Mismatched Decoding Rates for the Gaussian 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 study the Gaussian Gilbert-Elliott channel, which is a communication channel modeled as a two-state hidden Markov process with Gaussian observation variables. We derive the generalized mutual information (GMI) with a memoryless Gaussian mixture decoding metric and standard Gaussian inputs. We show that the GMI equals the mutual information of the additive Gaussian mixture noise channel. Furthermore, we derive the GMI using blockwise Gaussian mixture decoding metrics and show that it is monotonically increasing in the length of the decoding block memory, converging to the corresponding mutual information rate as the block length tends to infinity.**

I. Introduction

The Gilbert–Elliott channel [1], [2] has long served as a canonical model for bursty error behavior in communication systems. In its classical form, the channel is a binary-input binary-output channel with a binary input-independent additive noise process governed by the two-state Markov chain. Many practical channels encountered in modern communication systems produce outputs that take on real values. Examples include fading channels with real- or complex-valued gains, and/or additive noise processes with Gaussian or non-Gaussian statistics. In such settings, restricting the observation alphabet to be discrete oversimplifies the channel model and limits the scope of the information-theoretic analysis.

To address this limitation, we extend the Gilbert–Elliott model to the continuous setting. The hidden state process retains its Markovian structure, governing the evolution of channel conditions, while the output process is described by Gaussian probability density functions (PDF) associated with each state. If the transmitter and receiver have perfect channel state information, the channel capacity is known [3]. Without information on the channel state or its transition structure, the capacity reduces to that of an arbitrarily varying channel [4]. This paper explores the intermediate scenario where the channel transition structure is known.

Under independent and identically distributed (i.i.d.) Gaussian inputs, we first derive the generalized mutual information (GMI) using a naive additive white Gaussian noise (AWGN) decoder, also known as the nearest-neighbor decoding metric [5]. We then study the GMI under Gaussian mixture decoders. For memoryless decoding metrics, we prove that a Gaussian mixture decoder achieves a GMI that is no worse than that of the nearest-neighbor decoder. Moreover, by applying blocks of Gaussian mixture decoders, we show that the GMI increases monotonically with the blocklength and recovers the achievable rate asymptotically.

Notation: Continuous scalar random variables are denoted by capital letters, their realizations by respective lowercase letters, and their alphabets by corresponding calligraphic letters. For a random variable X taking values in $\mathcal{X} \in \mathbb{R}$, we denote its PDF by $p_X(\cdot)$. For continuous vector-valued random variables of length n, we write $\mathcal{X}^n$ by $x^n = (x_1, x_2, \cdots, x_n) \in \mathcal{X}^n$, and denote the joint PDF by p_{X^n}. We denote the Gaussian PDF with zero mean and variance σ^2 by $\mathcal{N}(x; 0, \sigma^2)$. A Gaussian mixture density with K components is written as $\sum_{k=1}^{K} w_k \mathcal{N}(x; 0, \sigma_k^2)$, where $w_k \geq 0$ and $\sum_{k=1}^{K} w_k = 1$. Information-theoretic quantities are denoted with the usual conventions, namely differential entropy $h(X)$ and mutual information $I(X;Y)$. The expectation operator is denoted by $\mathbb{E}(\cdot)$. Base-2 logarithms are assumed unless otherwise stated.

II. Preliminaries

A. Two-State AWGN Channel Model

The Gaussian Gilbert-Elliott channel is described by the two-state Markov chain in Fig. 1. When the channel is in the good state, transmission occurs over an AWGN channel with variance σ_{g}^2. Likewise, when the channel is in the bad state, transmission occurs over an AWGN channel with variance $\sigma_{\mathrm{b}}^2 > \sigma_{\mathrm{g}}^2$. In other words, the channel transition law $p(y|x,s)$ is determined by the AWGN channel corresponding to the state, and state transitions are governed by the Markov chain in Fig. 1. Analogously to the Gilbert-Elliott channel, this two-state Gaussian model is a finite-state Markov channel that is *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 [3, Sec. 5.9]. We denote the steady-state distribution of the Markov chain by $[\pi_{\mathrm{g}}, \pi_{\mathrm{b}}] = \left[\frac{g}{g+b}, \frac{b}{g+b}\right]$.

