Dispersion of the Gilbert-Elliott Channel

Yury Polyanskiy
Dept. of Electrical Engineering
Princeton University
Princeton, NJ 08544, USA
Email: ypolyans@princeton.edu

H. Vincent Poor
Dept. of Electrical Engineering
Princeton University
Princeton, NJ 08544, USA
Email: poor@princeton.edu

Sergio Verdú
Dept. of Electrical Engineering
Princeton University
Princeton, NJ 08544, USA
Email: verdu@princeton.edu

***Abstract*—Channel dispersion plays a fundamental role in assessing the backoff from capacity due to finite blocklength. This paper analyzes the channel dispersion for a simple channel with memory: the Gilbert-Elliott communication model in which the crossover probability of a binary symmetric channel evolves as a binary symmetric Markov chain, with and without side information at the receiver about the channel state. With side information, although capacity is invariant to the chain dynamics, dispersion is shown to be the sum of two terms: due to the Markov chain dynamics and due to the the randomness in the error generation, respectively.**

I. Introduction

The fundamental performance limit for a channel in the finite blocklength regime is $M^*(n,\epsilon)$, the maximum cardinality of a codebook of blocklength n which can be decoded with block error probability no greater than ϵ. Denoting the channel capacity by C, the approximation

$$\frac{\log M^*(n,\epsilon)}{n} \approx C \tag{1}$$

is asymptotically tight for channels that satisfy the strong converse. However for many channels, error rates and blocklength ranges of practical interest, (1) is too optimistic. It has been shown in [1] that a much tighter approximation can be obtained by defining a second parameter referred to as the channel dispersion:

Definition 1: The dispersion V (measured in squared information units per channel use) of a channel with capacity C is equal to

$$V \quad = \quad \lim_{\epsilon\to 0}\limsup_{n\to\infty} \frac{1}{n}\frac{(nC-\log M^*(n,\epsilon))^2}{2\ln\frac{1}{\epsilon}}\,. \tag{2}$$

In conjunction with the channel capacity C, channel dispersion emerges as a powerful analysis and design tool [1]. In order to achieve a given fraction η of capacity with a given error probability, the minimal required blocklength is proportional to V/C^2, namely[1],

$$n \gtrsim \left(\frac{Q^{-1}(\epsilon)}{1-\eta}\right)^2 \frac{V}{C^2}\,. \tag{3}$$

More specifically, [1] shows that for simple memoryless channels the two-term expansion

$$\log M^*(n,\epsilon) = nC - \sqrt{nV}Q^{-1}(\epsilon) + O(\log n)\,, \tag{4}$$

gives an excellent approximation (unless the blocklength is very small). The expansion (4) was first proved for a discrete memoryless channel by Strassen [2]. The new upper and lower bounds found in [1] allowed us to demonstrate the remarkable tightness of the approximation, prove (4) for the additive white Gaussian noise channels, refine Strassen's proof, and improve the bounds on the $O(\log n)$ term (see also [3], [4]).

In this paper, we initiate the study of the dispersion of channels subject to fading with memory. For coherent channels that behave ergodically, channel capacity is independent of the fading dynamics [5] since a sufficiently long codeword sees a channel realization whose empirical statistics have no randomness. In contrast, channel dispersion does depend on the extent of the fading memory since it determines the blocklength required to ride out not only the noise but the channel fluctuations due to fading. One of the simplest models that incorporates fading with memory is the Gilbert-Elliott channel (GEC): a binary symmetric channel where the crossover probability is a binary Markov chain [6], [7]. The results and required tools depend crucially on whether the channel state is known at the decoder.

II. Channel model

Let $\{S_j\}_{j=-\infty}^{\infty}$ be a homogeneous Markov process with states $\{1,2\}$, transition probabilities[2]

$$\mathbb{P}[S_2=1|S_1=1] = \mathbb{P}[S_2=2|S_1=2] = 1-\tau\,, \tag{5}$$
$$\mathbb{P}[S_2=2|S_1=1] = \mathbb{P}[S_2=1|S_1=2] = \tau\,, \tag{6}$$

and initial distribution

$$\mathbb{P}[S_1=1] = \mathbb{P}[S_1=2] = 1/2\,. \tag{7}$$

Now for $\delta_1,\delta_2\in[0,1]$ we define $\{Z_j\}_{j=-\infty}^{\infty}$ as conditionally independent given $S_{-\infty}^{\infty}$ and

$$\mathbb{P}[Z_j=0|S_j=s] \quad = \quad 1-\delta_s\,, \tag{8}$$
$$\mathbb{P}[Z_j=1|S_j=s] \quad = \quad \delta_s\,. \tag{9}$$

The research was supported by the National Science Foundation under Grants CCF-06-35154 and CNS-06-25637.

