Error Exponents of Block Codes for Finite-State Sources with Mismatch

Mehdi Dabirnia
Universitat Pompeu Fabra
mehdi.dabirnia@upf.edu

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

***Abstract*—We derive random-coding/binning error exponents for block source coding for finite-state sources. Specifically, our derivation accounts for a mismatch in the finite-state source model, recovers known special cases and provides an achievable rate, the generalized entropy rate, that quantifies the loss in rate with respect to the entropy rate induced by mismatch.**

I. Introduction

Consider a finite-state source with source alphabet $\mathcal{V}$, a finite set of states $\mathcal{S} = \{1, 2, \ldots, A\}$, A conditional probability measures $\{p(\cdot|s)\}_{s\in\mathcal{S}}$, and a next-state function $f : \mathcal{S} \times \mathcal{V} \to \mathcal{S}$. Given an initial state s_0, the conditional probability of a source sequence $v^n = v_1, \ldots, v_n \in \mathcal{V}^n$ is defined as

$$P_{V^n|S}(v^n|s_0) = \prod_{i=1}^{n} P_{V|S}(v_i|s_{i-1}) \tag{1}$$

where $s_i = f(s_{i-1}, v_i)$ for all $1 \le i \le n$. We define,

$$\bar{P}_{V^n}(v^n) = \sum_{s_0} \frac{1}{A} P_{V^n|S}(v^n|s_0). \tag{2}$$

With some abuse of notation we will use $\mathcal{S}$ to refer to this source model. A finite-state model $\mathcal{S}$ is said to be irreducible if and only if it is possible, with nonzero probability, to reach each state from any other state in a finite number of states.

A block source code $\mathcal{C}(n, R)$ is defined as a mapping $g : \mathcal{V}^n \to \mathcal{X}$ of source n-tuples $v^n \in \mathcal{V}^n$ to a set of indices/bins/codewords $\mathcal{X} = \{1, \ldots, M\}$ where $R = \frac{1}{n}\log M$ is the code rate. A decoder $\phi : \mathcal{X} \to \mathcal{V}^n$ maps each index/bin/codeword back into a source n-tuple $\hat{v}^n$. Typically the number of codewords is smaller than the number of n-tuples and the decoder makes an error whenever $\phi(x(v^n)) \neq v^n$.

Given the source model $\mathcal{S}$, the initial state s_0 and a block source code $\mathcal{C}(n, R)$, the decoder that minimizes the average probability of the error is the maximum-likelihood (ML) decoder which maps each codeword x to the most likely source sequence encoded into x, i.e.,

$$\hat{v}^n = \phi(x) = \operatorname*{arg\,max}_{v^n:\, g(v^n)=x} P_{V^n}(v^n|s_0). \tag{3}$$

Block source codes have been considered in a number of works in different contexts, see e.g. [1]–[4] and references therein.

This work has been funded in part by the European Research Council under ERC grant agreement 725411 and by the Spanish Ministry of Economy and Competitiveness under grant PID2020-116683GB-C22.

In practical systems, the exact source model is almost never known exactly and we are bound to fit some model to the source data and use the extracted model for designing an efficient code. Therefore it is meaningful to assume that in a block source coding setup, the decoder always uses a mismatched source model for decoding. Such a model can be generated during the encoding process and shared with the decoder to allow for correct decoding.

II. Main Results

Assume that instead of the real source model $\mathcal{S}$, we describe the source with a mismatched model with a finite set of states $\hat{\mathcal{S}} = \{1, 2, \ldots, \hat{A}\}$, $\hat{A}$ conditional probability measures $\{Q(\cdot|\hat{s})\}_{\hat{s}\in\hat{\mathcal{S}}}$, and a next-state function $\hat{f} : \hat{\mathcal{S}} \times \mathcal{V} \to \hat{\mathcal{S}}$.

We consider a maximum metric decoding based on mismatched model without the knowledge of initial state as follows,

$$\hat{v}^n = \hat{\phi}(x) = \operatorname*{arg\,max}_{v^n} q(v^n, x), \tag{4}$$

where

$$q(v^n, x) = \bar{Q}_{V^n}(v^n) \mathbb{1}\{g(v^n) = x\}, \tag{5}$$

and $\bar{Q}_{V^n}(v^n) = \frac{1}{\hat{A}} \sum_{\hat{s}_0} Q_{V^n|S}(v^n|\hat{s}_0)$.

In the following, we consider the random ensemble of (n, R) codes for alphabet $\mathcal{V}$ as the set of all (n, R) block codes where each source n-tuple is mapped randomly, independently and with equal probability $\frac{1}{M}$ into one of the M indices or codewords independent from the initial state s_0. In this paper, we study a random-coding error exponent for finite-state sources $\mathcal{S}$ with decoding based on a mismatched source model $\hat{\mathcal{S}}$.

