Single-Letter Mismatched Decoding Rates with Memory for the Gilbert-Elliott Channel

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

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

***Abstract*—We derive closed-form expressions of the generalized mutual information (GMI) for the Gilbert-Elliott channel by introducing memory in the decoding metric. We first study the simple case of block memory and then propose a unifilar decoder that explicitly tracks state memory. We show that the GMI for both decoders exhibits a monotonic improvement with the memory order, and that the rates achieved by the unifilar decoder of a given memory order are always higher than those achieved by the block decoder.**

I. INTRODUCTION

In high-frequency-band communication and cellular telephony, the transmitted signal arrives at the receiver via multiple paths, each of a different intensity, at several times. Channel memory effects need to be taken into account, and the assumption of uncorrelated noise between successive channel uses is no longer valid. A general framework for modeling such channels with memory is the class of finite-state channels (FSCs), in which the channel output depends on both the input and an underlying channel state. FSCs provide a practical and tractable way to model flat fading channels, intersymbol interference (ISI) channels, and channels that exhibit both fading and ISI [1].

If the transmitter and receiver have perfect channel state information, the FSC capacity is the statistical average across all states [1]. Without any information regarding the channel state or its transition structure, the capacity is reduced to that of an arbitrarily varying channel [2]. This paper explores the intermediate scenario where the channel transition structure of the FSC is known. Assuming the channel model meets some mild technical conditions [1, Th. 4.6.3], the capacity of FSCs is given by the multi-letter expression [1, Th. 4.6.4]

$$C = \lim_{n\to\infty} \sup_{Q_{X^n}} \frac{1}{n} I(X^n; Y^n) \tag{1}$$

where X^n, Y^n are the input and output sequences of the channel, and the optimization is over all choices of the input distribution Q_{X^n}. No closed-form finite-letter characterization exists, except for very specific cases. It was recently shown in [3] that the FSC capacity is not a computable function. Even for many simple FSC models, the capacity remains an open problem. In this paper, we study the Gilbert-Elliott channel [4], [5], for which only a single-letter upper and lower bounds are known [6]. The capacity can be simulated by the recursion in e.g. [7].

Mismatched decoding is the setting in which the decoder, instead of performing optimal maximum likelihood, maximizes a given decoding metric (see e.g., [8]). While typically employed to model channel uncertainty and low-complexity decoding, mismatched decoding can also be employed as a means to derive achievable rates by imposing relatively simple decoding metrics when the real channel description is complex.

This work mainly considers the Gilbert-Elliott channel under two types of decoders: block decoders and unifilar decoders. We prove that the optimized generalized mutual information (GMI) increases monotonically with the decoder memory for both decoders. Furthermore, we show that a unifilar Markov decoder with memory m achieves strictly better, or at least equal, performance compared to a block $(m+1)$ decoder. This improvement arises from the decoder's ability to leverage temporal dependencies in the sequence, whereas block decoders are constrained by fixed-length segmentation, leading to suboptimal use of historical information.

Notation: Discrete scalar random variables are denoted by capital letters, their realizations by respective lowercase letters, and their alphabets by corresponding calligraphic letters. The cardinality of a finite set $\mathcal{X}$ is denoted by $|\mathcal{X}|$. We represent vectors of length n in alphabet $\mathcal{X}^n$ by $x^n = (x_1, x_2, \cdots, x_n)$. For finite $\mathcal{X}$, the probability mass function (PMF) in $\mathcal{X}^n$ is denoted by P_{X^n} and the conditional PMF by $P_{Y^n|X^n}$.

Information-theoretic quantities are denoted following the usual conventions, namely entropy $H(X)$ and conditional entropy $H(Y|X)$. We represent the entropy rate as

$$H_\infty(X) = \lim_{n\to\infty} \frac{1}{n} H(X^n). \tag{2}$$

The expectation operator is represented by $\mathbb{E}(\cdot)$. All logarithms are taken to the base 2 unless otherwise stated.

