The Dispersion of Nearest-Neighbor Decoding for Additive Non-Gaussian Channels

The Dispersion of Nearest-Neighbor Decoding for Additive Non-Gaussian Channels

Jonathan Scarlett
Laboratory for Information and Inference Systems,
École Polytechnique Fédérale de Lausanne,
Email: jmscarlett@gmail.com

Vincent Y. F. Tan
Department of ECE,
National University of Singapore
Email: vtan@nus.edu.sg

Giuseppe Durisi
Department of Signals and Systems,
Chalmers University of Technology
Email: durisi@chalmers.se

***Abstract*—We study the second-order asymptotics of information transmission using random Gaussian codebooks and nearest neighbor (NN) decoding over a power-limited additive stationary memoryless non-Gaussian channel. We show that the dispersion term depends on the non-Gaussian noise only through its second and fourth moments. We also characterize the second-order performance of point-to-point codes over Gaussian interference networks. Specifically, we assume that each user's codebook is Gaussian and that NN decoding is employed, i.e., that interference from unintended users is treated as noise at each decoder.**

I. Introduction

In second-order asymptotic analyses, a line of work pioneered by Strassen [1] and extended by Hayashi [2] and Polyanskiy, Poor and Verdú [3] among others, one seeks a two-term approximation of the maximum possible number of messages $M^*(n,\varepsilon)$ that can be transmitted with average probability of error no larger than ε using a stationary and memoryless channel n times. In the real AWGN case, under the assumption that the noise has unit variance and that the codewords have power P, it is by now well-known that $M^*(n,\varepsilon)$ can be approximated as [2], [3]

$$\log M^*(n,\varepsilon) = n\mathsf{C}(P) - \sqrt{n\mathsf{V}(P)}\mathrm{Q}^{-1}(\varepsilon) + O(\log n) \quad (1)$$

where $\mathrm{Q}(a) := (1/\sqrt{2\pi})\int_a^\infty e^{-t^2/2}\,\mathrm{d}t$ is the complementary cumulative density function (CDF) of a standard Gaussian and $\mathrm{Q}^{-1}(\cdot)$ its inverse, $\mathsf{C}(P) = (1/2)\log(1+P)$ [nats per channel use] is the channel capacity, and

$$\mathsf{V}(P) = \frac{P(P+2)}{2(P+1)^2}, \quad \text{[nats}^2 \text{ per channel use]} \quad (2)$$

is the channel dispersion. *Gaussian codebooks*, where each codeword is uniformly distributed on a sphere of radius $\sqrt{nP}$, achieve (1). Furthermore, the *nearest-neighbor* (NN) or *minimum distance* decoding rule is optimal.

A natural question then beckons. What is the maximum rate we can achieve if the codebook is constrained to be Gaussian and the decoder uses the NN rule, but the noise is non-Gaussian? This question is relevant in situations when one knows how to combat Gaussian noise, and seeks to adopt a design accordingly despite the inherent mismatch.

Lapidoth [4] provided a "first-order-asymptotics" answer to this question by showing that the Gaussian capacity $\mathsf{C}(P)$ is also the largest rate achievable over unit-variance non-Gaussian stationary ergodic additive channels when Gaussian codebooks and NN decoding are used. In some sense, this result says that Gaussian codebooks and NN decoding, although possibly suboptimal for the case of non-Gaussian noise, form a robust communication scheme.

In this paper, we extend Lapidoth's result in the direction of second-order asymptotics by determining the analogue of $\mathsf{V}(P)$ in (1) when we use a Gaussian codebooks and a NN decoder. Specifically, we show that when the non-Gaussian noise is i.i.d. and has a finite fourth moment ξ and a finite sixth moment, the dispersion with Gaussian codebooks and NN decoder is given by

$$\mathsf{V}(P,\xi) := \frac{P^2(\xi-1)+4P}{4(P+1)^2}. \quad (3)$$

This means that the rate of convergence to $\mathsf{C}(P)$ depends on the fourth moment (or, more generally, on the *kurtosis*, i.e., the ratio between the fourth-moment and the square of the second moment) of the noise distribution: the higher the kurtosis, the slower the convergence.

Motivated by work by Baccelli, El Gamal and Tse [5] on communication over interference networks with point-to-point codes, we also establish the dispersion for the scenario where additional noise arises from unintended users equipped with Gaussian codebooks, and the NN decoding rule is used at the intended receiver, with the interference treated as noise. For this case, whereas the first term in the second-order expansion is simply the Gaussian channel capacity $\mathsf{C}(\cdot)$ with P replaced by the signal-to-interference-and-noise ratio (SINR), the expression for the channel dispersion is more involved. In particular, it depends on the individual power of each interferer and not just on the total interference power.

Our analysis relies on the *Berry-Esseen theorem for functions of random vectors* [6], [7] (also known as the *multivariate delta method* in statistics), which has previously been used to obtain inner bounds to the second-order rate regions for Gaussian multipleaccess and interference channels [6], [8].

