Channel Dispersion and Moderate Deviations Limits for Memoryless Channels

Yury Polyanskiy and Sergio Verdú

Abstract—Recently, Altug and Wagner [1] posed a question regarding the optimal behavior of the probability of error when channel coding rate converges to the capacity sufficiently slowly. They gave a sufficient condition for the discrete memoryless channel (DMC) to satisfy a moderate deviation property (MDP) with the constant equal to the channel dispersion. Their sufficient condition excludes some practically interesting channels, such as the binary erasure channel and the Z-channel. We extend their result in two directions. First, we show that a DMC satisfies MDP if and only if its channel dispersion is nonzero. Second, we prove that the AWGN channel also satisfies MDP with a constant equal to the channel dispersion. While the methods used by Altug and Wagner are based on the method of types and other DMC-specific ideas, our proofs (in both achievability and converse parts) rely on the tools from our recent work [2] on finite-blocklength regime that are equally applicable to non-discrete channels and channels with memory.

Index Terms—Shannon theory, channel capacity, channel dispersion, moderate deviations, discrete channels, AWGN channel, finite blocklength regime.

I. INTRODUCTION

A random transformation is defined by a pair of measurable spaces of inputs A and outputs B and a conditional probability measure $P_{Y|X} : A \mapsto B$. An (M, ϵ) *code (average probability of error)* for the random transformation $(A, B, P_{Y|X})$ is a pair of (possibly randomized) maps $f : \{1, \ldots, M\} \to A$ (the encoder) and $g : B \to \{1, \ldots, M\}$ (the decoder), satisfying

$$\frac{1}{M} \sum_{m=1}^{M} P[g(Y) \neq m | X = f(m)] \leq \epsilon. \tag{1}$$

Similarly, an (M, ϵ) *code (maximal probability of error)* is a pair of (possibly randomized) maps $f : \{1, \ldots, M\} \to A$ and $g : B \to \{1, \ldots, M\}$, satisfying

$$\max_{m \in \{1, \ldots, M\}} P[g(Y) \neq m | X = f(m)] \leq \epsilon. \tag{2}$$

In applications, we will take A and B to be n-fold Cartesian products of alphabets $\mathcal{A}$ and $\mathcal{B}$, and a channel to be a sequence of random transformations $\{P_{Y^n|X^n} : \mathcal{A}^n \to \mathcal{B}^n\}$ indexed by blocklength [3]. An (M, ϵ) code for $\{\mathcal{A}^n, \mathcal{B}^n, P_{Y^n|X^n}\}$ is called an (n, M, ϵ) code. For a chosen channel we define the

following non-asymptotic fundamental limits:

$$\epsilon^*(n, M)$$
$$= \inf\{\epsilon : \exists (n, M, \epsilon)\text{-code (maximal probab. of error)}\} \tag{3}$$
$$\epsilon^*_{avg}(n, M)$$
$$= \inf\{\epsilon : \exists (n, M, \epsilon)\text{-code (average probab. of error)}\}. \tag{4}$$

For several memoryless channels as well as some channels with memory it is known that

$$\lim_{n \to \infty} \epsilon^*(n, \exp\{nR\}) = \begin{cases} 0, & R < C \\ 1, & R > C, \end{cases} \tag{5}$$

where C is the capacity of the channel. The convergence in (5) is known to be exponential, but the precise evaluation of this exponent is generally an open problem even for the simplest DMCs.

If we replace $\exp\{nR\}$ with $\exp\{nC - A\sqrt{n}\}$ then the probability of error converges to a number between 0 and 1, as follows:[1]

$$\lim_{n \to \infty} \epsilon^*(n, \exp\{nC - A\sqrt{n}\}) = Q\left(\frac{A}{\sqrt{V}}\right), \tag{6}$$

where V is the channel dispersion, a fundamental characteristic of a channel, especially valuable in the finite blocklength analysis; see [2], [4].