Theorem 1. *For a finite-state source with irreducible model $\mathcal{S}$, there exists a block code with $M = \lceil e^{nR} \rceil$ codewords such that using a decoder based on mismatched source model $\hat{\mathcal{S}}$ for any initial state $\bar{s}_0$ we have*

$$p_e(\bar{s}_0) \le e^{-n\, e_r(R)} \tag{6}$$

where

$$e_r(R) = \sup_{\rho\in[0,1],\tau\ge 0} \rho R - E_{\mathrm{s}}(\rho, \tau), \tag{7}$$

$$E_{\mathrm{s}}(\rho, \tau) = \log \lambda(\rho, \tau) + \rho \log \hat{\lambda}(\tau) + o(n) \tag{8}$$

and $\lambda(\rho,\tau), \hat{\lambda}(\tau)$ are respectively the largest magnitude eigenvalues of the matrices $\Gamma_{\rho,\tau} \in \mathbb{R}^{A\hat{A}\times A\hat{A}}, \hat{\Gamma}_{\tau} \in \mathbb{R}^{\hat{A}\times\hat{A}}$ with entries

$$\gamma_{j\hat{j}k\hat{k}}(\rho,\tau) = \sum_{v} \frac{P_{V,S|S}(v,j|k)}{Q_{V,S|S}(v,\hat{j}|\hat{k})^{\tau\rho}} \tag{9}$$

$$\hat{\gamma}_{\hat{j}\hat{k}}(\tau) = \sum_{v} Q_{V,S|S}(v,\hat{j}|\hat{k})^{\tau}, \tag{10}$$

where $\gamma_{j\hat{j}k\hat{k}}(\rho,\tau)$ is the entry in row $(j-1)\hat{A}+\hat{j}$ and column $(k-1)\hat{A}+\hat{k}$ of matrix $\Gamma_{\rho,\tau}$.

Proof: Since the decoding metric is independent of the initial state we bound the random-coding error probability as

$$\bar{p}_e \le \sum_{v^n} \bar{P}_{V^n}(v^n)\mathbb{P}\Bigg[\bigcup_{\bar{v}^n\ne v^n}\{q(\bar{v}^n,X)\ge q(v^n,X)\}\Bigg] \tag{11}$$

For events $\{\mathcal{B}_i\}$ it can be shown that for any $0\le\rho\le 1$ we have $\mathbb{P}[\bigcup_i \mathcal{B}_i] \le (\sum_i \mathbb{P}[\mathcal{B}_i])^{\rho}$ [5, Ch. 5]. Using the random ensemble definition, for any sequences $v^n, \bar{v}^n$ we have

$$\mathbb{P}[q(\bar{v}^n,X)\ge q(v^n,X)] = \frac{1}{M}\mathbb{1}\left[\bar{Q}_{V^n}(\bar{v}^n)\ge \bar{Q}_{V^n}(v^n)\right] \tag{12}$$

$$\le \frac{1}{M}\frac{\bar{Q}_{V^n}(\bar{v}^n)^{\tau}}{\bar{Q}_{V^n}(v^n)^{\tau}} \tag{13}$$

for any $\tau \ge 0$. Therefore, the average error probability (averaged also over the initial state)

$$\bar{p}_e \le \frac{1}{M^{\rho}}\sum_{v^n}\bar{P}_{V^n}(v^n)\Bigg(\sum_{\bar{v}^n\ne v^n}\frac{\bar{Q}_{V^n}(\bar{v}^n)^{\tau}}{\bar{Q}_{V^n}(v^n)^{\tau}}\Bigg)^{\rho}. \tag{14}$$

Since the average error probability over the ensemble is upper bounded as in (14), there is at least one code in the ensemble that satisfies the above bound. Also, since the error probability for such a code is an average over A equally likely states, the conditional error probability given any particular initial state, can be no more than A times the average. This gives a bound on error probability which is valid for any initial state and no longer depends on the assumption of the equally likely states as per (2). Therefore, conditional on any initial state $\bar{s}_0 \in \mathcal{S}$ the average error probability is bounded as

$$\bar{p}_e(\bar{s}_0) \le \frac{A}{M^{\rho}}\sum_{v^n\in\mathcal{V}^n}\bar{P}_{V^n}(v^n)\Bigg(\sum_{\bar{v}^n\ne v^n}\frac{\bar{Q}_{V^n}(\bar{v}^n)^{\tau}}{\bar{Q}_{V^n}(v^n)^{\tau}}\Bigg)^{\rho}. \tag{15}$$

For any sequences $v^n, \bar{v}^n$ we have

$$\frac{\bar{Q}_{V^n}(\bar{v}^n)^{\tau}}{\bar{Q}_{V^n}(v^n)^{\tau}} \le \hat{A}^{|\tau-1|}\frac{\sum_{\hat{s}_0} Q_{V^n|S}(\bar{v}^n|\hat{s}_0)^{\tau}}{\sum_{\hat{s}_0'} Q_{V^n|S}(v^n|\hat{s}_0')^{\tau}}. \tag{16}$$