This channel can be interpreted as an additive-noise channel,

$$Y^n = X^n + Z^n \tag{1}$$

where x^n is the input sequence, Z^n being the signal-independent hidden Markov noise sequence, and Y^n is the corresponding channel output. Conditioned on the underlying state, the noise samples are i.i.d. white Gaussian noise, with

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

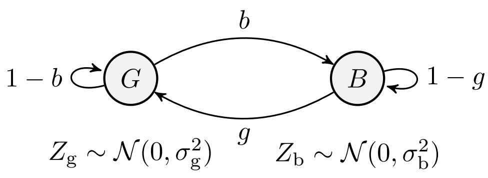

Fig. 1. Two-State Gaussian channel model.

variance depending on the state. We consider the case where input sequences X^n are subject to the average power constraint

$$\frac{1}{n}\mathbb{E}[\|X^n\|^2] \leq P. \tag{2}$$

We consider coded communication with 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 $R = \frac{1}{n}\log M$.

No closed-form single-letter expression for the entropy rate of the noise process is known. Consequently, the channel capacity admits only the multi-letter characterization

$$C = \lim_{n\to\infty} \max_{p_{X^n}: \frac{1}{n}\mathbb{E}[\|X^n\|^2]\leq P} \frac{1}{n} I(X^n; Y^n). \tag{3}$$

In principle, the capacity-achieving distribution for channels with memory generally requires input memory. In this work we adopt a memoryless i.i.d. Gaussian input with variance P because the absence of the channel state information at the transmitter precludes adaptive signaling that tracks the hidden Markov state. The penalty incurred by i.i.d. Gaussian inputs is often modest in practice, especially when the channel state dynamics are slow, or the noise variance ratio is moderate. For slowly varying fading channels, it is well-known that the loss from non-adaptive i.i.d. signaling is small [6].

B. Mismatched Decoding Framework

Mismatched decoding is the setting by which the decoder outputs the message $\hat{m}$ that maximizes a given decoding metric, not necessarily the channel likelihood (see e.g. [7])

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

While typically employed to model channel uncertainty and low-complexity decoding, mismatched decoding can also be employed as a means to derive achievable rates of channels that admit a complicated characterization by employing decoding metrics that are simpler to manipulate than the actual channel.

The GMI is a simple achievable rate [8], whose multi-letter version can be expressed as [7]

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

Under i.i.d. input and a memoryless decoding metric

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

we choose $\tau = 1$ and write (5) as

$$\begin{aligned} I_{\mathrm{gmi}}(P) &= \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^n \log \frac{q(X_i, Y_i)}{\int_{\mathbb{R}} p(\bar{x}) q(\bar{x}, Y_i)\, d\bar{x}}\right] \quad (7)\\ &= \lim_{n\to\infty} \frac{1}{n}\sum_{s^n} P(s^n) \sum_{i=1}^n \mathbb{E}\left[\log \frac{q(X_i, Y_i)}{\int_{\mathbb{R}} p(\bar{x}) q(\bar{x}, Y_i)\, d\bar{x}} \middle| S_i = s_i\right] \quad (8)\\ &= \pi_{\mathrm{g}} I_{\mathrm{gmi}}^{\mathrm{G}}(P) + \pi_{\mathrm{b}} I_{\mathrm{gmi}}^{\mathrm{B}}(P) \quad (9) \end{aligned}$$

where the last equality follows from the method of types and stationarity of the Markov chain [9], with

$$I_{\mathrm{gmi}}^{\mathrm{s}}(P) = \int_{\mathbb{R}} p(x) \int_{\mathbb{R}} p(y|x, s) \log \frac{q(x, y)}{\int_{\mathbb{R}} p(\bar{x}) q(\bar{x}, y)\, d\bar{x}}\, dy\, dx \tag{10}$$

for $s \in \{G, B\}$.

