frac{1}{A}$ in (13).

Taking the negative normalized logarithm of (18) we obtain a lower bound on the exponent of a typical code from an ensemble over a finite state channel with initial state $\bar{s}_0$ for a given n. That is,

$$-\frac{1}{n} \log P_{\mathrm{e}}(\mathcal{C}_n, \bar{s}_0) \geq E_{\mathrm{trc,x}}(R, Q^n), \tag{19}$$

where $E_{\mathrm{trc,x}}(R, Q^n)$ is the TRC lower bound given in (9) and (10), with the following expression of the expurgated function $E_{\mathrm{x}}^n(\hat{\lambda}_n, Q^n)$ for FSC,

$$E_{\mathrm{x}}^n(\lambda, Q^n) = -\frac{1}{n} \log \left(\sum_{\boldsymbol{x}} \sum_{\boldsymbol{x}'} Q^n(\boldsymbol{x}) Q^n(\boldsymbol{x}') \cdot \cdot \left(A Z_n(\boldsymbol{x}, \boldsymbol{x}')\right)^{\frac{1}{\lambda}} \right)^{\lambda} \tag{20}$$

Next we show that, under some conditions, the limit of (20) as n tends to infinity exists and is finite. We start with the following lemma (the proof of which is provided in [16]) which is the equivalent for TRC of [1, Lemma (5.9.1)].

Lemma 1. *For any finite-state channel the following holds:*

$$E_{\mathrm{x}}^n(\lambda, Q^n) \geq \frac{k}{n} E_{\mathrm{x}}^k(\lambda, Q^k) + \frac{l}{n} E_{\mathrm{x}}^l(\lambda, Q^l) \tag{21}$$

where k and l are positive integers and $k + l = n$.

The next accessory lemma is the equivalent for the TRC of [1, Lemma (5.9.2)].

Lemma 2. *Let us define:*

$$E_{\mathrm{x}}^{\infty}(\lambda, \boldsymbol{Q}) = \sup_n E_{\mathrm{x}}^n(\lambda, Q^n). \tag{22}$$

Let also $\zeta(\boldsymbol{x}, \boldsymbol{x}')$ be the tilted joint distribution

$$\zeta(\boldsymbol{x}, \boldsymbol{x}') = \frac{Q^n(\boldsymbol{x}) Q^n(\boldsymbol{x}') \left(A Z_n(\boldsymbol{x}, \boldsymbol{x}')\right)^{\frac{1}{\lambda}}}{\sum_{\bar{\boldsymbol{x}}} \sum_{\bar{\boldsymbol{x}}'} Q^n(\bar{\boldsymbol{x}}) Q^n(\bar{\boldsymbol{x}}') \left(A Z_n(\boldsymbol{x}, \boldsymbol{x}')\right)^{\frac{1}{\lambda}}}, \tag{23}$$

and consider the normalized relative entropy:

$$\frac{1}{n} D\left(\zeta^n(\boldsymbol{x}, \boldsymbol{x}') \| Q^n(\boldsymbol{x}) Q^n(\boldsymbol{x}')\right). \tag{24}$$

Then, for $\lambda \geq 1$, all pairwise-independent ensembles and all FSC for which (24) *exists and is finite from a certain n onwards, we have that $E_{\mathrm{x}}^n(\lambda, Q^n)$ converges to the limiting expurgated function $E_{\mathrm{x}}^{\infty}(\lambda, \boldsymbol{Q})$ in* (22)*, that is,*

$$\lim_{n \to \infty} E_{\mathrm{x}}^n(\lambda, Q^n) = E_{\mathrm{x}}^{\infty}(\lambda, \boldsymbol{Q}). \tag{25}$$

Furthermore, for $1 \leq \lambda < \infty$ the convergence is uniform in λ and $E_{\mathrm{x}}^{\infty}(\lambda, \boldsymbol{Q})$ is uniformly continuous in λ.