II. Point-To-Point Channels

Consider the point-to-point additive-noise channel

$$Y^n = X^n + Z^n, \quad (4)$$

where X^n is the input vector and Z^n is the noise vector over n scalar channel uses. Throughout, we shall focus exclusively on

Gaussian codebooks. More precisely, we consider *shell codes* for which X^n is uniformly distributed on a sphere with radius $\sqrt{nP}$, i.e.,

$$X^n \sim \mathrm{f}_{X^n}^{(\mathrm{shell})}(\mathbf{x}) := \delta(\|\mathbf{x}\|^2 - nP)/S_n(\sqrt{nP}). \quad (5)$$

Here, $\delta(\cdot)$ is the Dirac delta and $S_n(r) = 2\pi^{n/2} r^{n-1}/\Gamma(n/2)$ is the surface area of a radius-r sphere in $\mathbb{R}^n$. This random coding distribution is second- and third-order optimal for AWGN channels [9]. For comparison, we also consider *i.i.d. Gaussian codes*, in which each component of X^n is distributed according to a zero-mean, variance P normal distribution, i.e.,[1]

$$X^n \sim \mathrm{f}_{X^n}^{(\mathrm{iid})}(\mathbf{x}) := \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi P}} \exp\Big(-\frac{x_i^2}{2P}\Big). \quad (6)$$

This random coding distribution achieves $\mathsf{C}(P)$ but not $\mathsf{V}(P)$.

The noise Z^n is assumed to be a stationary and memoryless process that does not depend on the channel input:

$$Z^n \sim \mathrm{P}_{Z^n}(\mathbf{z}) = \prod_{i=1}^{n} \mathrm{P}_Z(z_i). \quad (7)$$

Hence, the nth extension of the channel is

$$\mathrm{P}_{Y^n|X^n}(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} \mathrm{P}_{Y|X}(y_i|x_i) = \prod_{i=1}^{n} \mathrm{P}_Z(y_i - x_i). \quad (8)$$

The distribution P_Z is non-Gaussian; the only assumptions are the following:

$$\mathbb{E}[Z^2] = 1, \quad \xi := \mathbb{E}[Z^4] < \infty, \quad \mathbb{E}[Z^6] < \infty. \quad (9)$$

The assumption that the second moment is unity is made for convenience. As we shall see, the assumption that the fourth moment of Z is finite is critical. The assumption that the sixth moment of Z is finite is made only for technical reasons.

Given either a shell or an i.i.d. codebook consisting of $M \in \mathbb{N}$ random codewords $\mathcal{C} := \{X^n(1), \dots, X^n(M)\}$, we consider an NN (or minimum distance) decoder that returns the message $\hat{W}$ whose corresponding codeword is closest in Euclidean distance to the channel output Y^n, i.e.,

$$\hat{W} := \operatorname*{arg\,min}_{w \in [1:M]} \|Y^n - X^n(w)\|. \quad (10)$$

This decoder is optimal if the noise is Gaussian, but may not be so in the more general setup considered here.

We also define the *average probability of error* as $\bar{p}_{\mathrm{e},n} := \Pr[\hat{W} \neq W]$. This probability is averaged over the uniformly distributed message W, the random codebook $\mathcal{C}$ and the channel noise Z^n. Note that in traditional channel-coding analyses [2], [3], the probability of error is averaged only over W and Z^n. Similar to [4], the additional averaging over the codebook $\mathcal{C}$ is required here to establish an ensemble converse for the Gaussian codebooks considered in this paper.

Let $M^*_{\mathrm{shell}}(n, \varepsilon, P; \mathrm{P}_Z)$ be the maximum number of messages that can be transmitted using a shell codebook over the channel (4) with average error probability no larger than $\varepsilon \in (0,1)$, when the noise is distributed according to P_Z. Let $M^*_{\mathrm{iid}}(n, \varepsilon, P; \mathrm{P}_Z)$ be the analogous quantity for the case of i.i.d. Gaussian codebooks. Lapidoth [4] showed that for all $\varepsilon \in (0,1)$ and $\dagger \in \{\mathrm{shell}, \mathrm{iid}\}$,

$$\lim_{n\to\infty} \frac{1}{n} \log M^*_{\dagger}(n, \varepsilon, P; \mathrm{P}_Z) = \mathsf{C}(P) \quad (11)$$

regardless of P_Z. Note that this result holds under milder conditions on the noise, not requiring the i.i.d. assumption.

Theorem 1. *Consider a noise distribution with statistics as in* (9)*. For shell codes,*

$$\log M^*_{\mathrm{shell}}(n, \varepsilon, P; \mathrm{P}_Z) = n\mathsf{C}(P) - \sqrt{n\mathsf{V}_{\mathrm{shell}}(P,\xi)}\,\mathrm{Q}^{-1}(\varepsilon) + O(\log n), \quad (12)$$

where the shell dispersion is

$$\mathsf{V}_{\mathrm{shell}}(P, \xi) := \frac{P^2(\xi - 1) + 4P}{4(P+1)^2}. \quad (13)$$

For i.i.d. codes,

$$\log M^*_{\mathrm{iid}}(n, \varepsilon, P; \mathrm{P}_Z) = n\mathsf{C}(P) - \sqrt{n\mathsf{V}_{\mathrm{iid}}(P,\xi)}\,\mathrm{Q}^{-1}(\varepsilon) + O(\log n), \quad (14)$$

where the i.i.d. dispersion is

$$\mathsf{V}_{\mathrm{iid}}(P, \xi) := \frac{P^2(\xi + 1) + 4P}{4(P+1)^2}. \quad (15)$$

A sketch of the proof is presented in Section IV.