This can be seen by considering separately the cases for $\tau\le 1$ and $\tau \ge 1$. For $\tau \le 1$ we use the inequality $(\sum a_i)^r \le \sum a_i^r$ for $0<r\le 1$ to upper bound the numerator and $(\sum P_i a_i)^r \ge \sum P_i a_i^r$ for $r\le 1$ to lower bound the denominator. For $\tau\ge 1$ we use $(\sum P_i a_i)^r \le \sum P_i a_i^r$ for $r \ge 1$ to upper bound the numerator and $(\sum a_i)^r \ge \sum a_i^r$ for $r\ge 1$ to lower bound the denominator. Substituting (16) in (15) we get

$$\bar{p}_e(\bar{s}_0) \le \frac{A\hat{A}^{\rho|\tau-1|}}{M^{\rho}} \times \sum_{v^n\in\mathcal{V}^n}\bar{P}_{V^n}(v^n)\left(\sum_{\bar{v}^n}\frac{\sum_{\hat{s}_0} Q_{V^n|S}(\bar{v}^n|\hat{s}_0)^{\tau}}{\sum_{\hat{s}_0'} Q_{V^n|S}(v^n|\hat{s}_0')^{\tau}}\right)^{\rho}. \tag{17}$$

We further bound the term in brackets by swapping the sums over $\bar{v}^n$ and $\hat{s}_0$ and upper bounding the numerator using $(\sum a_i)^r \le \sum a_i^r$ for $0<r\le 1$ and lower bounding the denominator using $(\sum P_i a_i)^r \ge \sum P_i a_i^r$ for $r \le 1$. This gives

$$\bar{p}_e(\bar{s}_0) \le \frac{A\hat{A}^{\rho|\tau-1|}}{M^{\rho}} \times \sum_{v^n\in\mathcal{V}^n}\bar{P}_{V^n}(v^n)\frac{\sum_{\hat{s}_0}\left(\sum_{\bar{v}^n} Q_{V^n|S}(\bar{v}^n|\hat{s}_0)^{\tau}\right)^{\rho}}{\hat{A}^{\rho-1}\sum_{\hat{s}_0'} Q_{V^n|S}(v^n|\hat{s}_0')^{\tau\rho}}. \tag{18}$$

Using (2), changing the order of sums over v^n and s_0 and upper bounding the sum over s_0 by A times maximum over s_0, and upper bounding the sum over $\hat{s}_0$ by $\hat{A}$ times maximum over $\hat{s}_0$ and upper bounding the sum over $\hat{s}_0'$ as

$$\frac{1}{\sum_{\hat{s}_0'} Q_{V^n|S}(v^n|\hat{s}_0')^{\tau\rho}} \le \max_{\hat{s}_0'}\frac{1}{\hat{A}Q_{V^n|S}(v^n|\hat{s}_0')^{\tau\rho}}$$

we obtain

$$\bar{p}_e(\bar{s}_0) \le \frac{A\hat{A}^{\rho|\tau-1|-\rho+1}}{M^{\rho}}\max_{s_0}\max_{\hat{s}_0'}\max_{\hat{s}_0} e^{nE_{\mathrm{s}}(\rho,\tau,s_0,\hat{s}_0)} \tag{19}$$

where

$$E_{\mathrm{s}}(\rho,\tau,s_0,\hat{s}_0',\hat{s}_0) = \frac{1}{n}\log\sum_{v^n\in\mathcal{V}^n}P_{V^n|S}(v^n|s_0)\Bigg(\sum_{\bar{v}^n}\frac{Q_{V^n|S}(\bar{v}^n|\hat{s}_0)^{\tau}}{Q_{V^n|S}(v^n|\hat{s}_0')^{\tau}}\Bigg)^{\rho} \tag{20}$$

We notice that the bound in (19) is valid for the general finite-state source without the deterministic state transition assumption. Similarly to the channel coding case with ML decoding [5, Sec. 5.9] it can be shown that for any $s_0, \hat{s}_0', \hat{s}_0$ the function $E_{\mathrm{s}}(\rho,\tau,s_0,\hat{s}_0',\hat{s}_0)$ is continuous, increasing and convex in ρ with $E_{\mathrm{s}}(0,s,s_0,\hat{s}_0',\hat{s}_0) = 0$. This will prove important to derive the corresponding achievable rate. In the following, we proceed to simplifying (19).

In the following, we split the maximization argument in (19) into two terms and work out the two terms separately

$$\bar{p}_e(\bar{s}_0) \le \frac{A\hat{A}^{\rho|\tau-1|-\rho+1}}{M^{\rho}}\left(\max_{s_0}\max_{\hat{s}_0}\sum_{v^n\in\mathcal{V}^n}\frac{P_{V^n|S}(v^n|s_0)}{Q_{V^n|S}(v^n|\hat{s}_0)^{\tau\rho}}\right) \times \max_{\hat{s}_0}\left(\sum_{\bar{v}^n} Q_{V^n|S}(\bar{v}^n|\hat{s}_0)^{\tau}\right)^{\rho}, \tag{21}$$

where in the first term we change the notation from $\hat{s}_0'$ to $\hat{s}_0$, since after splitting there is no more confusion between those.