Reference [1] raised the question of the best possible behavior of the probability of error when the coding rate approaches capacity slower than $1/\sqrt{n}$. If we assume that (6) holds uniformly in A, then we expect that

$$\epsilon^*(n, \exp\{nC - n\rho_n\}) \sim Q\left(\frac{\sqrt{n}\rho_n}{\sqrt{V}}\right) \sim e^{\frac{-n\rho_n^2}{2V}}. \tag{7}$$

This argument justifies the following definition:

Definition 1: A channel with capacity C is said to satisfy the moderate deviation property (MDP) with constant ν if for any sequence of integers M_n such that

$$\log M_n = nC - n\rho_n, \tag{8}$$

where $\rho_n > 0$, $\rho_n \to 0$ and $n\rho_n^2 \to \infty$, we have

$$\lim_{n \to \infty} \frac{1}{n\rho_n^2} \log \epsilon^*(n, M_n)$$
$$= \lim_{n \to \infty} \frac{1}{n\rho_n^2} \log \epsilon^*_{avg}(n, M_n) \tag{9}$$
$$= -\frac{\log e}{2\nu}. \tag{10}$$

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

The regime of rate slowly converging to capacity as in (8) falls between the central-limit theorem (CLT) regime (6) and the large deviations (or error-exponent) regime (5).

In [1] it was shown that MDP holds for a certain subset of the DMCs (which excludes, for example, the binary erasure channel (BEC) and the Z-channel). We show how a refinement of their result can be easily derived using methods developed in [2]. Namely, we show that a DMC satisfies MDP if and only if its channel dispersion is positive. Therefore, not only do we extend the subset of DMCs for which MDP is shown, but we show that this subset cannot be further extended. Additionally, we show that the additive white Gaussian noise (AWGN) channel satisfies MDP. We also show that for any channel (not necessarily stationary, memoryless, or even non-anticipatory) the constant ν in the MDP and the dispersion V in the central-limit (6) cannot differ.

One of the main tools in our treatment [2] is the performance of the optimal binary hypothesis test defined as follows. Consider a W-valued random variable W which can take probability measures P or Q. A randomized test between those two distributions is defined by a random transformation $P_{Z|W} : W \mapsto \{0, 1\}$ where 0 indicates that the test chooses Q. The best performance achievable among those randomized tests is given by[2]

$$\beta_\alpha(P, Q) = \min \sum_{w \in W} Q(w) P_{Z|W}(1|w), \qquad (11)$$

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

$$P_{Z|W} : \sum_{w \in W} P(w) P_{Z|W}(1|w) \geq \alpha. \qquad (12)$$

The minimum in (11) 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. [5]) for any $\gamma > 0$

$$\alpha \leq \mathbb{P}\left[\frac{dP}{dQ} \geq \gamma\right] + \gamma \beta_\alpha(P, Q). \qquad (13)$$

On the other hand,

$$\beta_\alpha(P, Q) \leq \frac{1}{\gamma_0}, \qquad (14)$$

for any γ_0 that satisfies

$$\mathbb{P}\left[\frac{dP}{dQ} \geq \gamma_0\right] \geq \alpha. \qquad (15)$$

II. ON THE MDP CONSTANT

For an arbitrary channel, neither (6) implies (10), nor vice versa. However, if both limits hold, then the respective constants must be equal:

Theorem 1: Consider an arbitrary channel (i.e. a sequence of random transformations) with capacity C and suppose that

---

central-limit (6) holds with dispersion V and MDP holds with constant ν (thus, $V > 0$ and $\nu > 0$). Then

$$V = \nu. \qquad (16)$$

Proof: Define a sequence of cumulative density functions (CDFs) as follows:

$$F_n(x) \triangleq \epsilon^*(n, \lfloor \exp\{nC + x\sqrt{nV}\} \rfloor). \qquad (17)$$

Then, on one hand the central-limit property (6) ensures that

$$F_n(x) \to \Phi(x), \qquad (18)$$

for all $x \in \mathbb{R}$, where Φ is the CDF of the standard Gaussian $\mathcal{N}(0, 1)$:

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy. \qquad (19)$$

On the other hand, the MDP property, cf. Definition 1, can be reformulated as follows: For every sequence $a_n > 0$ s.t. $1 \ll a_n \ll \sqrt{n}$ we have

$$\lim_{n \to \infty} \frac{1}{a_n^2} \log F_n(-a_n) = -\frac{\log e}{2\theta}, \qquad (20)$$

where $\theta = \frac{\nu}{V}$. We must show that $\theta = 1$.