The second-order terms in the asymptotic expansions of $\log M^*_{\mathrm{shell}}(n, \varepsilon, P; \mathrm{P}_Z)$ and $\log M^*_{\mathrm{iid}}(n, \varepsilon, P; \mathrm{P}_Z)$ depend on the distribution P_Z only through its second and fourth moments. If Z is standard Gaussian, then the fourth moment $\xi = 3$ and we recover from (13) the Gaussian dispersion (2). An expression of the same form as (2) was also derived by Shannon [10] in his study of the optimal asymptotic error probability of transmission over an AWGN channel at rates close to capacity. Comparing (13) with (2) we see that noise distributions P_Z with higher fourth moments than Gaussian (e.g., Laplace) result in a slower convergence to $\mathsf{C}(P)$. Conversely, P_Z with smaller fourth moment than Gaussian (e.g., Bernoulli) results in a faster convergence to $\mathsf{C}(P)$.

In the i.i.d. Gaussian case, if $Z \sim \mathcal{N}(0,1)$, we obtain

$$\mathsf{V}_{\mathrm{iid}}(P) = \frac{P}{P+1}. \quad (16)$$

An expression of the same form as (16) was derived by Rice [11], who used i.i.d. Gaussian codes to establish a lower bound on the error exponent for an AWGN channel at rates close to capacity. Note finally that $\mathsf{V}_{\mathrm{shell}}(P,\xi) \leq \mathsf{V}_{\mathrm{iid}}(P,\xi)$.

III. Interference Networks

We assume that K sender-receiver pairs operate concurrently over the same additive noise channel. Similarly to Section II, the additive noise Z^n is i.i.d. but possibly non-Gaussian. The senders use Gaussian codebooks with powers $\{P_j\}_{j=1}^{K}$ (as in Section II, we shall consider both the shell

[1] The quantity P in (5) can be interpreted as a power constraint in the sense of [3, Eq. (192)]; however, this does not hold for P in (6). There, P should instead be interpreted as the power averaged over the random codebook.

and i.i.d. cases) and all receivers use NN decoding. Hence, they treat the codewords from the unintended senders as additional noise. Note that for the case of shell codes, the resulting total additive noise is no longer i.i.d.. Although the communication strategy described above may be suboptimal rate-wise compared to more sophisticated strategies such as superposition or Han-Kobayashi coding [12], it is easy to implement since it relies exclusively on point-to-point channel codes [5]. Without loss of generality, we shall focus only on the signal at receiver 1, which is given by

$$Y^n = X_1^n + X_2^n + \ldots + X_K^n + Z^n. \tag{17}$$

Here, X_j^n denotes the codeword transmitted by the jth sender; X_1^n is the codeword intended for receiver 1. We consider two different models for the codewords. In the first, each codeword X_j^n follows a *shell distribution* as in (5) with power P_j. In the second, each codeword X_j^n follows an *i.i.d. Gaussian distribution* as in (6) with power P_j. We define $M^*_{\mathrm{shell}}(n, \varepsilon, \{P_j\}_{j=1}^K; \mathrm{P}_Z)$ to be the maximum number of messages that sender 1 can transmit using shell codes with error probability no larger than $\varepsilon \in (0,1)$ over the channel (17), when the receiver uses NN decoding, and, hence, treats $X_2^n, \ldots, X_K^n$ as noise. Similarly, let $M^*_{\mathrm{iid}}(n, \varepsilon, \{P_j\}_{j=1}^K; \mathrm{P}_Z)$ be the analogous quantity for the case of i.i.d. Gaussian codebooks. Let the SINR of the channel from sender 1 to receiver 1 be $\bar{P} := P_1/(1+\widetilde{P})$, where the total power of the interfering codewords is $\widetilde{P} := \sum_{j=2}^K P_j$.

In Theorem 2 below, we provide the second-order term in the asymptotic expansion of $\log M^*_\dagger(n, \varepsilon, \{P_j\}_{j=1}^K; \mathrm{P}_Z)$.

