A New Upper Bound on the Reliability Function of the Gaussian Channel

Alexei E. Ashikhmin, *Member, IEEE,* Alexander Barg, and Simon N. Litsyn, *Senior Member, IEEE*

***Abstract*—We derive a new upper bound on the exponent of error probability of decoding for the best possible codes in the Gaussian channel. This bound is tighter than the known upper bounds (the sphere-packing and minimum-distance bounds proved in Shannon's classical 1959 paper and their low-rate improvement by Kabatiansky and Levenshtein). The proof is accomplished by studying asymptotic properties of codes on the sphere $S^{n-1}(\mathbb{R})$. First we prove a general lower bound on the distance distribution of codes of large size. To derive specific estimates of the distance distribution, we study the asymptotic behavior of Jacobi polynomials $P_k^{ak,\,bk}$ as $k \to \infty$.**

Since on the average there are many code vectors in the vicinity of the transmitted vector $\boldsymbol{x}$, one can show that the probability of confusing $\boldsymbol{x}$ and one of these vectors cannot be too small. This proves a lower bound on the error probability of decoding and the upper bound announced in the title.

***Index Terms*—Distance distribution, error probability of decoding, Jacobi polynomials, spherical codes.**

I. Introduction

The classical model of communication over channels with noise, introduced by Shannon in 1948, assumes that messages are represented by vectors (points) in the n-dimensional Euclidean space. Under this model it is assumed that when a vector $\boldsymbol{x}$ is sent over the channel, the received signal is represented by a vector $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y}$, where $\boldsymbol{y}$ is a vector whose coordinates are independent Gaussian variables with mean zero and variance σ^2.

A consistent definition of capacity of such a channel is obtained if one assumes that the input signals satisfy some sort of energy constraints. Typically, one assumes that the energy, or the average energy, of input signals does not exceed a given number $A\sigma^2$ per dimension, where A is a positive number called the "signal-to-noise ratio." Shannon [31] has shown that for a set of input signals of sufficiently large size the study of the channel is reduced to considering signals of *constant* energy equal to $\sigma\sqrt{An}$, that is, points on the sphere of radius $\sigma\sqrt{An}$ in $\mathbb{R}^n$. In this paper we restrict ourselves to this communication model, which will be referred to as the Gaussian channel (with discrete time and continuous amplitude). It suffices to consider spheres of any fixed radius. Therefore, in a large part of the paper we study only codes on the *unit* sphere in $\mathbb{R}^n$, denoted by $S^{n-1} = S^{n-1}(\mathbb{R})$. A *code* is a finite subset of $S^{n-1}(\mathbb{R})$. To distinguish between codes on S^{n-1} and in the Hamming space, the former are often called spherical codes.

Analogously to the Hamming case, the most important parameters of spherical codes studied in geometry and coding and information theory are the minimum distance and the error probability of decoding as functions of the code size. A natural geometric motivation for the distance problem is studying the best possible packings of $S^{n-1}(\mathbb{R})$ with spherical caps. This and closely related problems of finding the best possible fillings of $\mathbb{R}^n$ with identical spheres and the kissing number were studied long before the emergence of coding theory (see a survey in [13]). Spherical codes in information theory were introduced by Slepian [34] (paper based on a 1951 Bell Labs report) and Shannon [31]. However, studies in geometry and coding theory developed independently of each other until the second half of the 1970s when important unifying steps were taken by Delsarte, Goethals, and Seidel [15], [16], and Kabatiansky and Levenshtein [21].

A. Parameters of Spherical Codes

Let $C \in S^{n-1}$ be a code and $R(C) = \frac{1}{n} \ln |C|$ its rate, $0 \le R(C) < \infty$. The distance $\operatorname{dist}(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|$ between two points in S^{n-1} can be also measured by the inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 1 - \frac{1}{2} \operatorname{dist}^2(\boldsymbol{x}, \boldsymbol{y})$ or by the geodesic distance on the sphere $\theta = \arccos \langle \boldsymbol{x}, \boldsymbol{y} \rangle$. Each of these measures is convenient in some coding-theoretic problems. Accordingly, let

$$d(C) := \min_{\substack{\boldsymbol{x}, \boldsymbol{y} \in C \\ \boldsymbol{x} \ne \boldsymbol{y}}} \operatorname{dist}(\boldsymbol{x}, \boldsymbol{y}) \qquad (0 \le d(C) \le 2)$$

$t(C) = \max \langle \boldsymbol{x}, \boldsymbol{y} \rangle = 1 - \frac{1}{2} d^2(C)$, and $\theta(C) = \arccos t(C)$.

Let

$$M(n, \theta) = \max_{C:\, \theta(C) \ge \theta} |C|$$

be the maximum size of a code on S^{n-1} of angular distance θ. Asymptotic properties of codes are characterized by the functions

$$\overline{R}(\theta) = \limsup_{n \to \infty} \frac{1}{n} \ln M(n, \theta)$$

$$\underline{R}(\theta) = \liminf_{n \to \infty} \frac{1}{n} \ln M(n, \theta).$$

Manuscript received February 9, 1999; revised March 27, 2000. This work was performed while S. N. Litsyn was visiting DIMACS Center, Rutgers University, Piscataway, NJ 08854 USA. The material in this paper was presented in part at the IEEE Information Theory Workshop, Kruger National Park, South Africa, June 20–25, 1999.

A. E. Ashikhmin was with Los Alamos National Laboratory, Los Alamos, NM 87545 USA. He is now with Bell Laboratories, Lucent Technologies, Rm. 2C-179, Murray Hill, NJ 07974 USA (e-mail: aea@research.bell-labs.com).

A. Barg is with Bell Laboratories, Lucent Technologies, Rm. 2C-375, Murray Hill, NJ 07974 USA (e-mail: abarg@research.bell-labs.com).

S. N. Litsyn is with the Department of Electrical Engineering–Systems, Tel-Aviv University, Ramat-Aviv 69978, Israel (e-mail: litsyn@eng.tau.ac.il).

Communicated by T. E. Fuja, Associate Editor At Large.

Publisher Item Identifier S 0018-9448(00)06993-5.

Usually one is interested in upper bounds on $\overline{R}(\theta)$ and lower bounds on $\underline{R}(\theta)$. Below we assume that these two sequences have a common limit and speak loosely of the maximum possible rate $R(\theta)$ of a code (sequence of codes C_n) of angular distance θ. Note that by [30], $R(\theta)=0$ for $\pi/2 \le \theta \le \pi$.

Likewise, define $d(R)$ to be the maximum distance of a code (sequence of codes) of rate $R(C_n) \ge R$.

Remark: Following the discrete case [1], it is not difficult to prove that $\overline{R}(\theta)$ is a continuous function of θ. Indeed, for any $\theta' > \theta$

$$(1-\cos\theta)^{(n-1)/2} M(n,\theta) \le \sqrt{2\pi n}(1-\cos\theta')^{(n-1)/2} M(n+1,\theta')$$

(Yaglom's inequality, see, for instance, [13].) Let ϵ be some small number such that $0 < \theta < \theta+\epsilon < \pi/2$. Apply Yaglom's inequality $\sqrt{n}$ times, putting each time $\theta' = \theta + \epsilon/\sqrt{n}$. We obtain

$$\frac{\ln M(n,\theta)}{n} \le \frac{1+n^{-1/2}}{2} \ln\left(\frac{1-\cos(\theta+\epsilon)}{1-\cos\theta}\right) + (1+n^{-1/2}) \frac{\ln M(n+n^{1/2}, \theta+\epsilon)}{n+n^{1/2}} + O\left(\frac{\ln n}{\sqrt{n}}\right).$$

Since $\overline{R}(\theta)$ is monotone, we have $\overline{R}(\theta) - \overline{R}(\theta+\epsilon) \ge 0$. Letting $\epsilon \to 0$ and $n \to \infty$, we see that $\overline{R}(\theta)$ is continuous.

The best known lower bound on $R(\theta)$ is Shannon's sphere-packing bound [31]

$$R(\theta) \ge -\ln \sin\theta \qquad (0 < \theta \le \pi/2) \qquad (1)$$

$$d^2(R) \ge 2(1-\sqrt{1-e^{-2R}}) \qquad (0 < R < \infty) \qquad (2)$$

which relies on the same type of argument that the Varshamov–Gilbert bound for the Hamming space.

Best known upper bounds on $R(\theta)$ were derived in [21]. One of the main results of [21] states that

$$R(\theta) \le \frac{1+\sin\theta}{2\sin\theta} H\left(\frac{1-\sin\theta}{1+\sin\theta}\right) \qquad (0 < \theta \le \pi/2) \quad (3)$$

where $H(x) = -x\ln x - (1-x)\ln(1-x)$ is the entropy function. This bound admits a small asymptotic improvement for $0 < \theta < 63^\circ$ [21]

$$R(\theta) \le -(1/2)\ln(1-\cos\theta) - 0.0686$$

(the real numbers here and below are approximate). These bounds can be also transformed to relate R and d (rather than θ). Indeed, let $\rho_{kl} = \rho_{kl}(R)$ be the root of the equation

$$R = (1+\rho) H\left(\frac{\rho}{1+\rho}\right) \qquad (4)$$

$\rho_{kl} = (1/2)(1-\sin\theta)/\sin\theta, 0 \le \rho_{kl} < \infty$. Then

$$d^2(R) \le d^2_{kl}(R) := \frac{2\left(\sqrt{1+\rho_{kl}} - \sqrt{\rho_{kl}}\right)^2}{1+2\rho_{kl}} \qquad (5)$$

$$d^2(R) \le 2e^{-2R-0.137}. \qquad (6)$$

For $0.234 < R < \infty$ bound (6) is better than (5). Some further details on the upper bounds will be provided in the next sections.

B. Error Probability of Decoding

A systematic study of the error probability of decoding for spherical codes was initiated by Shannon in [31]. Let W be a code on the sphere of radius $\sigma\sqrt{An}$. We assume that code vectors for transmission are chosen from W with equal probability. Then

$$P_e(W) = \frac{1}{|W|} \sum_{\boldsymbol{x} \in W} P_e(\boldsymbol{x})$$

is the average error probability, where $P_e(\boldsymbol{x})$ is the error probability of decoding provided that the transmitted vector is $\boldsymbol{x}$. Let

$$D(\boldsymbol{x}, W) = \{\boldsymbol{z} \in \mathbb{R}^n \colon \|\boldsymbol{z}-\boldsymbol{x}\| < \|\boldsymbol{z}-\boldsymbol{x}'\| \text{ for all } \boldsymbol{x}' \in W, \boldsymbol{x}' \ne \boldsymbol{x}\} \qquad (7)$$

be the Voronoi region of $\boldsymbol{x}$ with respect to the code W. Then the optimal decoding rule, i.e., the one minimizing the average error probability, associates to the received vector $\boldsymbol{z}$ a code vector $\boldsymbol{x}$ such that $\boldsymbol{z} \in D(\boldsymbol{x}, W)$. (The definition ignores vectors $\boldsymbol{z}$ at the same distance from two or more code vectors since their probability is 0.) Under this decoding, the error probability $P_e(W)$ equals

$$P_e(W) = \frac{1}{|W|} \sum_{\boldsymbol{x} \in W} \Pr\left\{\boldsymbol{z} \in \bigcup_{\substack{\boldsymbol{y} \in W \\ \boldsymbol{y} \ne \boldsymbol{x}}} D(\boldsymbol{y}, W) | x \text{transmitted}\right\} \qquad (8)$$

where the last probability equals the total probability, under the Gaussian distribution with mean at $\boldsymbol{x}$ and variance σ^2 along each coordinate, of the part of $\mathbb{R}^n$ complementary to the decoding region of $\boldsymbol{x}$. Further, let

$$P_e(R,A,n) = \min_{W \colon R(W) \ge R} P_e(W)$$

$$E(R,A,n) = -\frac{1}{n} \ln P_e(R,A,n)$$

$$\overline{E}(R,A) = \limsup_{n\to\infty} E(R,A,n)$$

$$\underline{E}(R,A) = \liminf_{n\to\infty} E(R,A,n).$$

Again we are interested in upper bounds on $\overline{E}(R,A)$ and lower bounds on $\underline{E}(R,A)$ as functions of R for given A. A common limit of these two functions, provided that it exists, is called the *reliability function* (or the error exponent) of the channel, denoted $E(R,A)$. By abuse of notation, below we speak of upper and lower bounds on $E(R,A)$.

Shannon [31] showed that $E(R,A) > 0$ for $R \in [0, \mathcal{C})$, where $\mathcal{C} = (1/2)\ln(1+A)$ is the capacity of the channel. In this interval $E(R,A)$ is bounded above by the sphere-packing bound [31]

$$E(R,A) \le E_{sp}(\theta(R), A) \qquad (9)$$

where

$$E_{sp}(\theta, A) = \frac{A}{2} - \frac{\sqrt{A} g(\theta, A)\cos\theta}{2} - \ln(g(\theta,A)\sin\theta)$$

$$g(\theta, A) = \frac{1}{2}(\sqrt{A}\cos\theta + \sqrt{A\cos^2\theta + 4})$$

and $\theta(R) = \arcsin e^{-R}$ is the "sphere-packing" angle; cf. (1). Further, $E(R, A)$ is bounded below as follows [31]:

$$E(R, A) \geq \begin{cases} \frac{A}{4}(1 - \cos\theta(R)), & 0 \leq R \leq R_1 \quad \text{(I)} \\ \frac{A}{4}(1 - \cos\theta(R_1)) + R_1 - R, & \\ & R_1 \leq R \leq R_2 \quad \text{(II)} \\ E_{sp}(\theta(R), A), & R_2 \leq R \leq \mathcal{C} \quad \text{(III)} \end{cases} \tag{10}$$

where

$$R_1 = \frac{1}{2}\ln\left(\frac{1}{2} + \frac{1}{2}\sqrt{1 + \frac{A^2}{4}}\right)$$

$$R_2 = \frac{1}{2}\ln\left(\frac{1}{2} + \frac{A}{4} + \frac{1}{2}\sqrt{1 + \frac{A^2}{4}}\right).$$

Bounds (9) and (10.III) show that $E(R, A)$ is known exactly for $R_2 \leq R \leq \mathcal{C}$. Shannon [31] also proved the inequality

$$E(R, A) \leq \frac{A}{4}, \qquad 0 \leq R \leq \mathcal{C} \tag{11}$$

which implies that bound (10.I) is tight for $R = 0$. The proof of (11) in [31] used the minimum-distance argument (the probability of confusing two code vectors at a minimum distance) together with a Plotkin-type bound on the size of the code. Independently and earlier, Plotkin-type bounds on $d(R)$ were proved by Rankin [30]. Later, it was realized that one can abstract from the Plotkin bound and use Shannon's argument to establish a general minimum-distance bound on $E(R, A)$ (see [32] for discrete channels and [21], [24] for the Gaussian channel). This bound has the form

$$E(R, A) \leq E_{md}(R, A) = \frac{A}{8}\, d^2(R). \tag{12}$$

Together with (5) this implies the best known bound on $E(R, A)$ for low code rates.

