

Article

Rate and Nearly-Lossless State over the Gilbert–Elliott Channel

Amos Lapidoth ^{*,†}  and Ligong Wang [†] 

Department of Information Technology and Electrical Engineering, ETH Zurich, 8092 Zurich, Switzerland

* Correspondence: lapidoth@isi.ee.ethz.ch

[†] These authors contributed equally to this work.

Abstract: The capacity of the Gilbert–Elliott channel is calculated for a setting in which the state sequence is revealed to the encoder and is, along with the transmitted message, to be conveyed to the receiver with a vanishing symbol error rate. Said capacity does not depend on whether the state sequence is provided to the encoder strictly causally, causally, or noncausally. It can be achieved using a Block-Markov coding scheme with backward decoding.

Keywords: causal; Gilbert–Elliott channel; noncausal; rate-and-state capacity; state information; strictly causal

1. Introduction

The Gilbert–Elliott channel [1–4] is a binary-input, binary-output finite-state channel [5–8], which was proposed as a simple model for digital communications with bursty errors. Its state sequence is a stationary two-state Markov process whose Time- i state S_i takes values in the set $\mathcal{S} = \{0, 1\}$. When the Time- i state S_i is 0, the Time- i channel output Y_i is the result of feeding the Time- i input x_i through a Binary Symmetric Channel (BSC) of crossover probability ϵ_0 ; when S_i is 1, the output Y_i is the result of feeding x_i through a BSC of crossover probability ϵ_1 . Alternatively, we can describe the output sequence as being the componentwise mod-2 addition of the input sequence with a noise sequence $\{Z_i\}$, where the latter is a binary two-state Hidden Markov process.

The capacity of the Gilbert–Elliott channel is achieved by having the input sequence comprise Independent and Identically Distributed (IID) random bits. It is given, in bits, by $1 - H(\{Z_i\})$, where $H(\{Z_i\})$ denotes the entropy rate of the noise $\{Z_i\}$ [3]. If the state sequence is revealed to the decoder, then capacity is still achieved by the above input distribution, but it now equals the weighted average of the capacities of the BSCs, namely, $\pi_0(1 - H_b(\epsilon_0)) + \pi_1(1 - H_b(\epsilon_1))$. Here, $\pi_1 = 1 - \pi_0$ denotes the probability that S_i is 1, and $H_b(\cdot)$ is the binary entropy function. Throughout this paper, logarithms are to base 2, and information is measured in bits.

Here, we study the case where the state sequence is revealed not to the decoder but to the encoder. We do not, however, seek the Shannon capacity but rather the “rate-and-nearly-lossless-state” capacity, where the decoder wishes to recover not only the transmitted message but also the state sequence. We thus require that the state sequence be conveyed to the decoder with a vanishing symbol error rate. This requirement is weaker than the requirement that the probability of the receiver correctly recovering the entire state sequence tends to one. Neither do we consider a general (nonzero) distortion constraint, as studied for memoryless channels in [9,10].



Academic Editor: Boris Ryabko

Received: 26 March 2025

Revised: 29 April 2025

Accepted: 30 April 2025

Published: 2 May 2025

Citation: Lapidoth, A.; Wang, L. Rate and Nearly-Lossless State over the Gilbert–Elliott Channel. *Entropy* **2025**, *27*, 494. <https://doi.org/10.3390/e27050494>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

We solve for this capacity and show that it does not depend on whether the state sequence is provided to the encoder strictly causally, causally, or noncausally. Moreover, it can be achieved using a Block-Markov coding scheme with backward decoding.

2. The Gilbert–Elliott Channel

The Gilbert–Elliott channel has binary input, output, and state alphabets: $\mathcal{X} = \mathcal{Y} = \mathcal{S} = \{0, 1\}$. The evolution of the state is unaffected by the channel inputs: irrespective of the channel inputs, it forms a stationary time-homogeneous Markov chain of kernel

$$\Pr[S_i = s_i | S^{i-1} = s^{i-1}] = \Pr[S_i = s_i | S_{i-1} = s_{i-1}] \quad (1)$$

and stationary distribution

$$P_S = (\pi_0, \pi_1) \quad (2)$$

by which we mean that

$$P_S(1) = 1 - P_S(0) = \pi_1 = 1 - \pi_0. \quad (3)$$

Above and throughout, we use A^i to denote $A_1, \dots, A_i$, and we use A_j^i to denote $A_j, \dots, A_i$. We denote the entropy rate of the state sequence $H(\{S_i\})$. It is given explicitly by [11]