[1] As usual, $Q(x)=\int_x^\infty \frac{1}{\sqrt{2\pi}}e^{-t^2/2}\,dt$.

[2] The results in this paper can be readily generalized at the expense of more cumbersome expressions to Gilbert-Elliott channels with asymmetric Markov chains.

The Gilbert-Elliott channel acts on an input binary vector X_1^n by adding (modulo 2) the vector Z_1^n:

$$Y_1^n = X_1^n + Z_1^n\,. \tag{10}$$

III. Capacity

The capacity C_1 of a Gilbert-Elliott channel with the state S^n known perfectly at the receiver depends only on the stationary distribution P_{S_1} and is given by

$$\begin{aligned} C_1 &= \log 2 - \mathbb{E}\left[h(\delta_{S_1})\right] & (11)\\ &= \log 2 - \mathbb{P}[S_1 = 1]h(\delta_1) - \mathbb{P}[S_1 = 2]h(\delta_2)\,, & (12)\end{aligned}$$

where $h(x) = -x\log x - (1-x)\log(1-x)$ is the binary entropy function. In the symmetric-chain special case considered in this paper,

$$C_1 = \log 2 - \frac{1}{2}h(\delta_1) - \frac{1}{2}h(\delta_2). \tag{13}$$

When the state S^n is not known at the receiver, the capacity is given by [8]

$$\begin{aligned} C_0 &= \log 2 - \mathbb{E}\left[h(\mathbb{P}[Z_0 = 1|Z_{-\infty}^{-1}])\right] & (14)\\ &= \log 2 - \lim_{n\to\infty}\mathbb{E}\left[h(\mathbb{P}[Z_0 = 1|Z_{-n}^{-1}])\right]\,. & (15)\end{aligned}$$

Throughout the paper we use subscripts 1 and 0 to denote when the state S^n is known and is not known, respectively.

IV. Main results

Before showing the expansion for the Gilbert-Elliott channel we recall the corresponding result for the binary symmetric channel (BSC) [1], [3].

Theorem 1: The dispersion of the BSC with crossover probability δ is

$$V(\delta) = \delta(1-\delta)\log^2\frac{1-\delta}{\delta}\,. \tag{16}$$

Furthermore, provided that $V(\delta) > 0$ and regardless of whether $\epsilon \in (0,1)$ is a maximal or average probability of error we have

$$\begin{aligned} \log M^*(n,\epsilon) &= n(\log 2 - h(\delta)) - \sqrt{nV(\delta)}Q^{-1}(\epsilon)\\ &\quad + \frac{1}{2}\log n + O(1)\,. & (17)\end{aligned}$$

Theorem 2: The dispersion of the Gilbert-Elliott channel with state S^n known at the receiver, and state transition probability $\tau \in (0,1)$ is

$$\begin{aligned} V_1 &= \frac{1}{2}(V(\delta_1) + V(\delta_2))\\ &\quad + \left(\frac{h(\delta_1) - h(\delta_2)}{2}\right)^2\left(\frac{1}{\tau} - 1\right)\,. & (18)\end{aligned}$$

Furthermore, provided that $V_1 > 0$ and regardless of whether $\epsilon \in (0,1)$ is a maximal or average probability of error we have

$$\log M^*(n,\epsilon) = nC_1 - \sqrt{nV_1}Q^{-1}(\epsilon) + O(\log n)\,, \tag{19}$$

where C_1 is given in (13). Moreover, (19) holds even if the transmitter knows the full state sequence S^n in advance (i.e., non-causally).

Note that the condition $V_1 > 0$ for (19) to hold excludes only some degenerate cases for which we have: $M^*(n,\epsilon) = 2^n$ (when both crossover probabilities are 0 or 1) or $M^*(n,\epsilon) = \lfloor\frac{1}{1-\epsilon}\rfloor$ (when $\delta_1 = \delta_2 = 1/2$).