Theorem 2. *Consider a noise distribution with statistics as in* (9)*. For shell codes,*

$$\begin{aligned}&\log M^*_{\mathrm{shell}}(n, \varepsilon, \{P_j\}_{j=1}^K; \mathrm{P}_Z)\\ &= n\mathsf{C}(\bar{P}) - \sqrt{n\mathsf{V}'_{\mathrm{shell}}(\{P_j\}_{j=1}^K, \xi)}\,\mathrm{Q}^{-1}(\varepsilon) + O(\log n),\end{aligned} \tag{18}$$

where the shell dispersion is

$$\begin{aligned}&\mathsf{V}'_{\mathrm{shell}}(\{P_j\}_{j=1}^K, \xi)\\ &:= \frac{P_1^2(\xi - 1 + 4\widetilde{P}) + 4P_1(1+\widetilde{P})^3 + 4P_1^2 \sum_{2\le i<j\le K} P_i P_j}{4(1+\widetilde{P})^2(1+P_1+\widetilde{P})^2}.\end{aligned} \tag{19}$$

For i.i.d. codes,

$$\begin{aligned}&\log M^*_{\mathrm{iid}}(n, \varepsilon, \{P_j\}_{j=1}^K; \mathrm{P}_Z)\\ &\quad= n\mathsf{C}(\bar{P}) - \sqrt{n\mathsf{V}_{\mathrm{iid}}(\bar{P}, \xi')}\,\mathrm{Q}^{-1}(\varepsilon) + O(\log n),\end{aligned} \tag{20}$$

where $\mathsf{V}_{\mathrm{iid}}(\cdot,\cdot)$ *is defined in* (15) *and* $\xi' := \frac{3\widetilde{P}^2 + 6\widetilde{P} + \xi}{(\widetilde{P}+1)^2}$.

The proof of this result can be found in [13] and is omitted due to space constraints.

A few comments are now in order. First, as expected, the first-order term in the asymptotic expansion is $\mathsf{C}(\bar{P})$, where $\bar{P}$ is the SINR. The second-order term for the i.i.d. case can be obtained straightforwardly from Theorem 1. This is because the effective noise is now $X_2^n + \ldots + X_K^n + Z^n$, which is also i.i.d. Gaussian, but of variance $\widetilde{P} + 1$ instead of 1. The second-order term for the shell case does not follow directly from Theorem 1, as the total noise $X_2^n + \ldots + X_K^n + Z^n$, which is the sum of $K-1$ shell random vectors and a single i.i.d. Gaussian random vector, is neither i.i.d. nor shell distributed.

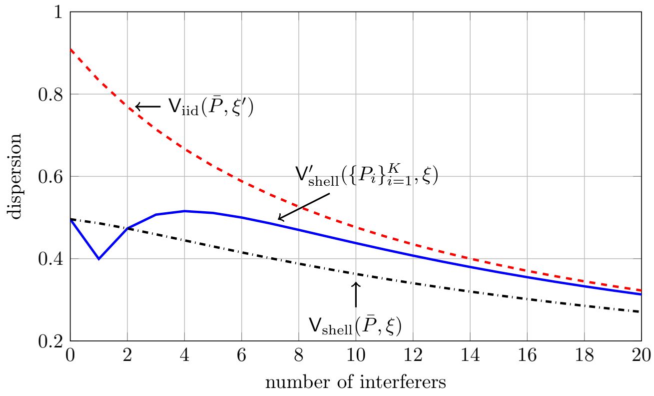

Fig. 1. Dispersion as a function of the number of interferers $K-1$ for both shell and i.i.d. codes. Here, $P_1 = 10$, $P_2 = \cdots = P_K = 1$ and $\bar{P} = P_1/(1+\sum_{i=2}^K P_i)$. The noise Z^n is i.i.d. Gaussian with unit variance.

Second, observe that the cross term in (19), namely $\sum_{2\le i<j\le K} P_i P_j = (\sum_{j=2}^K P_j)^2 - \sum_{j=2}^K P_j^2$ implies that the dispersion does not only depend on the sum $\widetilde{P}$ of the interferers' powers, but also on the individual power P_j of each interferer. This phenomenon is not present in the i.i.d. case. Since, by convexity, $\sum_{j=2}^K P_j^2 \le \widetilde{P}^2/(K-1)$, we conclude that the shell dispersion is maximized when all interferers have the same power. Through standard manipulation one can also show that $\mathsf{V}'_{\mathrm{shell}}(\{P_j\}_{j=1}^K, \xi) \le \mathsf{V}_{\mathrm{iid}}(\bar{P}, \xi')$ for fixed $\{P_j\}_{j=1}^K$.

In Fig. 1, we plot both $\mathsf{V}'_{\mathrm{shell}}(\{P_j\}_{j=1}^K, \xi)$ and $\mathsf{V}_{\mathrm{iid}}(\bar{P}, \xi')$ for the case $P_1 = 10$, $P_2 = \cdots = P_K = 1$. The noise Z^n is i.i.d. Gaussian with unit variance ($\xi = 3$). For comparison, we also plot $\mathsf{V}_{\mathrm{shell}}(\bar{P}, \xi)$, which corresponds to the dispersion for the case when the intended sender uses a shell code, whereas the interferers use i.i.d. codes. The resulting dispersion, which follows directly from Theorem 1, takes a particularly simple form: the Gaussian dispersion (2) computed at the SINR $\bar{P}$. As illustrated in Fig. 1, there is no ordering between $\mathsf{V}_{\mathrm{shell}}(\bar{P}, \xi)$ and $\mathsf{V}'_{\mathrm{shell}}(\{P_j\}_{j=1}^K, \xi)$. In this example, shell interference is preferable with a single interferer, but i.i.d. interference is preferable when there are three or more users.