$$H(\{S_i\}) = H(S_2 | S_1). \quad (4)$$

Given $S = 0$, the channel from X to Y is a BSC with crossover probability ϵ_0 , whereas, given $S = 1$, it is a BSC with crossover probability ϵ_1 . Here, $\epsilon_0, \epsilon_1 \in [0, 1]$ are arbitrary known constants. Thus, if we define

$$W(1 - x | x, 0) = 1 - W(x | x, 0) = \epsilon_0 \quad (5)$$

and

$$W(1 - x | x, 1) = 1 - W(x | x, 1) = \epsilon_1, \quad (6)$$

then we can express the behavior of the channel given the state as

$$\Pr[Y_i = y_i | X^i = x^i, Y^{i-1} = y^{i-1}, S^i = s^i] = W(y_i | x_i, s_i). \quad (7)$$

3. The Rate-and-Nearly-Lossless-State Capacity

When discussing rate- R blocklength- n communications over the Gilbert–Elliott channel, we consider the message set $\mathcal{M} = \{1, \dots, 2^{nR}\}$ with 2^{nR} messages, one of which is to be conveyed to the receiver. The latter observes the output sequence Y^n and attempts to recover the transmitted message m and the state sequence S^n . The decoder is thus specified by a function

$$\phi: \mathcal{Y}^n \rightarrow \mathcal{M} \times \mathcal{S}^n, \quad y^n \mapsto (\hat{m}, \hat{s}^n). \quad (8)$$

As to the encoder, its structure depends on the manner in which the state information is revealed to it. In the strictly causal setting, the Time- i channel input may depend, not only on the transmitted message, but also on the past states; it is therefore denoted $X_i(m, S^{i-1})$. The encoder is thus specified by n functions

$$f_i: \mathcal{M} \times \mathcal{S}^{i-1} \rightarrow \mathcal{X}, \quad i = 1, \dots, n, \quad (9)$$

with $X_i(m, S^{i-1})$ being $f_i(m, S^{i-1})$. In the causal case, the Time- i channel input is denoted $X_i(m, S^i)$, and the encoder is specified by n functions

$$f_i: \mathcal{M} \times \mathcal{S}^i \rightarrow \mathcal{X}, \quad i = 1, \dots, n, \quad (10)$$

with $X_i(m, S^i)$ being $f_i(m, S^i)$. Finally, in the noncausal case, the Time- i channel input is denoted $X_i(m, S^n)$, and the encoder is specified by one function

$$f: \mathcal{M} \times \mathcal{S}^n \rightarrow \mathcal{X}^n, \quad (11)$$

with $X_i(m, S^n)$ being the i -th component of $f(m, S^n)$.

We refer to an encoder/decoder pair as a “coding scheme”. The probability of error associated with a given coding scheme and a given message m is $\Pr[\hat{M} \neq m]$ calculated when the scheme’s encoder is used to transmit Message m , and the scheme’s decoder is used by the receiver. The average probability of error associated with a coding scheme is the arithmetic average of the probabilities of error associated with the different messages. It can be expressed as

$$P_e \triangleq \Pr[\hat{M} \neq M] \quad (12)$$

when the transmitted message M is drawn equiprobably from $\mathcal{M}$.

The symbol error rate, or (expected) Hamming distortion, in reconstructing the state sequence is

$$\beta \triangleq \frac{1}{n} \sum_{i=1}^n \Pr[\hat{S}_i \neq S_i] \quad (13)$$

again computed when the transmitted message M is drawn equiprobably from $\mathcal{M}$.

In all cases, we say that a rate R is achievable if there exists a sequence of coding schemes indexed by the blocklength for which P_e and β both tend to zero. The capacity is defined as the supremum of the achievable rates, with the understanding that, if no positive rate is achievable, then capacity is zero.

Our main result is the following theorem:

Theorem 1. *Irrespective of whether the state information is provided to the encoder strictly causally, causally, or noncausally, if*

$$1 - (\pi_0 H_b(\epsilon_0) + \pi_1 H_b(\epsilon_1)) - H(\{S_i\}) \quad (14)$$

is positive, then it equals the rate-and-nearly-lossless-state capacity of the Gilbert–Elliott channel; else, said capacity is zero.

The proof is provided in the next section. Expression (14) can be interpreted as the result of subtracting the optimal average description length of the state sequence from the capacity of the channel when the receiver is cognizant of the state.