To formulate the result for the case of no state information at the receiver, we define the stationary process:

$$F_j = -\log P_{Z_j|Z_{-\infty}^{j-1}}(Z_j|Z_{-\infty}^{j-1})\,. \tag{20}$$

Theorem 3: The dispersion of the Gilbert-Elliott channel with no state information and state transition probability $\tau \in (0,1)$ is

$$V_0 = \mathrm{Var}\left[F_0\right] + 2\sum_{i=1}^{\infty}\mathbb{E}\left[(F_i - \mathbb{E}\left[F_i\right])(F_0 - \mathbb{E}\left[F_0\right])\right]\,. \tag{21}$$

Furthermore, provided that $V_0 > 0$ and regardless of whether ϵ is a maximal or average probability of error, we have

$$\log M^*(n,\epsilon) = nC_0 - \sqrt{nV_0}Q^{-1}(\epsilon) + O(\log n)\,, \tag{22}$$

where C_0 is given by (14).

It can be shown that the process F_j has a smooth spectral density $S_F(f)$, and that

$$V_0 = S_F(0)\,, \tag{23}$$

which provides a way of computing V_0 by a Monte Carlo simulation paired with a spectral estimator. Another method is to notice that the terms in the infinite sum (21) decay as $(1-2\tau)^j$. Hence, given any prescribed precision it is sufficient to compute only finitely many terms in (21). Each term can in turn be computed with required precision by noting that $P_{Z_j|Z_{-\infty}^{j-1}}[1|Z_{-\infty}^{j-1}]$ is a Markov process with a simple transition kernel.

The proof of Theorem 2 is outlined in the appendix, while the proof of Theorem 3 being somewhat more technical can be found in [10].

V. Discussion

The natural application of expansion (4) is in approximating the maximal achievable rate. Unlike the BSC case (17), the prelog constant for the GEC is unknown and therefore a natural choice for the approximation would be $0\log n$. However, in view of robustness of the prelog in (17) to variation in crossover probability, we chose the following expression for numerical comparison

$$C_{0,1} - \sqrt{\frac{V_{0,1}}{n}}Q^{-1}(\epsilon) + \frac{1}{2n}\log n\,. \tag{24}$$

To demonstrate the tightness of (24) we have computed numerically the upper and lower bounds developed in the course of the proof of Theorems 2 and 3. The comparison is shown on Fig. 1. For the case of state known (left plot) the achievability bound is (45) and the converse bound is (62), see the appendix. For the state not known (right plot), we computed the capacity and dispersion:

$$\begin{aligned} C_0 &\approx 0.280 \text{ bit}\,, & (25)\\ V_0 &\approx 2.173 \text{ bit}^2\,. & (26)\end{aligned}$$

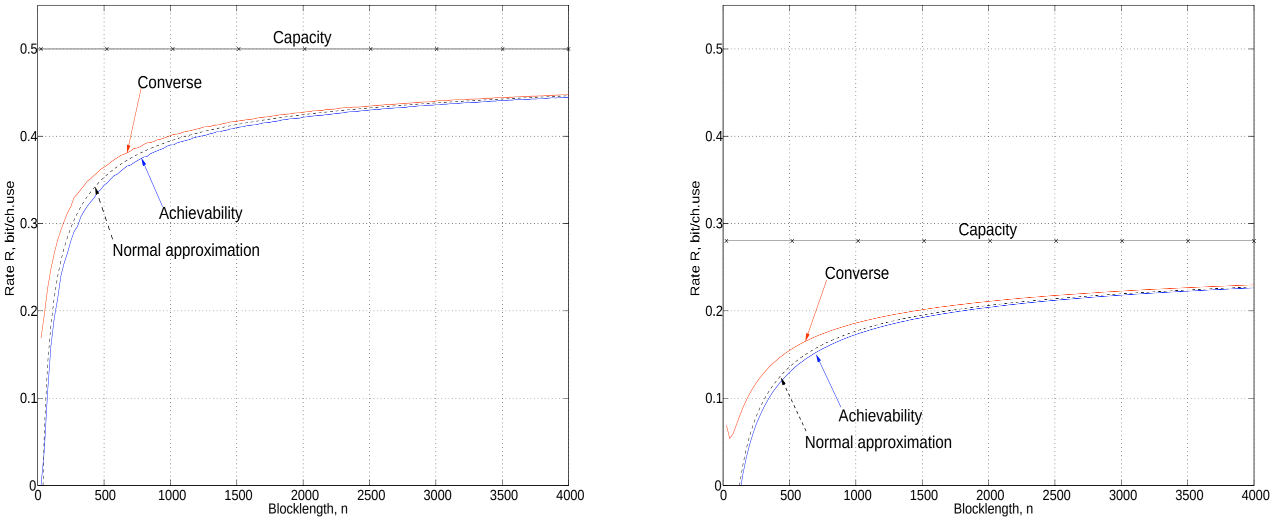

Fig. 1. Rate-blocklength tradeoff at block error rate $\epsilon = 10^{-2}$ for the Gilbert-Elliott channel with parameters $\delta_1 = 1/2$, $\delta_2 = 0$ and state transition probability $\tau = 0.1$. The left (right) plot is for the case when the state is known (not known) at the receiver.