In particular, we choose $q(x_i, y_i)$ to emulate a single AWGN channel that employs nearest-neighbor decoding [5], i.e.,

$$q(x_i, y_i) = \mathcal{N}(y_i - x_i; 0, \sigma_{\mathrm{d}}^2) \tag{11}$$

where σ_{d}^2 is a decoder parameter subject to optimization. The optimal decoding parameter that maximizes the GMI can be determined by setting the partial derivative to zero, yielding

$$\sigma_{\mathrm{d}}^{2^*} = \pi_{\mathrm{g}}\sigma_{\mathrm{g}}^2 + \pi_{\mathrm{b}}\sigma_{\mathrm{b}}^2. \tag{12}$$

Therefore, we conclude that the best achievable rate using a memoryless Gaussian decoding metric is

$$I_{\mathrm{gmi}}(P) = \frac{1}{2}\log\left[1 + \left(\pi_{\mathrm{g}}\mathrm{SNR}_{\mathrm{g}}{}^{-1} + \pi_{\mathrm{b}}\mathrm{SNR}_{\mathrm{b}}{}^{-1}\right)^{-1}\right], \tag{13}$$

which agrees with the result in [5] for i.i.d. noise. In other words, this GMI reduces to the capacity of an equivalent AWGN channel whose SNR is the harmonic mean of the SNRs of the individual state channels, with $\mathrm{SNR}_{\mathrm{g}} = \frac{P}{\sigma_{\mathrm{g}}^2}$ and $\mathrm{SNR}_{\mathrm{b}} = \frac{P}{\sigma_{\mathrm{b}}^2}$. Although this nearest-neighbor decoding metric leads to a single-letter expression of the GMI, it has to compromise between the low- and high-variance regimes. To overcome this limitation while retaining tractability, we consider a Gaussian mixture decoding model that preserves the bimodal nature of the noise distribution. We first consider a memoryless decoding metric based on a Gaussian mixture model. We then extend this approach to blockwise Gaussian mixture decoding, allowing the metric to partially capture the temporal memory of the channel.

III. Gaussian Mixture Decoding Metric

In this section, we choose i.i.d. Gaussian input distribution, and assume a Gaussian mixture decoding metric with K components that ignores the channel memory as in (6) with

$$q(x, y) = \sum_{k=1}^K w_k \mathcal{N}(y - x; 0, \sigma_k^2) \tag{14}$$

where $\sum_{k=1}^K w_k = 1$, and w_k and σ_k^2 are decoder parameters over which we allow optimization.

Building on the decomposition in (9) and (10), we next establish the optimal Gaussian mixture decoding parameters that maximize the GMI.

Theorem 1. *Under the memoryless Gaussian mixture decoding metric in* (14) *and fixed input distribution* $\mathcal{N}(x;0,P)$*, the GMI is maximized when the decoding parameters satisfy*

$$q^*(x,y) = \pi_{\mathrm{g}}\mathcal{N}(y-x;0,\sigma_{\mathrm{g}}^2) + \pi_{\mathrm{b}}\mathcal{N}(y-x;0,\sigma_{\mathrm{b}}^2). \quad (15)$$

Proof. For the denominator of (10), it follows from the convolution of Gaussian densities that

$$\int_{\mathbb{R}} p(\bar{x})q(\bar{x},y)\,d\bar{x} = \sum_{k=1}^{K} w_k\mathcal{N}(y;0,P+\sigma_k^2) \quad (16)$$

$$:= q(y) \quad (17)$$

where $q(y)$ denotes the output density induced by the Gaussian mixture decoding metric.

