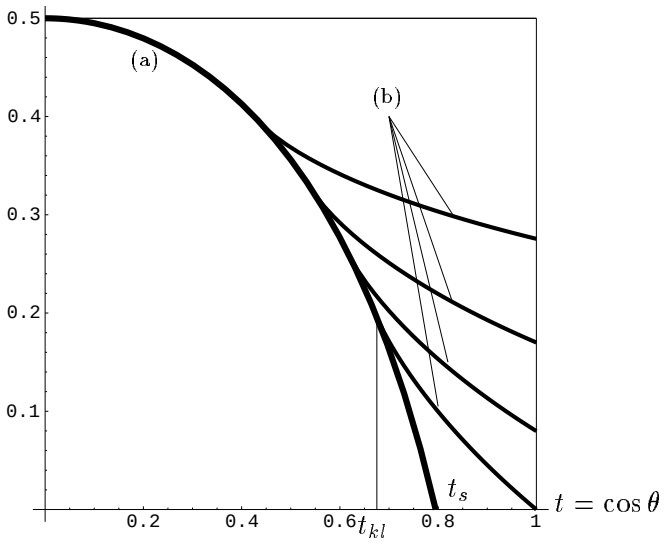


Figure 3: Bounds on the distance distribution of codes ($R = 0.5$): (a) "random" distance spectrum (46), (b) Bound (52) (The first of the four curves corresponds to $\rho_{kl} = 0.179$. The remaining 3 curves are drawn for $\rho = 0.139, 0.099, 0.059$, respectively); $t_s = 0.795$–the Shannon (Varshamov–Gilbert) distance; $t_{kl} = 0.675$–the KL distance.

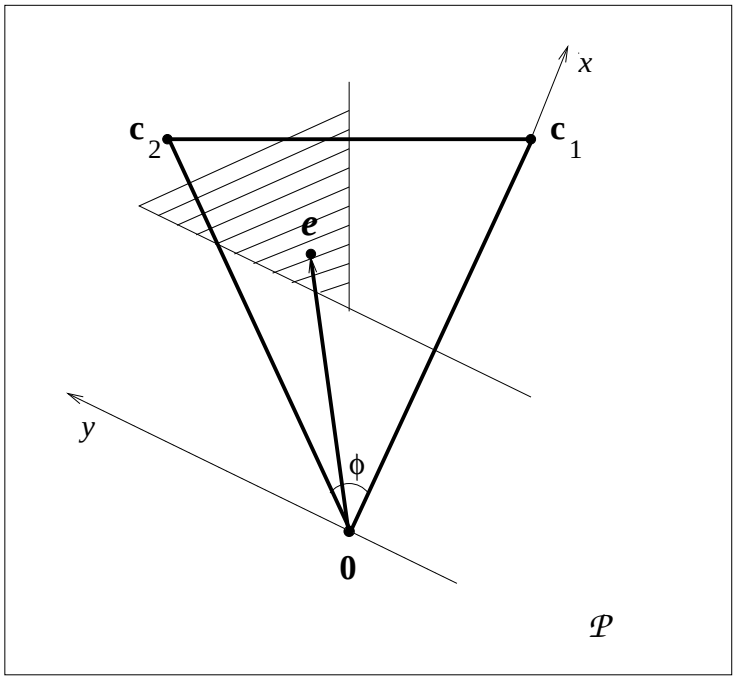


Figure 4: To the proof of the 3-point lemma: We need to find the probability that the received vector $\mathbf{e}$ is in the stroked area.

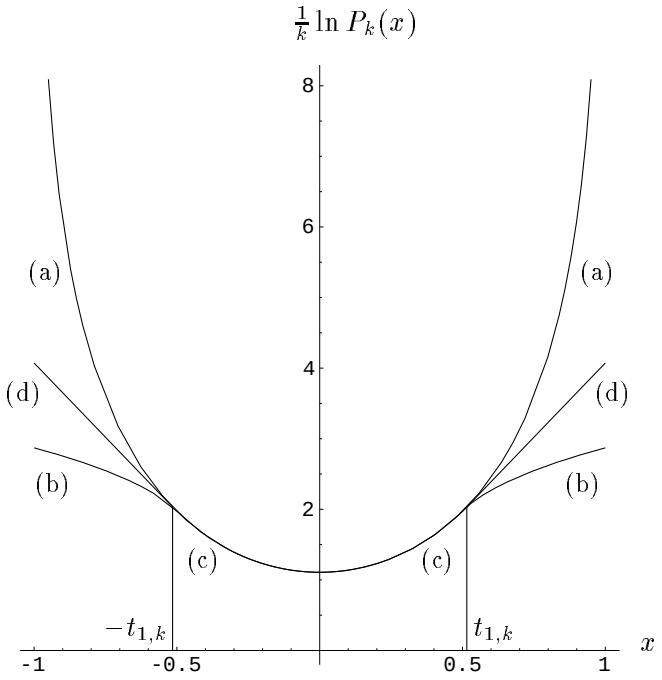


Figure 2: Exponent of the Jacobi polynomials $P_k^{ak,ak}(k \to \infty, a = 6)$: (a) upper bound (42), valid for $x \in [-1,1]$; (b) exact expression (41), valid for $t_{1,k} \leq |x| \leq 1$; (c) upper bound (40), valid for $x \in [-1,1]$; (d) upper bound (44).