Finally, as shown in [32], the reliability function of discrete channels is bounded above by the straight line connecting any point R_b of any upper bound on $E(R, A)$ with any point R_a, $R_a > R_b$, of the sphere-packing bound. Sheverdyaev [33] extended this result to the Gaussian channel, showing, in particular, that a segment of the common tangent to $E_{md}(R, A)$ and $E_{sp}(R, A)$ gives an upper bound on $E(R, A)$ (note that both E_{md} and E_{sp} are convex). Rather than writing out a cumbersome explicit expression for this bound, we simply denote it by $E_{st}(R, A)$.

Concluding, let us summarize the results on the upper bound on $E(R, A)$ known to date

$$E(R, A) \leq \begin{cases} E_{md}(R, A), & 0 \leq R \leq R' \quad \text{(I)} \\ E_{st}(R, A), & R' \leq R \leq R'' \quad \text{(II)} \\ E_{sp}(R, A), & R'' \leq R \leq \mathcal{C} \quad \text{(III)} \end{cases} \tag{13}$$

where R', R'' are certain numbers which are easier to compute for each given A than to write out in general.

In our paper, following [34] and [31], we assume that code vectors can be any points on S^{n-1}. In communication theory one also studies a restricted case of this problem, namely, transmission over the Gaussian channel with codes whose vectors have coordinates equal to (binary or nonbinary) roots of unity. Then it is possible [28] to obtain upper bounds on codes better in a certain region of rates than the Kabatiansky–Levenshtein bounds. For lower existence bounds on the reliability function of the Gaussian channel with *binary* codes see, e.g., [29].

C. Outline of the Paper

The goal of this paper is to prove a new upper bound on $E(R, A)$ given by the following theorem.

Theorem 1: The reliability function of the Gaussian channel with signal-to-noise ratio A satisfies the upper bound

$$E(R, A) \leq \min_{0 \leq \rho \leq \rho_{kl}} \max_{w,\, d} \left[\min\left(A\frac{d^2}{8},\, A\frac{w^2}{8} - \mathcal{L}(w, d, \rho)\right)\right] \tag{14}$$

where R is a value of the code rate

$$0 \leq d \leq \frac{\sqrt{2}\left(\sqrt{1+\rho_{kl}} - \sqrt{\rho_{kl}}\right)}{\sqrt{1+2\rho_{kl}}}$$

$$d \leq w \leq \frac{\sqrt{2}\left(\sqrt{1+\rho} - \sqrt{\rho}\right)}{\sqrt{1+2\rho}}$$

ρ_{kl} is the root of $R = (1+\rho)H(\frac{\rho}{1+\rho})$

$$\mathcal{L}(w, d, \rho) = \min\left\{\frac{Ad^2w^2}{8(4w^2 - d^2)},\, F\left(1 - \frac{1}{2}w^2,\, \rho\right)\right\}$$

$$\begin{aligned} F(x, \rho) = {} & R - (1+\rho)H\left(\frac{\rho}{1+\rho}\right) \\ & + \ln\left(\frac{1}{2}\left(x + \sqrt{(1+2\rho)^2x^2 - 4\rho(1+\rho)}\right)\right) \\ & - (1+2\rho) \\ & \cdot \ln\frac{(1+2\rho)x + \sqrt{(1+2\rho)^2x^2 - 4\rho(1+\rho)}}{2(1+\rho)}. \end{aligned}$$

Together with a segment of the common tangent to the curve on the right-hand side of (14) and the sphere-packing exponent (9) this theorem improves bounds (13.I)–(13.II) for all rates $R \in (0, R'']$. Indeed, observe that forgetting the second term inside the brackets in (14), we get (13.I), so (14) is at least as good as the minimum-distance bound. Now put in (14) $\rho = \rho_{kl}$. Suppose that $d = d_{kl}$, then the second term under the minimum in (14) is less that $Ad_{kl}^2/8$ since $\mathcal{L}(d_{kl}, d_{kl}, \rho_{kl}) > 0$, so in this case our bound is strictly less than (13.I). On the other hand, if $d < d_{kl}$, then already the first term in (14) is less than (13.I). Thus (14) is strictly less than (13.I) for all $R \in (0, R']$, so the straight-line bound associated with it is also strictly less than $E_{st}(R, A)$ in (13.II) and touches E_{sp} at some point between R'' and R_2 (see Fig. 1).

The proof combines geometric, analytic, and combinatorial arguments. Its idea is summarized as follows. It is well known that the error probability of decoding is determined not as much by code's minimum distance $d(C)$ as by its distance distribution. To take into account this influence one has to estimate the average number of neighbors of a code vector. This number affects the error exponent if it grows exponentially in n. Bounds of the type (12) only take into account the fact [31] that each code of a large size contains a large subcode in which every code vector has a neighbor at a minimum distance. In contrast, we use lower exponential estimates of the average distance distribution for all distances $d(C) \leq d \leq \sqrt{2}$ (i.e., $\theta(C) \leq \theta \leq \pi/2$). This accounts for a better estimate of $E(R, A)$ in the region of code rates where the best known bound was $E_{md}(R, A)$.

The paper is organized as follows. In Section II, we derive a general lower bound on the distance distribution of codes. This result is proved by a new application of Delsarte's polynomial

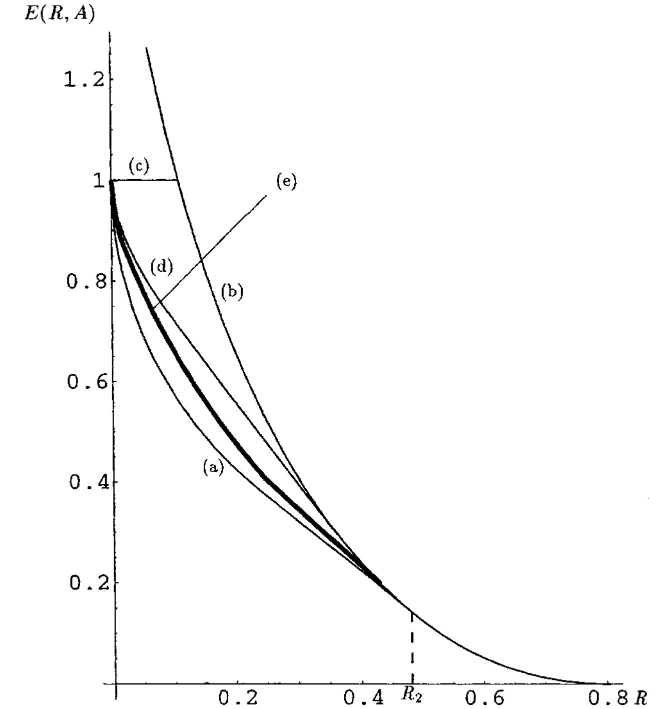

Fig. 1. Bounds on the reliability function $(A = 4)$. (a) "Random coding" exponent (10.I–III). (b) Sphere-packing bound (9). (c) Minimum-distance bound (11). (d) Minimum-distance bound (12). (e) The new bound (14). R_2 denotes the critical rate. Each of the curves (d)–(e) includes a segment of the common tangent to the curve and the sphere-packing exponent.

method in coding theory, discovered recently by the authors in [3] and [25]. As suggested by the Kabatiansky–Levenshtein (KL) approach [21], we prove these estimates simultaneously for codes in a number of metric spaces including $S^{n-1}(\mathbb{R})$. We believe that these estimates will find further use in coding theory as it happened with analogous results in the Hamming space [4], [6], [7].

To prove specific bounds, we need to establish the asymptotic behavior of Jacobi polynomials $P_k^{\alpha,\beta}$ as the degree $k \to \infty$ and $\frac{\alpha}{k} \to a$. This is the subject of a fairly technical Section III. By combination of classical and *ad hoc* methods we prove a number of asymptotic bounds on the exponent of $P_k^{\alpha,\beta}$ and, in a sense, give a definitive answer for the entire orthogonality segment. In this section we actually prove more than is needed to derive Theorem 1;[1] readers interested only in this theorem can skip everything except Theorem 6.

Section IV consists of two parts. In the first part, we use the estimates of Jacobi polynomials to derive exponential lower bounds on the distance distribution of spherical codes. In the second part, we establish some regularity properties of the distance distribution of spherical codes. This part is a technical aid for the proof of the lower bound on the error probability for spherical codes (Theorem 1). Here we prove that in any code one can isolate large subsets that in the asymptotics possess distance invariance properties similar to those of linear codes in the Hamming space, namely, that the distance spectrum with respect to any given vector in the subset is one and the same.

The remaining part of the proof of Theorem 1, given in Section V, is geometrically much more intuitive. It is accomplished by an argument analogous to the Hamming case in [25].

A few remarks on the asymptotic notation. Since in the paper we are interested only in logarithmic asymptotics of the reliability function and related parameters, we write $f(x)\overset{x}{\sim}g(x)$ to denote the fact that $(\ln f(x))/(\ln g(x)) \to 1$ as $x \to \infty$. For instance, the Stirling approximation for $\Gamma(x)$ gives $\ln \Gamma(a)\overset{a}{\sim}a(\ln a - 1)$. A short notation for

$$\lim_{x\to\infty} (f(x)/g(x)) = \text{const}$$

is $g = \Theta(f)$. Notation $f(x) \gtrsim g(x)$ means that

$$\liminf_{x\to\infty}(f(x)/g(x)) \geq 1.$$

II. Bounds on the Distance Distribution of Codes

In this section we prove a general bound on the distance distribution of codes. We take a somewhat broader view than in the rest of the paper since by one and the same method one can prove this bound for codes in many metric spaces simultaneously. The method is suggested by Kabatiansky and Levenshtein in [21] and applies to configurations in very general spaces (indeed, not necessarily metric). We restrict ourselves to a fragment of their theory, the unifying idea being to consider those spaces in which zonal spherical functions are given by Jacobi polynomials. (Incidentally, this covers all compact infinite spaces in which our results and the results of [21] are valid.)

Apart from our main example, the unit sphere $S^{n-1}(\mathbb{R})$, we also consider the $(n-1)$-dimensional projective spaces over $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ (the quaternions). Each of them can be realized as the set of lines through the origin in the corresponding n-dimensional linear space, or the sphere S^{n-1} with antipodal points identified. A code, again, is a *finite* subset of S^{n-1}. Let $\text{dist}\,(\boldsymbol{x}, \boldsymbol{y})$ be a certain metric and $\langle \boldsymbol{x}, \boldsymbol{y}\rangle$ be the usual (Hermitian) inner product on S^{n-1}. Let $\mathcal{D}$ be the set of all possible distances on S^{n-1} with respect to $\text{dist}(\boldsymbol{x}, \boldsymbol{y})$. For instance, for $S^{n-1}(\mathbb{R})$ with the Euclidean metric we have $\mathcal{D} = [0, 2]$. Let

$$t(\boldsymbol{x}, \boldsymbol{y}) = t(\text{dist}\,(\boldsymbol{x}, \boldsymbol{y})) \colon \mathcal{D} \to \mathcal{I} = [-1, 1]$$

be a monotone function that depends only on the distance between $\boldsymbol{x}$ and $\boldsymbol{y}$, such that $t(0) = 1$, $t(\max_{d\in\mathcal{D}}\, d) = -1$, and such that the zonal spherical functions are expressed as polynomials in t (see below). This substitution enables one to present results in a uniform way while not changing their analytic nature. For instance, for $S^{n-1}(\mathbb{R})$ with the Euclidean metric we can put

$$t(\boldsymbol{x}, \boldsymbol{y}) = \langle \boldsymbol{x}, \boldsymbol{y}\rangle = 1 - \frac{1}{2}\,\text{dist}^2(\boldsymbol{x}, \boldsymbol{y}).$$

Thus $t(0) = 1\,(\boldsymbol{x} = \boldsymbol{y})$ and $t(2) = -1$ (a pair of antipodal points).

Let $C \subset S^{n-1}$ be a code. Define the functions

$$b_{\boldsymbol{c}}(s, t) = |\{\boldsymbol{c}' \in C \colon s \leq t(\boldsymbol{c}, \boldsymbol{c}') \leq t\}| \qquad (15)$$

$$b(s, t) = \frac{1}{|C|} \sum_{\boldsymbol{c}\in C} b_{\boldsymbol{c}}(s, t). \qquad (16)$$

Typically, below we consider intervals $[s, t]$ of size $\Theta(\frac{1}{n})$ for growing n. In this case we keep only one of the two arguments, writing, for instance, $b(s)$ for $b(s, s + \Theta(\frac{1}{n}))$. Observe

[1] We believe that it is worthwhile to present these results in view of a prominent role that Jacobi polynomials play in coding theory. After this paper was submitted, we learned of related results [11]. Their results are given in the form that does not allow immediate use in our bounds.

that $b(s)$ can be thought of as the distance *density* of C (more precisely, the scalar products density). Further, let $B_{\boldsymbol{c}}(s) = b_{\boldsymbol{c}}(s, 1)$, $B(s) = b(s, 1)$ be the (local and average) distance distributions of C. To sum other functions according to $B(x)$, it is convenient to have a discrete measure associated with it. Observe that $B(s)$ is a nonincreasing function, so its jumps are negative. Therefore, let $\mu_B(I) = B(s+0) - B(t)$, where $I = (s, t)$, $\mu_B(I) = B(s) - B(t+0)$, where $I = [s, t]$, etc. Then

$$b(s, t) = \int_s^t dB(x) \qquad \int_{\mathcal{I}} dB(x) = |C| \tag{17}$$

where the integration is with respect to the measure μ_B.

One of the main results in [21] is that the distance distribution of codes on S^{n-1} satisfies certain positivity conditions. Recall that there is a natural way of associating with S^{n-1} a Fourier basis formed of zonal spherical functions Φ_k. A specific form of these functions depends both on the ground field and on the distance function on S^{n-1}. By [21]

$$\int_{\mathcal{I}} \Phi_k(x) dB(x) \ge 0, \qquad k = 0, 1, \ldots. \tag{18}$$

These inequalities follow from the fact that the action of the isometry group G of S^{n-1} is doubly transitive on it; hence zonal spherical functions form a complete system in $L_2(G)$, and Fourier coefficients of any positive-definite function in $L_2(G)$ with respect to this system are nonnegative. A particularly readable introduction in this part of harmonic analysis on compact groups is found in [37]; see also [19], [20], and [36]. In fact, inequalities (18) apply in a much more general context [21]. For Q-polynomial association schemes, they constitute the Delsarte inequalities [14].

We now derive a lower bound on the distance distribution of C.

Theorem 2: Let $B(x)$ be the distance distribution of a code $C \subset S^{n-1}$ and let m be an integer. Let $-1 \le u_0 < 1$ and suppose that $u_0 < u_1 < \ldots < u_{m-1} < u_m = t(d(C)) < 1$ are defining points of a partition of $[u_0, t(d(C))]$ into m segments $U_i = [u_i, u_{i+1}]$.