The state-dependent GMI in (10) simplifies into

$$I_{\mathrm{gmi}}^{\mathrm{s}}(P) = \mathbb{E}_{\mathcal{N}(x;0,P)\times\mathcal{N}(y-x;0,\sigma_{\mathrm{s}}^2)}\left[\log\frac{q(X,Y)}{q(Y)}\right] \quad (18)$$

$$= \mathbb{E}_{\mathcal{N}(0,\sigma_{\mathrm{s}}^2)}[\log q(Z)] - \mathbb{E}_{\mathcal{N}(0,P+\sigma_{\mathrm{s}}^2)}[\log q(Y)] \quad (19)$$

where the second equality follows by writing $Y = X + Z$ with $\mathcal{N}(z;0,\sigma_{\mathrm{s}}^2)$ independent of X, and by marginalizing over X.

By substituting the above expression into (9), we obtain

$$\begin{aligned} &I_{\mathrm{gmi}}(P) \\ &= \int_{\mathbb{R}} \left[\pi_{\mathrm{g}}\mathcal{N}(0,\sigma_{\mathrm{g}}^2) + \pi_{\mathrm{b}}\mathcal{N}(0,\sigma_{\mathrm{b}}^2)\right]\log q(z)\,dz \\ &\quad - \int_{\mathbb{R}} \left[\pi_{\mathrm{g}}\mathcal{N}(0,P+\sigma_{\mathrm{g}}^2) + \pi_{\mathrm{b}}\mathcal{N}(0,P+\sigma_{\mathrm{b}}^2)\right]\log q(y)\,dy \end{aligned} \quad (20)$$

$$= \mathbb{E}_p[\log q(Z)] - \mathbb{E}_p[\log q(Y)] \quad (21)$$

$$= h(p_Y) + D(p_Y\|q_Y) - h(p_Z) - D(p_Z\|q_Z) \quad (22)$$

where we denote

$$p(z) := \pi_{\mathrm{g}}\mathcal{N}(z;0,\sigma_{\mathrm{g}}^2) + \pi_{\mathrm{b}}\mathcal{N}(z;0,\sigma_{\mathrm{b}}^2) \quad (23)$$

$$p(y) := \pi_{\mathrm{g}}\mathcal{N}(y;0,P+\sigma_{\mathrm{g}}^2) + \pi_{\mathrm{b}}\mathcal{N}(y;0,P+\sigma_{\mathrm{b}}^2). \quad (24)$$

For the optimal GMI, we solve

$$I_{\mathrm{gmi}}^*(P) = h(p_Y) - h(p_Z) + \max_q\left[D(p_Y\|q_Y) - D(p_Z\|q_Z)\right] \quad (25)$$

$$= I^{\mathrm{gm}}(X;Y) + \max_q\left[D(p_Y\|q_Y) - D(p_Z\|q_Z)\right] \quad (26)$$

where $I^{\mathrm{gm}}(X;Y) = h(p_Y) - h(p_Z)$ is the mutual information of the additive Gaussian mixture noise channel with distribution (23) under Gaussian inputs. Since the output distributions can be written as convolution of the input and noise distributions, i.e.,

$$p(y) = p(z) * \mathcal{N}(0,P),\;\; q(y) = q(z) * \mathcal{N}(0,P), \quad (27)$$

it follows from the data-processing inequality that

$$D(p_Y\|q_Y) \le D(p_Z\|q_Z), \quad (28)$$

as the KL divergence cannot increase when both distributions pass through the same kernel. Equality, yielding a maximum of zero divergence difference in (26), is achieved by setting $p(z) = q(z)$ and $p(y) = q(y)$ and identifies the optimal decoding distribution in (15). □

This optimal choice of Gaussian mixture decoding metric in (15) achieves equality in (28). Thus, we conclude that the optimal GMI in (26) simplifies to the mutual information $I^{\mathrm{gm}}(X;Y)$, as stated in the following corollary.