To do so, define the following non-increasing sequence of numbers:

$$u_n \triangleq \sup_{m \geq n} \sup_{x \in \mathbb{R}} |F_m(x) - \Phi(x)|. \qquad (21)$$

Since the convergence in (18) is necessarily uniform, we have:

$$u_n \to 0. \qquad (22)$$

Notice that if $u_n = 0$ for all sufficiently large n, then the result follows automatically since for any sequence a_n satisfying conditions for (20) we have

$$\lim_{n \to \infty} \frac{1}{a_n^2} \log \Phi(-a_n) = -\frac{\log e}{2} \qquad (23)$$

and $\theta = 1$. Thus, in the remaining we assume that u_n does not vanish for any n.

First, suppose that

$$\limsup_{n \to \infty} \frac{1}{n} \log u_n < 0. \qquad (24)$$

Then for some $\delta > 0$ and all $n \geq N_1$ we should have

$$u_n \leq \exp\{-n\delta\} \qquad (25)$$

But then, for any admissible sequence a_n we have

$$\lim_{n \to \infty} \frac{1}{a_n^2} \log F_n(-a_n) = \lim_{n \to \infty} \frac{1}{a_n^2} \log \Phi(-a_n) \qquad (26)$$

because

$$|F_n(-a_n) - \Phi(-a_n)| \leq \exp\{-n\delta\} \qquad (27)$$

and by the conditions on a_n, $\Phi(-a_n) \gg \exp\{-n\delta\}$. Finally, application of (23) to (26) completes the proof in this case.

Second, suppose that

$$\limsup_{n \to \infty} \frac{1}{n} \log u_n = 0. \qquad (28)$$

Assume, for example, that $\theta > 1$. Fix any $\delta > 0$ such that

$$\frac{1}{2} + \delta > \frac{1}{2\theta} + 2\delta \,. \tag{29}$$

Choose the following sequence $a_n > 0$:

$$a_n = \sqrt{ -\left(\frac{1}{2\theta} + 2\delta\right)^{-1} \log_e u_n } \,. \tag{30}$$

The limit (23) implies that for all sufficiently large n we have

$$\Phi(-a_n) \le e^{-\left(\frac{1}{2} + \delta\right)a_n^2} = (u_n)^{r_1} \,, \tag{31}$$

where $r_1 > 1$ by (29). Condition (28) shows that sequence a_n satisfies conditions for (20), from which we find that for all sufficiently large n we have

$$F_n(-a_n) \ge e^{-\left(\frac{1}{2\theta} + \delta\right)a_n^2} = (u_n)^{r_2} \,, \tag{32}$$

where $r_2 < 1$. Consider the chain of inequalities:

$$u_n \ge |F_n(-a_n) - \Phi(-a_n)| \tag{33}$$
$$\ge F_n(-a_n) - \Phi(-a_n) \tag{34}$$
$$\ge (u_n)^{r_2} - (u_n)^{r_1} \tag{35}$$

where (33) is by the definition of u_n in (21) and (35) is by (31) and (32). Finally, (35) is a contradiction since $r_2 < 1 < r_1$. Similarly, one shows that $\theta < 1$ is also impossible. ∎

III. Discrete memoryless channels

In the sequel we use the notation of [2, Section IV.A]. In particular, the DMC has finite input alphabet $\mathcal{A}$, finite output alphabet $\mathcal{B}$, and conditional probabilities

$$P_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^{n} W(y_i|x_i) \,, \tag{36}$$

where $W(\cdot|x)$ is a conditional probability mass function on $\mathcal{B}$ for all $x \in \mathcal{A}$, which is abbreviated as W_x when notationally convenient. We denote the simplex of probability distributions on $\mathcal{A}$ by $\mathcal{P}$. It is useful to partition $\mathcal{P}$ into n-types:

$$\mathcal{P}_n = \{ P \in \mathcal{P} : \ nP(x) \in \mathbb{Z}_+ \ \forall x \in \mathcal{A} \} \,. \tag{37}$$

We use the following notation and terminology:

- *output distribution PW*

$$PW(y) \triangleq \sum_{x \in \mathcal{A}} P(x) W(y|x) \,. \tag{38}$$

- *information density*

$$i(x; y) = \log \frac{W(y|x)}{PW(y)} \,. \tag{39}$$

- *mutual information*

$$I(P, W) = \mathbb{E}[i(X; Y)] \tag{40}$$
$$= \sum_{x \in \mathcal{A}} \sum_{y \in \mathcal{B}} P(x) W(y|x) \log \frac{W(y|x)}{PW(y)} \tag{41}$$