Suppose that $f(x) = \sum_{k=0}^{l} f_k \Phi_k(x)$ is a polynomial of degree l such that

i) $f_k \ge 0$ for $0 \le k \le l$;
ii) $f(x) \le 0$ for $-1 \le x \le u_0 \quad f(x) \ge 0$ for $u_0 \le x \le 1$.

Then there exists a number i, $0 \le i \le m-1$, and a point $s \in U_i$ such that

$$b(u_i, u_{i+1}) \ge \frac{f_0|C| - f(1)}{m f(s)}. \tag{19}$$

Proof: We have

$$f_0|C| = f_0 \int_{\mathcal{I}} dB(x) \le \sum_{k=0}^{l} f_k \int_{\mathcal{I}} \Phi_k(x) dB(x)$$

where the first equality follows by (17) and the inequality is implied by i) and (18) and the fact that $\Phi_0 = 1$. Now let us interchange the sum and the integral and use ii)

$$\int_{\mathcal{I}} \sum_{k=0}^{l} f_k \Phi_k(x) dB(x) = \int_{\mathcal{I}} f(x) dB(x) \le \sum_{i=0}^{m-1} \int_{U_i} f(x) dB(x) + f(1)$$

(we have used the fact that $t(0) = 1$). Let

$$s_i = \arg\max_{x \in U_i} f(x).$$

Then we have

$$f_0|C| \le f(1) + \sum_{i=0}^{m-1} f(s_i) b(u_i, u_{i+1}).$$

Hence, there exists a number $j \in [0, m-1]$ such that the summation term satisfies the claimed inequality. □

Note that condition $f(x) \ge 0 (u_0 \le x \le 1)$ in the statement can be replaced by $f_0|C| > f(0)$. In the applications of this theorem below we usually choose the segments U_i to be of equal small length (of order $1/n$).

Theorem 2 *mutatis mutandis* is valid for all spaces covered by the KL theory (for instance, for all two-point homogeneous spaces with massive invariant subgroup). For codes in the Hamming space this theorem was proved in [25] (see [5] for an overview).

Let us specialize this theorem to the context of this paper, that is, to the unit sphere in $\mathbb{R}^n$.

a) $(S^{n-1}(\mathbb{R}), t(\boldsymbol{x}, \boldsymbol{y}) = \langle \boldsymbol{x}, \boldsymbol{y} \rangle)$. The zonal spherical functions were found by Cartan [10] (see also [37, Ch. 9]) in the form

$$\Phi_k(x) = C_k^{(n-2)/2}(x) / C_k^{(n-2)/2}(1)$$

where

$$C_k^{\lambda}(x) = \sum_{i=0}^{\lfloor k/2 \rfloor} (-1)^i \binom{k-i}{i} \binom{k-i+\lambda-1}{k-i} (2x)^{k-2i}$$

is the Gegenbauer, or ultraspherical, polynomial. It is known [35] that

$$C_k^{\lambda}(x) = \left(\binom{k+2\lambda-1}{k} \Big/ \binom{k+\lambda-(1/2)}{k} \right) \cdot P_k^{\lambda-(1/2), \lambda-(1/2)}(x)$$

where $P_k^{\alpha, \beta}(x)$ is the Jacobi polynomial.

Note that the inequalities above do not change if we divide out a positive constant. Therefore, in estimate (19), we can put $\Phi_k(x) = P_k^{(n-3)/2, (n-3)/2}(x)$.

b) *Projective Spaces.* Consider the projective spaces PX^{n-1}, where $X = \mathbb{R}$ or $\mathbb{C}$ or $\mathbb{H}$. The distance in PX^{n-1} can be expressed via the inner product $x = |\langle \boldsymbol{x}, \boldsymbol{y} \rangle|$. The substitution $t(x) = 2x^2 - 1$ maps $[0, 1]$ on $[-1, 1]$ and possesses the necessary properties. One can take [21]

$$\Phi_k(x) = P_k^{\delta(n-1)-1, \delta-1}(2x^2 - 1)$$

where $2\delta \in (1, 2, 4)$ is the dimension of X as a linear space over $\mathbb{R}$.

We see that to apply inequality (19) we need to study the asymptotic behavior of $f(s)$, $-1 \le s \le 1$. This question leads one to the study of values of Jacobi polynomials.

III. Asymptotic Estimates of Jacobi Polynomials

The subject of this section is the study of the asymptotic behavior of Jacobi polynomials $P_k^{\alpha,\beta}(x)$ for $k \to \infty$, $\alpha/k \to \mathrm{const}$. In order to derive asymptotic bounds on the distance distribution of spherical codes in the next section it suffices to consider the case $\alpha = \beta$. However, exponential estimates of general Jacobi polynomials, besides being of independent interest, are useful for constructing bounds on codes in different metric spaces (see the end of the previous section). Therefore, we begin with the case of general α, β especially since it is not much more difficult than the particular case mentioned.

Below we assume that $k \to \infty$, $\alpha/k \to a$, $\beta \le \alpha$. The general analytic situation that we treat in this section stems from the derivation of the so-called "linear programming (LP) bound" [26] and its extensions to other Q-polynomial spaces [21], [23], [2]. The bound on the rate of a code has a form of a certain function of the extremal zero t_1 of the corresponding family of zonal orthogonal polynomials. Asymptotics of the extremal zero for various systems of polynomials were studied in [26], [21], [23].

A more refined situation encountered in a number of problems that involve LP bounds [25], [3], [8] requires estimating the asymptotic behavior of the polynomials in the entire interval from the extremal zero to the end of the orthogonality segment. For a discrete space W, a simple uniform estimate follows from the identity $||P_k||^2 = m_k|W|$, where P_k is the Q-polynomial of degree k of the association scheme, m_k is the kth eigenvalue of the scheme, and $||P_k|| = \langle P_k, P_k\rangle^{1/2}$ is the corresponding L_2-norm (cf. our Theorem 5 which employs a different method to prove a similar result in the continuous case). For Krawtchouk polynomials K_k an extension of the method in [26] was employed in [22] to derive an exact expression for the main term of the exponent in the interval considered. The proof in [22] is based on the difference equation for K_k. In the continuous case one can rely on the distribution of zeros of the polynomial and derive, in a sense, a tight estimate for the entire orthogonality segment.

Properties of Jacobi polynomials $P_k^{\alpha,\beta}(x)$ are collected, for instance, in [35], [17, vol. II]. We need the following facts. The polynomials $P_k^{\alpha,\beta}(x)$ are orthogonal on $\mathcal{I} = [-1, 1]$ with weight $\mu(x) = (1-x)^\alpha(1+x)^\beta$

$$\langle P_k^{\alpha,\beta}(x), P_j^{\alpha,\beta}(x)\rangle_2 = \delta_{kj}\omega_k^{\alpha,\beta} \tag{20}$$

and

$$\omega_k^{\alpha,\beta} = \frac{2^{\alpha+\beta+1}\Gamma(k+\alpha+1)\Gamma(k+\beta+1)}{(2k+\alpha+\beta+1)k!\Gamma(k+\alpha+\beta+1)}. \tag{21}$$

The polynomial $P_k^{\alpha,\beta}(x)$ has k simple zeros on $\mathcal{I}$. Denote them by $t_{i,k} = t_{i,k}^{\alpha,\beta}$, $t_{k,k} < t_{k-1,k} < \cdots < t_{1,k}$. Thus we have

$$P_k^{\alpha,\beta}(x) = L_k \prod_{j=1}^{k}(x - t_{j,k}) \tag{22}$$

where

$$L_k = 2^{-k}\sum_{\nu=0}^{k}\binom{k+\alpha}{k-\nu}\binom{k+\beta}{\nu}.$$

Further

$$P_k^{\alpha,\beta}(x) = (-1)^k P_k^{\beta,\alpha}(-x)$$

so, in particular, $t_{k,k}^{\alpha,\beta} = -t_{1,k}^{\beta,\alpha}$. Zeros of $P_k^{\alpha,\alpha}(x)$ are symmetric with respect to 0, and $t_{k,k} = -t_{1,k}$. Zeros of the two adjacent Jacobi polynomials form two interlacing systems:

$$t_{i+1,k+1} < t_{i,k} < t_{i,k+1}$$

and so forth.

For k sufficiently large by [21] we have

$$\begin{aligned} t_{1,k} &\to q(a, b) := \frac{4\sqrt{(a+b+1)(a+1)(b+1)} - a^2 + b^2}{(a+b+2)^2} \\ t_{k,k} &\to -q(b, a) \end{aligned}$$

(an independent later proof was given in [27]). In particular, for $\alpha = \beta = ak$ this implies

$$t_{1,k} \to q(a, a) = \frac{\sqrt{1+2a}}{1+a}. \tag{23}$$

The zeros of the sequence $\{P_k^{\alpha,\beta}(x)\}_{k=0}^{\infty}$ fill the segment $[-q(b, a), q(a, b)]$ densely.

Let us proceed to bounds on $P_k^{\alpha,\beta}$. We present three results, each obtained by a different technique. The first result is obtained by transforming the differential equation (60) for the Jacobi polynomials to a form with (locally) constant coefficients and applying a method of Sturm–Liouville to estimate the distance between the consecutive zeros. This gives an exponential estimate for Jacobi polynomials in the entire segment $\mathcal{I}$. The second theorem extends the method of [26] and [22] and gives an exponentially tight estimate in the range $[-1, -q(b, a)) \cup (q(a, b), 1]$. In this range this estimate coincides with the first one, but is derived in a different manner, and has a totally different form.

The third theorem relies on the value of the L_2-norm of the Jacobi polynomial. It provides an estimate that passes through all the maxima of $|P_k^{\alpha,\beta}|$ and since the zeros of $P_k^{\alpha,\beta}$ are dense, can be thought of as the limiting envelope of the polynomials. It is asymptotically tight in the interval $[-q(b, a), q(a, b)]$ and in this interval coincides numerically with the first estimate. All the three estimates are equal at $x = -q(b, a)$, $x = q(a, b)$.

From now till the end of this section we denote zeros of P_k by t_i, omitting the degree. The next theorem, proved in the Appendix, gives the exact logarithmic asymptotics for $P_k^{\alpha,\beta}$.

Theorem 3: Let $P_k^{\alpha,\beta}(x)$ be the Jacobi polynomial. Suppose that $k \to \infty$, $\alpha/k \to a > 0$, $\beta/k \to b$, and $x \in [-1, 1]$. Then

$$\frac{1}{k}\ln|P_k^{\alpha,\beta}(x)| \le \frac{1}{k}\ln L_k + \frac{1}{\pi}\int_{-q(b,a)}^{q(a,b)} \sqrt{\frac{1+a+b+ab/2}{1-z^2} - \frac{a^2}{4(1-z)^2} - \frac{b^2}{4(1+z)^2}} \cdot \ln|x-z|\,dz + o(1). \tag{24}$$

Note that for $x \in [-q(b,a), q(a,b)]$ the integral in (24) has a singularity at the point $z = x$. However, its convergence is easily checked.

Theorem 4: Let $P_k^{\alpha,\beta}(x)$ be the Jacobi polynomial. Suppose that $k \to \infty$, $\alpha/k \to a > 0$, $\beta/k \to b$, and

$$x \in [-1, -q(b,a)) \cup (q(a,b), 1].$$

Then we get (25) at the bottom of the page, where the $-$ sign corresponds to $x > q(a,b)$ and the $+$ to $x < -q(b,a)$.

Proof: See the Appendix.

Note that (24) also gives the exact main term of the exponent of P_k. So for such $x \in [-1, -q(b,a)) \cup (q(a,b), 1]$, expressions (24) and (25) represent one and the same function.

Finally, let us derive an estimate which turns out to be tight in the subsegment from the first to the last zero of P_k.

Theorem 5: Let $P_k^{\alpha,\beta}(x)$ be the Jacobi polynomial. Suppose that $k \to \infty$, $\alpha/k \to a$, $\beta/k \to b$. Then

$$\frac{1}{k}\ln|P_k^{\alpha,\beta}(x)|^2 \lesssim -a\ln(1-x) - b\ln(1+x) + \ln\omega_k^{\alpha,\beta}, \qquad x \in [-1, 1]. \tag{26}$$

Proof: See the Appendix.

Remark: It can be proved that exponentially bound (26) is exact for $-q(b,a) \le x \le q(a,b)$. Hence in the oscillatory segment the right-hand side of (26) equals that of (24).

Since in the following section we deal with codes on $S^{n-1}(\mathbb{R})$, the case of $\alpha = \beta$ is of special interest to us. Recall that in this case $|P_k(x)|$ is even, in particular $t_k = -t_1$, and that $t_1 \to q(a,a)$. Below we abbreviate $q(a,a)$ to q. Estimates of Theorems 3–5 in this case are collected in the following theorem.

Theorem 6: Let $P_k^{\alpha,\alpha}$ be the Jacobi polynomial and suppose that $k \to \infty$, $\alpha/k \to a > 0$. Then up to $o(1)$ terms

a) $x \in [-1, 1]$

$$\frac{1}{k}\ln|P_k^{\alpha,\alpha}(x)| \le 2(a+1)H\left(\frac{1}{2(a+1)}\right) - \ln 2 + \frac{1}{\pi}\int_0^q \frac{\sqrt{2a+1-(a+1)z^2}}{1-z^2}\ln|x^2-z^2|\,dz. \tag{27}$$

b) $q < |x| \le 1$

$$\frac{1}{k}\ln|P_k^{\alpha,\alpha}(x)| = (1+a)H\left(\frac{a}{1+a}\right) \mp \int_x^1 \frac{(1+2a)\,dz}{az \pm \sqrt{a^2z^2-(1-z^2)(1+2a)}} \tag{28}$$

where the first choice of the signs corresponds to $x \in (q, 1]$ and the second to $x \in [-1, -q)$.

c) $x \in [-1, 1]$

$$\frac{1}{k}\ln|P_k^{\alpha,\alpha}(x)| \le (a+1)H\left(\frac{1}{2(a+1)}\right) - \frac{1}{2}a\ln(1-x^2) - \ln 2. \tag{29}$$

Proof: Standard asymptotic analysis, see the Appendix.

Obviously, remarks on the mutual relations and tightness of the bounds made in the general case are valid in the particular case considered in this theorem. For instance, (27) and (28) for $x \in [q, 1]$ represent one and the same function.

The integrals in (28) can be computed in a closed form using Mathematica. The answer is rather cumbersome, but can be transformed to a compact form (computed by the saddle point method from the integral representation of $P_k^{\alpha,\alpha}$ in [9]). For instance,

$$\int_x^1 \frac{(1+2a)\,dz}{az + \sqrt{a^2z^2-(1-z^2)(1+2a)}} = a\ln\frac{ax+\sqrt{(1+a)^2x^2-(1+2a)}}{2a} - (1+a)\ln\frac{(1+a)x+\sqrt{(1+a)^2x^2-(1+2a)}}{(1+2a)}. \tag{30}$$

The following simple corollary is of independent interest and may be useful in applications.