4. Proof of Theorem 1

4.1. Converse

We prove the converse part of the theorem under the assumption that the state sequence is provided to the encoder noncausally. Let M be uniform over $\mathcal{M}$. We have

$$H(S^n|M, Y^n) \leq H(S^n|\hat{S}^n) \quad (15)$$

$$\leq \sum_{i=1}^n H(S_i|\hat{S}_i) \quad (16)$$

$$\leq \sum_{i=1}^n H_b(\Pr[\hat{S}_i \neq S_i]) \quad (17)$$

$$\leq nH_b(\beta). \quad (18)$$

Here, (15) holds because $\hat{S}^n$ is a function of Y^n and hence also of (M, Y^n) ; (17) because $\hat{S}_i$ is binary and because conditioning reduces entropy; and (18) because $H_b(\cdot)$ is concave and by recalling the definition of β in (13). Note that $H_b(\beta)$ tends to zero when β tends to zero.

Since the decoder needs to recover M with high probability, Fano's inequality implies the existence of some sequence $\{\epsilon_n\}$ tending to zero with the blocklength such that

$$n(R - \epsilon_n) \leq I(M; Y^n). \quad (19)$$

From these two inequalities, we obtain

$$n(R - \epsilon_n - H_b(\beta)) \leq I(M; Y^n) - H(S^n|M, Y^n) \quad (20)$$

$$= I(M; Y^n, S^n) - I(M; S^n|Y^n) - H(S^n|M, Y^n) \quad (21)$$

$$= I(M; Y^n, S^n) - H(S^n|Y^n) \quad (22)$$

$$= I(M; Y^n|S^n) - H(S^n|Y^n) \quad (23)$$

$$= H(Y^n|S^n) - H(S^n|Y^n) - H(Y^n|S^n, M) \quad (24)$$

$$= (H(Y^n) - I(Y^n; S^n)) - (H(S^n) - I(Y^n; S^n)) - H(Y^n|S^n, M) \quad (25)$$

$$= H(Y^n) - H(S^n) - H(Y^n|S^n, M). \quad (26)$$

Note that, above, we only used the chain rule and the fact that $M \perp\!\!\!\perp S^n$ (i.e., that M and S^n are independent). The three terms in the last line can be bounded or simplified as follows:

$$H(Y^n) \leq n \quad (27)$$

$$H(S^n) = H(S_1) + (n-1)H(\{S_i\}) \quad (28)$$

$$H(Y^n|S^n, M) = H(Y^n|S^n, M, X^n) \quad (29)$$

$$= H(Y^n|S^n, X^n) \quad (30)$$

$$= \sum_{i=1}^n H(Y_i|S_i, X_i) \quad (31)$$

$$= n(\pi_0 H_b(\epsilon_0) + \pi_1 H_b(\epsilon_1)). \quad (32)$$

The converse now follows from (26), (27), (28), and (32).

4.2. Direct Part

For the direct part of the theorem, we assume that the state sequence is revealed to the encoder strictly causally and consider a Block-Markov coding scheme with backward decoding. Consider b blocks, each of k channel uses. Label the length- k typical state sequences $1, \dots, 2^{k(H(\{S_i\})+\epsilon)}$. For each block, randomly generate a codebook with $2^{k\tilde{R}}$ independent codewords (with $\tilde{R}$ to be specified later), each having k independent Bernoulli(1/2) components. In the first block, use the entire codebook to send a message of $k\tilde{R}$ bits. As to the transmission in Block i for $i \geq 2$, first check whether or not the state sequence in the previous block, namely, $(s^k)_{i-1}$, was typical. If it was, set ℓ_{i-1} to be the

label that is assigned to it; else, set $\ell_{i-1} = 1$. Use the generated codebook to send both ℓ_{i-1} and m_i , the latter consisting of $k(\tilde{R} - (H(\{S_i\}) + \epsilon))$ information bits, so that the total number of bits matches the size of the codebook.

After completing the transmission in all b blocks, add one extra block to transmit ℓ_b . To this end, the transmitter—ignoring any information about the states in this extra block—uses the Gilbert–Elliott channel to send ℓ_b (which we view as a message) and nothing else. The length of the extra block may be larger than k , but—as long as the Shannon capacity of our channel is positive—this will not affect the overall rate when we choose b to be very large. We will address this caveat shortly. However, first, we discuss the decoding.