- *conditional information variance*

$$V(P, W)$$
$$= \mathbb{E}[\mathrm{Var}(i(X; Y) \,|\, X)] \tag{42}$$
$$= \sum_{x \in \mathcal{A}} P(x) \Bigg\{ \sum_{y \in \mathcal{B}} W(y|x) \log^2 \frac{W(y|x)}{PW(y)}$$
$$- [D(W_x \| PW)]^2 \Bigg\} \tag{43}$$

- *third absolute moment of the information density*

$$T(P, W) = \sum_{x \in \mathcal{A}} \sum_{y \in \mathcal{B}} P(x) W(y|x) \left| \log \frac{W(y|x)}{PW(y)} \right.$$
$$\left. - D(W_x \| PW) \right|^3 \,. \tag{44}$$

Note that $I(P, W)$, $V(P, W)$ and $T(P, W)$ are continuous functions of $P \in \mathcal{P}$; see [2, Lemma 62].

- the compact subset of *capacity-achieving distributions* Π

$$\Pi \triangleq \{ P \in \mathcal{P} : \ I(P, W) = C \} \,. \tag{45}$$

where

$$C = \max_{P \in \mathcal{P}} I(P, W) \,. \tag{46}$$

- *channel dispersion*, which according to [2, Theorem 49] is equal to

$$V = \min_{P \in \Pi} V(P, W) \,. \tag{47}$$

Apart from analyzing the limit of ϵ_{avg}^* the result of [1] can be stated as follows:

Theorem 2 ([1]): Consider a DMC W. If $W(y|x) > 0$ for all $x \in \mathcal{A}, y \in \mathcal{B}$ and $V > 0$ then DMC W satisfies MDP with the constant V.

The main result of this section is:

Theorem 3: The DMC W satisfies MDP if and only if $V > 0$, in which case V is the MDP constant of the DMC.

Theorem 3 follows from Theorems 4 and 6 below.

Theorem 4: Consider a DMC W and a sequence ρ_n such that $\rho_n > 0$, $\rho_n \to 0$ and $\rho_n^2 n \to \infty$. If $V > 0$ then there exists a sequence of $(n, \exp\{nC - n\rho_n\}, \epsilon_n)$ codes (maximal probability of error) with

$$\limsup_{n \to \infty} \frac{1}{n\rho_n^2} \log \epsilon_n \le -\frac{1}{2V} \,. \tag{48}$$

On the other hand, when $V = 0$ there exists a sequence of $(n, \exp\{nC - n\rho_n\}, \epsilon_n)$ codes (maximal probability of error) with

$$\epsilon_n \le 2 \exp\{-n\rho_n\} \,, \tag{49}$$

so that the channel cannot satisfy MDP.

Proof: Denote by P the capacity achieving distribution that also achieves V in (47). According to the DT bound [2, Theorem 17], there exist an $(n, 2\exp\{nC - n\rho_n\}, \epsilon_n')$ code (average probability of error) such that

$$\epsilon_n' \le \mathbb{E}\left[\exp\left\{ -|i(X^n, Y^n) - nC + n\rho_n|^+ \right\} \right] \,, \tag{50}$$

where

$$i(x^n, y^n) \triangleq \sum_{j=1}^{n} \log \frac{W(y_j|x_j)}{PW(y_j)} \,. \tag{51}$$

And therefore, by a standard "purging" method, there also exists an $(n, \exp\{nC - n\rho_n\}, \epsilon_n)$ code (maximal probability of error) with $\epsilon_n = 2\epsilon'_n$, or

$$\epsilon_n \leq 2\mathbb{E}\left[\exp\left\{-|i(X^n, Y^n) - nC + n\rho_n|^+\right\}\right]. \tag{52}$$

If $V = 0$ then $i(X^n, Y^n) = nC$ and (49) readily follows.