Corollary 7: Let $P_k^{ak,ak}$ be the sequence of Jacobi polynomials $k \to \infty$. Then up to $o(1)$ terms

$$\frac{1}{k}\ln|P_k^{\alpha,\alpha}(x)| \le \frac{aq}{1-q^2}(x-q)+h(q), \qquad x \in [q, 1] \tag{31}$$

where

$$h(x) = (a+1)H\left(\frac{1}{2(a+1)}\right) - \frac{1}{2}a\ln(1-x^2) - \ln 2.$$

Proof: Note that the derivative of the function on the right-hand side of (28) equals $h'(x)$ at $x = q$. It can be checked that the function itself is concave for $x \in (q, 1]$. Therefore, the straight line drawn through the point $x = q$ with slope $h'(q)$ is an upper bound on the exponent of the polynomial in $x \in [q, 1]$. □

Clearly, a similar argument is valid for $x \in [-1, -q]$, i.e., the exponent of $|P_k^{\alpha,\alpha}(x)|$ is bounded above by a straight line symmetric to (31) with respect to the y-axis.

The behavior of the bounds is visualized in Fig. 2.

$$\frac{1}{k}\ln|P_k^{\alpha,\beta}(x)| = (1+a)H\left(\frac{a}{1+a}\right) \mp \int_x^1 \frac{(a+(a+b)z-b) \mp \sqrt{(a+(a+b)z-b)^2-4(1-z^2)(1+a+b)}}{2(1-z^2)}\,dz + o(1) \tag{25}$$

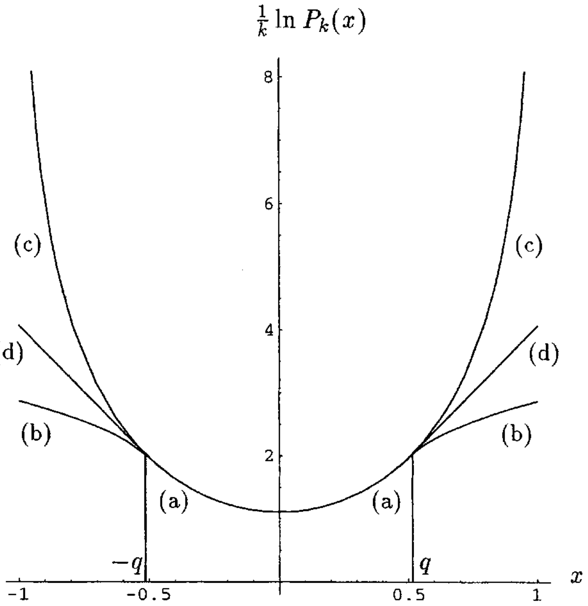

Fig. 2. Exponent of the Jacobi polynomials $|P_k^{ak,\, ak}|(k \to \infty,\ a = 6)$. (a) Exact expression (27), valid for $x \in [-1,\, 1]$. (b) Exact expression (28), valid for $q \le |x| \le 1$. (c) Upper bound (29), valid for $x \in [-1,\, 1]$. (d) Upper bound (31).

Remark: Observe that $h(x)$ is the function on the right-hand side of (29). The second derivative $h''(x) = a(1+x^2)/(1-x^2)$ is positive for all $x \in (-1,\, 1)$; so $h(x)$ is convex. So (31) is a *common* tangent to (28) and (29) at $x = q$ and separates these two curves. Bound (31) is useful because on the one hand is easy to work with, and on the other hand, it is much better than (29), in particular, it does not become infinite as $x \to \pm 1$.

IV. Asymptotic Bounds on the Distance Distribution of Spherical Codes

In this section, we return to the concrete setting of Section I and prove asymptotic bounds on the distance distribution of codes on the unit sphere $S = S^{n-1}(\mathbb{R})$ in n dimensions.

A. Absolute Bounds

Below we use the asymptotic expression for the distance density $\overline{b}(t)$ of a "random code," that is, the expectation of the distance density of a code C of a fixed size $M = e^{Rn}$ chosen on S in accordance with the uniform probability measure. Let $\boldsymbol{c} \in C$ be a code point and $b_{\boldsymbol{c}}(t)$ be the local distance density with respect to $\boldsymbol{c}$ (15). In other words, we are counting the number of code points in the spherical ring located on S between two cones with apex at the origin, "center" at $\boldsymbol{c}$ and half-angles $\theta = \arccos t$ and $\theta' = \arccos(t + \Theta(\frac{1}{n}))$, respectively. On the average, this number constitutes the same fraction of M as the area of the ring of the total area of S. Letting $\Omega(\theta)$ denote the $(n-1)$-dimensional area of a spherical cap on S with half-angle θ, we then obtain

$$\frac{\overline{b}_{\boldsymbol{c}}(\theta',\, \theta)}{M} = \frac{\Omega(\theta) - \Omega(\theta')}{\Omega(\pi)} \tag{32}$$

where the overbar refers to averaging over the ensemble of codes. By [31, p. 624]

$$\Omega(\theta) = (1 + o(1))\pi^{(n-1)/2}(\sin\theta)^n/\Gamma((n+1)/2)$$

so

$$\Omega(\theta) - \Omega(\theta') = \Omega(\theta)(1 - o(1)) \qquad (\theta \le \pi/2).$$

Thus the main term of the right-hand side of (32) is independent of θ'. Hence, we may as well put in (32) $\theta' = 0$. Thus we have

$$\frac{\overline{B}_{\boldsymbol{c}}(t)}{M} = \frac{\Omega(\theta)(1 - o(1))}{\Omega(\pi)}. \tag{33}$$

Further, since (33) holds with respect to every $\boldsymbol{c} \in C$, we can use the definition of the distance distribution of C (16) to conclude that for the random code

$$\overline{B}(t) = M\frac{\Omega(\theta)(1 - o(1))}{\Omega(\pi)}.$$

All told, we obtain

$$\overline{B}(t) \overset{n}{\sim} e^{nR}\sin^{n-1}\theta$$

or

$$\begin{aligned}\frac{1}{n}\ln\overline{B}(t) &= R + \ln\sin\theta + o(1)\\ &= R + \ln\sqrt{1 - t^2} + o(1).\end{aligned} \tag{34}$$

Below we call this expression the *random distance spectrum.*

Remark: The role of the random distance spectrum for spherical codes is much the same as that of the well-known "binomial" weight spectrum in the Hamming case

$$\overline{B}_i = \binom{n}{i}|C|2^{-n}$$

where this time n is the code length. Namely, both functions appear as the mean distance distribution of random codes chosen with uniform probability from their respective spaces. Moreover, for large n as long as $\frac{i}{n}$ is less than the sphere packing (Varshamov–Gilbert) bound on the code distance, $\overline{B}_i < 1$. Likewise, in the spherical case $\overline{B}(t) < 1$ as long as θ is less than the sphere packing (Shannon) bound (1). Finally, a code with the random distance distribution and minimum distance equal to the sphere packing bound asymptotically meets the "random coding exponent." This means that at rates below capacity the exponent of the error probability of decoding for such code asymptotically behaves as (10) for the spherical case or as its "discrete" counterpart [18].

In this section we use a shorthand notation $P_k(x)$ for $P_k^{\alpha,\,\alpha}(x)$. Let $\frac{k}{n} \to \rho$ (ρ is a constant). We will see below that ρ is the same parameter as appeared briefly in Section I. We need to take $\alpha = (n-3)/2$, where n is the dimension of the ambient space. Let

$$f^{(k)}(x) = P_k^2(s)\frac{(P_{k+1}(x) + P_k(x))^2}{x - s} \tag{35}$$

where $t_{1,\,k} < s < t_{1,\,k+1}$, and s is chosen so that $P_k(s) = -P_{k+1}(s)$ (so in particular, s depends on k). A polynomial of the same form as $f^{(k)}(x)$ was suggested in [26] for the Hamming and Johnson spaces and used there to derive upper bounds on the size of codes. $f^{(k)}(x)$ was used in [21] to derive bound (3)–(5).

We again assume that $\frac{\alpha}{k} \to a$, i.e., $a = \frac{1}{2\rho}$. Our task for now will be to specialize bound (19) to the case $f = f^{(k)}$. Let

$$f^{(k)}(x) = \sum_{i=0}^{2k+1} f_i P_k(x)$$

be the Gegenbauer expansion of $f^{(k)}(x)$. Then [21]

$$f_0 = -P_{k+1}(s)P_k(s)\frac{(2k+2\alpha+1)(2k+2\alpha+2)\omega_k^{\alpha,\,\alpha}}{2(k+1)(k+2\alpha+1)\omega_0^{\alpha,\,\alpha}}$$

$$f^{(k)}(1) = \frac{(P_k(s))^2}{1-s}\binom{k+\alpha}{k}^2\left(\frac{k+1+\alpha}{k+1} - \frac{P_{k+1}(s)}{P_k(s)}\right)^2.$$

The first equation follows by an application of the Christoffel–Darboux formula (61) and the orthogonality relations (20); the second one is straightforward. Then we have for $s \le x \le 1$

$$\frac{f_0}{f^{(k)}(x)} \overset{k}{\sim} \frac{\omega_k^{\alpha,\,\alpha}}{(P_{k+1}(x) + P_k(x))^2}$$

or, omitting small terms

$$\frac{1}{k}\ln\frac{f_0}{f^{(k)}(x)} \overset{(67)}{=} -2\ln 2 + 2(1+a)H\left(\frac{1}{2(1+a)}\right) - \frac{2}{k}\ln P_k(x) \tag{36}$$

and, in particular

$$\frac{1}{k}\ \ln\ \frac{f_0}{f^{(k)}(1)} = -(1+2a)H\left(\frac{1}{1+2a}\right). \tag{37}$$

In the last equation we have taken into account (59) and the identity

$$2\ \ln\ 2 - 2(1+a)H\left(\frac{1}{2(1+a)}\right) + 2(1+a)H\left(\frac{1}{1+a}\right) = (1+2a)H\left(\frac{1}{1+2a}\right).$$

Note in passing that (37) leads to the following result.

Proposition 8: Let $r \in \mathbb{R}$ be a number such that

$$r > (1+2a)H\left(\frac{1}{1+2a}\right). \tag{38}$$

Then $\ln(f^{(k)}(1)/e^{kr}f_0) < 0$, i.e., $e^{kr}f_0$ is exponentially greater than $f(1)$.

Remark: Note that

$$R_{kl} = (1+2a)H\left(\frac{1}{1+2a}\right)$$

is the KL bound (3), (4) renormalized to $k = \rho n$. So this proposition says that as long as r is greater than R_{kl}, bound (19) with $f = f^{(k)}(x)$ is dominated by the *first* term. For fixed R this holds as long as $\rho < \rho_{kl}$.

Let us return to the main topic of the section. Again by the Christoffel–Darboux, the coefficients f_i are nonnegative. Further, $f^{(k)}(x) \le 0$ for $x \in [-1, s]$. Finally, observe that $f^{(k)}(x)$ satisfies the conditions of Theorem 2, so we can apply the estimates of the previous section to derive concrete bounds on the possible distance distribution of spherical codes. Observe that as $k \to \infty$, $t_{1,\,k} \to t_{1,\,k+1}$, and so the number s approaches $t_{1,\,k}$.

Let C be a code of rate R and $b(x)$ its distance density. Let

$$\tau(\rho) = 2\frac{\sqrt{\rho(1+\rho)}}{1+2\rho}. \tag{39}$$

and let $t_{kl} = \tau(\rho_{kl})$, where ρ_{kl} is defined in (4). (Note that $t_{kl} = t(d_{kl})$, where $t(\cdot)$ is defined in Section II and d_{kl} is given by (5).) For $0 \le \rho \le \rho_{kl}$ we have $0 \le \tau(\rho) \le t_{kl}$.

Theorem 9: Let C be a code of rate R. Let ρ, $0 \le \rho < \rho_{kl}$, be a fixed number. Then there is a value x, $\tau(\rho) \le x \le 1$, such that

$$b(x) \ge J(n,\, x,\, \rho,\, |C|)$$
$$J(n,\, x,\, \rho,\, e^{nR}) = \exp\left(nj^*(x,\, \rho,\, R) - o(n)\right) \tag{40}$$

where

$$\begin{aligned} j^*&(x,\, \rho,\, R)\\ &:= R - (1+\rho)H\left(\frac{\rho}{1+\rho}\right)\\ &\quad + \ln\left(\frac{1}{2}\left(x + \sqrt{(1+2\rho)^2x^2 - 4\rho(1+\rho)}\right)\right)\\ &\quad - (1+2\rho)\ln\frac{(1+2\rho)x + \sqrt{(1+2\rho)^2x^2 - 4\rho(1+\rho)}}{2(1+\rho)}. \end{aligned} \tag{41}$$

Proof: Let $0 \le \rho \le \rho_{kl}$, $k = \rho n$, $m = \Theta(n^{1+\gamma})$, where $\gamma > 0$. By Theorem 2 and Proposition 8 we have, for some $U_i = [x,\, y]$ and $s \in U_i$

$$\ln b(x,\, y) \gtrsim \ln\frac{f_0|C|}{mf^{(k)}(s)}. \tag{42}$$

We plan to proceed by substituting in this estimate (36) together with (28). Note that since $y - x = o(n^{-1})$, the difference $\ln f^{(k)}(y) - \ln f^{(k)}(x) = o(k)$; hence inequality (42) is still asymptotically valid if we replace s with x. From the proof of Theorem 4 we see that the estimate in (28) is nothing else than $P_k(1)$ times $\exp\{-\int_x^1 \ldots dz\}$, where the integral is the same as in (28). Hence by (36)–(37) we get

$$\ln\, b(x) \ge 2k\int_x^1 \frac{(1+2a)\,dz}{az + \sqrt{a^2z^2 - (1-z^2)(1+2a)}} - k(1+2a)H\left(\frac{1}{1+2a}\right) + \ln\,|C|.$$

(Alternatively, to derive this use (36) together with (28) and the identity after (37).) To complete the proof it remains to change the scaling according to $2ak = n$, $2a\rho = 1$ and use the expression (30) for the resulting integral. □

Note that $J(n,\, x,\, \rho,\, e^{nR})$ is monotone increasing on R.

Remarks:

a) Let us see what happens with the lower bound on the distance distribution if in Theorem 2 we use estimate (29) instead of (28). The answer, not quite intuitive, is that instead of (41) we obtain in (40) the random distance spectrum (34). In other words, there is a point $\tau(\rho) \le x \le 1$ at which the logarithm of the distance distribution asymptotically at least equals $\overline{B}(x)$.