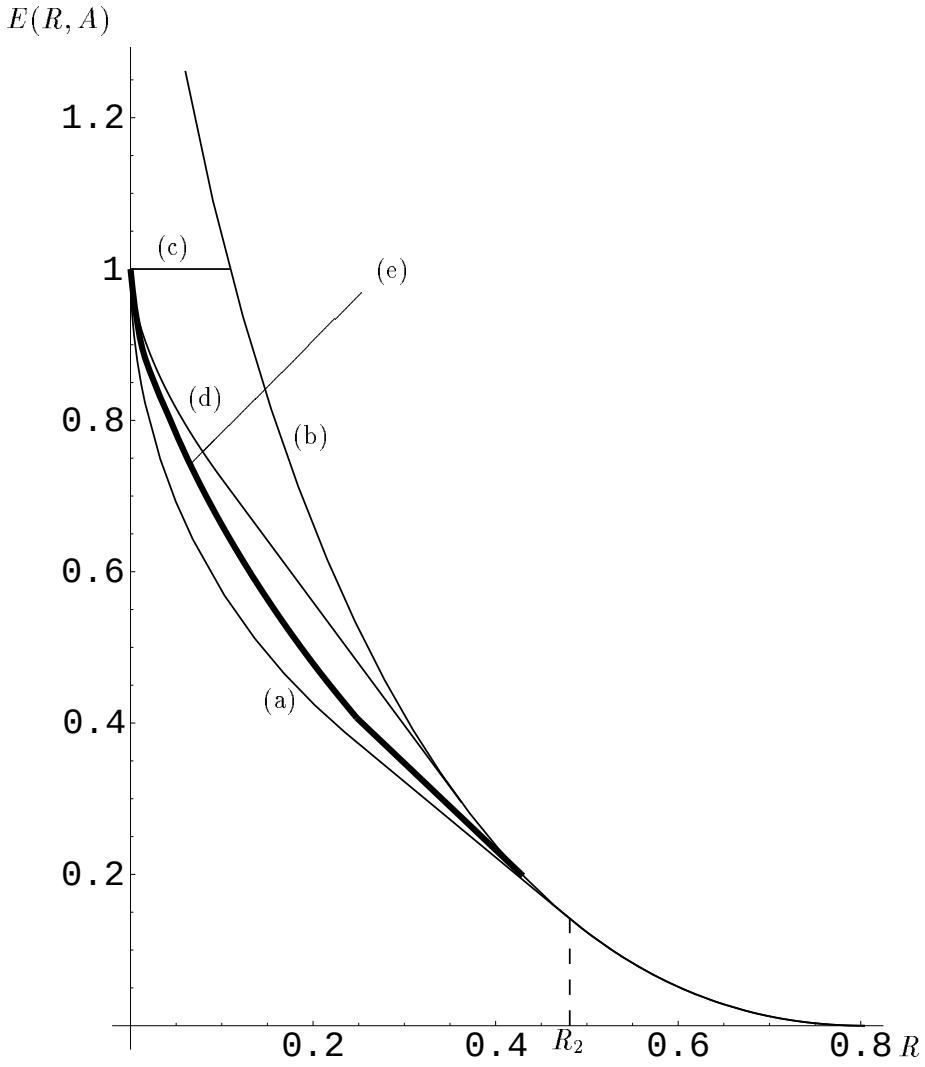


Figure 1: Bounds on the reliability function ($A = 4$): (a) "random coding" exponent (10.I-III), (b) sphere-packing bound (9), (c) minimum-distance bound (11), (d) minimum-distance bound (12), (e) the new bound (14). R_2 denotes the critical rate. Each of the curves (c)-(e) includes a segment of the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

[21] P. M. Piret, *Bounds for codes over the unit circle*, IEEE Trans. Inform. Theory **31** (1986), no. 6, 760–767.

[22] G. Sh. Poltyrev, *Bounds on the decoding error probability of binary linear codes via their spectra*, IEEE Trans. Inform. Theory **40** (1994), no. 4, 1284–1292.

[23] R. A. Rankin, *The closest packing of spherical caps in n dimensions*, Proc. Glasgow Math. Assoc. **2** (1955), 139–144.

[24] C. E. Shannon, *Probability of error for optimal codes in Gaussian channel*, Bell Syst. Techn. Journ. **38** (1959), no. 3, 611–656.

[25] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, *Lower bounds to error probability for codes on discrete memoryless channels, II*, Information and Control **10** (1967), 522–552.

[26] A. Yu. Sheverdyaev, *Decoding methods in channels with noise*, Moscow, 1971, Ph. D. Thesis (in Russian).

[27] D. Slepian, *Group codes for the Gaussian channel*, Bell Syst. Techn. Journ. **47** (1968), 575–602.

[28] G. Szegö, *Orthogonal polynomials*, Colloquium Publications, vol. 23, AMS, Providence, RI, 1975.

[29] N. Ya Vilenkin, *Special functions and the theory of group representations*, AMS, Providence, RI, 1968.

In this paper we worked out all the details of the spherical case. Clearly, the machinery introduced in it enables one to prove theorems concerning the distance distribution/reliability function for codes/channels in other metric spaces mentioned in the end of Section 2.

Acknowledgments. We are grateful to Vladimir Levenshtein for sending us a copy of [18]. The second named author would like to acknowledge enjoyable and useful conversations with Henry Landau on orthogonal polynomials.

References

[1] M. J. Aaltonen, *Linear programming bounds for tree codes*, IEEE Trans. Inform. Theory **25** (1977), 85–90.

[2] A. Ashikhmin and A. Barg, *Binomial moments of the distance distribution: Bounds and applications*, IEEE Trans. Inform. Theory **45** (1999), no. 2.

[3] A. Ashikhmin, A. Barg, E. Knill, and S. Litsyn, *Quantum error detection*, IEEE Trans. Inform. Theory, submitted.

[4] A. Ashikhmin, A. Barg, and S. Litsyn, *A new upper bound on codes decodable into size-2 lists*, Numbers, Information and Complexity, Kluwer Publishers.

[5] ———, *New bounds on generalized weights*, IEEE Trans. Inform. Theory (1999), to appear.

[6] A. Ashikhmin and S. Litsyn, *Upper bounds on the size of quantum codes*, IEEE Trans. Inform. Theory (1999).

[7] E. Cartan, *Sur la détermination d'un système orthogonal complet dans un espace de Riemann symmétrique clos*, Rend. Circ. Mat. Palermo **53** (1929), 217–252.

[8] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, *Packing lines, planes, etc.: Packings in grassmanian spaces*, Experimental Mathematics **5** (1996), no. 2, 139–159.

[9] J. H. Conway and N. J. A. Sloane, *Sphere packings, lattices and groups*, Springer-Verlag, New York-Berlin, 1988, With contributions by E. Bannai, J. Leech, S. P. Norton, A. M. Odlyzko, R. A. Parker, L. Queen and B. B. Venkov.

[10] P Delsarte, *An algebraic approach to the association schemes of coding theory*, Philips Research Repts Suppl. **10** (1973), 1–97.

[11] P. Delsarte, J. M. Goethals, and J. J. Seidel, *Bounds for systems of lines and Jacobi polynomials*, Philips Res. Repts. **30** (1975), 91*–105*.

[12] ———, *Spherical codes and designs*, Geometriae Dedicata **6** (1977), 363–388.