II. PRELIMINARIES

A. Gilbert-Elliott Channel

The Gilbert-Elliott channel is an elementary binary-input binary-output FSC described by the two-state Markov chain

This work has been funded in part by the European Research Council under ERC grant agreements 725411 and 101142747, by the Spanish Ministry of Economy and Competitiveness under grant PID2020-116683GB-C22, and the German Research Foundation (DFG) via the German-Israeli Project Cooperation (DIP) under project KR3517/13-1.

in Fig. 1. When the channel is in the good state, transmission occurs over a binary symmetric channel (BSC) with crossover probability δ_{g}. Similarly, when the channel is in the bad state, transmission occurs over a BSC with crossover probability $\delta_{\mathrm{b}} > \delta_{\mathrm{g}}$. In other words, the channel transition law $W(y|x,s)$ is determined by the BSC corresponding to the state. This channel is known to be *indecomposable*, i.e., the effect of the initial state dies away with time, and *non-anticipatory*, i.e., the current output is statistically independent of all future inputs [1, Sec. 5.9]. We denote the steady-state distribution of the Markov chain by $[\pi_{\mathrm{g}}, \pi_{\mathrm{b}}] = \left[\frac{g}{g+b}, \frac{b}{g+b}\right]$.

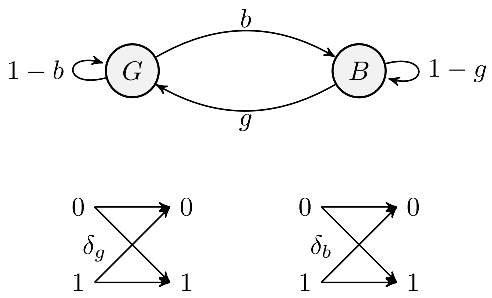

Fig. 1. Gilbert-Elliott channel model.

We consider coded communication over the Gilbert-Elliott channel and denote the codebook $\mathcal{C} = \{x^n(1), \ldots, x^n(M)\}$ as the set of all M codewords of length n, that is, $x^n(i) \in \mathcal{X}^n$. The rate of the code is defined as $R = \frac{1}{n}\log M$. The corresponding channel output is denoted by $y^n \in \mathcal{Y}^n$. Since the underlying channels are BSCs, the capacity is attained by the equiprobable input distribution. Since the Gilbert-Elliott channel can be interpreted as a binary additive-noise channel, its capacity is determined by the entropy rate of the noise process Z^n, namely

$$C = 1 - H_\infty(Z). \tag{3}$$

The capacity of the Gilbert-Elliott channel has been studied in several works. Mushkin and Bar-David [6] derived upper and lower bounds, but no single-letter expressions of the capacity have been derived – one can only simulate it (see e.g. [7]). The main reason is that the noise process Z^n is described by a hidden Markov model; single-letter expressions for the entropy rate of hidden Markov models are not available.

B. Mismatched Decoding Framework

Mismatched decoding is the setting by which the decoder, instead of performing optimal maximum likelihood, maximizes a given decoding metric

$$\hat{m} = \arg\max_{i \in \{1,\ldots,M\}} q^n(x^n(i), y^n). \tag{4}$$

While typically employed to model channel uncertainty and low-complexity decoding, mismatched decoding can also be employed as a means to derive achievable rates by imposing relatively simple decoding metrics when the real channel description is complex.