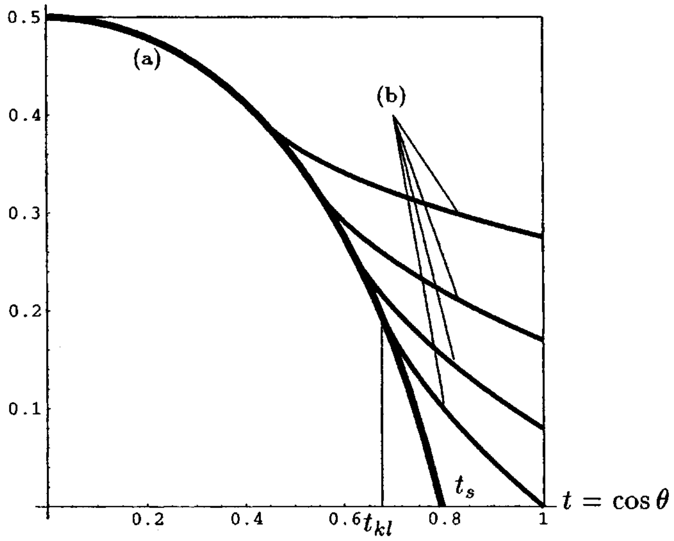

Fig. 3. Bounds on the distance distribution of codes ($R = 0.5$). (a) "Random" distance spectrum (34). (b) Bound (41). (The first curve from below corresponds to $\rho_{kl} = 0.179$.) The remaining three curves are drawn for $\rho = 0.139, 0.099, 0.059$, respectively); $t_s = 0.795$—the Shannon (Varshamov–Gilbert) distance; $t_{kl} = 0.675$—the KL distance.

Indeed, let us use (19) or its corollary (42) together with (29). Again taking into account (36), we obtain

$$\ln b(x, y) \gtrsim \ln |C| + ka \ln (1 - s^2).$$

Rescale the above inequality using $2ak = n$. This gives

$$\frac{1}{n} \ln b(x) \gtrsim \ln \sqrt{1 - x^2} + R.$$

Since estimate (28) is better than (29) except for $x = q$ where they are the same, Theorem 9 gives a lower bound (in fact, a family of lower bounds) on the distance distribution at least as strong as the average distribution. At this moment it is instructive to consider Fig. 3 which visualizes this remark.

b) *Projective spaces.* Codes in the complex projective space $P\mathbb{C}^{n-1}$ with distance measured by $|\langle \boldsymbol{x}, \boldsymbol{y}\rangle|$ are also known as families of sequences with small cross- and autocorrelation [23]. For this reason their properties are of interest to coding theory. It is possible to use polynomials of the form (35) to write out lower bounds on the distance distribution of codes in $P\mathbb{C}^{n-1}$ and other projective spaces mentioned in the end of Section II. Let us outline this derivation.

Let $C \subset PX^{n-1}$ be a code of rate R, where $X = \mathbb{R}$ or $\mathbb{C}$ or $\mathbb{H}$. Let $k = \rho n$. Together with the definitions of $a = \alpha/k$ and δ (see the end of Section II) we then obtain $a = \delta/\rho$. The maximal zero of $P_k^{\alpha, \beta}(y)$ converges to

$$q(a, 0) = \frac{4(a+1)^2 - a^2}{(a+2)^2} = \frac{4\rho(\delta+\rho) - \delta^2}{(\delta + 2\rho)^2}.$$

The bound on the rate R of a code C such that $|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| \le x$ for any two points $\boldsymbol{x}, \boldsymbol{y} \in C$, has the form [21]

$$R \le 2(\delta + \rho) H \left(\frac{\rho}{\delta + \rho}\right)$$

where $\rho = (\delta/2)((1-x^2)^{-1/2} - 1)$. The bound on the distance distribution has the following form. Let ρ_0 be the root of

$$R = 2(\delta + \rho) H(\rho/(\delta + \rho))$$

and let $\rho \in [0, \rho_0)$ be a fixed number. Then there exists a point

$$x, \, (4\rho(\delta + \rho) - \delta^2)/(\delta + 2\rho)^2 \le 2x^2 - 1 \le 1$$

such that

$$\begin{aligned}\ln b(x) &\ge \ln \frac{f_0 |C|}{f^{(k)}(2x^2 - 1)} \\ &= \ln \omega_k^{\alpha, \beta} - 2 \ln P_k^{\alpha, \beta}(2x^2 - 1) + Rn + o(n) \qquad (43)\end{aligned}$$

where α and β should be chosen as specified above. With $\alpha = ak$ and β constant, (21) yields $\omega_k^{\alpha, \beta} \overset{k}{\sim} 2^{ak}$. Next we substitite $b = 0$ in (25) and integrate (using Mathematica) to obtain the expression at the bottom of the page. Plugging all this into (43) and substituting $k = \rho n$, $a = \delta/\rho$, we obtain an exponential lower bound on the distance distribution of C.

Codes in the real Grassmann space $G_{k,n}$ recently attracted interest in geometry [12]. $G_{k,n}$ is the manifold of k-dimensional linear spaces in $\mathbb{R}^n$ passing through the origin. The case $k = 1$ corresponds to the real projective space $P\mathbb{R}^{n-1}$.

B. Regularity Properties

In this part of the section we prove a few results that hold uniformly for most local distance densities in the code. Together with Theorem 9 these theorems will be used in the next section.

Let us define the *effective distance* of the code as follows. Define a partition of the interval $[u_0 = 0, u_m = t(C)]$ into segments of equal length $\Theta(\frac{1}{n})$

$$U_0 = [u_0, u_1], \; U_1 = [u_1, u_2], \; U_{m-1} = [u_{m-1}, u_m]. \qquad (44)$$

For a code vector $\boldsymbol{c} \in C$ let

$$C(t, \boldsymbol{c}) = \left\{ \boldsymbol{c}' \in C \colon t \le \langle \boldsymbol{c}, \boldsymbol{c}' \rangle \le t + \Theta\left(\frac{1}{n}\right) \right\}. \qquad (45)$$

Further, let

$$C_i = \{\boldsymbol{c} \in C \colon (C(u_i, \boldsymbol{c}) \ne \emptyset) \text{ and } \forall_{j>i} (C(u_j, \boldsymbol{c}) = \emptyset)\}. \qquad (46)$$

$$\begin{aligned}\ln P_k^{\alpha,\beta}(y) = k \Bigg[&(1+a) H\left(\frac{a}{1+a}\right) - \frac{a}{2} \ln \frac{2(a+1)(y-1) + a^2(y+1) + a\sqrt{a^2(1+y)^2 - 4(1+a)(1-y^2)}}{4a^2} \\ &+ \frac{a+2}{2} \ln \frac{4y(1+a) + a^2(1+y) + (2+a)\sqrt{a^2(1+y)^2 - 4(1+a)(1-y^2)}}{4(a+1)^2} \Bigg] + o(k) \qquad (q(a, 0) < y \le 1)\end{aligned}$$

There are at most $\Theta(n)$ such subsets C_i; they all are pairwise-disjoint. The *effective distance* of C (measured in cosines) is defined as

$$\nu = \nu(C) := u_i \quad \Longleftrightarrow \quad |C_i| \geq |C_j|, \qquad j \neq i.$$

Let $C^{(1)} = C_i$ for this value of i. It is clear that $|C^{(1)}| \geq |C|/\Theta(n)$, i.e., exponentially this subcode has the same size. This proves the following lemma (below we omit the unessential constant in $\Theta(n)$).

Lemma 10: Let $C \subset S^{n-1}$ be a code. Then there exists a subset $C^{(1)} \subset C$ such that $t(C^{(1)}) = \nu(C)$ and every vector $\boldsymbol{c} \in C^{(1)}$ has a neighbor $\boldsymbol{c}' \in C(\nu, \boldsymbol{c})$. Moreover, $|C^{(1)}| \geq |C|/n$.

Since $|C^{(1)}| \overset{n}{\sim} |C|$, Theorem 9 implies that for some i there is a point in the partition (44) such that the exponent of the average distance density of $C^{(1)}$ is bounded below by the function $j^*(u_i, \rho, R)$. In the next theorem we isolate a subcode $C^{(2)} \subset C^{(1)}$ of the same exponential size as C with some additional properties. Namely, since we will have $R(C^{(2)}) = R$, Theorem 9 implies that for some i there is a point in the partition (44) such that the main term in the exponent of the average distance density of $C^{(2)}$ is bounded below by the function $j^*(u_i, \rho, R)$. We prove that on top of this $C^{(2)}$ can be chosen in such a way that all the *local* distance densities $b_{\boldsymbol{c}}(u_i)$, $\boldsymbol{c} \in C^{(2)}$, have at least the same exponential growth as $j^*(u_j, \rho, R)$. Of course, $t(C^{(2)}) \leq \nu(C)$.

Theorem 11: Let C be a code of rate R and distance $t = t(C)$. Let ρ, $0 \leq \rho \leq \rho_{kl}$, be a fixed number. There exists a subset $C^{(2)} \subset C^{(1)} \subset C$ such that $|C^{(2)}| \geq \frac{1}{2n^2}|C|$ and $t(C^{(2)}) \leq \nu(C)$. Moreover, there exists a number u, $\tau(\rho) \leq u \leq 1$, such that the average distance density in $C^{(2)}$ satisfies $b(u) \geq J(n, u, \rho, |C|/n)$ and for every vector $\boldsymbol{c} \in C^{(2)}$ the number of its neighbors in $C^{(1)}$

$$b_{\boldsymbol{c}}(u) \geq \frac{1}{2} J\left(n, u, \rho, \frac{|C|}{2n}\right). \tag{47}$$

Proof: Below we denote the average (local) distance density of a code D at a point $x \in [-1, 1]$ by $b(x, D)$ (resp., $b_{\boldsymbol{c}}(x, D)$).

We begin with the code $C^{(1)} \subset C$ constructed in Lemma 10 and show that it is possible to choose $C^{(2)}$ as a subset of $C^{(1)}$. Let

$$L_i = \left\{\boldsymbol{c} \in C^{(1)} \colon b_{\boldsymbol{c}}(u_i, C^{(1)}) \geq \frac{1}{2} J\left(n, u_i, \rho, \frac{|C^{(1)}|}{2}\right)\right\} \qquad (0 \leq i \leq m-1)$$

where $(u_i)_{i=0}^{m}$ is a defining sequence for partition of the form (44). If there is an i, $0 \leq i \leq m-1$, such that $|L_i| \geq \frac{1}{2m}|C^{(1)}|$, put $C^{(2)} = L_i$. This choice obviously satisfies the conditions of the theorem. Otherwise, consider the set

$$\tilde{C} = C^{(1)} \Big\backslash \bigcup_{i=0}^{m-1} L_i.$$

Clearly, $|\tilde{C}| \geq \frac{1}{2}|C^{(1)}|$ since the size $|L_i|$ for all i is by assumption at most $(1/2m)|C^{(1)}|$. By Theorem 9 there is an index $0 \leq i_0 \leq m-1$ such that

$$b(u_{i_0}, \tilde{C}) \geq J\left(n, u_{i_0}, \rho, \frac{|C^{(1)}|}{2}\right).$$

Hence in particular, there is a point $\tilde{\boldsymbol{c}} \in \tilde{C}$ such that

$$b_{\tilde{\boldsymbol{c}}}(u_{i_0}, \tilde{C}) \geq J\left(n, u_{i_0}, \rho, \frac{|C^{(1)}|}{2}\right).$$

Obviously $2b_{\tilde{\boldsymbol{c}}}(u_{i_0}, C^{(1)}) \geq b_{\tilde{\boldsymbol{c}}}(u_{i_0}, \tilde{C})$. Then by definition, $\tilde{\boldsymbol{c}} \in L_{i_0}$, contradicting the fact that all the subsets L_i, $0 \leq i \leq m-1$, are cast away from $C^{(1)}$. To complete the proof recall that $|C^{(1)}| \geq |C|/n$. □

This theorem establishes the existence in any code of a subcode with many neighbors in the vicinity of every code vector, i.e., some kind of distance invariance in the neighborhood of u_i.

V. Proof of Theorem 1

As in Section I, let W be a code on the sphere in $\mathbb{R}^n$ of radius $\sigma\sqrt{An}$, where σ^2 is the noise variance and A is the signal-to-noise ratio in the channel. In this section, we work with Euclidean distances in codes rather than with inner products. Distances on the sphere of radius $\sigma\sqrt{An}$ will be denoted by $\overline{d}$, $\overline{w}$, and so on, to distinguish them from distances on the unit sphere $(d, w, \ldots)$. Clearly, $\overline{d} = d\sigma\sqrt{An}$, and so on. By C we denote the projection of W on the concentric unit sphere S^{n-1}. Obviously, angular distances in C and W are equal. Let

$$W(\overline{w}, \boldsymbol{x}) = \left\{\boldsymbol{y} \in W \colon \overline{w} - \Theta\left(\frac{1}{\sqrt{n}}\right) \leq \|\boldsymbol{x} - \boldsymbol{y}\| \leq \overline{w}\right\}.$$

Note that the projection of $W(\overline{w}, \boldsymbol{x})$ on S^{n-1} gives the spherical ring $C(t, \boldsymbol{c})$ defined in (45), with the obvious relation between t and w. Distances in W and C are connected by the scaling $\overline{w} = w\sigma\sqrt{An}$.

For the rest of this section we assume that the rate R of W is between 0 and the channel capacity $\mathcal{C}$ and does not approach 0 as n grows. Let

$$\mathfrak{b}_{\boldsymbol{x}}(\overline{w}', \overline{w}) = \frac{1}{|W|}|\{(\boldsymbol{x}, \boldsymbol{y}) \in C^2 | \overline{w}' \leq \|\boldsymbol{x} - \boldsymbol{y}\| \leq \overline{w}\}|$$

be the local distance density related to $\boldsymbol{x} \in W$. As above, we omit one of the arguments and write $\mathfrak{b}_{\boldsymbol{x}}(\overline{w})$ if $\overline{w}' = \overline{w} - \Theta(\frac{1}{\sqrt{n}})$. The local distance distribution is by definition $\mathcal{B}_{\boldsymbol{x}}(\overline{w}) = \mathfrak{b}_{\boldsymbol{x}}(0, \overline{w})$. The average values of these functions over the code will be denoted $\mathfrak{b}(\overline{w})$, $\mathcal{B}(\overline{w})$. They are monotone nondecreasing functions of $\overline{w}$. We have

$$\mathfrak{b}(\overline{w}) = b\left(1 - \frac{\overline{w}^2}{2An\sigma}\right) = b\left(1 - \frac{w^2}{2}\right) \tag{48}$$

where $b(\cdot)$ is the distance function of C and w is the Euclidean distance. By virtue of this relation all our conventions and results of the previous section are readily translated into the present context. In particular, if $\nu(C)$ is the effective distance of C (measured in cosines), then $\overline{\nu}(W) = \sigma\sqrt{An}\sqrt{2 - 2\nu(C)}$ is the effective *Euclidean* distance of W.

Below we rely on two obvious facts which are worthwhile to isolate in a separate proposition.