The plots in Fig. 1 suggest that not only do our bounds tightly describe the value of $\frac{1}{n} \log M^*(n, \epsilon)$, but also that the simple expression (24) is precise enough for addressing many practical questions.

Let us discuss two such questions. First, for the state known case, capacity C_1 is independent of the state transition probability τ; see (13). However, as shown by Theorem 2, the channel dispersion V_1 does indeed depend on τ. But then (18) implies that this minimal blocklength behaves as $O\left(\frac{1}{\tau}\right)$ when $\tau \to 0$. This has an intuitive explanation: to achieve the full capacity of a Gilbert-Elliott channel we need to wait until the influence of the random initial state "washes away". Since transitions occur on average every $\frac{1}{\tau}$ channel uses, the blocklength should be $O\left(\frac{1}{\tau}\right)$ as $\tau \to 0$. Comparing (16) and (18) we can ascribe a meaning to each of the two terms in (18): the first one gives the dispersion due to the usual BSC noise, whereas the second one is due to memory in the channel.

Next, consider the case in which the state is not known at the decoder. As shown in [8], when the state transition probability τ decreases to 0 the capacity $C_0(\tau)$ increases to C_1. This is sometimes interpreted as implying that if the state is unknown at the receiver slower dynamics are advantageous. Our refined analysis, however, shows that this is true only up to a point.

Indeed, fix a rate $R < C_1$ and an $\epsilon > 0$. In view of the tightness of (24), the minimum blocklength, as a function of state transition probability τ needed to achieve rate R is approximately given by

$$N_0(\tau) \approx V_0(\tau) \left(\frac{Q^{-1}(\epsilon)}{C_0(\tau) - R} \right)^2 , \qquad (27)$$

provided that $C_0(\tau) > R$, of course.

Heuristically, there are two effects: when the state transition probability τ decreases we can predict the current state better and therefore, the capacity grows; on the other hand, as discussed above, when τ decreases we also have to wait longer until the chain "forgets" the initial state. The trade-off between these two effects is demonstrated on Fig. 2, where we plot $N_0(\tau)$ for the same parameters as in Fig. 1.

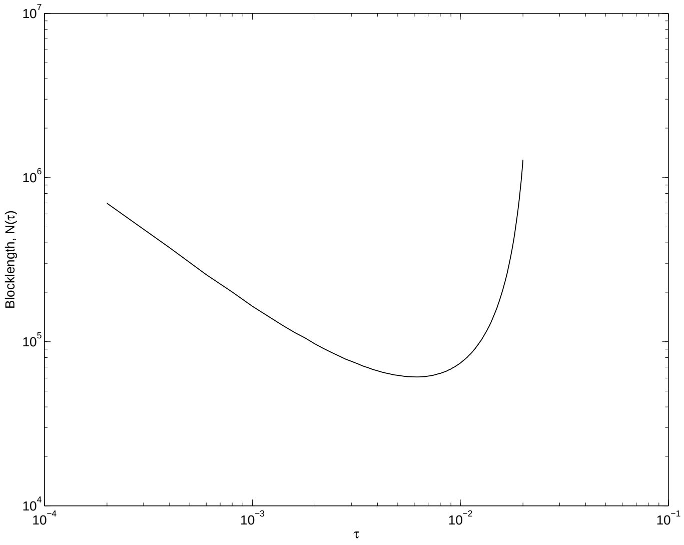

Fig. 2. Minimal blocklength, needed to achieve $R = 0.4$ bit and $\epsilon = 0.01$ as a function of state transition probability τ. The channel is the Gilbert-Elliott with no state information at the receiver, $\delta_1 = 1/2, \delta_2 = 0$.