Assume $V > 0$, fix arbitrary $\lambda < 1$ and consider a chain of elementary inequalities:

$$\exp\left\{-|i(X^n, Y^n) - nC + n\rho_n|^+\right\} \tag{53}$$
$$\leq \quad 1\{i(X^n, Y^n) \leq nC - \lambda n\rho_n\} \tag{54}$$
$$+ \exp\left\{-|i(X^n, Y^n) - nC + n\rho_n|^+\right\}$$
$$\times 1\{i(X^n, Y^n) > nC - \lambda n\rho_n\} \tag{55}$$
$$\leq \quad 1\{i(X^n, Y^n) \leq nC - \lambda n\rho_n\}$$
$$+ \exp\{-(1 - \lambda)n\rho_n\} \,. \tag{56}$$

By [6, Theorem 3.7.1] we have

$$\limsup_{n\to\infty} \frac{1}{n\rho_n^2} \log \mathbb{P}[i(X^n, Y^n) \leq nC - \lambda n\rho_n]$$
$$\leq \quad -\frac{\lambda^2 \log e}{2V} \,. \tag{57}$$

Therefore, by taking the expectation in (56) and by conditions on ρ_n the second term is asymptotically dominated by the first and we obtain:

$$\limsup_{n\to\infty} \frac{1}{n\rho_n^2} \log \mathbb{E}\left[\exp\left\{-|i(X^n, Y^n) - nC + n\rho_n|^+\right\}\right]$$
$$\leq \quad -\frac{\lambda^2 \log e}{2V} \,. \tag{58}$$

Since $\lambda < 1$ was arbitrary we can take $\lambda \to 1$ to obtain (48). ∎

The main analytic tool required in proving the converse bound in this section is a tight non-asymptotic lower bound for the probability of a large deviation of a random variable from its mean. This question has been addressed by many authors working in probability and statistics, starting from Kolmogorov [7]. Currently, one of the most general such results belongs to Rozovsky [8], [9]. The following is a weakening of [8, Theorem 1] which plays the same role as Berry-Esseen inequality in the analysis of (6); see [2].[3]

Theorem 5 (Rozovsky): There exist universal constants $A_1 > 0$ and $A_2 > 0$ with the following property. Let X_k, $k = 1, \ldots, n$ be independent with finite third moments:

$$\mu_k = \mathbb{E}[X_k], \ \sigma_k^2 = \text{Var}[X_k], \ \text{and } t_k = \mathbb{E}[|X_k - \mu_k|^3]. \tag{59}$$

---

[3]Similar to well-known extensions of the Berry-Esseen inequality to the case of random variables without a third absolute moment, Rozovsky does not require that $\mathbb{E}|X_k|^3$ be bounded. However, we only will need this weaker result.

Denote $S = \sum_{k=1}^{n} \sigma_k^2$ and $T = \sum_{k=1}^{n} t_k$. Whenever $x \geq 1$ we have

$$\mathbb{P}\left[\sum_{k=1}^{n}(X_k - \mu_k) > x\sqrt{S}\right] \geq Q(x)e^{-\frac{A_1 T}{S^{3/2}}x^3}\left(1 - \frac{A_2 T}{S^{3/2}}x\right). \tag{60}$$

Theorem 6: Consider a DMC W and a sequence of (n, M_n, ϵ_n) codes (average probability of error) with

$$\log M_n \geq nC - n\rho_n \,, \tag{61}$$

where $\rho_n > 0$, $\rho_n \to 0$ and $\rho_n^2 n \to \infty$. If $V > 0$ then we have

$$\liminf_{n\to\infty} \frac{1}{n\rho_n^2} \log \epsilon_n \geq -\frac{\log e}{2V} \,. \tag{62}$$

Proof: Replacing the encoder with an optimal deterministic one, we can only reduce the average probability of error. Next, if we have an (n, M_n, ϵ_n) code (average probability of error) with a deterministic encoder, then a standard argument shows that there exists an $(n, \frac{1}{2}M_n, 2\epsilon_n)$ subcode (maximal probability of error). Replacing $M_n \to \frac{1}{2}M_n$ and $\epsilon_n \to 2\epsilon_n$, without loss of generality we may assume the code to have a deterministic encoder and a maximal probability of error ϵ_n.

Now for each n denote by $P_n \in \mathcal{P}_n$ the n-type containing the largest number of codewords. A standard type-counting argument shows that then there exists an (n, M'_n, ϵ_n) constant composition P_n subcode with

$$\log M'_n \geq nC - n\rho_n - |\mathcal{A}| \log(n+1) \,. \tag{63}$$

By compactness of $\mathcal{P}$ the sequence P_n has an accumulation point P^*. Without loss of generality, we may assume $P_n \to P^*$.