Based on the state transition mechanism of source model $\mathcal{S}$ and mismatched model $\hat{\mathcal{S}}$ given the initial states s_0 and $\hat{s}_0$, the source sequence $v^n = (v_1,\dots,v_n)$ uniquely determines state sequences $\boldsymbol{s} = \boldsymbol{s}(v^n, s_0)$ and $\hat{\boldsymbol{s}} = \hat{\boldsymbol{s}}(v^n,\hat{s}_0)$. Therefore, similarly to [5, Eq. (5.9.31)] we can define

$$\frac{P_{V^n,\boldsymbol{S}|S}(v^n,\boldsymbol{s}|s_0)}{Q_{V^n,\boldsymbol{S}|S}(v^n,\hat{\boldsymbol{s}}|\hat{s}_0)^{\tau\rho}} = \begin{cases} \frac{P_{V^n|S}(v^n|s_0)}{Q_{V^n|S}(v^n|\hat{s}_0)^{\tau\rho}} & \text{for } \boldsymbol{s},\hat{\boldsymbol{s}} \\ 0 & \text{othewise,} \end{cases} \quad (22)$$

and

$$\frac{P_{V^n,\boldsymbol{S}|S}(v^n,\boldsymbol{s}|s_0)}{Q_{V^n,\boldsymbol{S}|S}(v^n,\hat{\boldsymbol{s}}|\hat{s}_0)^{\tau\rho}} = \prod_{i=1}^n \frac{P_{V,S|S}(v_i,s_i|s_{i-1})}{Q_{V,S|S}(v_i,\hat{s}_i|\hat{s}_{i-1})^{\tau\rho}} \quad (23)$$

where

$$P_{V,S|S}(v_i,s_i|s_{i-1}) = \begin{cases} P_{V|S}(v_i|s_{i-1}) & \text{for } s_i = f(s_{i-1},v_i) \\ 0 & \text{othewise.} \end{cases} \quad (24)$$

We thus write the first term in brackets in (21) as

$$\max_{s_0}\max_{\hat{s}_0}\sum_{\boldsymbol{s},\hat{\boldsymbol{s}}}\sum_{v^n}\prod_{i=1}^n \frac{P_{V,S|S}(v_i,s_i|s_{i-1})}{Q_{V,S|S}(v_i,\hat{s}_i|\hat{s}_{i-1})^{\tau\rho}} \quad (25)$$

$$= \max_{s_0}\max_{\hat{s}_0}\sum_{\boldsymbol{s},\hat{\boldsymbol{s}}}\prod_{i=1}^n\sum_{v_i} \frac{P_{V,S|S}(v_i,s_i|s_{i-1})}{Q_{V,S|S}(v_i,\hat{s}_i|\hat{s}_{i-1})^{\tau\rho}} \quad (26)$$

Similarly, we write the second term in brackets in (21) as

$$\max_{\hat{s}_0}\left(\sum_{\hat{\boldsymbol{s}}}\prod_{i=1}^n\sum_{\bar{v}_i} Q_{V,S|S}(\bar{v}_i,\hat{s}_i|\hat{s}_{i-1})^{\tau}\right)^{\rho}. \quad (27)$$

Now we define an $A\hat{A}\times A\hat{A}$ matrix $\Gamma_{\rho,\tau}$ with elements

$$\gamma_{j\hat{j}k\hat{k}}(\rho,\tau) = \sum_v \frac{P_{V,S|S}(v,j|k)}{Q_{V,S|S}(v,\hat{j}|\hat{k})^{\tau\rho}} \quad (28)$$

for $j=f(k,v)$, $\hat{j}=\hat{f}(\hat{k},v)$ and $\gamma_{j\hat{j}k\hat{k}}(\rho,\tau)=0$ otherwise. We also define an $\hat{A}\times\hat{A}$ matrix $\hat{\Gamma}_\tau$ with elements

$$\hat{\gamma}_{\hat{j}\hat{k}}(\tau) = \sum_v Q_{V,S|S}(v,\hat{j}|\hat{k})^{\tau}, \quad (29)$$

for $\hat{j}=\hat{f}(\hat{k},v)$ and $\hat{\gamma}_{\hat{j}\hat{k}}(\tau)=0$ otherwise. Observe that the matrices $\Gamma_{\rho,\tau}$ and $\hat{\Gamma}_\tau$, are not always stochastic matrices. We denote by $\mathbf{1}$ and $\hat{\mathbf{1}}$ respectively column vectors of length $A\hat{A}$ and $\hat{A}$ with all 1's and by $e(s_0\hat{s}_0)$ and $e(\hat{s}_0)$ respectively row vectors with a 1 in position corresponding to $(s_0-1)\hat{A}+\hat{s}_0$ and $\hat{s}_0$, and 0 in all other components. We rewrite the bound (21) as