VI. Conclusion

In this paper, we have proved an approximation of the form (4) for the Gilbert-Elliott channel. In Fig. 1, we have illustrated the relevance by comparing it numerically with bounds. As we have found previously in [1] and [3], expansions such as (4) have practical importance by providing tight approximations of the speed of convergence to capacity, and by allowing for estimation of the blocklength needed to achieve a given fraction of capacity, as given by (3).

References

[1] Y. Polyanskiy, H. V. Poor and S. Verdú, "Channel coding rate in the finite blocklength regime," submitted to *IEEE Trans. Inform. Theory*, Nov. 2008.

[2] V. Strassen, "Asymptotische Abschätzungen in Shannon's Informationstheorie," *Trans. Third Prague Conf. Information Theory*, 1962, Czechoslovak Academy of Sciences, Prague, pp. 689-723.

[3] Y. Polyanskiy, H. V. Poor and S. Verdú, "New channel coding achievability bounds," *Proc. IEEE Int. Symp. Information Theory (ISIT),* Toronto, Canada, 2008.

[4] Y. Polyanskiy, H. V. Poor and S. Verdú, "Dispersion of Gaussian channels," *Proc. IEEE Int. Symp. Information Theory (ISIT),* Seoul, Korea, 2009.

[5] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communication aspects," *IEEE Trans. Inform. Theory*, 50th Anniversary Issue, Vol. 44, No. 6, pp. 2619-2692, October 1998.

[6] E. N. Gilbert, "Capacity of burst-noise channels," *Bell Syst. Tech. J.*, Vol. 39, pp. 1253-1265, Sept. 1960.

[7] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *Bell Syst. Tech. J.*, Vol. 42, pp. 1977-1997, Sept. 1963

[8] M. Mushkin and I. Bar-David, "Capacity and coding for the Gilbert-Elliott channels," *IEEE Trans. Inform. Theory*, Vol. 35, No. 6, pp. 1277-1290, 1989.

[9] S. Verdú, *EE528–Information Theory, Lecture Notes,* Princeton University, Princeton, NJ, 2007.

[10] Y. Polyanskiy, H. V. Poor and S. Verdú, "Dispersion of the Gilbert-Elliott channel," *draft*, 2009.

[11] A. N. Tikhomirov, "On the convergence rate in the central limit theorem for weakly dependent random variables," *Theory of Probability and Its Applications*, Vol. XXV, No. 4, 1980.

Appendix
Proof of Theorem 2

For our analysis we need to invoke a few relevant results from [1] and [3]. Consider an abstract channel $P_{Y|X}$; for an arbitrary input distribution P_X define an (extended) random variable

$$i(X;Y) = \log \frac{dP_{Y|X}(Y|X)}{dP_Y(Y)}\,, \tag{28}$$

where $P_Y = \int dP_X P_{Y|X=x}$.

Theorem 4 (DT bound): For an arbitrary P_X there exists a code with M codewords and average probability of error ϵ satisfying

$$\epsilon \le \mathbb{E}\left[\exp\left\{-\left[i(X;Y) - \log\tfrac{M-1}{2}\right]^+\right\}\right]. \tag{29}$$

The optimal performance of binary hypothesis testing plays an important role in our development. Consider a random variable W taking values in a set W which can take probability measures P or Q. A randomized test between those two distributions is defined by a random transformation $P_{Z|W} : \mathsf{W} \mapsto \{0,1\}$ where 0 indicates that the test chooses Q. The best performance achievable among those randomized tests is given by[3]

$$\beta_\alpha(P,Q) = \min \sum_{w\in\mathsf{W}} Q(w) P_{Z|W}(1|w)\,, \tag{30}$$

where the minimum is taken over all probability distributions $P_{Z|W}$ satisfying

$$P_{Z|W} : \sum_{w\in\mathsf{W}} P(w) P_{Z|W}(1|w) \ge \alpha\,. \tag{31}$$

The minimum in (30) is guaranteed to be achieved by the Neyman-Pearson lemma. Thus, $\beta_\alpha(P,Q)$ gives the minimum probability of error under hypothesis Q if the probability of error under hypothesis P is not larger than $1-\alpha$. It is easy to show that (e.g. [9]) for any $\gamma > 0$

$$\alpha \le \mathbb{P}\left[\frac{dP}{dQ} \ge \gamma\right] + \gamma\beta_\alpha(P,Q). \tag{32}$$

On the other hand,

$$\beta_\alpha(P,Q) \le \frac{1}{\gamma_0}\,, \tag{33}$$

for any γ_0 that satisfies

$$\mathbb{P}\left[\frac{dP}{dQ} \ge \gamma_0\right] \ge \alpha\,. \tag{34}$$

Virtually all known converse results for channel coding follow from the next theorem by a judicious choice of $Q_{Y|X}$ and a lower bound on β, see [1].