[13] R. G. Gallager, *Information theory and reliable communication*, John Wiley & Sons, New York e.a., 1968.

[14] S. Helgason, *Differential geometry and symmetric spaces*, Academic Press, 1962.

[15] G. Kabatyansky and V. I. Levenshtein, *Bounds for packings on the sphere and in the space*, Problemy Peredachi Informatsii **14** (1978), no. 1, 3–25.

[16] G. Kalai and N. Linial, *On the distance distribution of codes*, IT **41** (1995), no. 5, 1467–1472.

[17] V. I. Levenshtein, *Bounds for packings of metric spaces and some of their applications*, Problemy Kibernet. (1983), no. 40, 43–110, (In Russian).

[18] V. I. Levenshtein, *Method of nonnegative definite functions in metric problems of coding theory*, Moscow, 1983, Doctor of Science Dissertation (In Russian).

[19] S. Litsyn, *New bounds on error exponents*, IEEE Trans. Inform. Theory **45** (1999), no. 2.

[20] 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 **23** (1977), no. 2, 157–166.

sphere. Then, taking into account (68), (69), we obtain

$$-\frac{1}{n}\ln P_e(W) \lesssim \min\Big(-\frac{\bar{d}^2}{8\sigma^2}, -\frac{\bar{w}^2}{8\sigma^2} + \ln L(\bar{w},\bar{d})\Big)$$
$$= \min\Big(-A\frac{d^2}{8}, -A\frac{w^2}{8} + \ln L(w,d)\Big),$$

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

$$L(w,d) = \min\Big(\exp\Big(\frac{Ad^2w^2}{8(4w^2-d^2)}\Big), b^*\big(1-\frac{1}{2}w^2,\rho,R\big)\Big),$$

w ($d \le w \le \sqrt{2}(\sqrt{1+\rho}-\sqrt{\rho})/\sqrt{1+2\rho}$) is the weight of "wrong" code vectors, b^* is defined in (52), and $\rho \in [0, \rho_{kl}(R)]$ is any fixed number. Inequality (14) is now immediate. ∎

6 Discussion

The polynomial ("linear programming") method in coding theory was defined in the founding works of Delsarte in 1972-73. Its applicability to bounding the size of codes (and designs) was extended by Delsarte, Goethals, and Seidel [11],[12] to include spherical codes by showing a possibility to use representations of the underlying isometry group. Kabatiansky and Levenshtein [15] used this idea in full generality to derive bounds on packings in a very broad class of Q-polynomial 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 [2],[19], where we studied the discrete case.

Applications in [2] 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. A way of doing this that takes into account the entire distance distribution of the code, was suggested in [19] (cf. "the three-point lemma" above). 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 [2],[19], and the present paper.

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

Now let $L(\bar{w},\bar{d})=\min(\mathfrak{b}(\bar{w}),\hat{\mathfrak{b}}(\bar{w},\bar{d}))$. By the above argument and Lemma 15 it is clear that $P_e(W)\gtrsim L(\bar{w},\bar{d})\Pr(\mathbf{z}\in D(\mathbf{y}_1))$ for any fixed $\mathbf{y}_1\in W(\bar{w},\mathbf{x})$. ∎

Lemma 17 gives a lower estimate of the second term on the right-hand side in (64). To estimate the first term observe that the probability to mix two code vectors at a distance $\bar{d}(W)$ asymptotically behaves as $\exp(-\bar{d}^2(W)/8\sigma^2)$. Therefore,

$$P_e(W)\gtrsim\max\Big(\exp\Big(-\frac{\bar{d}^2}{8\sigma^2}+o(n)\Big),L(\bar{w},\bar{d})\exp\Big(-\frac{\bar{w}^2}{8\sigma^2}+o(n)\Big)\Big),\tag{68}$$

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

However, generally W is not distance invariant. Therefore, we have to employ the asymptotic regularity results in the previous section. By Lemma 12, 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 12, 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 $\bar{d}(W^{(1)})=\bar{\nu}(W)$. Further, by Theorem 13 it is possible to isolate in $W^{(1)}$ a subset $W^{(2)}$ such that

(i) $(1/n)\ln|W^{(2)}|\to R$,

(ii) $\bar{d}(W^{(2)})\geq\bar{d}(W^{(1)})$,

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

(iii) its distance density is bounded below by $\frac{1}{q(n)}J(n,1-(1/2)w^2,\rho,R)$, where $q(n)$ is a subexponential function,

(iv) for each subset $W(\bar{w},\mathbf{x})$,

$$|W(\bar{w},\mathbf{x})|=\mathfrak{b}_{\mathbf{x}}(\bar{w})\geq\exp\big(nb^*(1-\frac{1}{2}w^2,\rho,R)-o(n)\big).$$

Here $\rho\in[0,\rho_{kl}(R)]$ is a fixed number and $\bar{w}$ depends on ρ.

We shall use subcodes $W^{(1)}$ and $W^{(2)}$ to estimate from below the first and the second terms in (63), respectively. In doing so, we rely on Proposition 14. Let $\bar{d}=\bar{d}(W^{(1)})\geq\bar{d}(W)$. Since $(1/n)\ln|W^{(1)}|\to R$, the first term in (63) is still bounded below by the first of the two exponents in (68). Let us estimate the second term. Let ρ be fixed and $\bar{w}$ be the same as in the definition of $W^{(2)}$. We assume that $\mathbf{x}$ ranges over $W^{(2)}$ and consider only decoding errors producing code vectors in $W^{(2)}(\mathbf{x},w)$. By virtue of Properties (i)-(ii),(iv) of $W^{(2)}$ and (62) we have

$$\begin{aligned}P_e(W)&\gtrsim L(\bar{w},\bar{d}(W^{(2)}))\exp\Big(-\frac{\bar{w}^2}{8\sigma^2}+o(n)\Big)\\&\geq L(\bar{w},\bar{d})\exp\Big(-\frac{\bar{w}^2}{8\sigma^2}+o(n)\Big).\end{aligned}\tag{69}$$

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

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

PROOF. According to the discussion before this lemma,

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

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 $||\mathbf{c}_1||$ equals $||\mathbf{c}_2||$ and both are exactly w^2. From Fig. 4 we have $\mathbf{c}_1 = (\bar{w}, 0)$, $\mathbf{c}_2 = (\bar{w}\cos\phi, \bar{w}\sin\phi)$, where $\phi = 2\arcsin\frac{\bar{d}}{2\bar{w}}$ is the angle between $\mathbf{c}_1$ and $\mathbf{c}_2$. Suppose that $\mathbf{e} = (x, y)$. Then the condition that $\mathbf{e}$ is closer to $\mathbf{c}_2$ than to $\mathbf{c}_1$ yields the inequality

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

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

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

where we have used the identity $G = 1 + \frac{(1-\cos\phi)^2}{\sin^2\phi}$. ∎

Note that (65) 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 the nest lemma $\bar{d}$ is a short notation for $\bar{d}(W)$.

