Coding for the Gilbert-Elliott Channel: A Mismatched Decoding Perspective

Yutong Han
Technical University of Munich
yutong.han@tum.de

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

***Abstract*—We derive a lower bound to the error exponent of the Gilbert-Elliott channel by means of mismatched decoding. The corresponding achievable rate, the generalized mutual information, is shown to coincide with the lower bound of Mushkin and Bar-David.**

I. Introduction

The Gilbert-Elliott channel [1], [2] is an elementary binary-input binary-output finite-state channel (FSC) described by the two-state Markov chain in Fig. 1. When the channel is in the 'good' state, transmission occurs over a 'good' binary-symmetric channel (BSC) with crossover probability δ_{g}. Similarly, when the channel is in the 'bad' state, transmission occurs over a 'bad' BSC with crossover probability δ_{b}. In other words, the channel transition law $W(y|x,s)$ is determined by the BSC corresponding to the state. Reference [3] defined the channel memory as $\mu \triangleq 1-g-b$. For $\mu > 0$, the channel has a persistent memory, whereas for $\mu < 0$ it has an oscillatory memory. When $\mu = 0$ the channel is said to be memoryless, i.e. the current state is independent of all previous states. The Gilbert-Elliott channel is known to be indecomposable, i.e., the effect of the initial state dies away with time [4, Sec. 4.6]. We denote the stochastic Markov transition matrix by $\Gamma \triangleq \begin{bmatrix} p_{GG} & p_{GB} \\ p_{BG} & p_{BB} \end{bmatrix} = \begin{bmatrix} 1-b & b \\ g & 1-g \end{bmatrix}$ and denote by $[\pi_G, \pi_B] = \left[\frac{g}{g+b}, \frac{b}{g+b}\right]$ the stationary distribution of the Markov chain that defines the channel (see Fig. 1).

We define the channel input and output sequences $x^n, y^n \in \{0,1\}^n$, where n is the length of the sequences. We consider reliable transmission of M messages over the Gilbert-Elliott channel described above. Each message is assigned a codeword from a codebook $\mathcal{C} = \{x_1^n, \ldots, x_M^n\}$. The rate of the code is defined as $R = \frac{1}{n}\log M$. The channel capacity of the Gilbert-Elliott channel has been studied in a number of works but no single-letter closed-form expression has yet been found. Reference [3] derived upper and lower bounds to the capacity, which was numerically evaluated in [5]. Since the underlying channels are BSCs, the capacity is attained by an equiprobable input distribution $Q(0) = Q(1) = \frac{1}{2}$.

In this paper, we develop a mismatched decoding (see e.g. [6] and references therein) approach to coding over the Gilbert-Elliott channel. Specifically, we derive a lower bound to the error exponent by means of mismatched decoding, employing a memoryless decoding metric corresponding to a single BSC. We show that the corresponding achievable rate, the generalized mutual information (GMI) [7], coincides with the bound derived by Mushkin and Bar-David [3].

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.

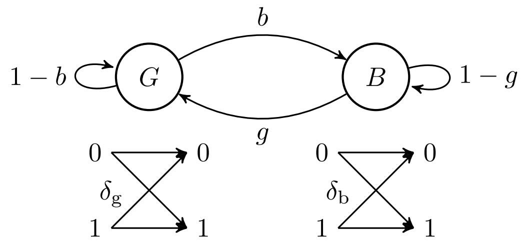

Fig. 1. Gilbert-Elliott channel model.

II. Mismatched Decoding

Mismatched decoding arises in situations where the decoder does not employ a maximum-likelihood decoder, but instead uses a maximum-metric decoder with a sub-optimal decoding metric $q^n(x^n, y^n)$ [6]. This occurs in a number of cases of relevance such as channel uncertainty, reduced-complexity decoding, bit-interleaved coded modulation, and zero-error communication [6]. In addition, mismatched decoding is employed to derive achievable information rates in situations where the channel capacity does not admit simple expressions. In these instances, a decoding metric that somehow simplifies the derivation is chosen. In this paper, although the channel has memory, we will assume a decoding metric that ignores this memory, i.e.,

$$q^n(x^n, y^n) = \prod_{i=1}^{n} q(x_i, y_i). \tag{1}$$

Specifically, we will assume that $q(x,y)$ is the channel transition probability of a single BSC with a crossover probability that depends on the Gilbert-Elliott channel parameters.