Corollary 1. *For the continuous Gilbert–Elliott channel with i.i.d. Gaussian input, the best GMI under a memoryless Gaussian mixture decoding metric is given by*

$$I_{\mathrm{gmi}}(P) = I^{\mathrm{gm}}(X;Y), \quad (29)$$

the mutual information with Gaussian inputs of the additive Gaussian mixture channel with noise distribution (23)*.*

Proposition 1. *The GMI under a memoryless Gaussian mixture decoding metric is upper bounded by*

$$I_{\mathrm{gmi}}(P) \le \pi_{\mathrm{g}}I(X;Y|S=G) + \pi_{\mathrm{b}}I(X;Y|S=B), \quad (30)$$

which coincides with the mutual information achievable under perfect channel state information at the receiver (CSIR).

Proof. For notational convenience in this proof, we write the mutual information as $I(p_X;p_{Y|X})$, which is equivalent to $I(X;Y)$. We know from Theorem 1 that

$$I_{\mathrm{gmi}}(P) = I\Big(p_X;\pi_{\mathrm{g}}\mathcal{N}(z;0,\sigma_{\mathrm{g}}^2) + \pi_{\mathrm{b}}\mathcal{N}(z;0,\sigma_{\mathrm{b}}^2)\Big) \quad (31)$$

$$\le \pi_{\mathrm{g}}I\Big(p_X;\mathcal{N}(z;0,\sigma_{\mathrm{g}}^2)\Big) + \pi_{\mathrm{b}}I\Big(p_X;\mathcal{N}(z;0,\sigma_{\mathrm{b}}^2)\Big) \quad (32)$$

where the inequality follows from Jensen's inequality since, for a fixed input distribution p_X, the mutual information is convex in the channel transition law $p_{Y|X}$. □

Proposition 2. *With i.i.d. Gaussian inputs, the GMI under a memoryless Gaussian mixture decoding metric is no worse than that under a nearest-neighbor decoding metric.*

Proof. For i.i.d. Gaussian inputs, the Gaussian noise maximizes differential entropy for a given variance, implying that it is the worst-case additive noise for point-to-point channels in terms of mutual information. The optimal nearest-neighbor decoder corresponds to an AWGN channel with variance $\pi_{\mathrm{g}}\sigma_{\mathrm{g}}^2 + \pi_{\mathrm{b}}\sigma_{\mathrm{b}}^2$, which equals the variance of the Gaussian mixture noise. By the worst-case property of Gaussian noise, the GMI cannot increase when replacing the mixture noise with a single Gaussian of the same variance. □

IV. Blockswise Gaussian Mixture Decoding

In this section, similarly to [10] for the standard Gilbert-Elliott channel, 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 of exposition, we decode in pairs of symbols. The approach extends naturally to arbitrary memory.

A block-2 decoding metric can be factorized as

$$q^n(x^n, y^n) = \prod_{\substack{i=1 \\ i \text{ even}}}^{n} q^2(x_{i-1}, x_i, y_{i-1}, y_i). \tag{33}$$

Since optimizing $q^2(x_{i-1}, x_i, y_{i-1}, y_i)^\tau$ is equivalent to fixing $\tau = 1$ and optimizing $q^2(x_{i-1}, x_i, y_{i-1}, y_i)$, we factorize (5) as

$$\begin{aligned} I_{\text{gmi}}(P) &= \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\Bigg[\sum_{\substack{i=1 \\ i \text{ even}}} \log \frac{q^2(X_{i-1}, X_i, Y_{i-1}, Y_i)}{\mathbb{E}_{\bar{X}_{i-1}, \bar{X}_i}[q^2(\bar{X}_{i-1}, \bar{X}_i, Y_{i-1}, Y_i)]} \Bigg] \quad (34) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{s^n} P(s^n) \sum_{\substack{i=1 \\ i \text{ even}}} \\ &\mathbb{E}\Bigg[\log \frac{q^2(X_{i-1}, X_i, Y_{i-1}, Y_i)}{\mathbb{E}_{\bar{X}_{i-1}, \bar{X}_i}[q^2(\bar{X}_{i-1}, \bar{X}_i, Y_{i-1}, Y_i)]} \Bigg| S_{i-1} = s_{i-1}, S_i = s_i \Bigg] \quad (35) \\ &= \frac{1}{2} \sum_{s_1, s_2} \pi(s_1, s_2) I^{\text{s}_1\text{s}_2}_{\text{gmi}}(P) \quad (36) \end{aligned}$$