Lemma 17 *Under the conditions of Lemma 15,*

$$P_e(W) \gtrsim L(\bar{w}, \bar{d})\exp\Big(-\frac{\bar{w}^2}{8\sigma^2} + o(n)\Big). \tag{67}$$

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

$$\hat{\mathfrak{b}}(\bar{w}.\bar{d}) = \exp\Big(\frac{\bar{d}^2\bar{w}^2}{8(4\bar{w}^2 - \bar{d}^2)}\Big).$$

PROOF. For the inequality in (65) to be nontrivial we need that

$$\mathfrak{b}(\bar{w}) \lesssim \frac{\Pr\{\mathbf{z} : ||\mathbf{x}-\mathbf{z}|| \geq ||\mathbf{y}_1-\mathbf{z}||\}}{\Pr\{\mathbf{z} : ||\mathbf{x}-\mathbf{z}|| \geq ||\mathbf{y}_1-\mathbf{z}|| \geq ||\mathbf{y}_2-\mathbf{z}|| \text{ and } ||\mathbf{y}_1-\mathbf{y}_2|| = \bar{d}(W)\}}.$$

Let $\hat{\mathfrak{b}}(\bar{w}, \bar{d})$ be equal to the right-hand side of the last inequality. Since the probability in the numerator behaves as $\exp(-\bar{w}^2/8\sigma^2)$, we obtain

$$\hat{\mathfrak{b}}(\bar{w}, \bar{d}) \sim \exp\Big(-\frac{\bar{w}^2}{2\sigma^2}(1-G)\Big) = \exp\Big(\frac{\bar{d}^2\bar{w}^2}{8(4\bar{w}^2 - \bar{d}^2)}\Big).$$

regularity properties of the previous section. This will prove that exponentially the bound established for "regular" codes W holds true for the general case. This course of action helps us to retain the geometric nature of the argument.

We begin with estimating the second term in (64). Let us assume that $W(\bar{w}, \mathbf{x})$ is nonempty and let $\mathfrak{b}(\bar{w}) = |W(\bar{w}, \mathbf{x})|$.

Lemma 15 *Let* $\mathbf{z}$ *be the received vector and* $\mathbf{y}_1$ *be any fixed vector in* $W(\bar{w}, \mathbf{x})$. *Then*

$$P_e(W) \geq \mathfrak{b}(\bar{w}) \Pr(\mathbf{z} \in D(\mathbf{y}_1)),$$

This probability can be bounded below by the following chain of inequalities, which is just a one-step inclusion-exclusion argument:

$$\begin{aligned}
\Pr(\mathbf{z} \in D(\mathbf{y}_1)) &= \Pr\{\mathbf{z} : ||\mathbf{x} - \mathbf{z}|| \geq ||\mathbf{y}_1 - \mathbf{z}||\} - \sum_{\substack{\mathbf{y}' \in W(\bar{w}, \mathbf{x}) \\ \mathbf{y}' \neq \mathbf{y}_1}} \Pr(\mathbf{z} \in D(\mathbf{y}')) \\
&= \Pr\{\mathbf{z} : ||\mathbf{x} - \mathbf{z}|| \geq ||\mathbf{y}_1 - \mathbf{z}||\} - \sum_{\substack{\mathbf{y}' \in W(\bar{w}, \mathbf{x}) \\ \mathbf{y}' \neq \mathbf{y}_1}} \Pr\Big\{\mathbf{z} : ||\mathbf{x} - \mathbf{z}|| \geq ||\mathbf{y}_1 - \mathbf{z}|| \geq ||\mathbf{y}' - \mathbf{z}||\Big\} \\
&\geq \Pr\{\mathbf{z} : ||\mathbf{x} - \mathbf{z}|| \geq ||\mathbf{y}_1 - \mathbf{z}||\} - \\
&\quad (\mathfrak{b}(\bar{w}) - 1) \Pr\Big\{\mathbf{z} : ||\mathbf{x} - \mathbf{z}|| \geq ||\mathbf{y}_1 - \mathbf{z}|| \geq ||\mathbf{y}_2 - \mathbf{z}|| \text{ and } ||\mathbf{y}_1 - \mathbf{y}_2|| = \bar{d}(W)\Big\}. \quad (65)
\end{aligned}$$

Observe that the distance between $\mathbf{x}$ and $\mathbf{y}_1, \mathbf{y}_2$ lies between $\bar{w} - \Theta(1/\sqrt{n})$ and $\bar{w}$. 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 $\bar{w}$. Then by the definition of the channel and properties of the normal distribution, the first probability in (65) is asymptotic to $\Phi(\bar{w}/2\sigma) \sim \exp(-\bar{w}^2/8\sigma^2)$, 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 (65). Let $\mathbf{y}_1, \mathbf{y}_2$ be two code points in $W(\bar{w}, \mathbf{x})$. As said above, the noise in the channel is a product of n independent i.i.d. Gaussian variables, each affecting the corresponding coordinate of $\mathbf{x}$. Let us have a closer look at the probabilities of error events that output $\mathbf{y}_1$ or $\mathbf{y}_2$ as a decoding result. These probabilities are completely determined by the pairwise distances in the triple $\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2$ and the relative distances between $\mathbf{z}$ and these three points. Therefore, we can restrict our attention to the secant plane $\mathcal{P}$ defined by $\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2$. Let us introduce affine coordinates in $\mathbf{R}^n$ in such a way that the origin is located at $\mathbf{x}$, the first coordinate vector is given by the direction $\mathbf{y}_1 - \mathbf{x}$, and the first two coordinates form an orthogonal basis for $\mathcal{P}$, making it into a linear space. Now let us write $\mathbf{y}_1, \mathbf{y}_2, \mathbf{z}$ in these coordinates and restrict our attention to the plane $\mathcal{P}$, ignoring the remaining $n-2$ coordinates (i.e., project $\mathbf{z}$ orthogonally on $\mathcal{P}$). Denote the corresponding two-dimensional vectors by $\mathbf{c}_1, \mathbf{c}_2, \mathbf{e}$. Note that $\mathbf{e}$ 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 $\mathbf{x}$.