$$\bar{p}_e(\bar{s}_0) \le \frac{A\hat{A}^{\rho|\tau-1|-\rho+1}}{M^\rho}\left(\max_{s_0}\max_{\hat{s}_0} e(s_0\hat{s}_0)\Gamma^n_{\rho,\tau}\mathbf{1}\right) \times\left(\max_{\hat{s}_0}\left(e(\hat{s}_0)\hat{\Gamma}^n_{\tau}\mathbf{1}\right)^{\rho}\right) \quad (30)$$

We note that if both actual and mismatched models are irreducible, the product model corresponding to the matrix $\Gamma_{\rho,\tau}$ will have a single irreducible subset and the rest of the product states will be transient states, namely their stationary probability will be zero. Therefore we can omit rows and columns corresponding to those transient states from $\Gamma_{\rho,\tau}$ matrix and obtain an irreducible matrix. Assuming that the matrices $\Gamma_{\rho,\tau}$ and $\hat{\Gamma}_\tau$ are irreducible, using Perron-Frobenius theorem we know that they have largest magnitude eigenvalues with real positive values. We denote these dominant eigenvalues by $\lambda(\rho,\tau)$ and $\hat{\lambda}(\tau)$ and their corresponding positive right eigenvectors by $\boldsymbol{u}(\rho,\tau)=\big(u_1(\rho,\tau),\dots,u_{A\hat{A}}(\rho,\tau)\big)$, $u_{j\hat{j}}(\rho,\tau)>0$ and $\hat{\boldsymbol{u}}(\tau)=\big(u_1(\tau),\dots,u_{\hat{A}}(\tau)\big)$, $u_{\hat{j}}(\tau)>0$ respectively, such that

$$\Gamma_{\rho,\tau}\boldsymbol{u}(\rho,\tau) = \lambda(\rho,\tau)\boldsymbol{u}(\rho,\tau) \quad (31)$$

$$\hat{\Gamma}_\tau\hat{\boldsymbol{u}}(\tau) = \hat{\lambda}(\tau)\hat{\boldsymbol{u}}(\tau). \quad (32)$$

The positive right eigenvectors $\boldsymbol{u}(\rho,\tau)$ and $\hat{\boldsymbol{u}}(\tau)$ are unique except for a multiplicative factor. To make them unique we assume that

$$\sum_{j\hat{j}} u_{j\hat{j}}(\rho,\tau)=1, \qquad \sum_{\hat{j}}\hat{u}_{\hat{j}}(\tau)=1. \quad (33)$$

If we denote by $u_{\min}(\rho,\tau)$ and $u_{\min}(\rho,\tau)$ the largest and smallest component of the positive right eigenvector $\boldsymbol{u}(\rho,\tau)$, then for any s_0 and $\hat{s}_0$ we have

$$\frac{u_{\min}(\rho,\tau)}{u_{\max}(\rho,\tau)}\lambda^n(\rho,\tau) \le e(s_0\hat{s}_0)\Gamma^n_{\rho,\tau}\mathbf{1} \le \frac{u_{\max}(\rho,\tau)}{u_{\min}(\rho,\tau)}\lambda^n(\rho,\tau). \quad (34)$$

Using a similar bound on the second term in (30) we obtain

$$\bar{p}_e(\bar{s}_0) \le \frac{A\hat{A}^{\rho|\tau-1|-\rho+1}}{M^\rho}\cdot\frac{u_{\max}(\rho,\tau)}{u_{\min}(\rho,\tau)}\lambda^n(\rho,\tau) \times\left(\frac{\hat{u}_{\max}(\tau)}{\hat{u}_{\min}(\tau)}\hat{\lambda}^n(\tau)\right)^{\rho} \quad (35)$$

$$= e^{-n(\rho R - E_{\mathrm{s}}(\rho,\tau))}, \quad (36)$$

where

$$E_{\mathrm{s}}(\rho,\tau) = \log\lambda(\rho,\tau)+\rho\log\hat{\lambda}(\tau)+\delta_n \quad (37)$$

and

$$\delta_n = \frac{1}{n}\log\left(\frac{u_{\max}(\rho,\tau)}{u_{\min}(\rho,\tau)}\cdot\frac{\hat{u}_{\max}(\tau)^\rho}{\hat{u}_{\min}(\tau)^\rho}A\hat{A}^{\rho|\tau-1|-\rho+1}\right). \quad (38)$$

Finally, observe that for any $s_0,\hat{s}_0$ using (34) we obtain

$$\left|E_{\mathrm{s}}(\rho,\tau,s_0,\hat{s}_0)-\log\lambda(\rho,\tau)-\rho\log\hat{\lambda}(\tau)\right| \le \frac{1}{n}\log\left(\frac{u_{\max}(\rho,\tau)}{u_{\min}(\rho,\tau)}\cdot\frac{\hat{u}_{\max}(\tau)^\rho}{\hat{u}_{\min}(\tau)^\rho}\right). \quad (39)$$