Proof: Let us start considering the case $1 \leq \lambda < \infty$. It can be easily shown that (24) is the derivative of $E_{\mathrm{x}}^n(\lambda, Q^n)$ with respect to λ. The boundedness and positiveness of the derivative, specialized to the case of FSC, for any finite λ together with the fact that $E_{\mathrm{x}}^n(1, Q^n)$ is finite (see proof of Lemma 3 below and [1, Lemma (5.9.2)]) implies that $E_{\mathrm{x}}^n(\lambda, Q^n)$ is positive and bounded. This fact together with Lemma 1 allows us to apply [1, Lemma (4A.2)], which implies (25). Furthermore, the finiteness of the derivative in λ for each n implies uniform convergence and uniform continuity. When λ is a sequence that diverges with n, that is $\lambda_n \to \infty$, we have that $\lim_{n\to\infty} E_{\mathrm{x}}^n(\lambda_n, Q^n)$ exists and is either finite or infinite. This follows from the fact that the derivative of $E_{\mathrm{x}}^n(\lambda, Q^n)$ with respect to λ exists for any n and is positive. ■

An important implication of Lemma 2 is that it guarantees that the limit of [1, Eq. (5.7.7)] for $n \to \infty$ exists. This means that Theorem 1 gives a result which is non trivial and in fact practically relevant. Yet, it remains to consider the important case that $\hat{\lambda}_n \to \infty$, since $E_{\mathrm{x}}^n(\lambda, Q^n)$ may diverge. Taking the limit of (20) as $\lambda \to \infty$, we obtain

$$\lim_{\lambda \to \infty} E_{\mathrm{x}}^n(\lambda, Q^n) = -\frac{1}{n} \sum_{\boldsymbol{x}} \sum_{\boldsymbol{x}'} Q^n(\boldsymbol{x}) Q^n(\boldsymbol{x}') \cdot \cdot \log\left(A Z_n(\boldsymbol{x}, \boldsymbol{x}')\right) \tag{26}$$

Let us consider expression of $\hat{\lambda}_n$ in (11). For some rate R, it can happen that the optimal $\hat{\lambda}_n$ grows unbounded as $n \to \infty$. While this observation is trivial for $R = 0$, since the derivative of $E_{\mathrm{x}}^n(\lambda, Q^n)$ is positive, for FSC there might exist a positive limiting threshold rate, denoted by $R_{\infty}(\boldsymbol{Q})$, such that $\hat{\lambda}_n \to \infty$ for any $R < R_{\infty}(\boldsymbol{Q})$. We next find expression for $R_{\infty}(\boldsymbol{Q})$ following a similar reasoning as in [1, Sec. 5.7].

Let us consider the function $E_{\mathrm{ex}}^n(2R, Q^n) + R$ in the right-hand side of (9). Using (10), we observe that $E_{\mathrm{ex}}^n(2R, Q^n) + R$ is a linear function in R, with R-axis intercept R_n given by

$$R_n = \frac{E_{\mathrm{x}}^n(\lambda_n, Q^n)}{2\lambda_n - 1}. \tag{27}$$

Let $\phi_{\boldsymbol{x},\boldsymbol{x}'}$ be the indicator function given by

$$\phi_{\boldsymbol{x},\boldsymbol{x}'} = \begin{cases} 1 & Z_n(\boldsymbol{x},\boldsymbol{x}') > 0 \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

Applying l'Hôpital's rule to (27) and using the expression of $E_{\mathrm{x}}^n(\lambda_n, Q^n)$ in (20), we obtain

$$\begin{aligned} R_\infty(\boldsymbol{Q}) &= \lim_{n\to\infty} R_n & (29) \\ &= \lim_{n\to\infty} -\frac{1}{2n}\log\sum_{\boldsymbol{x}}\sum_{\boldsymbol{x}'} Q^n(\boldsymbol{x})Q^n(\boldsymbol{x}')\phi_{\boldsymbol{x},\boldsymbol{x}'} & (30) \\ &= \lim_{n\to\infty} -\frac{1}{2n}\log\mathbb{P}\left[Z^n(\boldsymbol{x},\boldsymbol{x}') > 0\right]. & (31) \end{aligned}$$

For all rates smaller than $R_\infty(\boldsymbol{Q})$ as given in (31), the quantity $E_{\mathrm{x}}^n(\lambda_n, Q^n) - 2\lambda_n R$ goes to infinite. Notice that this is achieved using a $\hat{\lambda}_n$ that goes to infinity as n grows. In order to assess the behaviour of the TRC exponent we need to study the term δ_n in 9 as $\hat{\lambda}_n \to \infty$. Recalling that $\delta_n = \frac{\hat{\lambda}_n}{n}\log\gamma_n$ and repeating the derivation of $R_\infty(\boldsymbol{Q})$ including such term in the calculation, (27) now becomes

$$R_n = \frac{E_{\mathrm{x}}^n(\lambda_n, Q^n) - \lambda_n\frac{\log\gamma_n}{n}}{2\lambda_n - 1}. \tag{32}$$

Taking the limit in (32) we obtain $R_\infty(\boldsymbol{Q})$ for this case is again given by (31), since γ_n grows sub-exponentially.