Let us proceed to estimating the last probability in (65). This is done in the following lemma.

Lemma 16 (The "three-point lemma") *Let* $\mathbf{y}_1, \mathbf{y}_2 \in W(w, \mathbf{x})$ *be two code vectors at a distance* $||\mathbf{y}_1 - \mathbf{y}_2||^{1/2} = \bar{d}(W)$. *Then*

$$\Pr\Big\{\mathbf{z} : ||\mathbf{x} - \mathbf{z}|| \geq ||\mathbf{y}_1 - \mathbf{z}|| \geq ||\mathbf{y}_2 - \mathbf{z}||\Big\} \sim \exp\ \big(-G\frac{\bar{w}^2}{8\sigma^2}\big), \quad (66)$$

Note that the projection of $W(\bar{w}, \mathbf{x})$ on S^n gives the spherical ring $C(t, \mathbf{c})$ defined in (54), with the obvious relation between t and w. Distances in W and C are connected by the scaling $\bar{w} = w\sigma\sqrt{An}$.

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 $\mathcal{B}(\bar{w})$ be the average distribution of Euclidean distances in W, defined similarly to $B(t)$, see (16). This is a monotone nondecreasing function of $\bar{w}$. Denote the corresponding average distance density by $\mathfrak{b}(\bar{w}', \bar{w}) = \mathcal{B}(\bar{w}) - \mathcal{B}(\bar{w}')$. Similarly, let $\mathfrak{b}_{\mathbf{x}}(\bar{w}', \bar{w})$ denote the local distance densities. As above, if $\bar{w} - \bar{w}' = \Theta\big(\frac{1}{\sqrt{n}}\big)$, we keep only one argument and write $\mathfrak{b}(\bar{w})$. We have

$$\mathfrak{b}(\bar{w}) = b\big(1 - \frac{\bar{w}^2}{2An\sigma}\big) = b\big(1 - \frac{w^2}{2}\big), \tag{60}$$

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

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

Proposition 14 *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*

$$P_e(W) \geq \frac{1}{|W|} \sum_{\mathbf{x} \in W'} \Pr(\overline{D(\mathbf{x}, W)}) \tag{61}$$

$$\geq \frac{1}{|W|} \sum_{\mathbf{x} \in W'} \Pr(\overline{D(\mathbf{x}, W'')}) \tag{62}$$

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

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

Suppose for a while that W has the regularity properties discussed in the previous section. Namely, every vector in W has a neighbor at an effective distance $\bar{\nu}(W)$ equal to its minimum distance $\bar{d}(W)$ and if $\mathfrak{b}(\bar{w})$ is the average distance density of W for a certain value of $\bar{w}$ then for all the local densities we have $\mathfrak{b}_{\mathbf{x}}(\bar{w}) = \mathfrak{b}(\bar{w})$. Let $\mathcal{E}(\bar{w}, \mathbf{x})$ denote the error event consisting in decoding to a vector at a distance $\bar{w}$ from the transmitted vector $\mathbf{x}$. Then by (8) and Proposition 14 the error probability $P_e(W)$ satisfies

$$P_e(W) \geq \frac{1}{|W|} \max\Big(\sum_{\mathbf{x} \in W} \Pr(\mathcal{E}(\bar{\nu}(W), \mathbf{x})), \sum_{\mathbf{x} \in W} \mathfrak{b}_{\mathbf{x}}(\bar{w}) \Pr(\mathcal{E}(\bar{w}, \mathbf{x}))\Big) \tag{63}$$

$$= \max(\Pr(\mathcal{E}(\bar{d}(W), \mathbf{x})), \mathfrak{b}(\bar{w}) \Pr(\mathcal{E}(\bar{w}, \mathbf{x}))), \tag{64}$$

where in the last formula $\mathbf{x}$ is an arbitrary fixed vector from W. We shall now prove several lemmas giving lower estimates on the quantities in this inequality still assuming that W is distance invariant, and then return to the general case of arbitrary W by relying on the asymptotic

The last inequality follows by (16) and the fact that $|C^{(1,i)}| < 2|C^{(1)}|$, which follows by (57) since the size $|I|$ is at most linear in n. We conclude that after each step the size of I decreases by at least one.

This argument implies that after all the indices in I are exhausted, in the remaining subset $K := C^{(1,|I|)}$ for all j, $1 \le j \le s$, we have

$$b_{\mathbf{c}}(u_j, K) \le q(n) J(n, u_j, \rho, R) \quad (\forall \mathbf{c} \in K), \tag{58}$$

where $q(n)$ is at most a subexponential function. On the other hand, since the rate of K is still asymptotically equal to R, this subcode satisfies the conditions of Theorem 11. So for a certain j, $1 \le j \le s$, we have

$$(1/n) \ln b(u_j, K) \ge b^*(u_j, \rho, R). \tag{59}$$

Together (58) and (59) suffice to show that one can isolate in K a subset of vectors of exponentially the same size, for which the local densities satisfy (56). Indeed, assume the contrary, i.e., that the subset $M \subset K$ for which (56) holds true, has size $|M| = |K|e^{-\mu n}$ for some $\mu > 0$. Then by (16)

$$\begin{aligned} b(u_j, K) &= \frac{1}{|K|} \sum_{\mathbf{c} \in K} b_{\mathbf{c}}(u_j, K) \\ &\overset{(58)}{\le} J(n, u_j, \rho, R) \Big[\frac{1}{q(n)} (1 - e^{-\mu n}) + q(n) e^{-\mu n} \Big]. \end{aligned}$$

Since $|K| = e^{nR - o(n)}$, this would violate (59) beginning from some large value of n. Therefore, the size of M differs from $|K|$ by at most a subexponential factor, or $|M| = e^{Rn - o(n)}$. Put $C^{(2)} = M$. We conclude that the average distance density in M satisfies (59) for some u_j and all local densities at this point satisfy (56). ∎

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. It generalizes a regularity theorem in [24] and its strengthening in [18].