IV. Proof of (12) in Theorem 1

Because of space constraints we will only provide the proof of (12). Our analysis makes use of the "mismatched" information density

$$\tilde{\imath}(x,y) := \log \frac{\mathcal{N}(y; x, 1)}{\mathcal{N}(y; 0, P+1)} = \mathsf{C}(P) + \frac{y^2}{2(P+1)} - \frac{(y-x)^2}{2}. \tag{21}$$

This is the information density of the Gaussian channel $\mathcal{N}(y; x, 1)$; indeed, the denominator in (21) is its capacity-achieving output distribution $\mathcal{N}(y; 0, P+1)$. We set $\tilde{\imath}^n(\mathbf{x}, \mathbf{y}) := \sum_{i=1}^n \tilde{\imath}(x_i, y_i)$.

A. Proof of the Direct Part of (12)

By (21), the NN rule is equivalent to maximizing $\tilde{\imath}^n(\mathbf{x}(w),\mathbf{y})$ over w. Hence by the random coding union bound [3], the ensemble error probability can be bounded as

$$\bar{p}_{\mathrm{e},n} \le \mathbb{E}\big[\min\{1, M\Pr(\tilde{\imath}^n(\bar{X}^n,Y) \ge \tilde{\imath}^n(X^n,Y)|X^n,Y^n)\}\big], \quad (22)$$

where $(\bar{X}^n, X^n, Y^n) \sim \mathrm{f}^{(\mathrm{shell})}_{X^n}(\bar{\mathbf{x}})\mathrm{f}^{(\mathrm{shell})}_{X^n}(\mathbf{x})\mathrm{P}_{Y^n|X^n}(\mathbf{y}|\mathbf{x})$. Let $g(t,\mathbf{y}) := \Pr\big[\tilde{\imath}^n(\bar{X}^n,\mathbf{y}) \ge t\big]$. With this definition,

$$\bar{p}_{\mathrm{e},n} \le \mathbb{E}\big[\min\big\{1, Mg(\tilde{\imath}^n(X^n,Y^n),Y^n)\big\}\big]. \quad (23)$$

Using [9, Eq. (58)], $g(t,\mathbf{y})$ can be bounded as follows

$$g(t,\mathbf{y}) \le \frac{K_0 e^{-t}}{\sqrt{n}} \le K_0 e^{-t}, \quad (24)$$

where K_0 is a finite constant independent of $\mathbf{y}$.

Substituting (24) into (23), using the fact that for every real-valued random variable J and every positive integer n $\mathbb{E}\big[\min\{1,J\}\big] \le \Pr[J > 1/\sqrt{n}] + 1/\sqrt{n}$, and after some algebra, we can further upper-bound the right-hand-side of (23) as follows:

$$\bar{p}_{\mathrm{e},n} \le \Pr\Big[P\|Z^n\|^2 - nP - 2\langle X^n, Z^n\rangle \ge 2(P+1)\Big(n\mathsf{C}(P) - \log M - \log(K_0\sqrt{n})\Big)\Big] + \frac{1}{\sqrt{n}}. \quad (25)$$

Here, we also used the fact that $\|X^n\|^2 = nP$ almost surely. Now we make use of the Berry-Esseen Theorem for functions of random vectors (see [6, Prop. 1] and [7, Prop. 1]), by proceeding similarly as in [6, Sec. IV.D]. A shell codeword can be written as

$$X^n = \sqrt{nP}\tilde{X}^n / \|\tilde{X}^n\|, \quad (26)$$

where $\tilde{X}^n \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$. We may write $P\|Z^n\|^2 - nP - 2\langle X^n, Z^n\rangle$ in terms of the i.i.d. random variables

$$A_{1i} := 1 - Z_i^2, \quad A_{2i} := \sqrt{P}\tilde{X}_i Z_i, \quad A_{3i} := \tilde{X}_i^2 - 1, \quad (27)$$

$i = 1, \dots, n$, and of the smooth function

$$f(a_1, a_2, a_3) = Pa_1 + 2a_2/\sqrt{1+a_3}. \quad (28)$$

It is easy to verify that $-nf(\frac{1}{n}\sum_{i=1}^n [A_{1i}\ A_{2i}\ A_{3i}]) = P\|Z^n\|^2 - nP - 2\langle X^n, Z^n\rangle$. Now, the Jacobian matrix of f evaluated at $\mathbf{0}$ is $\mathbf{J} = [P\ 2\ 0]$ and the covariance matrix $\mathbf{V}$ of the vector $[A_{11}\ A_{21}\ A_{31}]$ is $\mathbf{V} = \mathrm{diag}([\xi - 1\ P\ 2]^T)$. Hence, it follows from [6, Prop. 1] and [7, Prop. 1] that $(P\|Z^n\|^2 - nP - 2\langle X^n, Z^n\rangle)/\sqrt{n}$ converges in distribution to a zero-mean normal random variable with variance $\mathbf{J}\mathbf{V}\mathbf{J}^T = P^2(\xi-1) + 4P$. Thus, the probability in (25) is upper-bounded by [7, Prop. 1]

$$\mathrm{Q}\left(\frac{2(P+1)\big(n\mathsf{C}(P) - \log M - \log(K_0\sqrt{n})\big)}{\sqrt{n(P^2(\xi-1)+4P)}}\right) + O\Big(\frac{1}{\sqrt{n}}\Big). \quad (29)$$

Equating this to ε, solving for $\log M$, and Taylor-expanding $\mathrm{Q}^{-1}(\cdot)$, we establish the desired lower bound to (13).