To sum up, if $R_\infty(\boldsymbol{Q}) > 0$, the TRC exponent is infinite for all rates $R < R_\infty(\boldsymbol{Q})$. The cases in which $R_\infty(\boldsymbol{Q}) = 0$ and the analysis of the optimal $\hat{\lambda}$ for $R = 0$ can be derived in a similar fashion as done here and in [15], respectively, and are not reported here for a matter of space. As a final remark, we complement Lemma 2 by noticing that from the discussion above it follows that $\lim_{n\to\infty} E_{\mathrm{x}}^n(\lambda_n, Q^n)$ is always ∞ for $R > 0$ (the dependency on R is hidden in $\hat{\lambda}_n$, which is a function of the rate), since $\hat{\lambda}_n \to \infty$ for a positive rate R only if $R < R_\infty(\boldsymbol{Q})$, while it can be finite in $R = 0$.

B. Lower Bound $E_{\mathrm{trc,r}}(R, Q^n)$

The second lower bound used in (7), meaningful at high rates, is the random-coding error exponent with an asymptotically vanishing back-off $\iota_n = \frac{1}{n}\log\gamma_n \to 0$, namely

$$E_{\mathrm{trc,r}}(R, Q^n) = E_{\mathrm{r}}^n(R, Q^n) - \iota_n \tag{33}$$

where

$$E_{\mathrm{r}}^n(R, Q^n) = E_0^n(\hat{\rho}_n, Q^n) - \hat{\rho}_n R \tag{34}$$

is the multi-letter version of Gallager's random coding exponent [1, Eq. (5.6.16)],

$$\hat{\rho}_n = \arg\max_{0\le\rho\le1}\left\{E_0^n(\rho, Q^n) - \rho R\right\} \tag{35}$$

is the bound parameter that yields the highest exponent. The specialization of $E_{\mathrm{r}}^n(R, Q^n)$ to the case of finite-state channels has been carried out in [1, Section 5.9], where the following bound is derived:

$$E_{\mathrm{r}}(R, \boldsymbol{Q}) = \max_{0\le\rho\le1} F_0^\infty(\rho, \boldsymbol{Q}) - \rho R \tag{36}$$

where

$$F_0^\infty(\rho, \boldsymbol{Q}) = -\frac{\rho\log A}{n} + \min_{s_0} E_0^\infty(\rho, \boldsymbol{Q}, s_0) \tag{37}$$

and $E_0^\infty(\rho, \boldsymbol{Q}, s_0)$ is the limit of [1, Eq. (5.9.8)] for $n \to \infty$.

C. Maximization Step in Theorem 1

In the following we show that there exists a rate R^* below which our TRC lower bound is dominated by the expurgated lower bound $E_{\mathrm{trc,x}}(R, Q^n)$, while dominated by the random-coding lower bound $E_{\mathrm{trc,r}}(R, Q^n)$ for larger rates.

We start by pointing out that $E_{\mathrm{r}}(R, \boldsymbol{Q})$ is convex and strictly decreasing in R for $R < C$, as shown in [1], and the optimal ρ is a decreasing function of R so that for R that goes to C the optimal ρ gets close to 0 while it takes value 1 in a continuous interval that includes the point $R = 0$. As for $E_{\mathrm{x}}^n(\lambda_n, Q^n)$, we can see that R and λ play the same roles as in $E_x(\lambda, (Q))$ as defined in [1, Section 5.7], and thus $E_{\mathrm{x}}^n(\lambda_n, Q^n) - \lambda_n R$ is convex and strictly decreasing in R and the optimal λ is also a decreasing function in R. Note, however, that such quantity (which is a bound on the expurgated exponent for finite-state channels) can reach 0 at a rate which is much smaller than the capacity. Now let us call R_{cr} the largest rate at which the maximum of $E_{\mathrm{r}}(R, \boldsymbol{Q})$ is obtained for $\rho = 1$.

The following lemma (see [16] for a proof) is instrumental to prove Theorem 2, where the TRC exponent for all rates is presented.

Lemma 3. *For finite-state channels:*

$$E_0^n(1, Q^n) - \frac{\log A}{n} \le E_{\mathrm{x}}^n(1, Q^n) \le E_0^n(1, Q^n) + \frac{\log A}{n}. \tag{38}$$

Furthermore, if $E_0^n(1, Q^n)$ is finite $E_{\mathrm{x}}^n(1, Q^n)$ is also finite.