Following the footsteps of Gallager [4, Sec. 5.9], it can be shown that there exists a code of rate R and length n such that the error probability for a given message m, given the initial state s_0, can be bounded by

$$P_{e,m}(s_0) \le \mathrm{e}^{-n(E_r(R)-\epsilon)} \tag{2}$$

for any $\epsilon > 0$, and sufficiently large n, where

$$E_r(R) = \max_{0\le\rho\le 1} \sup_{\tau\ge 0} F_\infty(\rho,\tau) - \rho R \quad (3)$$

with $F_\infty(\rho,\tau) = \lim_{n\to\infty} F_n(\rho,\tau)$,

$$F_n(\rho,\tau) = \max_{Q_{X^n}} \min_{s_0} E_{0,n}(\rho,\tau,Q_{X^n},s_0) \quad (4)$$

$$E_{0,n}(\rho,\tau,Q_{X^n},s_0) = -\frac{1}{n}\log \mathbb{E}\left[\left(\sum_{\bar{x}^n} Q(\bar{x}^n)\frac{q^n(\bar{x}^n,Y^n)^\tau}{q^n(X^n,Y^n)^\tau}\right)^\rho\right] \quad (5)$$

where $\mathbb{E}[\cdot]$ denotes the expectation over the joint distribution given the initial state $P(x^n,y^n|s_0)$. The E_0 function will be denoted by $E_{0,n}(\rho,\tau)$ by leaving the dependencies on input distribution and initial state implicit. This exponent naturally leads to the following generalized mutual information rate

$$I_{\text{gmi}} = \sup_{\tau\ge 0}\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\log\frac{q^n(X^n,Y^n)^\tau}{\sum_{\bar{x}^n} Q(\bar{x}^n)q(\bar{x}^n,Y^n)^\tau}\right]. \quad (6)$$

III. Gilbert-Elliott Error Exponent

Let s^n be a binary state sequence of length n and let n_G be the number of good states in s^n. Define the set of all sequences with n_G good states as $\mathcal{T}^n_{n_G}$; this is the binary type class of type $\frac{n_G}{n}$. Also, let $E_0^{\mathrm{g}}(\rho,\tau)$ and $E_0^{\mathrm{b}}(\rho,\tau)$ be the mismatched E_0 functions corresponding to the good and bad BSCs with decoding crossover probabilities δ_q, and define $\Delta E_0(\rho,\tau) \triangleq E_0^{\mathrm{b}}(\rho,\tau) - E_0^{\mathrm{g}}(\rho,\tau)$. By spelling out the expectation in (5) and marginalizing over state sequences s^n, we have that

$$\sum_{s^n} P(s^n)\sum_{x^n,y^n} P(x^n,y^n|s^n)\left(\frac{\sum_{\bar{x}^n} Q(\bar{x}^n)q^n(\bar{x}^n,y^n)^\tau}{q^n(x^n,y^n)^\tau}\right)^\rho \quad (7)$$

$$= \sum_{s^n} P(s^n)\prod_{i=1}^n\sum_{x_i,y_i} P(x_i,y_i|s_i)\left(\frac{\sum_{\bar{x}} Q(\bar{x})q(\bar{x},y_i)^\tau}{q(x_i,y_i)^\tau}\right)^\rho \quad (8)$$

$$= \sum_{n_G=0}^n\sum_{\bar{s}^n\in\mathcal{T}^n_{n_G}} P(\bar{s}^n)e^{-n_G E_0^{\mathrm{g}}(\rho,\tau)}e^{-(n-n_G)E_0^{\mathrm{b}}(\rho,\tau)} \quad (9)$$