■

We observe that the term (38) is the only term depending on the block length n is decreasing with n since the argument of the log function is greater than or equal to 1. This in turn shows that the corresponding achievable rate, termed the generalized entropy rate, is non-increasing in n and thus, it is attained in the limit for $n\to\infty$. The generalized entropy rate is defined as

$$H_{\mathrm{ger}}(\mathcal{V}) = \inf_{\tau\ge 0}\sum_{k\hat{k}} u_{k\hat{k}}(0)\sum_v -P_{V|S}(v|k) \times\log\left(\frac{Q_{V|S}(v|\hat{k})^\tau}{\sum_{\hat{k}'}\hat{u}_{\hat{k}'}(\tau)\sum_{\bar{v}}Q_{V|S}(\bar{v}|\hat{k}')^\tau}\right), \quad (40)$$

where $u_{k\hat{k}}(0)$ is the stationary probability of the product state $k\hat{k}$ and $\hat{u}_{\hat{k}'}(\tau)$ is the $\hat{k}'$-th element of the eigenvector $\hat{\boldsymbol{u}}(\tau)$.

Theorem 2. *For a finite-state source with irreducible model $\mathcal{S}$ using block code and a decoder based on a mismatched source model $\hat{\mathcal{S}}$, the generalized entropy rate $H_{\rm ger}(\mathcal{V})$ is an achievable rate.*

Proof: For any $\tau \geq 0$ the generalized entropy rate is obtained as $n \to \infty$ and thus it is given by

$$H_{\rm ger}(\mathcal{V},\tau) = \lim_{n\to\infty} \frac{\partial}{\partial\rho} E_{\rm s}(\rho,\tau,s_0,\hat{s}_0)|_{\rho=0} = \frac{\partial}{\partial\rho}\log\lambda(\rho,\tau)|_{\rho=0} + \log\hat{\lambda}(\tau). \quad (41)$$

The next steps to obtain the generalized entropy rate follow similar lines as [6] where Vašek derived the error exponent and entropy rate of an ergodic Markov source. Using (33), from (31) and (32) we have

$$\sum_{j\hat{j}}\sum_{k\hat{k}} \gamma_{j\hat{j}k\hat{k}}(\rho,\tau) u_{k\hat{k}}(\rho,\tau) = \sum_{j\hat{j}} \lambda(\rho,\tau) u_{j\hat{j}}(\rho,\tau) = \lambda(\rho,\tau), \quad (42)$$

and

$$\sum_{\hat{j}}\sum_{\hat{k}} \hat{\gamma}_{\hat{j}\hat{k}}(\tau)\hat{u}_k(\tau) = \sum_{\hat{j}} \hat{\lambda}(\tau)\hat{u}_{\hat{j}}(\tau) = \hat{\lambda}(\tau). \quad (43)$$

In the following, in order to simplify the notation, we define $p_{jk}(v) = P_{V,S|S}(v,j|k)$ and $q_{\hat{j}\hat{k}}(v) = Q_{V,S|S}(v,\hat{j}|\hat{k})$. Taking the derivative of $\log\lambda(\rho,\tau)$ with respect to ρ using (42) and simplifying it we obtain

$$\frac{\partial}{\partial\rho}\log\lambda(\rho,\tau) = -\sum_{j\hat{j}}\sum_{k\hat{k}}\sum_{v} \frac{p_{jk}(v)}{q_{\hat{j}\hat{k}}(v)^{\tau\rho}\lambda(\rho,\tau)} \times \log(q_{\hat{j}\hat{k}}(v)^{\tau}) u_{k\hat{k}}(\rho,\tau) + \sum_{j\hat{j}}\sum_{k\hat{k}} \frac{\gamma_{j\hat{j}k\hat{k}}(\rho,\tau)}{\lambda(\rho,\tau)} u'_{k\hat{k}}(\rho,\tau). \quad (44)$$

Since for $\rho = 0$ from (28) the entries of the matrix $\Gamma_{0,\tau}$ do not depend on τ, we omit the dependence on τ. We observe that the resulting matrix denoted by Γ_0 is a stochastic matrix with column sums equal to 1, i.e., $\sum_{j\hat{j}}\gamma_{j\hat{j}k\hat{k}}(0) = \sum_j p_{jk}(v) = 1$. Therefore it has a largest magnitude eigenvalue $\lambda(0) = 1$ with positive right eigenvector $\boldsymbol{u}(0)$ which is the stationary state distribution of the product finite-state model.