The generalized mutual information (GMI) is a simple achievable rate, whose multi-letter version can be expressed as (see e.g. [8])

$$I_{\mathrm{gmi}} = \sup_{\tau \geq 0} \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left[\log \frac{q^n(X^n, Y^n)^\tau}{\sum_{\bar{x}^n} Q(\bar{x}^n) q^n(\bar{x}^n, Y^n)^\tau}\right]. \tag{5}$$

For a memoryless decoding metric of the form

$$q^n(x^n, y^n) = \prod_{i=1}^{n} q(x_i, y_i), \tag{6}$$

it is established in [9] that selecting $q(x_i, y_i)$ to emulate a BSC with crossover probability $\delta^* = \pi_{\mathrm{g}}\delta_{\mathrm{g}} + \pi_{\mathrm{b}}\delta_{\mathrm{b}}$ maximizes the GMI for the Gilbert-Elliott channel. The resulting maximal GMI is given by the single-letter expression

$$I_{\mathrm{gmi}} = 1 - h_2(\delta^*), \tag{7}$$

where $h_2(p) = -p\log_2 p - (1-p)\log_2(1-p)$ denotes the binary entropy function. This expression recovers the lower bound by Mushkin and Bar-David [6] using mismatched decoding.

The channel's inherent infinite memory makes the memoryless decoding metric overly restrictive, as it ignores temporal dependencies introduced by the channel. To reconcile this limitation while retaining tractability, in this paper, we develop two ways of adding memory to the decoding metric that lend themselves to a single-letter characterization of the corresponding GMI. In both cases, we observe that when the memory order of the decoding metric increases, both methods monotonically approach the capacity of the channel.

III. Block Decoding Metrics

In this section, we develop block-memory, i.e., a decoding metric in which memorylessness is enforced only across blocks (i.e., dependencies are localized within fixed-length segments), approximating the channel's infinite-memory behavior without incurring the full computational complexity of maximum likelihood decoding. In the following, for simplicity, we use a block decoder with memory 2 as an illustrative example.

A block-2 decoding metric can be factorized as

$$q^n(x^n, y^n) = \prod_{\substack{i=1 \\ i \text{ even}}}^{n} q^2(x_{i-1}, y_{i-1}, x_i, y_i). \tag{8}$$

In other words, the block-2 decoder's operation induces an equivalent quaternary channel used for decoding. In this work, we consider a quaternary symmetric channel (QSC) model, with transition matrix

$$\Gamma_{\mathrm{QSC}} = \begin{bmatrix} \delta_{00} & \delta_{01} & \delta_{10} & \delta_{11} \\ \delta_{01} & \delta_{00} & \delta_{11} & \delta_{10} \\ \delta_{10} & \delta_{11} & \delta_{00} & \delta_{01} \\ \delta_{11} & \delta_{10} & \delta_{01} & \delta_{00} \end{bmatrix} \tag{9}$$

where we have assumed the QSC employed at the decoder is memoryless and $\delta_{00} + \delta_{01} + \delta_{10} + \delta_{11} = 1$.

Since optimizing $q^2(x_{i-1}, y_{i-1}, x_i, y_i)^\tau$ is equivalent to setting $\tau = 1$ and optimizing $q^2(x_{i-1}, y_{i-1}, x_i, y_i)$, we substitute (8) into (5) to obtain