$$= e^{-nE_0^{\mathrm{b}}(\rho,\tau)}\sum_{n_G=0}^n\sum_{\bar{s}^n\in\mathcal{T}^n_{n_G}} P(\bar{s}^n)e^{n_G\Delta E_0(\rho,\tau)} \quad (10)$$

where (8) holds since we assume a product input distribution, a memoryless decoding metric, and the fact that $P(x^n,y^n|s^n) = \prod_{i=1}^n P(x_i,y_i|s_i)$, and (9) follows from re-writing as a function of n_G. We rewrite (10) as the expectation over the random variable N_G with $\Pr\{N_G = n_G\} = \sum_{\bar{s}^n\in\mathcal{T}^n_{n_G}} P(\bar{s}^n)$ as

$$F_\infty(\rho,\tau) = E_0^{\mathrm{b}}(\rho,\tau) - \lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\left[e^{N_G\Delta E_0(\rho,\tau)}\right]. \quad (11)$$

In order to calculate the error exponent of the Gilbert-Elliott channel we need to find $\lim_{n\to\infty} E_{0,n}(\rho,\tau)$. To this end, define the Markov composition of a given sequence s^n as $A(s^n) \triangleq \begin{bmatrix} n_{GG} & n_{GB} \\ n_{BG} & n_{BB}\end{bmatrix}$, where n_{jk} stands for the number of transitions from state j to state k, $j,k \in \{G,B\}$. By normalizing this matrix we get $\Phi(s^n) \triangleq \begin{bmatrix} f_{GG} & f_{GB} \\ f_{BG} & f_{BB}\end{bmatrix}$ where $f_{jk} = n_{jk}/n_j$ with n_j representing the number of state j in sequence s^n. Similarly we define the empirical distribution as $F(s^n) \triangleq [f_G, f_B]$ where $f_j = n_j/n$, and by construction $F(s^n)$ is the stationary distribution for $\Phi(s^n)$.

The probability of a given sequence s^n with Markov composition $A(s^n)$ can be expressed as [8]

$$P(s^n) = p_{GG}^{n_{GG}}p_{GB}^{n_{GB}}p_{BG}^{n_{BG}}p_{BB}^{n_{BB}} \quad (12)$$

$$= \exp\left[\sum_{j,k\in\{G,B\}} n_{jk}\log p_{jk}\right] \quad (13)$$

$$= \exp\left[n\sum_{j\in\{G,B\}} f_j\sum_{k\in\{G,B\}} f_{jk}\log p_{jk}\right] \quad (14)$$

$$= \exp\left[n\sum_{j\in\{G,B\}} f_j\sum_{k\in\{G,B\}} f_{jk}\log\frac{p_{jk}}{f_{jk}} + f_{jk}\log f_{jk}\right] \quad (15)$$

$$= \exp\left[-n\sum_{j\in\{G,B\}} f_j\Big(D(\Phi^{(j)}\|\Gamma^{(j)}) + H(\Phi^{(j)})\Big)\right] \quad (16)$$

$$= \exp\left[-n\big(D(\Phi\|\Gamma|F) + H(\Phi|F)\big)\right] \quad (17)$$

where (14) follows directly from the previous definition; (17) holds since $D(\Phi\|\Gamma|F)$ is the conditional relative entropy between the rows of Φ and those of Γ, that is

$$\begin{aligned} D(\Phi\|\Gamma|F) = f_G\bigg(f_{GG}\log\frac{f_{GG}}{p_{GG}} + f_{GB}\log\frac{f_{GB}}{p_{GB}}\bigg) \\ + f_B\bigg(f_{BG}\log\frac{f_{BG}}{p_{BG}} + f_{BB}\log\frac{f_{BB}}{p_{BB}}\bigg). \end{aligned} \quad (18)$$