B. Proof of the Ensemble Tightness Part of (12)

Since the probability of ties for the NN rule in (10) is zero, the exact random coding probability can be written as

$$\bar{p}_{\mathrm{e},n} = \mathbb{E}[\bar{p}_{\mathrm{e},n}(X^n, Y^n)] \quad (30)$$

where $(X^n, \bar{X}^n, Y^n)$ are distributed as in Section IV-A, and

$$\bar{p}_{\mathrm{e},n}(\mathbf{x},\mathbf{y}) = 1 - \big(1 - \Pr[\|\mathbf{y} - \bar{X}^n\| \le \|\mathbf{y} - \mathbf{x}\|]\big)^{M-1}. \quad (31)$$

Let $\mathbf{z} := \mathbf{y} - \mathbf{x}$. By symmetry (see [4]), the probability in (31) depends on $(\mathbf{x},\mathbf{y})$ only through the powers $\hat{P}_Y = \|\mathbf{y}\|^2/n$ and $\hat{P}_Z = \|\mathbf{z}\|^2/n$. We denote the inner probability in (31) as $\Psi(\hat{P}_Y, \hat{P}_Z)$. Because of (11), we shall assume without loss of generality that $\liminf_{n\to\infty}(\log M)/n > 0$. Let $c_z := \xi - 1$ and $c_y := \xi - 1 + 4P$. Let $\eta > 0$ be a small constant. We define the typical sets

$$\mathcal{P}_Y := \left\{\hat{p}_Y \in \mathbb{R} : |\hat{p}_Y - (P+1)| \le \sqrt{c_z(\log n)/n}\right\}, \quad (32)$$

$$\mathcal{P}_Z := \left\{\hat{p}_Z \in \mathbb{R} : |\hat{p}_Z - 1| \le \sqrt{c_y(\log n)/n}\right\}. \quad (33)$$

$$\mathcal{Q} := \big\{(\hat{p}_Y, \hat{p}_Z) \in \mathbb{R} : \hat{p}_Y + P - \hat{p}_Z > \eta\big\}, \quad (34)$$

and let $\mathcal{T} := (\mathcal{P}_Y \times \mathcal{P}_Z) \cap \mathcal{Q}$. Using [7, Prop. 1] and the assumption that $\mathbb{E}[Z^6]$ is finite, we conclude that $\Pr[(\|Y^n\|^2/n, \|Z^n\|^2/n) \notin \mathcal{T}] = O(1/\sqrt{n})$ (see [13]).

Now consider an arbitrary pair of powers $(\hat{P}_Y, \hat{P}_Z)$ such that $\Psi(\hat{P}_Y, \hat{P}_Z) \ge n/(M-1)$. In this case

$$\begin{aligned}
1 - \Big(1 - \Psi(\hat{P}_Y, \hat{P}_Z)\Big)^{M-1} &\ge 1 - \Big(1 - \frac{n}{M-1}\Big)^{M-1} \\
&= 1 - \bigg(\Big(1 - \frac{n}{M-1}\Big)^{\frac{M-1}{n}}\bigg)^n = 1 - \big(e^{-1}(1+o(1))\big)^n \\
&= 1 - e^{-n\gamma} \quad (35)
\end{aligned}$$

where we used that $(1-\zeta^{-1})^\zeta \to e^{-1}$ as $\zeta \to \infty$ and the last line holds for any $\gamma \in (0,1)$ provided that n is large enough.

Combining the analyses in the previous two paragraphs, we obtain from (31) that

$$\bar{p}_{\mathrm{e},n} \ge \Pr\bigg[\Psi(\hat{P}_Y, \hat{P}_Z) \ge \frac{n}{M-1} \cap (\hat{P}_Y, \hat{P}_Z) \in \mathcal{T}\bigg] + O\Big(1/\sqrt{n}\Big). \quad (36)$$