$$
\begin{aligned}
I_{\mathrm{gmi}} &= \lim_{n\to\infty} \frac{1}{n} \sum_{s^n} P(s^n) \sum_{x^n}\sum_{y^n} Q(x^n) W(y^n|x^n, s^n) \times \\
&\sum_{\substack{i=1 \\ i \text{ even}}}^{n} \log \frac{q^2(x_{i-1}, y_{i-1}, x_i, y_i)}{\sum_{\bar{x}_{k-1},\bar{x}_k} Q(\bar{x}_{k-1}, \bar{x}_k) q^2(\bar{x}_{k-1}, y_{i-1}, \bar{x}_k, y_i)} && (10)\\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{s^n} P(s^n) \sum_{\substack{i=1 \\ i \text{ even}}}^{n} \sum_{\substack{x_{i-1},x_i \\ y_{i-1},y_i}} P(x_{i-1}^i, y_{i-1}^i | s_{i-1}^i) \times \\
&\log \frac{q^2(x_{i-1}, y_{i-1}, x_i, y_i)}{\sum_{\bar{x}_{k-1},\bar{x}_k} Q(\bar{x}_{k-1}, \bar{x}_k) q^2(\bar{x}_{k-1}, y_{i-1}, \bar{x}_k, y_i)} && (11)\\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{s^n} P(s^n) \sum_{\substack{i=1 \\ i \text{ even}}}^{n} I_{\mathrm{gmi}}(s_{i-1}, s_i) && (12)\\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{\substack{(n_{ss'}) \\ \sum n_{ss'}=n/2}} \sum_{\bar{s}^n \in \mathcal{T}^n_{n_{ss'}}} P(\bar{s}^n) \sum_{s,s'} n_{ss'} I_{\mathrm{gmi}}(s, s') && (13)\\
&= \sum_{s,s' \in \{G,B\}} I_{\mathrm{gmi}}(s, s') \lim_{n\to\infty} \frac{1}{n} \mathbb{E}[N_{ss'}] && (14)\\
&= \frac{1}{2} \sum_{s,s'} \pi(s, s') I_{\mathrm{gmi}}(s, s') && (15)
\end{aligned}
$$

where

$$
\begin{aligned}
I_{\mathrm{gmi}}(s, s') = &\sum_{x,x'} Q(x)Q(x') \sum_{y,y'} W(y|x, s) W(y'|x', s') \\
&\times \log \frac{q^2(x, y, x', y')}{\sum_{\bar{x},\bar{x}'} Q(\bar{x})Q(\bar{x}') q^2(\bar{x}, y, \bar{x}', y')}, \qquad (16)
\end{aligned}
$$

(11) follows from 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; $n_{ss'}$ is the number of transitions from state s to state s in a given state sequence s^n; $\mathcal{T}^n_{n_{ss'}}$ represents the set of state sequences with transition counts $n_{ss'} \in \{n_{\mathrm{gg}}, n_{\mathrm{gb}}, n_{\mathrm{bg}}, n_{\mathrm{bb}}\}$; and the last equality follows from the the stationarity of the chain.

For the Gilbert-Elliott channel with decoding metric (9) under equiprobable input distribution, the GMI in (15) satisfies

$$
\begin{aligned}
I_{\mathrm{gmi}} &= 1 + \frac{1}{2} \sum_{z_1,z_2} P(z_1, z_2) \log \delta_{z_1 z_2} && (17)\\
&\leq 1 + \frac{1}{2} \sum_{z_1,z_2} P(z_1, z_2) \log P(z_1, z_2) && (18)
\end{aligned}
$$

where $\delta_{z_1 z_2}$ are the parameters of the QSC used for decoding and where the inequality follows from the non-negativity of KL divergence, i.e., the cross entropy bound. This suggests that the optimal decoding parameters are

$$
\delta^*_{z_1 z_2} = P(z_1, z_2) = \sum_{s_1,s_2} \pi(s_1) P(s_2|s_1) P(z_1|s_1) P(z_2|s_2). \qquad (19)
$$

Therefore, we conclude that the best GMI rate under a block decoder of memory order $m = 2$ satisfies

$$
I^{\mathrm{block}}_{\mathrm{gmi}} = 1 - \frac{1}{2} H(Z^2). \qquad (20)
$$

The preceding analysis naturally extends to decoders with arbitrary memory order m, yielding Theorem 1.

Theorem 1. *The optimal GMI of the Gilbert-Elliott channel using a mismatched block decoder with memory order m is*

$$
I^{\mathrm{block}}_{\mathrm{gmi}}(m) = 1 - \frac{1}{m} H(Z^m) \qquad (21)
$$

where $H(Z^m)$ is the joint steady-state noise entropy.