Now for each n define the following probability distribution Q_{Y^n} on $\mathcal{B}^n$:

$$Q_{Y^n}(y^n) = \prod_{j=1}^{n} P_n W(y_j) \,. \tag{64}$$

According to [2, Theorem 31] we have

$$\beta_{1-\epsilon_n}(P_{X^n Y^n}, P_{X^n} Q_{Y^n}) \leq \frac{1}{M'_n} \,, \tag{65}$$

where here and below P_{X^n} is the distribution induced by the encoder on $\mathcal{A}^n$.

Applying (13) we get that for any γ we have:

$$\epsilon_n \geq \mathbb{P}\left[\log \frac{W(Y^n|X^n)}{Q_{Y^n}(Y^n)} < \gamma\right] - \exp\{\gamma - \log M'_n\} \,. \tag{66}$$

We now fix arbitrary $\lambda > 1$ and take $\gamma = nC - \lambda n\rho_n$ to obtain:

$$\epsilon_n \geq \mathbb{P}\left[\log \frac{W(Y^n|X^n)}{Q_{Y^n}(Y^n)} < nC - \lambda n\rho_n\right]$$
$$- \exp\{-n\rho_n(\lambda - 1) + |\mathcal{A}| \log(n+1)\} \,. \tag{67}$$

Notice that since the code has constant composition P_n, the distribution of $\log \frac{W(Y^n|X^n)}{Q_{Y^n}(Y^n)}$ given $X^n = x^n$ is the same for all x^n. Therefore, assuming such conditioning we have

$$\log \frac{W(Y^n|X^n)}{Q_{Y^n}(Y^n)} \sim \sum_{j=1}^{n} Z_j \,, \tag{68}$$

where Z_j are independent and

$$\sum_{j=1}^{n} \mathbb{E}\left[Z_j\right] = nI(P_n, W), \qquad (69)$$

$$\sum_{j=1}^{n} \text{Var}[Z_j] = nV(P_n, W), \qquad (70)$$

$$\sum_{j=1}^{n} \mathbb{E}\left[|Z_j - \mathbb{E}\left[Z_j\right]|^3\right] = nT(P_n, W). \qquad (71)$$

In terms of Z_j the bound in (67) asserts

$$\epsilon_n \geq \mathbb{P}\left[\sum_{j=1}^{n} Z_j < nC - \lambda n \rho_n\right]$$
$$- \exp\{-n\rho_n(\lambda - 1) + |\mathcal{A}|\log(n+1)\}. \quad (72)$$

First, suppose that $I(P^*, W) < C$. Then a simple Chernoff bound implies that the right-hand side of (67) converges to 1 and (62) holds.

Next, assume $I(P^*, W) = C$. Since $I(P_n, W) \leq C$ we have from (72):

$$\epsilon_n \geq \mathbb{P}\left[\sum_{j=1}^{n} Z_j - nI(P_n, W) < -\lambda n \rho_n\right]$$
$$- \exp\{-n\rho_n(\lambda - 1) + |\mathcal{A}|\log(n+1)\}. \quad (73)$$

Note that by continuity of $V(P, W)$ we have

$$V(P_n, W) \to V(P^*, W) \geq V > 0, \qquad (74)$$

where $V(P^*, W) \geq V$ since P^* is capacity-achieving. Therefore, by Theorem 5 we obtain:

$$\mathbb{P}\left[\sum_{j=1}^{n} Z_j - nI(P_n, W) < -\lambda n \rho_n\right]$$
$$\geq Q\left(\frac{\lambda}{\sqrt{V(P_n, W)}}\sqrt{n\rho_n^2}\right) e^{-\frac{\lambda^3 A_1 T(P_n, W)}{V^3(P_n, W)} n\rho_n^3}$$
$$\times \left(1 - \frac{\lambda A_2 T(P_n, W)}{V^2(P_n, W)}\rho_n\right), \qquad (75)$$

since $T(P_n, W)$ is continuous and, thus, bounded on $\mathcal{P}$, we see that the term in parentheses is $1 + o(1)$ because of the conditions on ρ_n. Therefore,