Ensemble tightness is established if we can show that for all typical powers $(\hat{P}_Y, \hat{P}_Z) \in \mathcal{T}$,

$$\Psi(\hat{P}_Y, \hat{P}_Z) \ge q(n)\exp\bigg(-n\Big(\mathsf{C}(P) + \frac{\hat{P}_Y}{2(P+1)} - \frac{\hat{P}_Z}{2}\Big)\bigg) \quad (37)$$

for some $q(n)$ satisfying $\log q(n) = O(\log n)$. This is because we can then further lower-bound (36) as

$$\begin{aligned}
\bar{p}_{\mathrm{e},n} &\ge \Pr\bigg[-n\Big(\mathsf{C}(P) + \frac{\hat{P}_Y}{2(P+1)} - \frac{\hat{P}_Z}{2}\Big) \ge -\log M + O(\log n) \\
&\qquad \cap (\hat{P}_Y, \hat{P}_Z) \in \mathcal{T}\bigg] + O\Big(\frac{1}{\sqrt{n}}\Big) \quad (38) \\
&= \Pr\bigg[\mathsf{C}(P) + \frac{\hat{P}_Y}{2(P+1)} - \frac{\hat{P}_Z}{2} \le \frac{\log M}{n} + O\Big(\frac{\log n}{n}\Big)\bigg] \\
&\qquad + O\Big(\frac{1}{\sqrt{n}}\Big) \quad (39)
\end{aligned}$$

where the last step follows from the fact that $\Pr[(\|Y^n\|^2/n, \|Z^n\|^2/n) \notin \mathcal{T}] = O(1/\sqrt{n})$. To complete the proof (*sans* the justification of (37)), we recall the definitions of $\hat{P}_Y$ and $\hat{P}_Z$, and follow steps (25)–(29) in the direct part, which are all tight in the second-order sense.

Now we prove (37). By the definitions of Ψ and the powers $\hat{P}_Y$ and $\hat{P}_Z$, and the fact that $\|X^n\|^2 = nP$, we have

$$\Psi(\hat{P}_Y, \hat{P}_Z) = \Pr\Big[2\langle \bar{X}^n, \mathbf{y}\rangle \ge n\hat{P}_Y + nP - n\hat{P}_Z\Big] \quad (40)$$

for any $\mathbf{y}$ with $\|\mathbf{y}\|_2^2 = n\hat{P}_Y$. We choose $\mathbf{y} = (\sqrt{n\hat{P}_Y}, 0, \dots, 0)$, from which we obtain for any $\delta > 0$ that

$$\Psi(\hat{P}_Y, \hat{P}_Z) = \Pr\left[\bar{X}_1 \ge \frac{n\hat{P}_Y + nP - n\hat{P}_Z}{2\sqrt{n\hat{P}_Y}}\right] \quad (41)$$

$$\ge \Pr\left[\frac{n\hat{P}_Y + nP - n\hat{P}_Z}{2\sqrt{n\hat{P}_Y}} \le \bar{X}_1 \le \frac{n\hat{P}_Y + nP - n\hat{P}_Z}{2\sqrt{n\hat{P}_Y}}\Big(1 + \frac{\delta}{n}\Big)\right] \quad (42)$$

$$\ge \frac{\delta(\hat{P}_Y + P - \hat{P}_Z)}{2n\sqrt{n\hat{P}_Y}} \min_x f_{\bar{X}_1}(x), \quad (43)$$

where $\bar{X}_1$ is the first symbol in the vector $\bar{X}^n$, $f_{\bar{X}_1}$ is its probability density function, which is given by [14, Eq. (4)]

$$f_{\bar{X}_1}(x) = \frac{1}{\sqrt{\pi n P}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}\left(1 - \frac{x^2}{nP}\right)^{\frac{n-3}{2}} \mathbf{1}\{x^2 \le nP\}, \quad (44)$$

and the minimization in (43) is over the interval in (42). Note that since the right-hand side of (44) is a decreasing function of $|x|$, we can further lower-bound (43) (for sufficiently large n) by replacing x with the right-hand term in (42). Since $\hat{P}_Y$ and $\hat{P}_Z$ are bounded within the typical sets in (32)–(33), we have that for all sufficiently large n and for some constant K_1 (depending only on P and δ),

$$\left(\frac{n\hat{P}_Y + nP - n\hat{P}_Z}{2\sqrt{n\hat{P}_Y}}\Big(1 + \frac{\delta}{n}\Big)\right)^2 \le \frac{n(\hat{P}_Y + P - \hat{P}_Z)^2}{4\hat{P}_Y} + K_1. \quad (45)$$

Moreover, $\Gamma(\frac{n}{2})/\Gamma(\frac{n-1}{2})$ behaves as $\Theta(\sqrt{n})$, and thus the prefactors in (43) and (44) (which is bounded away from zero by (34)) can be combined into a single prefactor $p_0'(n)$ satisfying $\log p_0'(n) = O(\log n)$:

$$\Psi(\hat{P}_Y, \hat{P}_Z)$$

$$\ge p_0'(n)\left(1 - \frac{(\hat{P}_Y + P - \hat{P}_Z)^2}{4P\hat{P}_Y} - \frac{K_1}{nP}\right)^{\frac{n-3}{2}} \quad (46)$$

$$= p_0'(n)\left(\left(1 - \frac{(\hat{P}_Y + P - \hat{P}_Z)^2}{4P\hat{P}_Y}\right)\left(1 - \frac{K_2}{nP}\right)\right)^{\frac{n-3}{2}} \quad (47)$$

$$\ge p_0''(n)\left(1 - \frac{(\hat{P}_Y + P - \hat{P}_Z)^2}{4P\hat{P}_Y}\right)^{\frac{n}{2}} \quad (48)$$

$$= p_0''(n)\exp\left(\frac{n}{2}\log\left(1 - \frac{(\hat{P}_Y + P - \hat{P}_Z)^2}{4P\hat{P}_Y}\right)\right) \quad (49)$$

where (47) holds for some finite constant K_2, and (48) follows by factoring further terms into the prefactor and calling the result $p_0''(n)$, which still satisfies $\log p_0''(n) = O(\log n)$. In particular, this step uses that $\lim_{n\to\infty}\big(1 - K_2/(nP)\big)^{n/2} = e^{-K_2/(2P)}$, which is a constant.