where (34) follows from the use of i.i.d. inputs and the block-memory decoder (33), and the last equality follows from the method of types and stationarity of the Markov chain, with

$$I^{\text{s}_1\text{s}_2}_{\text{gmi}}(P) = \int_{\mathbb{R}^2} p(x_1)p(x_2) \int_{\mathbb{R}^2} p(y_1|x_1, s_1)p(y_2|x_2, s_2) \times \log \frac{q^2(x_1, x_2, y_1, y_2)}{\int_{\mathbb{R}^2} p(\bar{x}_1)p(\bar{x}_2)q^2(\bar{x}_1, \bar{x}_2, y_1, y_2)\, d\bar{x}_1 d\bar{x}_2}\, dx_1 dx_2 dy_1 dy_2 \tag{37}$$

for $s_1, s_2 \in \{G, B\}$.

Choose a two-dimensional Gaussian mixture decoding metric as in (33) with $K_1 \times K_2$ components of the form

$$q^2(x_1, x_2, y_1, y_2) = \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} w(k_1, k_2)\mathcal{N}(y_1 - x_1; 0, \sigma^2_{k_1}) \times \mathcal{N}(y_2 - x_2; 0, \sigma^2_{k_2}) \tag{38}$$

where $\sum_{k_1, k_2} w(k_1, k_2) = 1$.

Using the same argument as in Theorem 1, the block-2 GMI is maximized when the decoding parameters satisfy

$$\begin{aligned} &q^{2*}(x_1, x_2, y_1, y_2) \\ &= \sum_{s_1, s_2} \pi(s_1, s_2)\mathcal{N}(y_1 - x_1; 0, \sigma^2_{\text{s}_1})\mathcal{N}(y_2 - x_2; 0, \sigma^2_{\text{s}_2}) \quad (39) \\ &= \sum_{s_1, s_2} \pi(s_1, s_2)\mathcal{N}\left(\begin{bmatrix} y_1 - x_1 \\ y_2 - x_2 \end{bmatrix}; \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\text{s}_1} & 0 \\ 0 & \sigma^2_{\text{s}_2} \end{bmatrix} \right) \quad (40) \end{aligned}$$

where the second equality follows from the fact that the product of independent Gaussian densities can be expressed as a multivariate Gaussian distribution with diagonal covariance matrix.

Therefore, we conclude that the optimal GMI rate attainable under a block-2 decoder satisfies

$$\begin{aligned} I^*_{\text{gmi}}(P) &= h(p_{Y^2}) - h(p_{Z^2}) \quad (41) \\ &= \frac{1}{2} I^{\text{gm}}(X_1 X_2; Y_1 Y_2) \quad (42) \end{aligned}$$

where

$$p_{Z^2}(z_1, z_2) = \sum_{s_1, s_2} \pi(s_1, s_2)\mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\text{s}_1} & 0 \\ 0 & \sigma^2_{\text{s}_2} \end{bmatrix} \right) \tag{43}$$

$$p_{Y^2}(y_1, y_2) = \sum_{s_1, s_2} \pi(s_1, s_2)\mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} P + \sigma^2_{\text{s}_1} & 0 \\ 0 & P + \sigma^2_{\text{s}_2} \end{bmatrix} \right). \tag{44}$$

The preceding analysis naturally extends to decoders with arbitrary blocklength, yielding the following theorem.