Proposition 12: Let W be a code and $P_e(W)$ its error probability of decoding (8). Let $W'' \subseteq W' \subseteq W$ be two subsets of W. Then

$$P_e(W) \geq \frac{1}{|W|} \sum_{\boldsymbol{x} \in W''} \Pr(\mathbb{R}^n \setminus D(\boldsymbol{x}, W)) \tag{49}$$

$$\geq \frac{1}{|W|} \sum_{\boldsymbol{x} \in W''} \Pr(\mathbb{R}^n \setminus D(\boldsymbol{x}, W')) \tag{50}$$

where $D(\boldsymbol{x}, \cdot)$ is the Voronoi region (7).

Proof: The first inequality follows by putting in (8) $\Pr(D(\boldsymbol{x}, W)) = 0$ for $\boldsymbol{x} \in W \setminus W''$. The second one holds true since $D(\boldsymbol{x}, W) \subseteq D(\boldsymbol{x}, W')$. □

Suppose for a while that W possesses the regularity properties discussed in the previous section. Namely, every vector in W has a neighbor at an effective distance $\overline{\nu}(W)$ equal to its minimum distance $\overline{d}(W)$, and if $\mathfrak{b}(\overline{w})$ is the average distance density of W for a certain value of $\overline{w}$ then for all the local densities we have $\mathfrak{b}_{\boldsymbol{x}}(\overline{w}) = \mathfrak{b}(\overline{w})$.

Let us assume that $\boldsymbol{x} \in W$ is the transmitted vector (this condition is omitted from our notation below; it should be always kept in mind) and $\boldsymbol{z} \in \mathbb{R}^n$ is the received vector. Suppose that for some $\overline{w}$ the set $W(\overline{w}, \boldsymbol{x})$ is nonempty and let $\mathfrak{b}(\overline{w}) = |W(\overline{w}, \boldsymbol{x})|$. By (8) and Proposition 12 we have

$$P_e(W) \geq \mathfrak{b}(\overline{w}) \min_{\boldsymbol{y} \in W(\overline{w}, \boldsymbol{x})} \Pr(\boldsymbol{z} \in D(\boldsymbol{y}, (W(\overline{w}, \boldsymbol{x}) \cup \boldsymbol{x}))). \tag{51}$$

Let $\boldsymbol{y}_1 \in W(\overline{w}, \boldsymbol{x})$ be any code vector and let us bound the last probability below. Let

$$R_{\boldsymbol{x}}(\boldsymbol{y}) = \{\boldsymbol{z} \in \mathbb{R}^n: ||\boldsymbol{y} - \boldsymbol{z}|| < ||\boldsymbol{x} - \boldsymbol{z}||\}$$

be the half-space of points closer to $\boldsymbol{y}$ than to $\boldsymbol{x}$. We have the following chain of (in)equalities, which is just a one-step inclusion-exclusion argument:

$$\begin{aligned}
&\Pr(\boldsymbol{z} \in D(\boldsymbol{y}_1, (W(\overline{w}, \boldsymbol{x}) \cup \boldsymbol{x})))\\
&\quad= \Pr(\boldsymbol{z} \in R_{\boldsymbol{x}}(\boldsymbol{y}_1))\\
&\qquad - \sum_{\substack{\boldsymbol{y}_2' \in W(\overline{w}, \boldsymbol{x})\\ \boldsymbol{y}_2 \neq \boldsymbol{y}_1}} \Pr(\boldsymbol{z} \in [R_{\boldsymbol{x}}(\boldsymbol{y}_1) \cap (D(\boldsymbol{y}_2, (W(\overline{w}, \boldsymbol{x}) \cup \boldsymbol{x})))])\\
&\quad= \Pr(\boldsymbol{z} \in R_{\boldsymbol{x}}(\boldsymbol{y}_1))\\
&\qquad - \sum_{\substack{\boldsymbol{y}_2 \in W(\overline{w}, \boldsymbol{x})\\ \boldsymbol{y}_2 \neq \boldsymbol{y}_1}} \Pr\left\{\boldsymbol{z} \in R_{\boldsymbol{x}}(\boldsymbol{y}_1) \cap R_{\boldsymbol{y}_1}(\boldsymbol{y}_2)\right\}\\
&\quad\geq \Pr(\boldsymbol{z} \in R_{\boldsymbol{x}}(\boldsymbol{y}_1)) - (\mathfrak{b}(\overline{w}) - 1)\\
&\qquad \cdot \Pr\{\boldsymbol{z}: ||\boldsymbol{x} - \boldsymbol{z}|| \geq ||\boldsymbol{y}_1 - \boldsymbol{z}|| \geq ||\boldsymbol{y}_2 - \boldsymbol{z}|| \text{ and}\\
&\qquad\qquad ||\boldsymbol{y}_1 - \boldsymbol{y}_2|| = \overline{d}(W)\}
\end{aligned} \tag{52}$$

where the last inequality follows since the case of

$$\operatorname{dist}(\boldsymbol{y}_1, \boldsymbol{y}_2) = \overline{d}^2(W)$$

is the worst one for our estimate.

Observe that the distance between $\boldsymbol{x}$ and $\boldsymbol{y}_1$, $\boldsymbol{y}_2$ lies between $\overline{w} - \Theta(1/\sqrt{n})$ and $\overline{w}$. Since we are deriving lower estimates on $P_e(W)$ (and since the rate is below the capacity of the channel), we can assume that all these distances equal $\overline{w}$. Then by the definition of the channel and properties of the normal distribution, the first probability in (52) is asymptotic to $\Phi(\overline{w}/2\sigma) \overset{n}{\sim} \exp(-\overline{w}^2/8\sigma^2)$, where $\Phi(\cdot)$ is the $(0, 1)$ Gaussian distribution function. To bound $P_e(W)$ below we still need to compute the last probability in (52). Let $\boldsymbol{y}_1, \boldsymbol{y}_2$ be two code points in $W(\overline{w}, \boldsymbol{x})$. As mentioned above, the noise in the channel is a product of n independent and identically distributed (i.i.d.) Gaussian variables, each affecting the corresponding coordinate of $\boldsymbol{x}$. Let us have a closer look at the probabilities of error events that output $\boldsymbol{y}_1$ or $\boldsymbol{y}_2$ as a decoding result. These probabilities are completely determined by the pairwise distances in the triple $\boldsymbol{x}, \boldsymbol{y}_1, \boldsymbol{y}_2$ and the relative distances between $\boldsymbol{z}$ and these three points. Therefore, we can restrict our attention to the secant plane $\mathcal{P}$ defined by $\boldsymbol{x}, \boldsymbol{y}_1, \boldsymbol{y}_2$. Let us introduce affine coordinates in $\mathbb{R}^n$ in such a way that the origin is located at $\boldsymbol{x}$, the first coordinate vector is given by the direction $\boldsymbol{y}_1 - \boldsymbol{x}$, and the first two coordinates form an orthogonal basis for $\mathcal{P}$, making it into a linear space. Now let us write $\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{z}$ in these coordinates and restrict our attention to the plane $\mathcal{P}$, ignoring the remaining $n - 2$ coordinates (i.e., project $\boldsymbol{z}$ orthogonally on $\mathcal{P}$). Denote the corresponding two-dimensional vectors by $\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{e}$. Note that $\boldsymbol{e}$ is a random vector in $\mathcal{P}$ whose coordinates are i.i.d. Gaussian with mean zero and variance σ^2. The resulting picture is shown in Fig. 4, where $\mathbf{0}$ corresponds to $\boldsymbol{x}$.

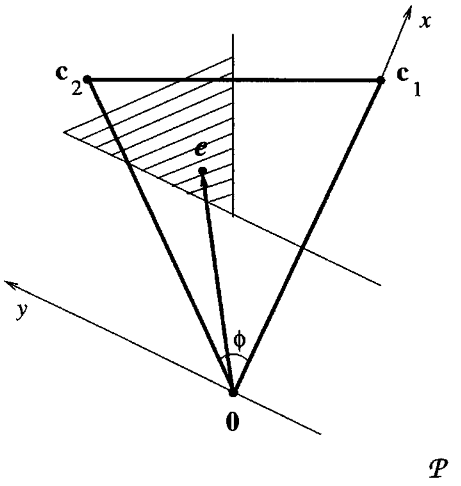

Fig. 4. To the proof of Lemma 13: We need to find the probability that the received vector $\boldsymbol{e}$ is in the stroked area.

Let us proceed to estimate the last probability in (52). This is done in the following lemma.

Lemma 13: Let $\boldsymbol{y}_1, \underline{\boldsymbol{y}_2} \in W(\overline{w}, \boldsymbol{x})$ be two code vectors at a distance $||\boldsymbol{y}_1 - \boldsymbol{y}_2|| = \overline{d}(W)$. Then

$$\Pr\{\boldsymbol{z}: ||\boldsymbol{x} - \boldsymbol{z}|| \geq ||\boldsymbol{y}_1 - \boldsymbol{z}|| \geq ||\boldsymbol{y}_2 - \boldsymbol{z}||\} \overset{n}{\sim} \exp\left(-G\frac{\overline{w}^2}{8\sigma^2}\right) \tag{53}$$

where $G = 4\overline{w}^2/(4\overline{w}^2 - \overline{d}^2(W))$.

Proof: According to the discussion before this lemma

$$\Pr\{\boldsymbol{z}: ||\boldsymbol{x} - \boldsymbol{z}|| \geq ||\boldsymbol{y}_1 - \boldsymbol{z}|| \geq ||\boldsymbol{y}_2 - \boldsymbol{z}||\} = \Pr\{\boldsymbol{e}: ||\boldsymbol{e}|| \geq ||\boldsymbol{c}_1 - \boldsymbol{e}|| \geq ||\boldsymbol{c}_2 - \boldsymbol{e}||\}$$

where the last probability is computed under a two-dimensional Gaussian noise centered at 0 with variance σ^2 along each coordinate. Further, we assume that $||\boldsymbol{c}_1||$ equals $||\boldsymbol{c}_2||$ and both are exactly w^2. From Fig. 4 we have $\boldsymbol{c}_1 = (\overline{w}, 0)$, $\boldsymbol{c}_2 = (\overline{w} \cos \phi, \overline{w} \sin \phi)$, where $\phi = 2 \arcsin \frac{\overline{d}}{2\overline{w}}$ is the angle between $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$. Suppose that $\boldsymbol{e} = (x, y)$. Then the condition that $\boldsymbol{e}$ is closer to $\boldsymbol{c}_2$ than to $\boldsymbol{c}_1$ yields the inequality

$$(x - \overline{w} \cos \phi)^2 + (y - \overline{w} \sin \phi)^2 < (x - \overline{w})^2 + y^2$$

or $x \frac{1-\cos \phi}{\sin \phi} < y < \infty$. Further, since $\boldsymbol{e}$ is closer to $\boldsymbol{c}_1$ than to 0, we also have $\frac{\overline{w}}{2} < x < \infty$. Denote the required probability by p. Then as $\overline{w}, \overline{d}(W) \to \infty$

$$\begin{aligned} p &= \frac{1}{2\pi\sigma^2} \int_{\overline{w}/2}^{\infty} e^{-\frac{x^2}{2\sigma^2}} \int_{\frac{1-\cos \phi}{\sin \phi} x}^{\infty} e^{-\frac{y^2}{2\sigma^2}}\, dx\, dy \\ &\overset{n}{\sim} \frac{\sin \phi}{2\pi\sigma(1+\cos \phi)} \int_{\overline{w}/2}^{\infty} e^{-\frac{x^2 G}{2\sigma^2}} \frac{dx}{x} \\ &= \frac{\sin \phi}{2\pi\sigma(1+\cos \phi)} \int_{\overline{w}\sqrt{G}/2}^{\infty} e^{-\frac{y^2}{2\sigma^2}} \frac{dy}{y} \\ &\overset{n}{\sim} \exp\left(-G\, \frac{\overline{w}^2}{8\sigma^2}\right), \end{aligned}$$

where we have used the identity

$$G = 1 + \frac{(1-\cos \phi)^2}{\sin^2 \phi}. \qquad \square$$

Note that (52) is only nontrivial if the second term is smaller than the first one. This restricts the number of code points that can be taken into account in this estimate. In other words, in some situations in our estimate we can only rely on a subcode of $W(\overline{w}, \boldsymbol{x})$. In the next lemma $\overline{d}$ is a short notation for $\overline{d}(W)$.

Lemma 14:

$$P_e(W) \geq L(\overline{w}, \overline{d}) \exp\left(-\frac{\overline{w}^2}{8\sigma^2} + o(n)\right) \tag{54}$$

where $L(\overline{w}, \overline{d}) = \min(\mathfrak{b}(\overline{w}), \hat{\mathfrak{b}}(\overline{w}, \overline{d}))$, and

$$\hat{\mathfrak{b}}(\overline{w}, \overline{d}) = \exp\left(\frac{\overline{d}^2 \overline{w}^2}{8\sigma^2(4\overline{w}^2 - \overline{d}^2)}\right).$$

Proof: For the inequality in (52) to be nontrivial we need the inequality at the bottom of this page. Let $\hat{\mathfrak{b}}(\overline{w}, \overline{d})$ be equal to the right-hand side of this inequality. If $\mathfrak{b}(\overline{w}) \leq \hat{\mathfrak{b}}(\overline{w}, \overline{d})$, then we can substitute $\mathfrak{b}(\overline{w})$ in (52). Otherwise, in this estimate we only take into account $\hat{\mathfrak{b}}(\overline{w}, \overline{d})$ code points from $W(\overline{w}, \boldsymbol{x})$, which is possible by Proposition 12.

Since

$$\Pr(\boldsymbol{z} \in R_{\boldsymbol{x}}(\boldsymbol{y}_1)) \overset{n}{\sim} \exp(-\overline{w}^2/8\sigma^2)$$

we obtain

$$\hat{\mathfrak{b}}(\overline{w}, \overline{d}) \overset{n}{\sim} \exp\left(-\frac{\overline{w}^2}{8\sigma^2}(1-G)\right) = \exp\left(\frac{\overline{d}^2 \overline{w}^2}{8\sigma^2(4\overline{w}^2 - \overline{d}^2)}\right).$$

Observe that $\hat{\mathfrak{b}}(\overline{w}, \overline{d})$ is a growing function of $\overline{d}$. Therefore, in the worst (for our estimates) case we must assume that every pair of vectors in $W(\overline{w}, \boldsymbol{x})$ is at a distance $\overline{d}$ apart.

Now let $L(\overline{w}, \overline{d}) = \min(\mathfrak{b}(\overline{w}), \hat{\mathfrak{b}}(\overline{w}, \overline{d}))$. By the above argument it is clear that

$$P_e(W) \gtrsim L(\overline{w}, \overline{d}) \Pr(\boldsymbol{z} \in D(\boldsymbol{y}_1, (W(\overline{w}, \boldsymbol{x}) \cup \boldsymbol{x})))$$

for any fixed $\boldsymbol{y}_1 \in W(\overline{w}, \boldsymbol{x})$. $\square$

Lemma 14 gives a lower estimate for the expression on the right-hand side in (51). Recall that by Shannon's minimum-distance bound (12) the error probability is bounded below by the the probability of confusing two code vectors at a distance $\overline{d}(W)$. Therefore,

$$P_e(W) \geq \max\left(\exp\left(-\frac{\overline{d}^2}{8\sigma^2} + o(n)\right), L(\overline{w}, \overline{d}) \exp\left(-\frac{\overline{w}^2}{8\sigma^2} + o(n)\right)\right) \tag{55}$$

provided that W is a distance-invariant code (here $\overline{d} = \overline{d}(W)$).