5 Proof of Theorem 1

As in Section 1, let W be a code on the sphere in $\mathbf{R}^n$ of radius $\sigma\sqrt{An}$, where σ 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 $\bar{d}, \bar{w}$, and so on, to distinguish them from distances on the unit sphere $(d, w, \ldots)$. Clearly, $\bar{d} = d\sigma\sqrt{An}$, and so on. By C we denote the projection of W on the concentric unit sphere S^n. Obviously, angular distances in C and W are equal. Let

$$W(\bar{w}, \mathbf{x}) = \{\mathbf{y} \in W : \bar{w} - \Theta\big(\frac{1}{\sqrt{n}}\big) \le ||\mathbf{x} - \mathbf{y}|| \le \bar{w}\}.$$

The plan of the proof is, beginning with the subset $C^{(1)}$ constructed in the previous lemma, to compare its average distance densities $b(u_i)$, $1 \le i \le s$, to $J(n, u_i, \rho, R)$. Since the rate of $C^{(1)}$ asymptotically equals R, there exists a point u_j at which $b(u_j, C^{(1)})$ is exponentially greater than or equal to (52). For each u_j at which this holds true, we examine local densities of $C^{(1)}$. If for some j the set V_j of vectors $\mathbf{c} \in C^{(1)}$ for which $(1/n)\ln b_{\mathbf{c}}(u_j, C^{(1)}) > b^*(u_j, \rho, R)$ for $n \to \infty$, has exponentially the same size as $C^{(1)}$, we put $C^{(2)} = V_j$. Otherwise, for each j at which $(1/n)\ln b_{\mathbf{c}}(u_j, C^{(1)}) > b^*(u_j, \rho, R)$, the size of V_j is exponentially smaller than that of $C^{(1)}$. Then it proves possible to discard all the vectors in V_j from $C^{(1)}$ without decreasing its asymptotic rate. By virtue of Theorem 11, in the remaining subset K there exists a point u_j at which $(1/n)\ln b(u_j, K) \ge b^*(u_j, \rho, R)$. Further, by the construction we know that $b(u_j, K)$ does not exponentially exceed $J(n, u_j, \rho, R)$. Therefore, one can prove that the number of vectors $\mathbf{c}$ for which $(1/n)\ln b_{\mathbf{c}}(u_j, K) < b^*(u_j, \rho, R)$ (i.e., the local distance density is exponentially smaller than $J(n, u_j, \rho, R)$) does not exhaust the entire subset K. This proves that the remaining vectors in K form a code $C^{(2)}$ of exponential size $e^{Rn - o(n)}$ with the properties claimed in the theorem.

Let $C^{(1)}$ be a subset constructed in (55). Since $C^{(1)}$ has rate R, for each $\rho \in [0, \rho_{kl}]$ there exists a point u_j in the partition (53) such that

$$b(u_i, C^{(1)}) \ge J(n, u_i, \rho, R).$$

Let $I \subset \{1, 2, \ldots, s\}$ be the set of such points. We give a constructive definition of $C^{(2)}$ by performing the following expurgation procedure. Without loss of generality we assume that $I = \{1, 2, \ldots, |I|\}$ is a subset of consecutive integers.

The procedure is repeated for each $i \in I$. Step i begins with a subset $C^{(1,i-1)}$. For $i = 1$ we take $C^{(1,0)} = C^{(1)}$. Let $V_i \subset C^{(1,i-1)}$ be a subset of those vectors $\mathbf{c}$ for which $b_{\mathbf{c}}(u_i, C^{(1,i-1)})$ is exponentially greater than $J(n, u_i, \rho, R)$. Then either

$$|V_i| \sim R$$

or

$$|V_i| \sim R - \gamma_i$$

for some $\gamma_i > 0$. In the first case we put $C^{(2)} = V_i$. Clearly, this subset satisfies the properties claimed in the theorem. Otherwise, discard all the vectors in V_i from $C^{(1,i-1)}$ and put $C^{(1,i)} = C^{(1,i-1)} \setminus V_i$. Note that

$$|C^{(1,i)}| = |C^{(1,i-1)}| - |V_i| = |C^{(1)}|\Big(1 - \frac{1}{e^{\gamma_i n}}\Big), \tag{57}$$

i.e., asymptotically the rate of $C^{(1,i)}$ is the same as the rate of $C^{(1)}$. Further, we have

$$b_{\mathbf{c}}(u_i, C^{(1,i)}) \le b_{\mathbf{c}}(u_i, C^{(1,i-1)})$$

and

$$b(u_i, C^{(1,i)}) < 2b(u_i, C^{(1)}).$$

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

$$U_0 = [u_0, u_1],\ U_1 = [u_1, u_2],\ U_s = [u_{s-1}, u_s] \tag{53}$$

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

$$C(t, \mathbf{c}) = \{\mathbf{c}' \in C : t \le \langle \mathbf{c}, \mathbf{c}' \rangle \le t + \Theta(\frac{1}{n})\}. \tag{54}$$

Further, let

$$C_i = \{c \in C : (C(u_i, \mathbf{c}) \neq \emptyset) \text{ and } \forall_{j<i}(C(u_j, \mathbf{c}) = \emptyset)\}. \tag{55}$$

There are at most $\Theta(n)$ such subsets C_i. The effective distance of C is defined as

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

Let $C^{(1)} = C_i$ for this value of i. It is clear that $|C^{(1)}| \ge |C|/\Theta(n)$, i.e., asymptotically this subcode has the same size. This proves the following lemma.

Lemma 12 *Let $C \subset S^n$ be a code. Then there exists a subset $C^{(1)} \subset C$ such that $t(C^{(1)}) = \nu(C)$ and every vector $\mathbf{c} \in C^{(1)}$ has a neighbor $\mathbf{c}' \in C(\nu, \mathbf{c})$. Moreover, $|C| \sim |C^{(1)}|$.*

In the next theorem we intend to prove that a code C contains a subcode $C^{(2)}$ with the following properties:

(i) $C^{(2)}$ asymptotically has the same rate as C,

(ii) $t(C^{(2)}) \le \nu(C)$,

(iii) for a certain i the exponent of its average distance density $b(u_i)$ is bounded below by the function $b^*(u_j, \rho, R)$ from Theorem 11, and

(iv) all the local distance densities $b_{\mathbf{c}}(u_i)$, $\mathbf{c} \in C^{(2)}$ have at least the same exponential growth as $b^*(u_j, \rho, R)$.