Proposition 1. *$I^{\mathrm{block}}_{\mathrm{gmi}}(m)$ is non-decreasing with m.*

Proof. For a stationary stochastic noise process Z^m, the normalized entropy rate $\frac{H(Z^m)}{m}$ is non-increasing in m. ☐

Proposition 2. *The channel capacity is recovered by choosing an infinitely long block, i.e.,*

$$
\lim_{m\to\infty} I^{\mathrm{block}}_{\mathrm{gmi}}(m) = 1 - \lim_{m\to\infty} \frac{1}{m} H(Z^m) = 1 - H_\infty(Z). \qquad (22)
$$

IV. Unifilar Decoding Metrics

In the previous section, we studied block decoders, which partially account for memory effects compared to strictly memoryless decoders. However, since the channel itself possesses infinite memory, the assumption of memorylessness across blocks remains restrictive. To address this limitation, we now turn to unifilar decoders, which explicitly incorporate state-dependent decoding.

Specifically, we consider a memory-1 unifilar decoder with decoding states $\{0, 1\}$, where the next state s_{i+1} is deterministic given the current input x_i, output y_i, and state s_i. Namely, the state update rule is

$$
s_{i+1} = \begin{cases} 0 & \text{if } x_i \oplus y_i = 0, \\ 1 & \text{otherwise.} \end{cases} \qquad (23)
$$

Each state s is associated with a BSC with crossover probability δ_{s}, as summarized in Fig. 2. While this framework generalizes to arbitrary order memories, we focus here on the memory-1 case for clarity.

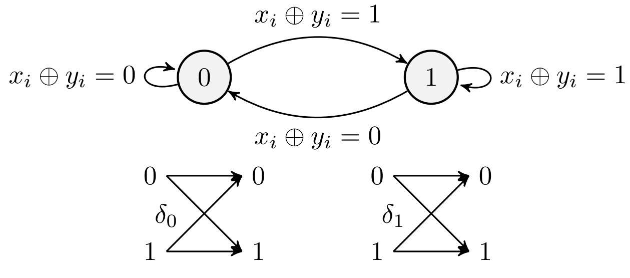

Fig. 2. Unifilar decoder with memory-1.

Given the state-dependent decoding process described above, we formulate the decoding metric for the memory-1 unifilar decoder as

$$q^n(x^n, y^n) = \prod_{i=1}^{n} q(x_i, y_i | x_{i-1}, y_{i-1}), \tag{24}$$

where (x_0, y_0) gives the initial decoding state and is assumed to be known and fixed. Choosing $\tau = 1$ and equiprobable inputs, we can express the denominator of the GMI as

$$\sum_{\bar{x}^n} Q(\bar{x}^n) q^n(\bar{x}^n, y^n)$$
$$= \frac{1}{2^n} \sum_{\bar{x}_1} q(\bar{x}_1, y_1 | x_0, y_0) \cdots \sum_{\bar{x}_n} q(\bar{x}_n, y_n | \bar{x}_{n-1}, y_{n-1}) \tag{25}$$
$$= \frac{1}{2^n} \tag{26}$$

where the last equality follows from the fact that for any fixed decoding state s, $\sum_{\bar{x}} q(\bar{x}, y|s) = 1$ regardless of the value of y. Substituting this expression for the denominator into (5) yields:

$$I_{\text{gmi}} = 1 + \lim_{n\to\infty} \frac{1}{n} \sum_{s^n} P(s^n) \sum_{x^n} \sum_{y^n} Q(x^n) W(y^n | x^n, s^n) \times \sum_{i=1}^{n} \log q(x_i, y_i | x_{i-1}, y_{i-1}) \tag{27}$$
$$= 1 + \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{s^n} P(s^n) \sum_{\substack{x_{i-1}, x_i \\ y_{i-1}, y_i}} Q(x_{i-1}, x_i) \times W(y_{i-1} | x_{i-1}, s_{i-1}) W(y_i | x_i, s_i) \log q(x_i, y_i | x_{i-1}, y_{i-1}) \tag{28}$$
$$= 1 + \sum_{s_1, s_2} \pi(s_1, s_2) \sum_{\substack{x_1, x_2 \\ y_1, y_2}} Q(x_1, x_2) W(y_1 | x_1, s_1) \times W(y_2 | x_2, s_2) \log q(x_2, y_2 | x_1, y_1) \tag{29}$$
$$= 1 + \sum_{z_1, z_2} P(z_1, z_2) \log q(z_2 | z_1) \tag{30}$$
$$\leq 1 + \sum_{z_1, z_2} P(z_1, z_2) \log P(z_2 | z_1) \tag{31}$$