However, generally W is not distance-invariant. Therefore, we have to employ the asymptotic regularity results in the previous section. By Lemma 10, starting from W, one can construct a subset of code vectors $W^{(1)}$ of rate R in which the minimum distance equals the effective Euclidean distance. This is done by taking the code C, isolating in it the subset $C^{(1)}$ whose existence is proved in Lemma 10, and lifting it back to the sphere of radius $\sigma\sqrt{An}$. In this way, each vector in $W^{(1)}$ will have a neighbor at a distance $\overline{d}(W^{(1)}) = \overline{\nu}(W)$. Further, by Theorem 11, it is possible to isolate in $W^{(1)}$ a subset $W^{(2)}$ such that

i) $(1/n) \ln |W^{(2)}| \to R$;
ii) $\overline{d}(W^{(2)}) \geq \overline{d}(W^{(1)})$;

and, for a certain $\overline{w} = w\sigma\sqrt{An}$,

iii) its distance density $\mathfrak{b}(\overline{w})$ is bounded below by $J(n, 1-(1/2)w^2, \rho, |C|/(2n^2))$;
iv) for each $\boldsymbol{x} \in W^{(2)}(\overline{w}, \boldsymbol{x})$ the subset $W^{(1)}(\overline{w}, \boldsymbol{x})$ satisfies
$$e^{-nR}|W^{(1)}(\overline{w}, \boldsymbol{x})| \geq \exp(nj^*(1-\frac{1}{2}w^2, \rho, R) - o(n)).$$

Here $\rho \in [0, \rho_{kl}(R)]$ is a fixed number and $\overline{w}$ depends on ρ. By Property iii), the average density $\mathfrak{b}(\overline{w})$ is bounded below by $J(\cdot)$. If it is exponentially greater than J, this only improves our estimate, so the case of equality assumed below is the worst one. Moreover, by Property iv), for this $\overline{w}$ all the local densities have at least the same growth. Again the equality $\mathfrak{b}_{\boldsymbol{x}}(\overline{w}) = J$ assumed in the derivation of this section is the worst case, so our course of action is legitimate. The same applies to the equality $\overline{d}(W) = \overline{\nu}(W)$ of the minimum and effective distances assumed above.

We will use subcode $W^{(2)}$ to estimate from below the first and the second terms in (55), respectively. In doing so, we rely on Proposition 12. Let $\overline{d} = \overline{d}(W) \leq \overline{d}(W^{(1)}) \leq \overline{d}(W^{(2)})$. Since $(1/n) \ln |W^{(2)}| \to R$, the minimum-distance bound (12)

$$\mathfrak{b}(\overline{w}) \lesssim \frac{\Pr(\boldsymbol{z} \in R_{\boldsymbol{x}}(\boldsymbol{y}_1))}{\Pr\{\boldsymbol{z}\colon \|\boldsymbol{x} - \boldsymbol{z}\| \geq \|\boldsymbol{y}_1 - \boldsymbol{z}\| \geq \|\boldsymbol{y}_2 - \boldsymbol{z}\| \text{ and } \|\boldsymbol{y}_1 - \boldsymbol{y}_2\| = \overline{d}(W)\}}$$

still gives the first of the two terms in (55). Let us estimate the second term. Let ρ be fixed and $\overline{w}$ be the same as in the definition of $W^{(2)}$. We assume that $\boldsymbol{x}$ ranges over $W^{(2)}$ and consider only decoding errors producing code vectors in $W^{(1)}(\boldsymbol{x}, w)$. By virtue of Properties i)–ii), iv) of $W^{(2)}$ and (50) we have

$$\begin{aligned} P_e(W) &\geq L(\overline{w}, \overline{d}(W^{(2)})) \exp\left(-\frac{\overline{w}^2}{8\sigma^2} + o(n)\right) \\ &\geq L(\overline{w}, \overline{d}) \exp\left(-\frac{\overline{w}^2}{8\sigma^2} + o(n)\right). \end{aligned} \tag{56}$$

To complete the proof of Theorem 1 it remains to write everything in terms of distances on the unit sphere. Again let C be the code obtained by projecting W on the concentric unit sphere. Then, taking into account (55) and (56), we obtain

$$\begin{aligned} -\frac{1}{n} \ln P_e(W) &\lesssim \min\left(-\frac{\overline{d}^2}{8\sigma^2}, -\frac{\overline{w}^2}{8\sigma^2} + \ln L(\overline{w}, \overline{d})\right) \\ &= \min\left(-A\frac{d^2}{8}, -A\frac{w^2}{8} + \ln L(w, d)\right), \end{aligned}$$

where $0 \leq d \leq d_{kl}$,

$$L(w, d) = \min\left(\exp\left(\frac{Ad^2w^2}{8(4w^2 - d^2)}\right), j^*\left(1 - \frac{1}{2}w^2, \rho, R\right)\right);$$

$w, d \leq w \leq \sqrt{2}(\sqrt{1+\rho} - \sqrt{\rho})/\sqrt{1+2\rho}$, is the weight of "wrong" code vectors, j^* is defined in (41), and $\rho \in [0, \rho_{kl}(R)]$ is any fixed number. Inequality (14) is now immediate. □

VI. Discussion

The polynomial ("linear programming") method in coding theory was introduced in the founding works of Delsarte in 1972–1973. Its applicability to bounding the size of codes (and designs) was extended by Delsarte, Goethals, and Seidel [15], [16] to include the spherical case. Kabatiansky and Levenshtein [21] developed a general approach, based on harmonic analysis of noncommutative compact groups, to deriving bounds on packings in a very broad class of homogeneous spaces. Our paper further extends the scope of the polynomial method. Though technically speaking it is devoted to the proof of Theorem 1, on a more conceptual level it involves a large circle of ideas some of whose consequences are yet to be realized. A particular case of the linear programming method studied in coding theory hitherto relies upon the equality $|C| = \int dB(x)$, where $B(x)$ is the average distance distribution, and positivity conditions of the form (18). However, many other functionals of primary interest to coding/information theory, notably, the error probability of decoding, can be written as, or are related to, linear forms of the distance coefficients. This enables one to study bounds on these quantities in the same fashion as bounds on the size of codes and shows that many information-theoretic problems have their natural place in the geometric context of coding theory. Curiously, this possibility has been overlooked for about 25 years until having been explored recently in [3], [25], where we studied the discrete case.

Applications in [3] include bounds on the undetected error exponent, which is directly expressible via the distance coefficients. The same holds true for error probability of decoding up to any radius t for which the spheres around code vectors are disjoint. For larger t, and in particular, for maximum-likelihood decoding (t equals the covering radius of the code) it is not possible to write the error probability as a function of the distance distribution, but is possible to estimate it. This is in contrast to previous works which relied only on the minimum distance of the code, and explains improvements in upper bounds on error exponents in [3], [25], and the present paper.

Appendix
Jacobi Polynomials

A. Further Properties

The explicit expression for $P_k^{\alpha,\beta}(x)$ has the form

$$P_k^{\alpha,\beta}(x) = 2^{-k} \sum_{i=0}^{k} \binom{k+\alpha}{i} \binom{k+\beta}{k-i} (x+1)^i (x-1)^{k-i}. \tag{57}$$

This implies the following useful relation (the forward shift operator):

$$\frac{d}{dx} P_k^{\alpha,\beta}(x) = \frac{1}{2}(k+\alpha+\beta+1) P_{k-1}^{\alpha+1,\beta+1}(x). \tag{58}$$

From (57) one obtains

$$P_k^{\alpha,\beta}(1) = \binom{k+\alpha}{\alpha}. \tag{59}$$

It is known that $P_k^{\alpha,\beta}(x)$ satisfies the equation

$$(1-x^2)y'' + (\beta-\alpha-(\alpha+\beta+2)x)y' + k(k+\alpha+\beta+1)y = 0 \tag{60}$$

Polynomials $P_k^{\alpha,\beta}$ satisfy the Christoffel–Darboux formula

$$\begin{aligned} &\frac{\omega_k^{\alpha,\beta} L_{k+1}}{L_k} \sum_{i=0}^{k} \frac{1}{\omega_i^{\alpha,\beta}} P_i^{\alpha,\beta}(x) P_i^{\alpha,\beta}(y) \\ &\qquad = \frac{P_{k+1}^{\alpha,\beta}(x) P_k^{\alpha,\beta}(y) - P_k^{\alpha,\beta}(x) P_{k+1}^{\alpha,\beta}(y)}{x-y}. \end{aligned} \tag{61}$$

B. Proof of Theorem 3

We need a result from Sturm's comparison theory [35, p. 19].

Theorem 15: Let $f(x)$ and $F(x)$ be functions continuous in $x_0 < x < X_0$ with $f(x) \leq F(x)$. Let the functions $y(x)$ and $Y(x)$ satisfy the differential equations

$$y'' + f(x)y = 0 \quad \text{and} \quad Y'' + F(x)Y = 0,$$

respectively. Let x' and x'', $x' < x''$, be two consecutive zeros of $y(x)$. Then the function $Y(x)$ has at least one root in the interval $x' < x < x''$.

From (22) we have

$$\frac{1}{k} \ln |P_k^{\alpha,\beta}(x)| = \frac{1}{k} \ln L_k + \frac{1}{k} \sum_{i=1}^{k} \ln |x - t_i|, \qquad x \neq t_i. \tag{62}$$

Rewrite the second term in the last expression as

$$\sum_{i=1}^{k-1} \frac{\ln |x - t_i|}{k(t_i - t_{i+1})} (t_i - t_{i+1}). \tag{63}$$

(Note that the segment containing x gives rise to two terms; we have omitted one of them since it does not affect the main term of the answer. Similarly, we omit the last term in the sum.)

To estimate the distance $t_i - t_{i+1}$, apply a transformation of (60) discussed in [35, Ch. 1,4, esp. p. 67]. It can be checked that the function

$$u(x) = (1-x)^{(\alpha+1)/2}(1+x)^{(\beta+1)/2} P_k^{\alpha,\beta}(x)$$

satisfies the following equation:

$$u'' + N(x)u = 0 \tag{64}$$

where

$$N(x) = \frac{1}{4}\,\frac{1-\alpha^2}{(1-x)^2} + \frac{1}{4}\,\frac{1-\beta^2}{(1+x)^2} + \frac{k(k+\alpha+\beta+1)+(\alpha+1)(\beta+1)/2}{1-x^2}.$$

The set of zeros of $u(x)$ is the same as of $P_k^{\alpha,\beta}(x)$. Since $k \to \infty$, the quantity $t_i - t_{i+1}$ is $o(1)$; so we also have

$$\frac{1}{k^2}N(t_i) = \frac{1}{k^2}N(t_{i+1})\left(1 \pm O\left(\frac{1}{k}\right)\right).$$

Note that $(1/k^2)N(q(a,b))$ and $(1/k^2)N(-q(b,a))$ both vanish as $k \to \infty$ and that in the limit $N(x) > 0$ for $x \in (-q(b,a), q(a,b))$ and $N(x) < 0$ otherwise. The idea is to replace equation (64) in the segment $[t_{i+1}, t_i]$ by an equation with a constant coefficient N; its solution $\sin\sqrt{N}x$ will approximate the required distance.

More specifically, let

$$N_{\blacktriangle i} = \max_{x\in[t_{i+1},\,t_i]} N(x)$$
$$N_{\blacktriangledown i} = \min_{x\in[t_{i+1},\,t_i]} N(x)$$

(typically, the extrema are attained at the endpoints of the segment since $N'(x)$ has at most a constant number of sign changes). Applying Theorem 15, we obtain

$$\frac{\pi}{\sqrt{N_{\blacktriangle i}}} \le t_i - t_{i+1} \le \frac{\pi}{\sqrt{N_{\blacktriangledown i}}}.$$

In particular, this implies that the distance between consecutive zeros falls as k^{-1}. Let $z \in (t_{i+1}, t_i)$ be a point such that $t_i - t_{i+1} = \pi/\sqrt{N(z)}$. Then we have

$$\begin{aligned} k(t_i - t_{i+1}) &= \frac{\pi k}{\sqrt{N(z)}} \\ &= \frac{\pi}{\sqrt{\frac{1+a+b+ab/2}{1-t_i^2} + \frac{a^2}{4(1-t_i^2)} - \frac{b^2}{4(1+t_i^2)} + O\left(\frac{1}{k}\right)}} \\ &= \frac{\pi}{\sqrt{\frac{1+a+b+ab/2}{1-t_i^2} + \frac{a^2}{4(1-t_i^2)} - \frac{b^2}{4(1+t_i^2)}}} \\ &\quad\cdot\left(1 + O\left(\frac{1}{k}\right)\right). \end{aligned}$$

Substituting this into (63) and letting $k \to \infty$, we observe that the sum in (63) converges to the integral in (24). Using this in (62), we obtain the required expression.

Note that this argument gives the exact value for the main term of the exponent of $P_k^{\alpha,\beta}(x)$ for all $x \ne t_i$, $1 \le i \le k$.

C. *Proof of Theorem 4*

As above, let $t_1 > \cdots > t_k$ be zeros of $P_k^{\alpha,\beta}$. Let $\epsilon_k = k^{-\gamma}$, $0 < \gamma < 1/2$. We consider only the case $x \in [q(a,b)+\epsilon_k, 1]$ since the case $x \in [-1, -q(b,a)-\epsilon_k]$ is almost identical. In the computations below we have suppressed superscripts α, β. We have

$$P_k(x) = L_k \prod_{i=1}^{k}(x - t_i)$$

so

$$\frac{P_k'(x)}{P_k(x)} = \sum_{i=1}^{k}\frac{1}{x - t_i}$$

and

$$\frac{d}{dx}\,\frac{P_k'}{P_k} = \frac{P_k''P_k - (P_k')^2}{P_k^2} = -\sum_{i=1}^{k}\frac{1}{(x-t_i)^2} = -kO(\epsilon_k^{-2}). \tag{65}$$

This gives

$$\frac{P_k''}{P_k} = \left(\frac{P_k'}{P_k}\right)^2 - kO(\epsilon_k^{-2}).$$

Now consider (60) in a segment where y has no zeros. Then it can be rewritten as

$$(1-x^2)\frac{y''}{y} + (\beta-\alpha-(\alpha+\beta+2)x)\frac{y'}{y} + k(k+\alpha+\beta+1) = 0.$$

Let $y = P_k$ and $u = \frac{y'}{y}$. Then we obtain

$$\begin{aligned} &(1-x^2)u^2 + (\beta-\alpha-(\alpha+\beta+2)x)u \\ &\qquad + [k(k+\alpha+\beta+1) - kO(\epsilon_k^{-2})(1-x^2)] = 0. \end{aligned}$$

This is a quadratic with respect to u. Let us use (58) and (59) to compute

$$\begin{aligned} \frac{(P_k^{\alpha,\beta}(1))'}{P_k^{\alpha,\beta}(1)} &= \frac{k+\alpha+\beta+1}{2}\,\frac{P_{k-1}^{\alpha+1,\beta+1}(1)}{P_k^{\alpha,\beta}(1)} \\ &= \frac{k+\alpha+\beta+1}{2}\,\frac{\binom{k+\alpha}{\alpha+1}}{\binom{k+\alpha}{\alpha}} \\ &= \frac{k(k+\alpha+\beta+1)}{2(\alpha+1)}. \end{aligned}$$

On the other hand, taking u in the form of the equation shown at the bottom of this page, we compute

$$u(1) = \frac{k(k+\alpha+\beta+1)}{2(\alpha+1)}.$$

To establish that the choice of the sign in the solution of the quadratic equation is uniform over $x \in (q(a,b), 1]$, observe that the only zero of the expression under the square root in this interval converges to $q(a,b)$. So the second term in the formula for $u(x)$ does not (for large k) become zero for $x \in (q(a,b), 1]$; hence by continuity of $u(x)$ the choice of the sign is uniform.