To decode, we begin with the last block to decode ℓ_b and thus recover $(s^k)_b$. This task will be accomplished successfully provided that the last extra block is sufficiently long. With the state information $(s^k)_b$ at hand, we then decode both ℓ_{b-1} and m_b from Block b . Both can be decoded correctly with high probability (as k grows large) provided that

$$\tilde{R} < I(X_i; Y_i | S_i) = 1 - (\pi_0 H_b(\epsilon_0) + \pi_1 H_b(\epsilon_1)). \quad (33)$$

We continue this procedure backwards: By the time we get to decoding Block i , we will have already reliably recovered $(s^k)_i$ in Block $(i + 1)$, so we could use it as decoder-side information. The overall information rate—when b tends to infinity—approaches

$$\tilde{R} - H(\{S_i\}) - \epsilon, \quad (34)$$

which can indeed be made arbitrarily close to (14).

We have now established that, provided that $\tilde{R}$ is appropriately chosen, as k grows large, the probability of the decoder correctly decoding both the message and s^{kb} tends to one. It only remains to note that this also means that the symbol error rate β is small. Indeed, having $\hat{S}^{kb} = S^{kb}$ with high probability guarantees that the symbol error rate in the first b blocks is close to zero, whereas the influence of the extra block (whose symbol error rate will typically be high, since we do not attempt to decode the states there) becomes negligible when we choose b to be large.

We now return to the caveat regarding the Shannon capacity and the extra block. We need to show that the Shannon capacity is positive whenever (14) is positive. (If (14) is not positive, there is no need for a direct part.) Recall that the Shannon capacity of the Gilbert–Elliott channel without any state information is $1 - H(\{Z_i\})$, with $\{Z_i\}$ denoting the noise random variables, which form a Hidden Markov process. This Shannon capacity is positive unless $H(\{Z_i\}) = 1$ bit. Our concern regarding the caveat is thus only when $H(\{Z_i\})$ is 1.

In Appendix A, we show the following:

Proposition 1. *If $H(\{Z_i\}) = 1$, then*

1. *$\{Z_i\}$ are IID Bernoulli(1/2), and*
2. *either $\{S_i\}$ are IID or $\epsilon_0 = \epsilon_1 = 1/2$.*

If $\{Z_i\}$ are IID Bernoulli(1/2) and $\{S_i\}$ are IID, then (14) cannot be positive because, in this case, it equals

$$H(Z_i) - H(Z_i | S_i) - H(S_i) = -H(S_i | Z_i). \quad (35)$$

If $\epsilon_0 = \epsilon_1 = 1/2$, then (14) is also (trivially) not positive. Hence, indeed, the Shannon capacity is positive whenever (14) is positive.

Author Contributions: Writing—original draft preparation, A.L. and L.W.; writing—review and editing, A.L. and L.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Swiss National Science Foundation (SNSF) under Grant 200021-215090.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BSC Binary Symmetric Channel
IID Independent and Identically Distributed

Appendix A. Proof of Proposition 1

In this appendix, we prove Proposition 1. We start with the first part, which we formally state as follows:

Lemma A1. *If $\{Z_i\}$ is a stationary process taking values in the binary set $\{0, 1\}$, and if its entropy rate is 1, then it must be IID Bernoulli(1/2).*

Proof. For any positive integers ν and ℓ ,

$$\frac{1}{\ell\nu} H(Z_1^{\ell\nu}) = \frac{1}{\ell\nu} H(Z_1^\nu, \dots, Z_{(\ell-1)\nu+1}^{\ell\nu}) \quad (\text{A1})$$

$$\leq \frac{1}{\ell\nu} \sum_{i=1}^{\ell} H(Z_{(i-1)\nu+1}^{\nu}) \quad (\text{A2})$$

$$= \frac{1}{\ell\nu} \ell H(Z_1^\nu) \quad (\text{A3})$$

$$= \frac{1}{\nu} H(Z_1^\nu) \quad (\text{A4})$$

where (A3) follows from the stationarity hypothesis. Letting ℓ tend to infinity so that the left-hand side of (A1) converges to the entropy rate, i.e., to 1, we obtain

$$1 \leq \frac{1}{\nu} H(Z_1^\nu). \quad (\text{A5})$$

We obtain the reverse inequality by noting that Z_1^ν takes values in a set of cardinality 2^ν . We thus conclude that Z_1^ν is equiprobably distributed, and its components are thus IID Bernoulli(1/2). Since ν can be any positive integer, we conclude that $\{Z_i\}$ is an IID Bernoulli(1/2) process. $\square$

We next show the remaining part of Proposition 1:

Lemma A2. *If $\{Z_i\}$ are IID Bernoulli(1/2), and if ϵ_0 and ϵ_1 are not both equal to 1/2, then $\{S_i\}$ must be IID.*