The following theorem shows that, for finite-state channels, the limits in Theorem 1 exist and, if $\delta_n \to 0$, the result of the theorem is non-trivial, i.e., the typical exponent exists and is strictly positive. Furthermore, it provides the result to the maximization in Theorem 1 for each rate below capacity.

Theorem 2. *For all FSC for which $\delta_n \to 0$ the following holds:*

$$\liminf_{n\to\infty} \max\left\{E_{\mathrm{trc,x}}(R, Q^n), E_{\mathrm{trc,r}}(R, Q^n)\right\} = \begin{cases} E_{\mathrm{ex}}(2R, \boldsymbol{Q}) + R & R \le R^* \\ E_{\mathrm{r}}(R, \boldsymbol{Q}) & R > R^* \end{cases} \tag{39}$$

where $R^* = \frac{R_{\mathrm{cr}}}{2}$. Furthermore, $E_{\mathrm{ex}}(2R, \boldsymbol{Q}) + R$ is positive for the indicated range while $E_{\mathrm{r}}(R, \boldsymbol{Q})$ is positive in $R^* < R < C$ if a maximization over $\boldsymbol{Q}$ is carried out.

Proof: Using Lemma 3, the fact that both $E_{\mathrm{x}}^n(\lambda, Q^n)$ and $E_0^n(\rho, Q^n)$ are increasing in λ and ρ, respectively, and the fact that $\lambda \ge 1$ while $0 < \rho \le 1$, we have that for enough large n, $E_{\mathrm{ex}}^n(R, Q^n) \ge E_{\mathrm{r}}(R, Q^n)$ when $R \le R_{\mathrm{cr}}$, while $E_{\mathrm{ex}}^n(R, Q^n) < E_{\mathrm{r}}(R, Q^n)$ when $R > R_{\mathrm{cr}}$, where R_{cr} is the critical rate [1, Eq. (5.6.30)]. Finally, using $E_{\mathrm{trc,x}}(R, Q^n)$

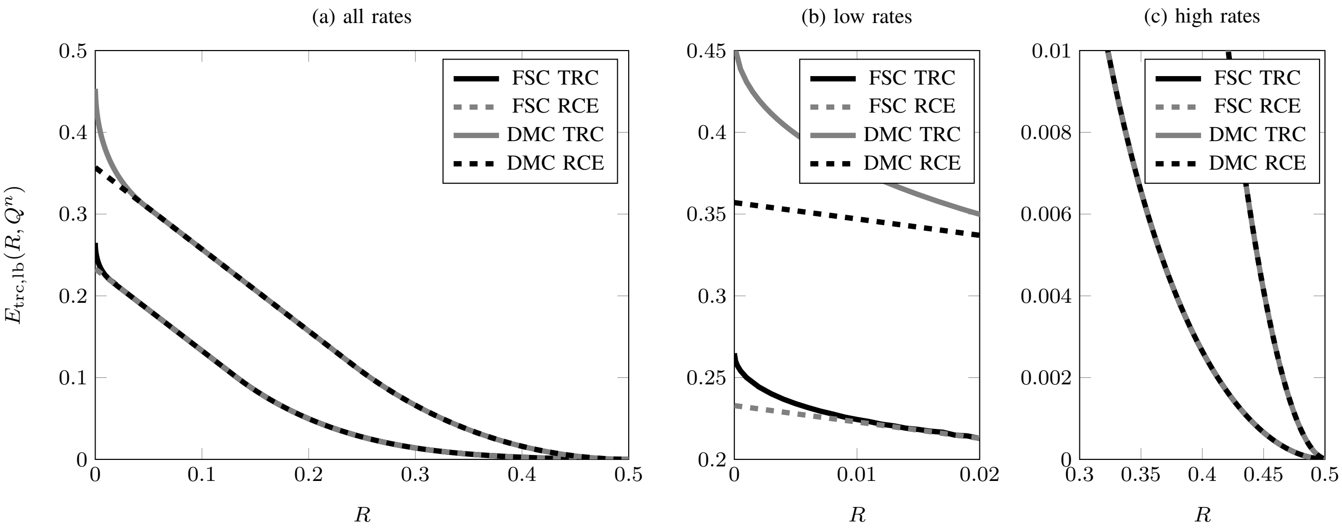

Fig. 1. TRC and RCE exponents for FSC described in [1, Figure 5.9.2] obtained with Monte Carlo for $n = 200$ and 10^6 iterations. The TRC and the RCE exponents over memoryless channel are also shown for comparison. The latter is obtained by setting the state transition probability in [1, Fig. 5.9.2] to $1/2$.