Theorem 2. *For the continuous Gilbert–Elliott channel with i.i.d. Gaussian input, the best GMI under a block-m Gaussian mixture decoding metric is given by*

$$I_{\text{gmi}}(m, P) = \frac{1}{m} I^{\text{gm}}(X^m; Y^m) \tag{45}$$

where

$$p_{Z^m}(z^m) = \sum_{s^m} \pi(s^m) \prod_{i=1}^{m} \mathcal{N}(z_i; 0, \sigma^2_{\text{s}_i}) \tag{46}$$

$$p_{Y^m}(y^m) = \sum_{s^m} \pi(s^m) \prod_{i=1}^{m} \mathcal{N}(y_i; 0, P + \sigma^2_{\text{s}_i}). \tag{47}$$

In other words, we observe that this again transforms into the mutual information of an equivalent Gaussian mixture channel with noise distribution (46).

Proposition 3. *$I_{\text{gmi}}(m, P)$ is non-decreasing in m.*

Proof. We firstly prove that $h(X_{m+1}|X^m Y^{m+1})$ is non-increasing with m since

$$\begin{aligned} h(X_{m+1}|X^m Y^{m+1}) &\le h(X_{m+1}|X^m_2 Y^{m+1}) \quad (48) \\ &= h(X_m|X^{m-1} Y^{m+1}) \quad (49) \\ &= h(X_m|X^{m-1} Y^m) \quad (50) \end{aligned}$$

where (49) follows from stationarity (time-invariance). Using this property, we have

$$\begin{aligned} m h(X_m|X^{m-1} Y^m) &\le \sum_{i=1}^{m} h(X_i|X^{i-1} Y^i) \quad (51) \\ &= h(X^m|Y^m). \quad (52) \end{aligned}$$

It follows from the chain rule and non-anticipatory property that

$$h(X^{m+1}|Y^{m+1}) = h(X^m|Y^m) + h(X_{m+1}|X^m Y^{m+1}) \quad (53)$$
$$\leq h(X^m|Y^m) + h(X_m|X^{m-1}Y^m) \quad (54)$$
$$\leq h(X^m|Y^m) + \frac{1}{m}h(X^m|Y^m) \quad (55)$$
$$= \frac{m+1}{m}h(X^m|Y^m) \quad (56)$$

where the first inequality is due to (50), and the second is a result of (52).

By rearranging (56), we obtain

$$\frac{1}{m+1}h(X^{m+1}|Y^{m+1}) \leq \frac{1}{m}h(X^m|Y^m), \quad (57)$$

which implies that

$$\frac{1}{m+1}I(X^{m+1};Y^{m+1}) \geq \frac{1}{m}I(X^m;Y^m), \quad (58)$$

since the entropy rate for i.i.d. input sequence equals $h(X)$. □

Proposition 4. *The mutual information rate with i.i.d. Gaussian inputs is recovered by choosing a Gaussian mixture decoder with $m \to \infty$, i.e.,*

$$\lim_{m\to\infty} I_{\text{gmi}}(m, P) = \lim_{n\to\infty} \frac{1}{n}I(X^n;Y^n). \quad (59)$$

Proof. For the continuous Gilbert-Elliott channel model, the noise distribution can be computed via marginalization as

$$p_{Z^m}(z^m) = \sum_{s^m} P(s^m) \prod_{i=1}^{m} p(z_i|s_i) \quad (60)$$
$$= \sum_{s^m} P(s^m) \prod_{i=1}^{m} \mathcal{N}(z_i; 0, \sigma^2_{s_i}), \quad (61)$$

which asymptotically converges to (46) for $m \to \infty$ for ergodic Markov chains. Analogous argument applies to p_{Y^m}. □

Although the analysis is specific to a two-state hidden Markov AWGN channel, the result extends straightforwardly to channels with an arbitrary number of hidden states. However, given the difficulty of computing the achievable rates for Gaussian mixture channels, it becomes computationally demanding for large blocklengths.