Evaluating (44) at $\rho = 0$ we obtain

$$\frac{\partial}{\partial\rho}\log\lambda(\rho,\tau)|_{\rho=0} = -\sum_{j\hat{j}}\sum_{k\hat{k}}\sum_{v} p_{jk}(v)\log(q_{\hat{j}\hat{k}}(v)^{\tau})u_{k\hat{k}}(0) + \sum_{k\hat{k}} u'_{k\hat{k}}(0). \quad (45)$$

Taking the derivative of both sides in (42) with respect to ρ we obtain

$$\sum_{j\hat{j}} \left(\lambda'(\rho,\tau)u_{j\hat{j}}(\rho,\tau) + \lambda(\rho,\tau)u'_{j\hat{j}}(\rho,\tau)\right) = \lambda'(\rho,\tau). \quad (46)$$

Simplifying the left hand side and canceling $\lambda'(\rho,\tau)$ from both sides we have

$$\lambda(\rho,\tau)\sum_{j\hat{j}} u'_{j\hat{j}}(\rho,\tau) = 0. \quad (47)$$

Since $\lambda(\rho,\tau)$ is strictly positive, we get $\sum_{j\hat{j}} u'_{j\hat{j}}(\rho,\tau) = 0$ and therefore, the second term in (45) is cancelled. Introducing (45) and (43) in (41) we obtain

$$H_{\rm ger}(\mathcal{V},\tau) = -\sum_{j\hat{j}}\sum_{k\hat{k}}\sum_{v} p_{jk}(v)\log(q_{\hat{j}\hat{k}}(v)^{\tau})u_{k\hat{k}}(0) + \log\left(\sum_{\hat{j}}\sum_{\hat{k}}\sum_{v} q_{\hat{j}\hat{k}}(v)^{\tau}\hat{u}_{\hat{k}}(\tau)\right). \quad (48)$$

Noting that $p_{jk}(v) = 0$ if $j \neq f(k,v)$ and similarly $q_{\hat{j}\hat{k}}(v) = 0$ if $\hat{j} \neq \hat{f}(\hat{k},v)$, we merge the sums over $j\hat{j}$ and $k\hat{k}$ and also sums over $\hat{j}$ and $\hat{k}$ in (48) obtaining

$$H_{\rm ger}(\mathcal{V},\tau) = -\sum_{k\hat{k}} u_{k\hat{k}}(0)\sum_{v} P_{V|S}(v|k)\log(Q_{V|S}(v|\hat{k})^{\tau}) + \log\left(\sum_{\hat{k}}\hat{u}_{\hat{k}}(\tau)\sum_{v} Q_{V|S}(v|\hat{k})^{\tau}\right). \quad (49)$$

Noticing that $\sum_{k\hat{k}} u_{k\hat{k}}(0)\sum_v P_{V|S}(v|k) = 1$, combining the two terms in (49) we obtain (40). ∎

III. Special cases

Using our general result from Section II, we recover special cases of a memoryless mismatched model and a matched finite-state model.

Theorem 3. *For a finite-state source with irreducible model $\mathcal{S}$, there exists a block code with $M = \lceil e^{nR}\rceil$ codewords such that using a decoder based on a memoryless mismatched source model for any initial state $\bar{s}_0$ we have*

$$\bar{p}_e(\bar{s}_0) \leq e^{-n\, e_r(R)}$$

where

$$e_r(R) = \sup_{\rho\in[0,1],\tau\geq 0} \rho R - E_{\rm s}(\rho,\tau),$$

and

$$E_{\rm s}(\rho,\tau) = \log\lambda(\rho,\tau) + \rho\log\sum_{v} Q_V(v)^{\tau} \quad (50)$$

$$+ \frac{1}{n}\log\left(\frac{u_{\max}(\rho,\tau)}{u_{\min}(\rho,\tau)}A\right), \quad (51)$$

with achievable rate

$$H_{\rm ger}(V) = \inf_{\tau\geq 0} -\sum_{v} P_V(v)\log\frac{Q_V(v)^{\tau}}{\sum_{\bar{v}} Q_V(\bar{v})^{\tau}}, \quad (52)$$

where $P_V(v) = \sum_k u_k(0)P_{V|S}(v|s=k)$.

In the case where the source is also memoryless, (51) reduces to

$$E_{\rm s}(\rho,s) = \log\sum_{v} P_V(v)\left(\frac{\sum_{\bar{v}} Q_V(\bar{v})^{\tau}}{Q_V(v)^{\tau}}\right)^{\rho}. \quad (53)$$

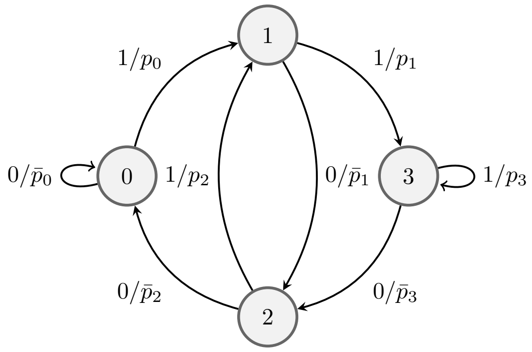

Fig. 1: Binary source with 4 states.