Since $n_j = \sum_k n_{jk} = \sum_k n_{kj}$, symmetry property $n_{GB} = n_{BG}$ holds. Thus, the Markov type of a length-n sequence is determined if n_G and n_{GG} are known. Thus,

$$\sum_{\bar{s}^n\in\mathcal{T}^n_{n_G}} P(\bar{s}^n) = \sum_{n_{GG}=0}^{n_G}\sum_{\bar{s}^n\in\mathcal{A}^n_{n_G,n_{GG}}} p_{GG}^{n_{GG}}p_{GB}^{n_{GB}}p_{BG}^{n_{BG}}p_{BB}^{n_{BB}} \quad (19)$$

$$= \sum_{n_{GG}=0}^{n_G}\left|\mathcal{A}^n_{n_G,n_{GG}}\right| e^{-n[D(\Phi\|\Gamma|F)+H(\Phi|F)]} \quad (20)$$

where $\mathcal{A}^n_{n_G,n_{GG}}$ is the set of sequences with Markov type described by n_G and n_{GG}. Davisson *et al.* [8] showed that for a two-state Markov transition,

$$\left|\mathcal{A}^n_{n_G,n_{GG}}\right| \doteq e^{nH(\Phi|F)} \quad (21)$$

where the notation $a_n \doteq b_n$ means that $\lim_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0$. Substituting this into (20), we get

$$\sum_{\bar{s}^n \in \mathcal{T}^n_{n_G}} P(\bar{s}^n) \doteq \sum_{n_{GG}=0}^{n_G} e^{-nD(\Phi\|\Gamma|F)} \tag{22}$$

$$\doteq \max_{n_{GG}\in[0,n_G]} e^{-nD(\Phi\|\Gamma|F)} \tag{23}$$

$$= e^{-nD^*(\Phi\|\Gamma|F)} \tag{24}$$

where we denote $D^*(\Phi\|\Gamma|F) \triangleq \min_{f_{GG}\in[0,f_G]} D(\Phi\|\Gamma|F)$.

Lemma 1. *The minimum relative entropy in* (24) *is given by*

$$D^*(\Phi\|\Gamma|F) = \begin{cases} f_G \log \frac{bg-\beta+2\mu f_G}{2(1-b)\mu f_G} + f_B \log \frac{bg-\beta+2\mu f_B}{2(1-g)\mu f_B} & (\mu \neq 0) \\ f_G \log \frac{f_G}{g} + f_B \log \frac{f_B}{b} & (\mu = 0) \end{cases} \tag{25}$$

and it is achieved when

$$f^*_{GG} = \begin{cases} \frac{bg+2\mu f_G-\beta}{2\mu f_G} & \mu \neq 0, \\ f_G & \mu = 0. \end{cases} \tag{26}$$

where $\beta = \sqrt{b^2g^2 + 4bg\mu f_G f_B}$.

Observe that $D^*(\Phi\|\Gamma|F)$ is a function of f_G and the channel parameters. The divergence becomes zero if and only if $\Phi = \Gamma$. In other words, the empirical distribution is exactly equal to the stationary distribution, namely

$$\min_{f_G\in[0,1]} D^*(\Phi\|\Gamma|F) = D^*(\Phi\|\Gamma|F)\big|_{f_G=\pi_G} = 0. \tag{27}$$

To see this, we differentiate w.r.t. f_G, which gives

$$\frac{\partial D^*(\Phi\|\Gamma|F)}{\partial f_G} = \log \frac{(1-g)f_B f^*_{GG}}{(1-b)(f_B - f_G + f_G f^*_{GG})} = 0 \tag{28}$$

for $\mu \neq 0$. We find that $f_G = \pi_G$ and we write $f^*_{GG} = 1-b$, which by substituting back to (25), we get $D^*(\phi\|\Gamma|F) = 0$, which is achieved uniquely at $f_G = \pi_G$.