Now observe that

$$u = \frac{d}{dx}\ln y(x)$$

$$u(x) = \frac{(\alpha+(\alpha+\beta+2)x-\beta)}{2(1-x^2)} - \frac{\sqrt{(\alpha+(\alpha+\beta+2)x-\beta)^2 - 4(1-x^2)[k(k+\alpha+\beta+1) - kO(\epsilon_k^{-2})(1-x^2)]}}{2(1-x^2)}$$

whence we get for $q(a, b) + \epsilon_k \le x \le 1$, ${\epsilon_k}^{-1} = o(\sqrt{k})$

$$\frac{1}{k}\int_x^1 d\,\ln P_k(z) = \int_x^1 \left[\frac{(a+(a+b)z-b)}{2(1-z^2)} - \frac{\sqrt{(a+(a+b)z-b)^2 - 4(1-z^2)(1+a+b)}}{2(1-z^2)}\right] dz + o(1). \tag{66}$$

The final answer is obtained by invoking the boundary condition (59) and collecting the $o(1)$ terms.

D. Proof of Theorem 5

First let $q(a, b) \le x \le 1$. We know that in this segment $|P_k^{\alpha,\beta}|$ is monotone increasing. Let us partition the segment $[q(a, b), 1]$ into m equal subsegments U_i and denote their endpoints by $u_0 = q(a, b), u_1, \ldots, u_{m-1}, u_m = 1$. Note that in this segment $\mu(x)$ is a falling function (recall that $\alpha \ge \beta$). For $x \in [u_{i-1}, u_i]$ we have

$$\begin{aligned}\frac{1-q(a, b)}{m}\,&|P_k^{\alpha,\beta}(x)|^2 \mu(u_{i+1}) \\ &< \sum_{i=0}^{m-1} \frac{1-q(a, b)}{m} |P_k^{\alpha,\beta}(u_i)|^2 \mu(u_{i+1}) \\ &< \int_{q(a, b)}^1 |P_k^{\alpha,\beta}(x)|^2 \mu(x)\,dx < \omega_k^{\alpha,\beta}\end{aligned}$$

the last inequality by (20). Now assume that m grows, for instance linearly, in k. Then if there is a point $x \in [q(a, b), 1]$ at which (26) is violated, the lower Darboux sum will also exponentially exceed $\omega_k^{\alpha,\beta}$, a contradiction.

The argument for $-1 \le x \le -q(b, a)$ is a slight variation of the above since the segment may contain the maximum of $\mu(x)$. Then one must be careful in choosing the points u_i to substitute in the Darboux sum above; otherwise, the logic is the same.

Now let $-q(b, a) \le x \le q(a, b)$ and let $t_k < t_{k-1} < \cdots < t_1$ be the zeros of $P_k^{\alpha,\beta}(x)$. Likewise, let t_i', $1 \le i \le k-1$, be the zeros of $P_{k-1}^{\alpha+1,\beta+1}(x)$. Let $m_i = |P_k^{\alpha,\beta}(t_i')|$ be the values of the maxima of $|P_k^{\alpha,\beta}(x)|$ (58). Obviously, $t_{i+1} < t_i' < t_i$, $1 \le i \le k-1$.

From (60) we have for $x = t_i$

$$\frac{y'}{y''} = \frac{1-x^2}{\alpha + (\alpha+\beta+2) - \beta}$$

i.e., at these points y' and y'' have the same sign. Hence for every interval $[t_{i+1}, t_i]$ the function $|P_k^{\alpha,\beta}(x)|$ is concave either for $x \in [t_i', t_i]$ or for $x \in [t_{i+1}, t_i']$. We will treat only the first case; the second one is analogous.

By Theorem 3, the function $(1/k)\ln|P_k(x)|$ converges pointwise to the limit function (24), which is continuous for all $x \in [-1, 1]$. Therefore, the quotient $(1/k)\ln m_{i+1}/m_i$ tends to 0 as k grows. Then by (22) we see that $(t_i - t_i')^{-1}$ grows slower that any exponential function in k. This is sufficient to bound m_i above.

Indeed, consider the function

$$\psi_i(x) = \frac{m_i}{t_i' - t_i}(x - t_i), \qquad x \in [t_i', t_i]$$

i.e., a segment of the straight line connecting the points (t_i', m_i) and $(t_i, 0)$. Since by assumption $|P_k^{\alpha,\beta}|$ for $x \in [t_i', t_i]$ is concave, in this segment we have $|P_k^{\alpha,\beta}(x)| \ge \psi_i(x)$. So letting

$$I_{k,i} = \int_{t_i'}^{t_i} |P_k^{\alpha,\beta}(x)|^2 (1-x)^\alpha (1+x)^\beta dx$$

we obtain for $x \in [t_i', t_i]$

$$\begin{aligned}\omega_k^{\alpha,\beta} &\ge I_{k,i} \ge (1-t_i)^\alpha (1+t_{i+1})^\beta \int_{t_i'}^{t_i} \psi_i^2(x)\,dx \\ &= (1-x-o(1))^\alpha (1+x-o(1))^\beta \frac{m_i^2}{3}(t_i - t_i')\end{aligned}$$

where $o(1)$ is of order k^{-1} or less. This gives

$$m_i^2 \le 3(1-x-o(1))^{-\alpha}(1+x-o(1))^{-\beta}\omega_k^{\alpha,\beta}(t_i - t_i')^{-1}$$

proving the theorem.

E. Proof of Theorem 6

To obtain part a) from (27) let us break the integral in (24) into two terms

$$\int_{-q}^q = \int_{-q}^0 + \int_0^q.$$

Changing the variable $z \to -z$ in the integral over $[-q, 0]$ and simplifying, we arrive at

$$\int_{-q}^0 \ldots dz = \int_0^q \frac{\ln(x+z)\sqrt{2a+1-(a+1)z^2}}{1-z^2}\,dz.$$

This leads to the integral in (27). To complete the proof of part a) we have to derive an asymptotic expression for the leading coefficient L_k. For $a = b$ it is easy to see that $L_k \overset{k}{\sim} 2^{-k}\binom{k+\alpha}{k/2}^2$, so

$$\frac{1}{k}\ln L_k = -\ln 2 + 2(a+1)H\left(\frac{1}{2(a+1)}\right) + o(1).$$

Part b) follows directly from (25) upon substituting $b = a$. In (28), we have moved the $\sqrt{\cdot}$ to the denominator to underline that the integral does not have a singularity at $z = 1$.

Part c) follows upon writing the asymptotic expression for $\omega_k^{\alpha,\alpha}$. From (21) we obtain

$$\omega_k^{\alpha,\alpha} \overset{k}{\sim} \frac{2^{2\alpha}\Gamma^2(k+\alpha)}{\Gamma(k+1)\Gamma(k+2\alpha)}.$$

From the Stirling formula we obtain the following asymptotic equality ($p \to \infty$):

$$2\ln(2^p\Gamma(p)) = \ln\Gamma(2p) + o(1).$$

So, neglecting the vanishing terms

$$\begin{aligned}\frac{1}{k}\ln\omega_k^{\alpha,\alpha} &= \frac{1}{k}\ln\frac{2^{-2k}\Gamma(2(k+\alpha))}{\Gamma(k)\Gamma(k+2\alpha)} \\ &= -2\ln 2 + 2(1+a)H\left(\frac{1}{2(1+a)}\right).\end{aligned} \tag{67}$$

This completes the proof.

ACKNOWLEDGMENT

The authors wish to thank V. Levenshtein for sending them a copy of [24] and G. Kabatiansky for careful reading of an earlier draft. In particular, he has simplified considerably Theorem 11. A. B. would like to acknowledge enjoyable and useful conversations with H. Landau on orthogonal polynomials.

REFERENCES

[1] M. Aaltonen, “Notes on the asymptotic behavior of the information rate of block codes,” *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 84–85, Jan. 1984.

[2] M. J. Aaltonen, “Linear programming bounds for tree codes,” *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 85–90, Jan. 1979.

[3] A. Ashikhmin and A. Barg, “Binomial moments of the distance distribution: Bounds and applications,” *IEEE Trans. Inform. Theory*, vol. 45, pp. 438–452, Mar. 1999.

[4] A. Ashikhmin, A. Barg, E. Knill, and S. Litsyn, “Quantum error detection, II,” *IEEE Trans. Inform. Theory*, vol. 45, pp. 789–800, May 2000.

[5] A. Ashikhmin, A. Barg, and S. Litsyn, “Polynomial method in coding and information theory,” Lanl e-print, math.CO/9910175, 1999.

[6] ——, “New bounds on generalized weights,” *IEEE Trans. Inform. Theory*, vol. 45, pp. 1258–1263, May 1999.

[7] ——, “A new upper bound on codes decodable into size-2 lists,” in *Numbers, Information and Complexity*, I. Althöfer, Ed. Boston, MA: Kluwer, 2000, pp. 239–244.

[8] A. Ashikhmin and S. Litsyn, “Upper bounds on the size of quantum codes,” *IEEE Trans. Inform. Theory*, vol. 45, pp. 1206–1215, May 1999.

[9] M. V. Burnashev, “On relation between code spectrum and decoding error probability,” *Probl. Inform. Transm.*, to be published.

[10] E. Cartan, “Sur la détermination d’un système orthogonal complet dans un espace de Riemann symmétrique clos,” *Rend. Circ. Mat. Palermo*, vol. 53, pp. 217–252, 1929.

[11] L.-C. Chen and M. E. H. Ismail, “On asymptotics of Jacobi polynomials,” *SIAM J. Math. Anal.*, vol. 22, no. 5, pp. 1442–1449, 1991.

[12] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, “Packing lines, planes, etc.: Packings in Grassmannian spaces,” *Exper. Math.*, vol. 5, no. 2, pp. 139–159, 1996.

[13] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*. New York, NY/Berlin, Germany: Springer-Verlag, 1988.

[14] P. Delsarte, “An algebraic approach to the association schemes of coding theory,” *Philips Res. Repts. Suppl.*, vol. 10, pp. 1–97, 1973.

[15] P. Delsarte, J. M. Goethals, and J. J. Seidel, “Bounds for systems of lines and Jacobi polynomials,” *Philips Res. Repts.*, vol. 30, pp. 91*–105*, 1975.

[16] ——, “Spherical codes and designs,” *Geometriae Dedicata*, vol. 6, pp. 363–388, 1977.

[17] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions. Vols. I—III (Bateman Manuscript Project)*. New York, NY/Toronto, Canada/London, U.K.: McGraw-Hill, 1953–1955.

[18] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[19] S. Helgason, *Differential Geometry and Symmetric Spaces*. New York: Academic, 1962.

[20] E. Hewitt and K. A. Ross, *Abstract Harmonic Analysis, Vol. II*. New York: Springer, 1970.

[21] G. Kabatyansky and V. I. Levenshtein, “Bounds for packings on the sphere and in the space,” *Probl. Pered. Inform.*, vol. 14, no. 1, pp. 3–25, 1978.

[22] G. Kalai and N. Linial, “On the distance distribution of codes,” *IEEE Trans. Inform. Theory*, vol. 41, pp. 1467–1472, Sept. 1995.

[23] V. I. Levenshtein, “Bounds for packings of metric spaces and some of their applications” (in Russian), *Probl. Kibernet.*, pp. 43–110, 1983.

[24] ——, “Method of nonnegative definite functions in metric problems of coding theory,” Doctor of Science Dissertation (in Russian), Moscow, USSR, 1983.

[25] S. Litsyn, “New upper bounds on error exponents,” *IEEE Trans. Inform. Theory*, vol. 45, pp. 385–398, Mar. 1999.

[26] R. J. McEliece, E. R. Rodemich, H. Rumsey, and L. R. Welch, “New upper bound on the rate of a code via the Delsarte–MacWilliams inequalities,” *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 157–166, Mar. 1977.

[27] D. S. Moak, E. B. Saff, and R. S. Varga, “On the zeros of Jacobi polynomials $P_n^{(\alpha_n, \beta_n)}(x)$,” *Trans. Amer. Math. Soc.*, vol. 249, no. 1, pp. 159–162, 1979.

[28] P. M. Piret, “Bounds for codes over the unit circle,” *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 760–767, Nov. 1986.

[29] G. Sh. Poltyrev, “Bounds on the decoding error probability of binary linear codes via their spectra,” *IEEE Trans. Inform. Theory*, vol. 40, pp. 1284–1292, July 1994.

[30] R. A. Rankin, “The closest packing of spherical caps in n dimensions,” *Proc. Glasgow Math. Assoc.*, vol. 2, pp. 139–144, 1955.

[31] C. E. Shannon, “Probability of error for optimal codes in a Gaussian channel,” *Bell Syst. Tech. J.*, vol. 38, no. 3, pp. 611–656, 1959.

[32] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, “Lower bounds to error probability for codes on discrete memoryless channels—II,” *Inform. Contr.*, vol. 10, pp. 522–552, 1967.

[33] A. Yu. Sheverdyaev, “Decoding methods in channels with noise,” Ph.D. dissertation (in Russian), Moscow, USSR, 1971.

[34] D. Slepian, “Group codes for the Gaussian channel,” *Bell Syst. Tech. J.*, vol. 47, pp. 575–602, 1968.

[35] G. Szegö, “Orthogonal polynomials,” in *Colloquium Publications*. Providence, RI: Amer. Math. Soc., 1975.

[36] N. Ja. Vilenkin and A. U. Klimyk, *Representations of Lie Groups and Special Functions*. Dordrecht, The Netherlands: Kluwer, 1991.

[37] N. Ya. Vilenkin, *Special Functions and the Theory of Group Representations*. Providence, RI: Amer. Math. Soc., 1968.