V. Numerical Results

For stationary ergodic channels, the normalized information density converges almost surely to the mutual information rate. Consequently, the achievable rate can be simulated by evaluating the log-likelihood ratio $\frac{1}{n}\log\frac{p(y^n|x^n)}{p(y^n)}$ along a sufficiently long sample path, with likelihoods computed via Monte Carlo simulation using a BCJR forward recursion [11]. For a given i.i.d. Gaussian input sequence with power P, the channel output is generated by propagating the Markov state process and adding Gaussian noise conditioned on the state. The conditional likelihood $p(y^n|x^n)$ is computed using a forward recursion that propagates the joint quantities $p(y^k, S_k = s|x^k)$, while the marginal likelihood $p(y^n)$ is obtained by propagating $p(y^k, S_k = s)$. These densities are initialized using the stationary distribution of the channel state; however, due to ergodicity, the resulting normalized log-likelihood is asymptotically insensitive to the initialization.

We now compare the GMI achievable by different mismatched decoders for a Gaussian Gilbert-Elliott channel with parameters $b = 0.1$, $g = 0.3$, $P = 10$ dB and $\text{SNR}_\text{b} = -10$ dB, as shown in Fig. 2. The solid and dotted black curves depict the simulated mutual information rates and the optimal GMI for memoryless Gaussian decoders, respectively. Although the rate can be made arbitrarily large by choosing SNR_g sufficiently large, the GMI converges to an SNR_b-dependent limit as a consequence of the harmonic mean in (13). The optimal GMIs for memoryless and block-2 Gaussian mixture decoders appear as dashed blue and dash-dotted red curves. We observe that the memoryless Gaussian mixture decoder attains a substantially higher GMI than the nearest-neighbor decoder, which is consistent with Proposition 2. In accordance with Proposition 3, the block-2 decoder outperforms its memoryless counterpart.

This behavior aligns with our theoretical predictions, as the Gaussian mixture model reflects the marginal distribution of the channel output, thereby offering a more accurate representation of the channel statistics while preserving decoding tractability. As a result, the corresponding decoder assigns higher likelihood to observations consistent with either state, leading to improved reliability compared to the single-Gaussian approximation.

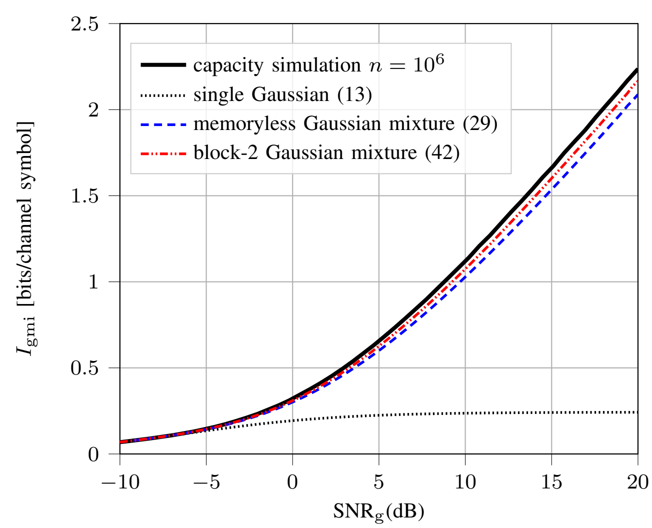

Fig. 2. GMI comparison for $b = 0.1, g = 0.3, P = 10$ dB, $\text{SNR}_\text{b} = -10$ dB.

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

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

[5] A. Lapidoth, “Nearest neighbor decoding for additive non-Gaussian noise channels,” *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1520–1529, Sept. 1996.

[6] A. Goldsmith and P. Varaiya, “Capacity of fading channels with channel side information,” *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1986–1992, 1997.

[7] 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

[8] 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.

[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.

[10] ——, “Single-letter mismatched decoding rates with memory for the Gilbert-Elliott channel,” in *2025 IEEE Inf. Theory Workshop, Sydney, Australia*, Sep. 2025.

[11] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, “Optimal decoding of linear codes for minimizing symbol error rate (corresp.),” *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 284–287, 1974.