Applying the LogSumExp(LSE) inequality $\max_i a_i \leq \log\sum_{i=1}^n \exp(a_i) \leq \log n + \max_i a_i$ and substituting (24) into (11), 67we obtain

$$F_\infty(\rho,\tau) = E^{\mathrm{b}}_0(\rho,\tau) - \lim_{n\to\infty}\frac{1}{n}\max_{n_G\in[0,n]}\left[n_G\Delta E_0(\rho,\tau) - nD^*(\Phi\|\Gamma|F)\right] \tag{29}$$

$$= E^{\mathrm{b}}_0(\rho,\tau) - \max_{\lambda\in[0,1]}\left[\lambda\Delta E_0(\rho,\tau) - D^*(\Phi\|\Gamma|F)\big|_{f_G=\lambda}\right] \tag{30}$$

where the last equation holds by interchanging the maximum and limit as a result of the function inside the square bracket of (29) being uniformly continuous, and we denote $\lambda \triangleq \lim_{n\to\infty}\frac{n_G}{n}$.

It can be shown that the function $\lambda\Delta E_0(\rho,\tau) - D^*(\Phi\|\Gamma|F)\big|_{f_G=\lambda}$ is concave in λ. Thus, the maximum value can be found by equating the partial derivative to zero. which leads to the optimizing λ^*

$$\lambda^* = \frac{\sqrt{\alpha} - 1 + g + (1-b)e^{\Delta E_0(\rho,\tau)}}{2\sqrt{\alpha}} \tag{31}$$

with $\alpha = (1-g)^2 + 2(bg-\mu)e^{\Delta E_0(\rho,\tau)} + (1-b)^2 e^{2\Delta E_0(\rho,\tau)}$. By construction λ^* is independent of the blocklength, and substituting λ^* into (30) gives rise to the following theorem.

Theorem 1. *The mismatched Gilbert-Elliott F_∞ function is equal to*

$$F_\infty(\rho,\tau) = \lambda^* E^{\mathrm{g}}_0(\rho,\tau) + (1-\lambda^*)E^{\mathrm{b}}_0(\rho,\tau) + D^*(\Phi\|\Gamma|F)|_{f_G=\lambda^*} \tag{32}$$

with λ^* given in (31).

For memoryless channels with $\mu = 0$, i.e., the current state is independent of all previous states, we have

$$\lambda^* = \frac{ge^{\Delta E_0(\rho,\tau)}}{b + ge^{\Delta E_0(\rho,\tau)}}. \tag{33}$$

Together with (25), Theorem 1 can be written in the form

$$F_\infty(\rho,\tau) = E^{\mathrm{b}}_0(\rho,\tau) - \log\left(b + ge^{\Delta E_0(\rho,\tau)}\right). \tag{34}$$

Observe that applying Jensen's inequality to (11) yields

$$E_{0,n}(\rho,\tau) \leq E^{\mathrm{b}}_0(\rho,\tau) - \frac{\mathbb{E}[N_G]}{n}\Delta E_0(\rho,\tau). \tag{35}$$

Since the definition of stationarity implies

$$\lim_{n\to\infty}\frac{\mathbb{E}[N_G]}{n} = \pi_G \tag{36}$$

this gives the simple upper bound

$$F_\infty(\rho,\tau) \leq \pi_G E^{\mathrm{g}}_0(\rho,\tau) + \pi_B E^{\mathrm{b}}_0(\rho,\tau). \tag{37}$$

IV. Generalized Mutual Information

In this section, we study the GMI of the Gilbert-Elliott channel. We write the GMI as

$$I_{\mathrm{gmi}} = \sup_{\tau\geq 0} I_{\mathrm{gmi}}(\tau). \tag{38}$$

We rewrite (6) using the assumption of memoryless decoding metric as stated in (1)

$$I_{\mathrm{gmi}}(\tau) = \lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^n \log\frac{q(X_i,Y_i)^\tau}{\sum_{\bar{x}} Q(\bar{x})q(\bar{x},Y_i)^\tau}\right] \tag{39}$$

$$= \lim_{n\to\infty}\frac{1}{n}\sum_{s^n} P(s^n)\sum_{x^n,y^n} P(x^n,y^n|s^n,s_0) \times \sum_{i=1}^n \log\frac{q(x_i,y_i)^\tau}{\sum_{\bar{x}} Q(\bar{x})q(\bar{x},y_i)^\tau} \tag{40}$$