Theorem 5 (meta-converse): Consider two different abstract channels $P_{Y|X}$ and $Q_{Y|X}$ defined on the same input and output spaces. For a given code (possibly randomized encoder and decoder pair), let

$$\begin{aligned} \epsilon &= \text{average error probability with } P_{Y|X} \\ \epsilon' &= \text{average error probability with } Q_{Y|X} \\ P_X = Q_X &= \text{encoder output distribution with} \\ & \text{equiprobable codewords.} \end{aligned}$$

Then,

$$\beta_{1-\epsilon}(P_{XY}, Q_{XY}) \le 1 - \epsilon'\,, \tag{35}$$

where $P_{XY} = P_X P_{Y|X}$ and $Q_{XY} = Q_X Q_{Y|X}$.

Proof of Theorem 2: Achievability: We choose P_{X^n} equiprobable. Since the output of the channel is (Y^n, S^n) we need to write down the expression for $i(X^n; Y^n S^n)$. To do that we define an operation on $\mathbb{R} \times \{0,1\}$:

$$a^{[b]} = \begin{cases} 0\,, & b = 0\,, \\ a\,, & b = 1 \end{cases}. \tag{36}$$

Then we obtain

$$\begin{aligned} i(X^n; Y^n S^n) &= \log \frac{P_{Y^n|X^n S^n}}{P_{Y^n|S^n}} & (37) \\ &= n\log 2 + \sum_{j=1}^n \log\left(\delta_{S_j}^{[Z_j]} + \bar\delta_{S_j}^{[1-Z_j]}\right) & (38) \end{aligned}$$

where (37) is by independence of X^n and S^n and (38) is because under equiprobable X^n we have that $P_{Y^n|S^n}$ is also equiprobable. Using (38) we can find

$$\begin{aligned} \mathbb{E}\left[i(X^n; Y^n S^n)\right] &= nC_1 \quad \text{and} & (39) \\ \mathrm{Var}[i(X^n; Y^n S^n)] &= nV_1 + O(1)\,. & (40) \end{aligned}$$

[3] We write summations over alphabets for simplicity; however, all of our general results hold for arbitrary probability spaces.

By (38) we see that $i(X^n;Y^nS^n)$ is a sum of weakly dependent random variables. For such a process, Tikhomirov's theorem [11] provides an extension of the Berry-Esseen inequality, namely:

$$\left|\mathbb{P}\left[i(X^n;Y^nS^n) > nC_1 + \sqrt{nV_1}\lambda\right] - Q(\lambda)\right| \le \frac{B_2 \log n}{\sqrt{n}}\,. \tag{41}$$

In addition, similarly to Lemma 45, Appendix G of [1], we can show for arbitrary A:

$$\mathbb{E}\left[\exp\{-i(X^n;Y^nS^n) + A\} \cdot 1\{i(X^n;Y^nS^n) \ge A\}\right] \le \frac{B_1 \log n}{\sqrt{n}}\,. \tag{42}$$

In (41) and (42), B_1 and B_2 are fixed constants, which do not depend on A or λ.

Finally, we set

$$\log \tfrac{M-1}{2} = nC_1 - \sqrt{nV_1}Q^{-1}(\epsilon_n)\,, \tag{43}$$

where

$$\epsilon_n = \epsilon - \frac{(B_1+B_2)\log n}{\sqrt{n}}\,. \tag{44}$$

Then, by Theorem 4 we know that there exists a code with M codewords and average probability of error p_e bounded by[4]

$$\begin{aligned}
p_e &\le \mathbb{E}\left[\exp\left\{-\left[i(X^n;Y^nS^n) - \log\tfrac{M-1}{2}\right]^+\right\}\right] && (45)\\
&\le \mathbb{P}\left[i(X^n;Y^nS^n) \le \log\tfrac{M-1}{2}\right] + \frac{B_1 \log n}{\sqrt{n}} && (46)\\
&\le \epsilon_n + \frac{(B_1+B_2)\log n}{\sqrt{n}} && (47)\\
&\le \epsilon\,, && (48)
\end{aligned}$$

where (46) is by (42) with $A = \log\frac{M-1}{2}$, (47) is by (41) and definition of $\log\frac{M-1}{2}$ and (48) is by (44). Therefore, by invoking Taylor's expansion in (43) we have

$$\log M^*(n,\epsilon) \ge \log M \ge nC_1 - \sqrt{nV_1}Q^{-1}(\epsilon) + O(\log n)\,. \tag{49}$$