where (28) follows from an analogous derivation to that used in (11), (29) results again from the stationarity of Markov chains, (30) is obtained by rewriting in terms of noise symbols and marginalizing over the states, and the last inequality can be shown using the non-negativity of the KL divergence.

This implies that the optimal decoding parameters are given by the conditional stationary transition probabilities of the noise process, i.e., $q(z_2|z_1) = P(z_2|z_1)$. In other words,

$$\begin{cases} \delta_0^* &= P_{Z_2|Z_1}(1|0) \\ \delta_1^* &= P_{Z_2|Z_1}(1|1). \end{cases} \tag{32}$$

Thus, the maximum GMI rate under a unifilar decoder with memory order $m = 1$ is given by

$$I_{\text{gmi}}^{\text{unifilar}} = 1 - H(Z_2 | Z_1). \tag{33}$$

The preceeding analysis applies to any finite memory m, as seen in Theorem 2.

Theorem 2. *The optimal GMI of the Gilbert-Elliott channel using a mismatched unifilar decoder with memory m is*

$$I_{\text{gmi}}^{\text{unifilar}}(m) = 1 - H(Z_{m+1} | Z^m) \tag{34}$$

where $H(Z_{m+1}|Z^m)$ is the conditional steady-state noise entropy.

Proposition 3. *$I_{\text{gmi}}^{\text{unifilar}}(m)$ is non-decreasing with m.*

Proof. For a stationary stochastic process Z^{m+1}, the conditional entropy $H(Z_{m+1}|Z^m)$ is non-increasing in m. □

Proposition 4. *The optimal GMI under a memory-m unifilar decoder is no less than that under a block-$(m+1)$ decoder.*

Proof. For a stationary stochastic process Z^{m+1}, it is known that $H(Z_{m+1}|Z^m) \leq \frac{H(Z^{m+1})}{m+1}$ [1]. □

Combining Proposition 4 with the limiting behavior of block decoders established in Proposition 2 yields the following result.

Proposition 5. *The channel capacity is recovered by choosing a unifilar decoder with infinite memory, i.e.,*

$$\lim_{m\to\infty} I_{\text{gmi}}^{\text{unifilar}}(m) = 1 - \lim_{m\to\infty} H(Z_{m+1} | Z^m) \tag{35}$$
$$= 1 - H_\infty(Z). \tag{36}$$

V. Numerical Results

We now compare the GMI achievable by block decoders and unifilar decoders for a Gilbert-Elliott channel with parameters $b = 0.1$, $g = 0.3$, $\delta_{\text{g}} = 0.1$ and $\delta_{\text{b}} = 0.4$, as shown in Fig. 3. The black line depicts the simulated capacity obtained via the coin-tossing method [7] with a blocklength of $n = 10^8$. Optimal GMIs for block and unifilar decoders appear as blue and magenta curves, respectively. The shaded green and red regions highlight rates within 0.5% and 0.1% of the simulated capacity, providing visual quantification of the decoders' proximity to the fundamental limit. We observe that both decoders exhibit monotonic convergence with increasing m, though the unifilar decoder consistently achieves higher rates due to its ability to exploit the state memory, as stated in Proposition 4.