Moreover, $C^{(2)}$ can be chosen a subset of $C^{(1)}$ from Lemma 12.

Theorem 13 *Let C be a code of rate R and distance $t = t(C)$. Let ρ, $0 \le \rho \le \rho_{kl}$, be a fixed number. There exists a subset $C^{(2)} \subset C$ and a number u, $\tau(\rho) \le u \le 1$, such that $(1/n) \ln b(u) \gtrsim b^*(u, \rho, R)$, and for every vector $\mathbf{c} \in C^{(2)}$*

$$b_{\mathbf{c}}(u) \ge \frac{1}{q(n)} b(u), \tag{56}$$

where $q(n)$ is a function of subexponential growth ($q(n) = o(2^{\xi n})$ for any $\xi > 0$). Further, $t(C^{(2)}) \le \nu(C)$ and $|C^{(2)}| \sim |C|$.

PROOF. Below we refer to distance densities of several different codes at a time. To avoid ambiguity, we denote the average (local) distance density of a code C at a point $x \in [-1, 1]$ by $b(x, C)$ (resp., $b_{\mathbf{c}}(x, C)$).

Theorem 10 *Let C be a code of rate R. Let ρ, $0 \le \rho \le \rho_{kl}$, be a fixed number. Then there is a point x, $\tau(\rho) \le x \le 1$, such that*

$$b(x) \ge \exp\big(n(\sqrt{1-x^2} + R) - o(n)\big).$$

PROOF. By Theorem 2 and Proposition 9 we have, for some $U = [x, y] \subset U$ and $s \in U$,

$$b(x, y) \gtrsim \frac{f_0|C|}{f^{(k)}(s)},$$

where $k = \rho n$. Substituting (42) and (47), we obtain

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

Now take $m = (\text{const})\, n$ in (19). Then since $y - x = \Theta(1/n) = o(1)$, the last inequality is still asymptotically valid if we replace s with x. To complete the proof it remains to change the scaling according to $2ak = n$. This gives

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

Thus, already estimate (42) shows that for a putative code that meets the KL bound, there is a small segment in which the distance distribution asymptotically equals at some point $x \ge t_{kl}$ the average distribution (46). The better estimate (34) gives a stronger result.

Theorem 11 *Let C be a code of rate R. Let ρ, x have the same meaning and satisfy the same conditions as in Theorem* 10. *Then*

$$b(x) \ge J(n, x, \rho, R) \tag{51}$$
$$J(n, x, \rho, R) = \exp(n b^*(x, \rho, R) - o(n))$$

where

$$b^*(x, \rho, R) := 4\rho(1+\rho) \int_x^1 \frac{dz}{z + \sqrt{z^2 - 4(1-z^2)\rho(1+\rho)}} - (1+\rho) H\Big(\frac{\rho}{1+\rho}\Big) + R, \tag{52}$$

PROOF. The argument is the same as in the previous theorem except that now in (19) we use bound (41). First, for $k = \rho n$ by Proposition 9 we again need to consider only the first term in (19). Further, from the proof of Theorem 5 we see that the estimate in (41) is nothing else than $P_k(1)$ times $\exp\big\{ - \int_x^1 \ldots dz\big\}$, where the integral is the same as in (41). Now the proof is completed by invoking (48) as follows:

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

Again we need to rescale this using $2ak = n$. $\blacksquare$

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 11 these theorems will be used in the next section.

Let $f^{(k)}(x) = \sum_{i=0}^{2k+1} f_i P_k(x)$ be the Gegenbauer expansion of $f^{(k)}(x)$. Then

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

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

The first equation follows by an application of the Christoffel-Darboux formula (27) 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,

$$\frac{1}{k}\ln\frac{f_0}{f^{(k)}(x)} \overset{(43)}{=} -2\ln 2 + 2(1+a)H\Big(\frac{1}{2(1+a)}\Big) - 2\ln P_k(x), \tag{47}$$

and, in particular,

$$\frac{1}{k}\ln\frac{f_0}{f^{(k)}(1)} = -(1+2a)H\Big(\frac{1}{1+2a}\Big). \tag{48}$$

In the last equation we have taken into account (25) and the identity

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

Note in passing that (48) leads to the following result.

Proposition 9 *Let $r \in \mathbf{R}$ be a number such that*

$$r \ge (1+2a)H\Big(\frac{1}{1+2a}\Big) \tag{49}$$

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

Remark. Note that $R_{kl} = (1+2a)H\big(\frac{1}{1+2a}\big)$ 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.

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 are satisfied, so we can apply the estimates of the previous section to derive concrete bounds on the possible distance distribution of spherical codes. Observe that $k \to \infty$, the number s approaches $t_{1,k}$.

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

$$\tau(\rho) = 2\frac{\sqrt{\rho(1+\rho)}}{1+2\rho}. \tag{50}$$

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 Sect. 2 and d_{kl} is given by (5).) For $0 \le \rho \le \rho_{kl}$ we have $0 \le \tau(\rho) \le t_{kl}$.

Further, since (45) holds with respect to every $\mathbf{c} \in C$, we can use the definition of the distance distribution of C (16) to conclude that for the random code,

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

Since the surface area of S equals $\Omega(\pi) = n\pi^{n/2}/\Gamma(n/2+1)$, we obtain

$$\overline{B}(t) = \frac{e^{nR}(n-1)\pi^{(n-1)/2}\Gamma\left(\frac{n}{2}+1\right)}{n\pi^{n/2}\Gamma\left(\frac{n+1}{2}\right)} \int_0^\theta \sin^{n-2} t \, dt.$$

The asymptotic expression for the right-hand side ($n \to \infty$) was computed in Shannon [24]. Namely, we have

$$\overline{B}(t) \sim e^{nR} \frac{\Gamma\left(\frac{n}{2}+1\right) \sin^{n-1}\theta}{n\Gamma\left(\frac{n+1}{2}\right)\pi^{1/2}\cos\theta},$$

or, using the Stirling formula,

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

Taking logarithms, we obtain

$$\begin{aligned} \frac{1}{n} \ln \overline{B}(t) &= R + \ln \sin\theta + o(1) \\ &= R + \ln \sqrt{1-t^2} + o(1). \end{aligned} \tag{46}$$

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 the role 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 distance distribution of random codes chosen with uniform probability from their respective spaces. Moreover, a code with binomial weight spectrum in the Hamming case asymptotically meets the sphere packing (Varshamov–Gilbert) bound on the code distance (this is the minimal i for which $\overline{B}_i \geq 1$). Likewise, in the spherical case a code with the random distance spectrum meets the sphere packing (Shannon) bound (1). Finally, a code with the binomial (average) distance distribution 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 [13].