Converse: In the converse part we will assume that the transmitter has access to the full state sequence S^n and then generates X^n based on both the input message and S^n. Take the best such code with $M^*(n,\epsilon)$ codewords and average probability of error no greater than ϵ. We now propose to treat the pair (X^n,S^n) as a combined input to the channel and the pair (Y^n,S^n) as a combined output, available to the decoder. Note that in this situation, the encoder induces a distribution $P_{X^nS^n}$ and is necessarily randomized because the distribution of S^n is given by the output of the Markov chain and is independent of the transmitted message W.

To apply Theorem 5 we choose the auxiliary channel that passes S^n unchanged and generates Y^n equiprobably:

$$Q_{Y^n|X^nS^n}(y^n,s^n|x^n) = 2^{-n}\,. \tag{50}$$

Note that by the constraint on the encoder, S^n is independent of the message W. Moreover, under the Q-channel the Y^n is also independent of W and we clearly have

$$\epsilon' \ge 1 - \frac{1}{M^*}\,. \tag{51}$$

Therefore by Theorem 5 we obtain

$$\beta_{1-\epsilon}\left(P_{X^nY^nS^n}, Q_{X^nY^nS^n}\right) \le \frac{1}{M^*}\,. \tag{52}$$

To lower bound $\beta_{1-\epsilon}\left(P_{X^nY^nS^n}, Q_{X^nY^nS^n}\right)$ via (32) we notice that

$$\begin{aligned}
\log\frac{P_{X^nY^nS^n}}{Q_{X^nY^nS^n}} &= \log\frac{P_{Y^n|X^nS^n}P_{X^nS^n}}{Q_{Y^n|X^nS^n}Q_{X^nS^n}} && (53)\\
&= \log\frac{P_{Y^n|X^nS^n}}{Q_{Y^n|X^nS^n}} && (54)\\
&= i(X^n;Y^nS^n)\,, && (55)
\end{aligned}$$

where (53) is because $P_{X^nS^n} = Q_{X^nS^n}$ and (55) follows simply by noting that $P_{Y^n|S^n}$ in (37) is also equiprobable. Now set

$$\log\gamma = nC_1 - \sqrt{nV_1}Q^{-1}(\epsilon_n)\,, \tag{56}$$

where this time

$$\epsilon_n = \epsilon + \frac{B_2\log n}{\sqrt{n}} + \frac{1}{\sqrt{n}}\,. \tag{57}$$

By (32) we have for $\alpha = 1-\epsilon$:

$$\begin{aligned}
\beta_{1-\epsilon} &\ge \frac{1}{\gamma}\left(1-\epsilon-\mathbb{P}\left[\log\frac{P_{X^nY^nS^n}}{Q_{X^nY^nS^n}} \ge \log\gamma\right]\right) && (58)\\
&= \frac{1}{\gamma}\left(1-\epsilon-\mathbb{P}\left[i(X^n;Y^nS^n) \ge \log\gamma\right]\right) && (59)\\
&\ge \frac{1}{\gamma}\left(1-\epsilon-(1-\epsilon_n)-\frac{B_2\log n}{\sqrt{n}}\right) && (60)\\
&= \frac{1}{\sqrt{n}\gamma}\,, && (61)
\end{aligned}$$

where (59) is by (55), (60) is by (41) and (61) is by (57).

Finally,

$$\begin{aligned}
\log M^*(n,\epsilon) &\le -\log\beta_{1-\epsilon} && (62)\\
&\le \log\gamma + \frac{1}{2}\log n && (63)\\
&= nC_1 - \sqrt{nV_1}Q^{-1}(\epsilon_n) + \frac{1}{2}\log n && (64)\\
&= nC_1 - \sqrt{nV_1}Q^{-1}(\epsilon) + O(\log n), && (65)
\end{aligned}$$

where (62) is just (52), (63) is by (61), (64) is by (56) and (65) is by Taylor's formula applied to Q^{-1} using (57) for ϵ_n. ■

[4] In the statement of the theorem we claimed a stronger result about *maximal* probability of error. Its proof is only slightly different and is omitted.