We also examine a more extreme channel scenario with parameters $b = 0.01$, $g = 0.02$, $\delta_g = 0.05$, $\delta_b = 0.4$, as shown in Fig. 4. These parameters induce stronger channel memory, evidenced by the more persistent transition probabilities. As expected, the unifilar decoder again outperforms the block decoder across all m, though convergence is significantly slower, which is a direct consequence of the channel's more extreme memory effects. This behavior aligns with our theoretical predictions, as the increased autocorrelation in channel states requires longer observation windows to reliably estimate the underlying Markov structure of the channel state transitions.

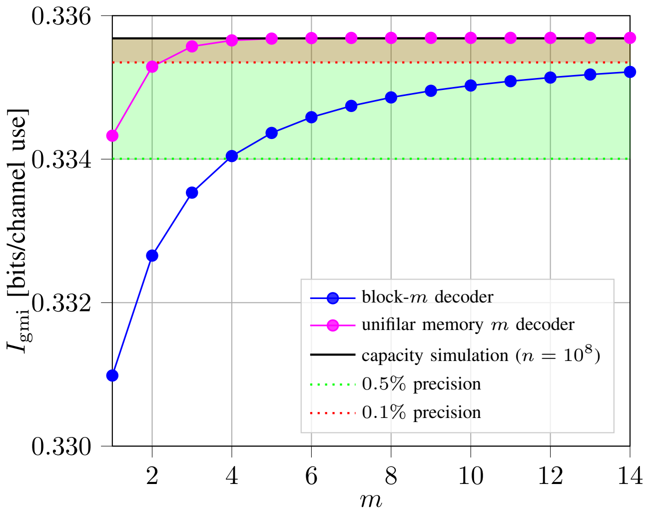

Fig. 3. GMI comparison for $b = 0.1,\ g = 0.3,\ \delta_g = 0.1,\ \delta_b = 0.4$.

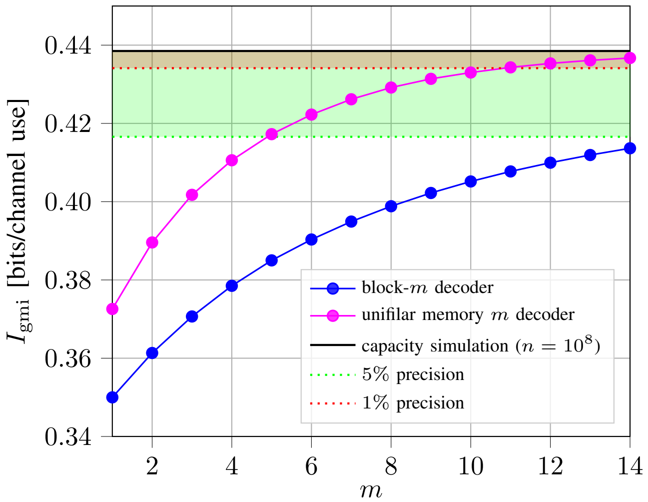

Fig. 4. GMI comparison for $b = 0.01,\ g = 0.02,\ \delta_g = 0.05,\ \delta_b = 0.4$.

References

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

[2] I. Csiszár and J. Körner, "Information theory: Coding theorems for discrete memoryless channels," *New York: Academic Press*, 1981.

[3] H. Boche, R. F. Schaefer, and H. V. Poor, "Shannon meets Turing: Non-computability and non-approximability of the finite state channel capacity," *Communications in Information and Systems*, vol. 20, no. 2, pp. 81–116, 2020.

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

[5] 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.

[6] 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.

[7] M. Rezaeian, "Computation of capacity for Gilbert-Elliott channels, using a statistical method," in *2005 Australian Commun. Theory Workshop*, 2005, pp. 56–61.

[8] 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

[9] Y. Han and A. Guillén i Fàbregas, "Coding for the Gilbert-Elliott channel: A mismatched decoding perspective," in *Proc. 2024 International Zürich Seminar, Zürich, Switzerland, Mar. 2024*, 2024.