$$= \lim_{n\to\infty}\frac{1}{n}\sum_{s^n} P(s^n)\prod_{i=1}^n\sum_{x_i,y_i} P(x_i,y_i|s_i) \times \sum_{j=1}^n \log\frac{q(x_j,y_j)^\tau}{\sum_{\bar{x}} Q(\bar{x})q(\bar{x},y_j)^\tau} \tag{41}$$

where (40) follows by marginalizing over the state sequence s^n and (41) uses the fact that the state sequence is independent of the input sequence. Using the distributive law of multiplication and the fact that the term inside the logarithm only selects the corresponding joint probability while the rest will sum up to one, we can express the GMI in terms of conditional expectation as

$$I_{\mathrm{gmi}}(\tau) = \lim_{n\to\infty} \frac{1}{n} \sum_{s^n} P(s^n) \sum_{i=1}^{n} \mathbb{E}\left[\log \frac{q(X_i, Y_i)^\tau}{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, Y_i)^\tau} \middle| S_i \right] \tag{42}$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{s^n} P(s^n) \sum_{i=1}^{n} I_{\mathrm{gmi}}^{s_i}(\tau) \tag{43}$$

which is a general expression for FSCs with state transition being independent of the input sequence under any memoryless decoding metric. For the Gilbert-Elliott channel using the same argument as before, we can write

$$I_{\mathrm{gmi}}(\tau)$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{n_G=0}^{n} \sum_{\bar{s}^n \in \mathcal{T}_{n_G}^n} P(\bar{s}^n) \left[n_G I_{\mathrm{gmi}}^{\mathrm{g}}(\tau) + n_B I_{\mathrm{gmi}}^{\mathrm{b}}(\tau) \right] \tag{44}$$

$$= I_{\mathrm{gmi}}^{\mathrm{b}}(\tau) + \left[I_{\mathrm{gmi}}^{\mathrm{g}}(\tau) - I_{\mathrm{gmi}}^{\mathrm{b}}(\tau) \right] \lim_{n\to\infty} \frac{\mathbb{E}(N_G)}{n} \tag{45}$$

which using (36) yields

$$I_{\mathrm{gmi}} = \sup_{\tau \geq 0} \pi_G I_{\mathrm{gmi}}^{\mathrm{g}}(\tau) + \pi_B I_{\mathrm{gmi}}^{\mathrm{b}}(\tau), \tag{46}$$

the weighted sum of the GMIs per channel, weighted by the stationary distribution.

Given that the memoryless decoding metric $q(x, y)$ can be chosen arbitrarily, we select it as a BSC with crossover probability δ_q with $0 < \delta_q < 0.5$. In this case, we have that

$$I_{\mathrm{gmi}}^{\mathrm{g}}(\tau) = \log \frac{2}{\delta_q^\tau + (1-\delta_q)^\tau} + (1-\delta_{\mathrm{g}}) \log(1-\delta_q)^\tau + \delta_{\mathrm{g}} \log \delta_q^\tau \tag{47}$$

and $I_{\mathrm{gmi}}^{\mathrm{b}}(\tau)$ has a similar form with δ_{g} replaced by δ_{b}. It was shown in [6, Sec. 2] that for a given $q(x, y)$, the GMI is a concave maximization problem (38).

For the optimal τ, the I_{gmi} can be shown to be concave in δ_q. Then we can determine the optimal value of τ and δ_q that maximize the GMI by setting the partial derivatives to zero, yielding $\tau = 1$ and $\delta_q = \pi_G \delta_{\mathrm{g}} + \pi_B \delta_{\mathrm{b}}$.