$$\liminf_{n \to \infty} \frac{1}{n\rho_n^2} \log \mathbb{P}\left[\sum_{j=1}^{n} Z_j - nI(P_n, W) < -\lambda n \rho_n\right]$$
$$\geq \lim_{n \to \infty} \frac{1}{n\rho_n^2} \log Q\left(\frac{\lambda}{\sqrt{V(P_n, W)}}\sqrt{n\rho_n^2}\right)$$
$$+ \lim_{n \to \infty} \frac{1}{n\rho_n^2}\left(-\frac{\lambda^3 A_1 T(P_n, W)}{V^3(P_n, W)}n\rho_n^3\right) \qquad (76)$$

$$= -\frac{\lambda^2 \log e}{2V(P_*, W)} \qquad (77)$$

$$\geq -\frac{\lambda^2 \log e}{2V}. \qquad (78)$$

Finally, it is easy to see that the second term in (73) is asymptotically dominated by the first term according to (78) and $n\rho_n \gg n\rho_n^2$. Thus, from (78) we conclude that

$$\liminf_{n \to \infty} \frac{1}{n\rho_n^2} \log \epsilon_n \geq -\frac{\log e}{2V}. \qquad (79)$$

■

IV. THE AWGN CHANNEL

The AWGN channel with signal-to-noise ratio (SNR) equal to P is defined for each blocklength n as follows: the input space is a subset of vectors of $\mathbb{R}^n$ satisfying

$$||x^n||^2 \leq nP, \qquad (80)$$

the output space is $\mathbb{R}^n$ and the channel acts by adding a white Gaussian noise of variance 1:

$$Y^n = X^n + Z^n, \qquad (81)$$

where $Z^n \sim \mathcal{N}(0, \mathbf{I}_n)$.

The channel dispersion of the AWGN channel is given by [2, Theorem 54]

$$V(P) = \frac{\log^2 e}{2}\left(1 - \frac{1}{(1+P)^2}\right). \qquad (82)$$

Theorem 7: The AWGN channel with SNR P satisfies MDP with constant $V(P)$.

Proof: We rely heavily on the notation and results of [2, Section III.J].

Converse: Consider a sequence of (n, M_n, ϵ_n) codes (average probability of error) with

$$M_n = \exp\{nC - n\rho_n\}, \qquad (83)$$

where $\rho_n > 0$, $\rho_n \to 0$ and $\rho_n^2 n \to \infty$. Following the method of [10] and [2, Lemma 39] we can assume without loss of generality that every codeword $\mathbf{C}_j \in \mathbb{R}^n, j = 1, \ldots, M_n$ lies on a power-sphere:

$$||\mathbf{C}_j||^2 = nP. \qquad (84)$$

We apply the meta-converse bound [2, Theorem 27] with Q_{Y^n} chosen as

$$Q_{Y^n} = \prod_{j=1}^{n} \mathcal{N}(0, 1+P), \qquad (85)$$

to obtain

$$\beta_{1-\epsilon_n}(P_{X^nY^n}, P_{X^n}Q_{Y^n}) \leq \exp\{-nC + n\rho_n\}, \qquad (86)$$

where P_{X^n} is the distribution induced by the encoder on $\mathbb{R}^n$. As explained in [2, Section III.J] we have the equality

$$\beta_{1-\epsilon_n}(P_{X^nY^n}, P_{X^n}Q_{Y^n}) = \beta_{1-\epsilon_n}(P_{Y^n|X^n=x}, Q_{Y^n}), \quad (87)$$

where $x = [\sqrt{P}, \ldots, \sqrt{P}]^T$. Now applying (13) $\beta_{1-\epsilon_n}(P_{Y^n|X^n=x}, Q_{Y^n})$ with $\gamma = nC - \lambda n\rho_n$, where $\lambda > 1$ is arbitrary we obtain

$$\epsilon_n \geq \mathbb{P}\left[\frac{\log e}{2(1+P)}\sum_{i=1}^{n} P(1 - Z_i^2) + 2\sqrt{P}Z_i < -\lambda n\rho_n\right]$$
$$- \exp\{-n\rho_n(\lambda - 1)\}, \qquad (88)$$

where we have written the distribution of $\log \frac{P_{Y^n|X^n=x}}{Q_{Y^n}}$ explicitly in terms of the i.i.d. random variables $Z_j \sim \mathcal{N}(0,1)$; see [2, (205)]. According to [6, Theorem 3.7.1], the first term in the right-hand side of (88) dominates the second one and we have

$$\liminf_{n\to\infty} \frac{1}{n\rho_n^2} \log \epsilon_n \geq -\frac{\lambda^2}{2V(P)}, \qquad (89)$$

and taking $\lambda \searrow 1$ we obtain

$$\liminf_{n\to\infty} \frac{1}{n\rho_n^2} \log \epsilon_n \geq -\frac{1}{2V(P)}. \qquad (90)$$