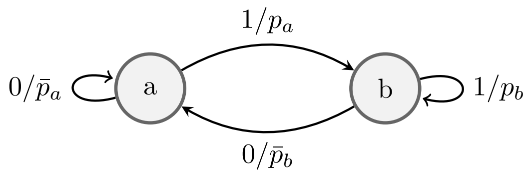

Fig. 2: Mismatched model with 2 states.

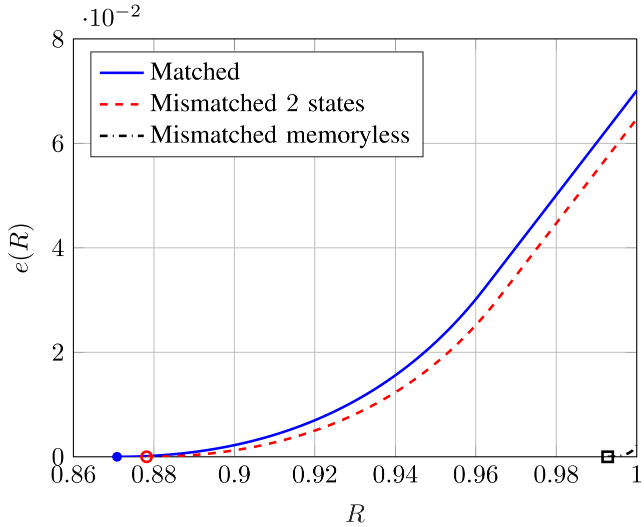

Fig. 3: Error exponents for Example 1.

with achievable rate given by (52).

Theorem 4. *For a finite-state source with irreducible model $\mathcal{S}$, there exists a block code with $M = \lceil e^{nR} \rceil$ codewords such that using a matched decoder for any initial state $\bar{s}_0$ we have*

$$\bar{p}_e(\bar{s}_0) \leq e^{-n\, e_r(R)}$$

where

$$e_r(R) = \max_{\rho \in [0,1]} \rho R - E_s(\rho),$$

and

$$E_s(\rho) = (1+\rho) \log \lambda(\rho) + \frac{1+\rho}{n} \log \left(\frac{u_{\max}(\rho)}{u_{\min}(\rho)} A \right), \tag{54}$$

with achievable rate

$$\begin{aligned} H_{\text{ger}}(\tau) &= \inf_{\tau \geq 0} -\sum_k u_k(0) \sum_v P_{V|S}(v|k) \log P_{V|S}(v|k), \\ &= \inf_{\tau \geq 0} \sum_k u_k(0) H(V|k). \end{aligned} \tag{55}$$

Example 1. *Consider a binary source with 4 states given in Fig. 1. The entropy rate of this source is given by $H(\mathcal{V}) = \sum_i \pi_i H(V|s_i)$ where π_i is the stationary probability of being in state s_i. Assuming the conditional distributions of the source as $\{p_0, p_1, p_2, p_3\} = \{0.3, 0.6, 0.2, 0.7\}$ we can calculate the stationary probabilities as $\{\pi_0, \pi_1, \pi_2, \pi_3\} = \{0.4, 0.15, 0.15, 0.3\}$ and the entropy rate of the source as $H(\mathcal{V}) = 0.8708$ bits. Assume that at the decoder we attempt to describe this source with three different models: i) a matched model, ii) a mismatched model with 2 states as shown in Fig. 2 with conditional probabilities $\{p_a, p_b\}$ as $p_a = \sum_{i \in \{0,2\}} \frac{\pi_i}{\pi_0 + \pi_2} p_i = 0.2727$ and $p_b = \sum_{i \in \{1,3\}} \frac{\pi_i}{\pi_1 + \pi_3} p_i = 0.6667$, iii) a memoryless model with distribution $\{1-p, p\}$ $p = \sum_i \pi_i p_i = 0.45$. The generalized entropy rate of the models are $H^{(ii)}_{\text{ger}}(\mathcal{V}) = 0.8782$ and $H^{(iii)}_{\text{ger}}(\mathcal{V}) = 0.9928$ bits, respectively. Fig. 3 illustrates the error exponent and entropy rate losses due mismatch.*

References

[1] L. Davisson, "Universal noiseless coding," *IEEE Trans. Inf. Theory*, vol. 19, no. 6, pp. 783–795, 1973.

[2] T. Ancheta, "Syndrome-source-coding and its universal generalization," *IEEE Trans. Inf. Theory*, vol. 22, no. 4, pp. 432–436, 1976.

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

[4] G. Caire, S. Shamai, and S. Verdú, "Noiseless data compression with low-density parity-check codes," *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 66, pp. 263–284, 2004.

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

[6] K. Vašek, "On the error exponent for ergodic Markov source," *Kybernetika*, vol. 16, no. 4, pp. 318–329, 1980.