Theorem 2. *The GMI of the Gilbert-Elliott channel using a mismatched BSC with crossover probability $\delta_q = \pi_G \delta_{\mathrm{g}} + \pi_B \delta_{\mathrm{b}}$ for decoding is*

$$I_{\mathrm{gmi}} = \log 2 - h_2(\pi_G \delta_{\mathrm{g}} + \pi_B \delta_{\mathrm{b}}) \tag{48}$$

where $h_2(p) \triangleq -p \log p - (1-p) \log(1-p)$ denotes the binary entropy function. Equality in (48) *is attained if and only if $\mu = 0$, i.e., in the memoryless case.*

We illustrate Theorems 1 and 2 by means of an example for a persistent Gilbert-Elliott channel with parameters $b = 0.1,\ g = 0.4,\ \delta_{\mathrm{g}} = 0.05$ and $\delta_{\mathrm{b}} = 0.2$. The F_∞ function given in (32) and the Jensen's inequality upper bound in (37) are depicted in Fig. 2 together with those for the good and

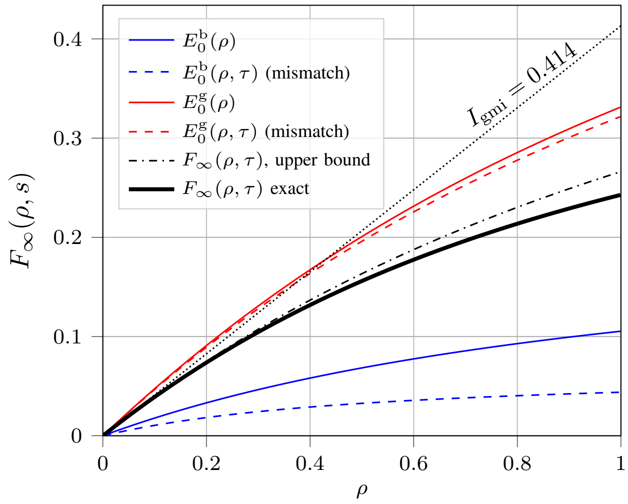

Fig. 2. Function F_∞ for a persistent Gilbert-Elliott channel with parameters $b = 0.1,\ g = 0.4,\ \delta_{\mathrm{g}} = 0.05$ and $\delta_{\mathrm{b}} = 0.2$.

The capacity lower bound (48) coincides with the lower bound Mushkin and Bar-David [3, eq. (2.31)].
bad states using a mismatched BSC with crossover probability $\delta_q = \pi_G \delta_{\mathrm{g}} + \pi_B \delta_{\mathrm{b}}$. The parameter τ has been optimized in all curves. We use (48) to compute

$$I_{\mathrm{gmi}} = \log 2 - h_2\left(\frac{0.4}{0.4 + 0.1} \times 0.05 + \frac{0.1}{0.4 + 0.1} \times 0.2 \right) \tag{49}$$

$$= 0.414 \text{ nat/channel use} \tag{50}$$

which coincides with the gradient of $F_\infty(\rho, \tau)$ at $\rho = 0$. As we observe, the upper bound is very close to F_∞ for small values of ρ.

References

[1] E. N. Gilbert, "Capacity of a burst-noise channel," *The Bell System Technical Journal*, vol. 39, no. 5, pp. 1253–1265, 1960.

[2] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *The Bell System Technical Journal*, vol. 42, no. 5, pp. 1977–1997, 1963.

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

[5] M. Rezaeian, "Computation of capacity for gilbert-elliott channels, using a statistical method," in *2005 Australian Communications Theory Workshop*, 2005, pp. 56–61.

[6] J. Scarlett, A. Guillén i Fàbregas, A. Somekh-Baruch, and A. Martinez, "Information-theoretic foundations of mismatched decoding," *Found. Trends Commun. Inf. Theory*, vol. 17, no. 2-3, pp. 149–401, 2020. [Online]. Available: https://doi.org/10.1561/0100000101

[7] G. Kaplan and S. Shamai, "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *AEU. Archiv für Elektronik und Übertragungstechnik*, vol. 47, no. 4, pp. 228–239, 1993.

[8] L. Davisson, G. Longo, and A. Sgarro, "The error exponent for the noiseless encoding of finite ergodic Markov sources," *IEEE Trans. Inf. Theory*, vol. 27, no. 4, pp. 431–438, 1981.