in (9) and assuming that $\delta_n \to 0$ as $n \to \infty$, the limit in the left-hand side of the inequality in (6) satisfies

$$\liminf_{n\to\infty} \max\{E_{\mathrm{trc,x}}(R, Q^n), E_{\mathrm{trc,r}}(R, Q^n)\} = \begin{cases} \liminf\limits_{n\to\infty} E_{\mathrm{trc,x}}(R, Q^n) = E_{\mathrm{ex}}(2R, \boldsymbol{Q}) + R & 0 \le R \le R^* \\ \liminf\limits_{n\to\infty} E_{\mathrm{trc,r}}(R, Q^n) = E_{\mathrm{r}}(R, \boldsymbol{Q}) & R > R^*. \end{cases} \tag{40}$$

We obtained (40) by equating $E_{\mathrm{ex}}(2R, \boldsymbol{Q}) + R$ and $E_{\mathrm{r}}(R, \boldsymbol{Q})$, from the definition of R_{cr} and setting $R^* = \frac{R_{\mathrm{cr}}}{2}$. At the right-hand side of (40), the quantity $E_{\mathrm{ex}}(2R, \boldsymbol{Q}) + R$ is positive in the indicated range, while Gallager's exponent $E_{\mathrm{r}}(R, \boldsymbol{Q})$ can be made positive up to capacity by maximizing over the input distribution, which gives the theorem statement. ∎

A discussion about the condition $\delta_n \to 0$ of Theorem 2 is provided in [16].

We conclude the paper with a numerical example of the TRC exponent for a FSC. We consider the two-state model presented in [1, Figure 5.9.1]. As in [1, Figure 5.9.1], we assume a uniform i.i.d. input distribution. The bound in (39) is plotted against the rate in Fig. 1. The term $E_{\mathrm{r}}(R, \boldsymbol{Q})$ given in (36) is evaluated in closed form as in [1, Figure 5.9.2], while $E_{\mathrm{trc,x}}(R, Q^n)$, given in (9), is evaluated using the Monte Carlo method with 10^6 iterations for a codeword length of $n = 200$. Specifically, Monte Carlo was used to evaluate the statistical mean inside square brackets in Eq. (18). The TRC for the DMC channel derived from this FSC (see [1, Figure 5.9.3]) is also shown for comparison. The result is consistent with the behaviour of the random coding exponent (RCE) presented in [1]. Note that, for this channel, the memory does not decrease the capacity with respect to its i.i.d. counterpart. This can be seen in Fig. 1 (c) noting that the smallest rates for which the two curves are zero coincide.

References

[1] R. G. Gallager, *Information Theory and Reliable Communication*. USA: John Wiley & Sons, Inc., 1968.

[2] N. Merhav, "On zero-rate error exponents of finite-state channels with input-dependent states," *IEEE Trans. Inf. Theory*, vol. 61, no. 2, pp. 741–750, 2015.

[3] M. Mushkin and I. Bar-David, "Capacity and coding for the Gilbert-Elliott channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 6, pp. 1277–1290, 1989.

[4] D. Blackwell, "Exponential error bounds for finite–state channels," in *Berkeley Symp. on Math. Stat. and Prob.*, Berkeley, California, June, 20-30 1960.

[5] H. Yudkin, "On the exponential error bound and capacity for finite-state channels," in *Int. Symp. Inf. Theory*, San Remo, Italy, Sep. 4-9 1967.

[6] D. Blackwell, L. Breiman, and A. Thomasian, "Proof of Shannon's transmission theorem for finite-state indecomposable channels," *The Annals of Mathematical Statistics*, vol. 29, no. 4, pp. 1209–1220, 1958. [Online]. Available: https://doi.org/10.1214/aoms/1177706452

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

[8] D. Arnold, H.-A. Loeliger, P. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, 2006.

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

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

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

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

[13] L. Truong, G. Cocco, J. Font-Segura, and A. Guillén i Fàbregas, "Concentration of random-coding error exponents," in *IEEE Information Theory Workshop (ITW)*, Kanazawa, Japan, Oct. 17-21 2021.

[14] R. Tamir, 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, 2020.

[15] G. Cocco, A. Guillén i Fàbregas, and J. Font-Segura, "A dual-domain achievability of the typical error exponent," in *IEEE Int. Symp. Inf. Theory*, Melbourne, Australia, July 12-20 2021.

[16] G. Cocco, A. Guillén i Fàbregas, and J. Font-Segura, "Typical error exponents: A dual domain derivation," 2022. [Online]. Available: https://arxiv.org/abs/2203.15607v2