We prove (37) by performing a Taylor expansion of $\frac{1}{2}\log\Big(1 - \frac{(\hat{P}_Y + P - \hat{P}_Z)^2}{4P\hat{P}_Y}\Big)$ about $(\hat{P}_Y, \hat{P}_Z) = (1+P, 1)$:

$$-\frac{1}{2}\log\Big(1 - \frac{(\hat{P}_Y + P - \hat{P}_Z)^2}{4P\hat{P}_Y}\Big)$$

$$= \frac{1}{2}\log(1+P) + \frac{1}{2(1+P)}\Big(\hat{P}_Y - (1+P)\Big) - \frac{1}{2}\Big(\hat{P}_Z - 1\Big) + O\Big(\big|\hat{P}_Y - (1+P)\big|^2 + \big|\hat{P}_Z - 1\big|^2\Big) \quad (50)$$

$$= \frac{1}{2}\log(1+P) + \frac{\hat{P}_Y}{2(1+P)} - \frac{\hat{P}_Z}{2} + O\Big(\frac{\log n}{n}\Big). \quad (51)$$

Here, the remainder term is $O\big(\frac{\log n}{n}\big)$ due to the definitions of the typical sets in (32)–(33). This remainder term can be factored into the prefactor in (49), yielding $q(n)$ in (37). The proof of (37) is thus complete.

Acknowledgments: The authors would like to thank Albert Guillén i Fàbregas, Alfonso Martinez, Yury Polyanskiy and Wei Yang for useful discussions. VT is supported in part by a Ministry of Education Tier 2 grant (R-263-000-B61-112). GD is supported by the Swedish Research Council (VR) under grant no. 2012-4571. The support of the Swedish foundation for international cooperation in research and higher education (STINT, grant 2013-019) is gratefully acknowledged.

REFERENCES

[1] V. Strassen. Asymptotische Abschätzungen in Shannons Informationstheorie. In *Trans. Third Prague Conf. Inf. Theory*, pages 689–723, Prague, 1962. http://www.math.cornell.edu/∼pmlut/strassen.pdf.

[2] M. Hayashi. Information spectrum approach to second-order coding rate in channel coding. *IEEE Trans. on Inf. Th.*, 55(11):4947–4966, 2009.

[3] Y. Polyanskiy, H. V. Poor, and S. Verdú. Channel coding rate in the finite blocklength regime. *IEEE Trans. on Inf. Th.*, 56(5):2307–2359, 2010.

[4] A. Lapidoth. Nearest neighbor decoding for additive non-Gaussian noise channels. *IEEE Trans. on Inf. Th.*, 42(5):1520–1529, 1996.

[5] F. Baccelli, A. El Gamal, and D. N. C. Tse. Interference networks with point-to-point codes. *IEEE Trans. on Inf. Th.*, 57(5):2582–2596, 2011.

[6] E. MolavianJazi and J. N. Laneman. A second-order achievable rate region for Gaussian multi-access channels via a central limit theorem for functions. *IEEE Trans. on Inf. Th.*, 61(12):6719–6733, Dec 2015.

[7] N. Iri and O. Kosut. Third-order coding rate for universal compression of Markov sources. In *Proc. of ISIT*, pages 1996–2000, 2015.

[8] S.-Q. Le, V. Y. F. Tan, and M. Motani. A case where interference does not affect the channel dispersion. *IEEE Trans. on Inf. Th.*, 61(5):2439–2453, 2015.

[9] V. Y. F. Tan and M. Tomamichel. The third-order term in the normal approximation for the AWGN channel. *IEEE Trans. on Inf. Th.*, 61(5):2430–2438, 2015.

[10] C. E. Shannon. Probability of error for optimal codes in a Gaussian channel. *The Bell System Technical Journal*, 38:611–656, 1959.

[11] S. O. Rice. Communication in the presence of noise–probability of error for two encoding schemes. *The Bell System Technical Journal*, 29:60, 1950.

[12] A. El Gamal and Y.-H. Kim. *Network Information Theory*. Cambridge University Press, Cambridge, U.K., 2012.

[13] J. Scarlett, V. Y. F. Tan, and G. Durisi. The dispersion of nearest-neighbor decoding for additive non-Gaussian channels. arXiv:1512.06618 [cs.IT], Dec 2015.

[14] A. J. Stam. Limit theorems for uniform distributions on spheres in high-dimensional Euclidean spaces. *Journal of Applied Probability*, 19(1):221–228, 1982.