Achievability: Similar to [2, Section III.J] we apply the $\kappa\beta$ bound [2, Theorem 25], with F chosen to be the power sphere

$$\mathsf{F} = \{x^n \in \mathbb{R}^n : ||x^n||^2 = nP\} \qquad (91)$$

and Q_{Y^n} as in (85). Using the identity (87) and the lower bound on $\kappa_\tau(\mathsf{F}, Q_{Y^n})$ given by [2, Lemma 61] we show that for all $0 < \epsilon < 1$ and $0 < \tau < \epsilon$ there exists an (n, M, ϵ) code (maximal probability of error) with

$$M \geq \frac{1}{C_1} \frac{\tau - e^{-C_2 n}}{\beta_{1-\epsilon+\tau}(P_{Y^n|X^n=x}, Q_{Y^n})}, \qquad (92)$$

where $x = [\sqrt{P}, \ldots, \sqrt{P}]^n \in \mathbb{R}^n$ is a vector on the the power sphere, and C_1 and C_2 are some positive constants. We now take $\tau = \frac{\epsilon}{2}$ and apply the upper bound on β from (14) to obtain the statement: For any γ there exists and (n, M, ϵ) code (maximal probability of error) with

$$M \geq \frac{\epsilon - 2e^{-C_2 n}}{2C_1} \exp\{\gamma\} \qquad (93)$$

and

$$\epsilon = 2\mathbb{P}\left[\log \frac{dP_{Y^n|X^n=x}}{Q_{Y^n}} \leq \gamma\right]. \qquad (94)$$

Now take $\gamma_n = nC - \lambda n\rho_n$, where $\lambda < 1$ is arbitrary. By [6, Theorem 3.7.1] we obtain a sequence of codes with

$$\log M_n \geq nC - n\rho_n \qquad (95)$$

for all n sufficiently large and

$$\limsup_{n\to\infty} \frac{1}{n\rho_n^2} \log \epsilon_n \leq -\frac{\lambda^2}{2V(P)}. \qquad (96)$$

In particular,

$$\limsup_{n\to\infty} \frac{1}{n\rho_n^2} \log \epsilon^*(n, \exp\{nC - n\rho_n\}) \leq \frac{-\lambda^2}{2V(P)}, \quad (97)$$

and since $\lambda < 1$ is arbitrary we can take $\lambda \nearrow 1$ to finish the proof. ∎

REFERENCES

[1] Y. Altug and A. B. Wagner, "Moderate deviation analysis of channel coding: Discrete memoryless case," in *Proc. 2010 IEEE Int. Symp. Inf. Theory (ISIT)*, Austin, TX, USA, Jun. 2010.

[2] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[3] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, 1994.

[4] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Dispersion of the Gilbert-Elliott channel," *IEEE Trans. Inf. Theory*, to appear.

[5] S. Verdú, *EE528–Information Theory, Lecture Notes*. Princeton, NJ: Princeton Univ., 2007.

[6] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, 2nd ed. New York: Springer Verlag, 2009.

[7] A. N. Kolmogorov, "Über das Gesets des iterierten Logarithmus," *Math. Ann.*, vol. 101, pp. 126–135, 1929.

[8] L. V. Rozovsky, "Estimate from below for large-deviation probabilities of a sum of independent random variables with finite variances (in Russian)," *Zapiski Nauchn. Sem. POMI*, vol. 260, pp. 218–239, 1999.

[9] ——, "Estimate from below for large-deviation probabilities of a sum of independent random variables with finite variances," *J. Math. Sci.*, vol. 109, no. 6, pp. 2192–2209, May 2002.

[10] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, pp. 611–656, 1959.