Proof. We assume $\pi_0, \pi_1 \in (0, 1)$; otherwise, $\{S_i\}$ are all deterministic and trivially IID. We also assume, without loss of generality, that

$$\epsilon_0 < \frac{1}{2}. \quad (\text{A6})$$

For Z_i to be Bernoulli(1/2), we must then have

$$\epsilon_1 > \frac{1}{2}. \quad (\text{A7})$$

It then follows that the equation

$$\pi\epsilon_0 + (1 - \pi)\epsilon_1 = \frac{1}{2} \quad (\text{A8})$$

has a unique solution in π , which must be π_0 as in the stationary distribution (2).

Because Z_i is Bernoulli(1/2) and S_i is Bernoulli(π_1), we can use Bayes's rule to express the conditional distribution of S_i given $Z_i = 1$ as

$$P_{S_i|Z_i=1} = (2\epsilon_0\pi_0, 2\epsilon_1\pi_1) \quad (\text{A9})$$

$$\neq (\pi_0, \pi_1), \quad (\text{A10})$$

where the inequality follows from (A6) and (A7). Let M denote the (matrix form of the) Markov kernel $P_{S_{i+1}|S_i}$. We next assume that M is nonsingular and show that it leads to a contradiction. If M is nonsingular, then

$$P_{S_{i+1}|Z_i=1} = (2\epsilon_0\pi_0, 2\epsilon_1\pi_1) M \quad (\text{A11})$$

$$\neq (\pi_0, \pi_1) M \quad (\text{A12})$$

$$= (\pi_0, \pi_1), \quad (\text{A13})$$

where the inequality follows from (A10) and the nonsingularity of M , and the last equality holds because (π_0, π_1) is the stationary distribution corresponding to M .

Recalling that (A8) is only satisfied when π equals π_0 , we conclude that

$$P_{Z_{i+1}|Z_i=1}(1) = P_{S_{i+1}|Z_i=1}(0)\epsilon_0 + P_{S_{i+1}|Z_i=1}(1)\epsilon_1 \neq \frac{1}{2}, \quad (\text{A14})$$

which contradicts our assumption that $\{Z_i\}$ are IID Bernoulli(1/2).

We thus conclude that the Markov kernel M must be singular, which, in the 2×2 case, means that its two rows are identical, and that $\{S_i\}$ are hence IID. $\square$

References

1. Gilbert, E.N. Capacity of a burst-noise channel. *Bell Syst. Tech. J.* **1960**, *39*, 1253–1265. [\[CrossRef\]](#)
2. Elliott, E.O. Estimates of error rates for codes on burst-noise channels. *Bell Syst. Tech. J.* **1963**, *42*, 1977–1997. [\[CrossRef\]](#)
3. Mushkin, M.; Bar-David, I. Capacity and coding for the Gilbert-Elliott channels. *IEEE Trans. Inf. Theory* **1989**, *35*, 1277–1290. [\[CrossRef\]](#)
4. Han, Y.; Guillén i Fàbregas, A. Fixed-memory capacity bounds for the Gilbert-Elliott channel. In Proceedings of the 2024 IEEE International Symposium on Information Theory (ISIT), Athens, Greece, 7–12 July 2024; pp. 155–159.
5. Gallager, R.G. *Information Theory and Reliable Communication*; John Wiley & Sons: Hoboken, NJ, USA, 1968.
6. Goldsmith, A.; Varaiya, P. Capacity, mutual information, and coding for finite-state Markov channels. *IEEE Trans. Inf. Theory* **1996**, *42*, 868–886. [\[CrossRef\]](#)
7. Permuter, H.H.; Weissman, T.; Goldsmith, A.J. Finite state channels with time-invariant deterministic feedback. *IEEE Trans. Inf. Theory* **2006**, *55*, 644–662. [\[CrossRef\]](#)
8. Shrader, B.; Permuter, H. Feedback capacity of the compound channel. *IEEE Trans. Inf. Theory* **2009**, *55*, 3629–3644. [\[CrossRef\]](#)
9. Choudhuri, C.; Kim, Y.H.; Mitra, U. Causal state communication. *IEEE Trans. Inf. Theory* **2012**, *59*, 3709–3719. [\[CrossRef\]](#)

10. Bross, S.I.; Lapidoth, A. The rate-and-state capacity with feedback. *IEEE Trans. Inf. Theory* **2018**, *64*, 1893–1918. [[CrossRef](#)]
11. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2006.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.