In this section we use a shorthand notation $P_k(x)$ for $P_k^{\alpha,\alpha}(x)$. We need to take $\alpha = (n-3)/2$, where n is the dimension of the sphere. Let

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

where $\frac{k}{n} \to \rho$ (ρ is a constant), $t_{1,k} < s < t_{1,k+1}$, and s is chosen so that $P_k(s) = -P_{k+1}(s)$. We again assume that $\frac{\alpha}{k} \to a$, i.e., $a = \frac{1}{2\rho}$. In the remaining part of the section we specialize bound (19) to the case $f = f^{(k)}$.

PROOF. Note that $h(x)$ is the function on the right-hand side of (42). 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. Therefore, the straight line drawn through the point $x = t_1$ with slope $h'(t_1)$ is an upper bound on the exponent of the polynomial in $x \in [t_1, 1]$. ∎

Clearly, a similar argument is valid for $x \in [-1, -t_1]$, i.e., the exponent of $|P_k^{\alpha,\beta}(x)|$ is bounded above by a straight line symmetric to (44) with respect to the y-axis.

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

Remarks. (a) Interestingly, the derivative of the function on the right-hand side of (41) equals $h'(x)$ at $x = t_1$ (the function itself is concave for $x \in [t_1, 1]$). So (44) is a *common* tangent to (41) and (42) at $x = t_1$ and separates these two curves. Bound (44) is useful because on the one hand is is easy to work with, and on the other hand, it is much better than (42), in particular, it does not become infinite as $x \to \pm 1$.

(b) Note that the integrand in (41) is real when

$$x \geq x_0 = \frac{a\sqrt{2a+1} - 1 - 2a}{a^2 - 2a - 1}.$$

From (28) it is easy to see that $x_0 \leq t_1$ for all $a \geq 0$ with the equality only for $a = 0$. Thus, the function in (41) is well-defined in the entire segment $[t_1, 1]$.

4 Asymptotic bounds on the distance distribution of spherical codes

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

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 $\mathbf{c} \in C$ be a code point and $b_{\mathbf{c}}(t)$ be the local distance density with respect to $\mathbf{c}$ (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 $\mathbf{c}$ and half-angles $\theta = \arccos t$ and $\theta' = \arccos\left(t - \Theta\left(\frac{1}{n}\right)\right)$, respectively. On the average this number constitutes the same fraction of M as the area of the ring of the total area of S. Since the area of this ring grows exponentially in n and the exponent is monotone increasing in θ, we may as well put $\theta' = 0$ and consider the area of the entire spherical cap, i.e., the distance distribution function $B_{\mathbf{c}}(t)$. Thus, we have

$$\frac{\overline{B}_{\mathbf{c}}(t)}{M} = \frac{\Omega(\theta)}{\Omega(\pi)}, \tag{45}$$

where by $\Omega(\theta)$ we denote the $(n-1)$-dimensional area of a spherical cap on S with half-angle θ,

$$\Omega(\theta) = \frac{(n-1)\pi^{(n-1)/2}}{\Gamma\left(\frac{n+1}{2}\right)} \int_0^\theta \sin^{n-2} t \, dt.$$

where the first choice of the signs corresponds to $x \in (t_{1,k}, 1]$ *and the second to* $x \in [-1, t_{1,k})$.
(*c*) $x \in [-1, 1]$

$$\frac{1}{k} \ln |P_k^{\alpha,\alpha}(x)| \le (a+1) H\Big(\frac{1}{2(a+1)}\Big) - \frac{1}{2} a \ln(1-x^2) - \ln 2. \tag{42}$$

PROOF. To obtain part (a) from (29), 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

$$\frac{1}{k} \ln L_k = -\ln 2 + 2(a+1) H\Big(\frac{1}{2(a+1)}\Big)$$

The remaining part of the expression is straightforward simplifications.

Part (b) follows directly from (34) upon substituting $b = a$. In (41) 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 classical formula

$$\Gamma(p)\Gamma\Big(p+\frac{1}{2}\Big) = \frac{\sqrt{\pi}}{2^{2p-1}} \Gamma(2p).$$

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,

$$\frac{1}{k} \ln \omega_k^{\alpha,\alpha} = \frac{1}{k} \ln \frac{2^{-2k} \Gamma(2(k+\alpha))}{\Gamma(k)\Gamma(k+2\alpha)} = -2 \ln 2 + 2(1+a) H\Big(\frac{1}{2(1+a)}\Big). \tag{43}$$

This completes the proof. ∎

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, expressions (40) and (41) for $x \in [t_1, 1]$ represent one and the same function.

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

Corollary 8 *Let* $P_k^{\alpha,\alpha}$ *be the Jacobi polynomial and suppose that* $k \to \infty$, $\alpha/k \to a$. *Let* t_1 *be the largest zero of* $P_k^{\alpha,\alpha}$. *Then up to* $o(1)$ *terms,*

$$\frac{1}{k} \ln |P_k^{\alpha,\alpha}(x)| \le \frac{a t_1}{1-t_1^2}(x - t_1) + h(t_1), \quad x \in [t_1, 1], \tag{44}$$

where

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

Theorem 6 *Let $P_k^{\alpha,\beta}(x)$ be the Jacobi polynomial. Suppose that $k \to \infty$, $\alpha/k \to a, \beta/k \to b$. Then*

$$\frac{1}{k}\ln|P_k^{\alpha,\beta}(x)|^2 \lesssim -a\ln(1-x) - b\ln(1+x) + \ln\omega_k^{\alpha,\beta}, \quad x \in [-1,1]. \tag{38}$$

PROOF. First let $t_1 \le x \le 1$ (or $-1 \le x \le -t_1$). We already know (Thm. 5) that in this segment $P_k^{\alpha,\beta}$ is monotone increasing. From (20) we have $||P_k^{\alpha,\beta}||_2^2 = \omega_k^{\alpha,\beta}$. Obviously,

$$I := \int_{t_1}^1 |P_k^{\alpha,\beta}(x)|^2(1-x)^\alpha(1+x)^\beta dx < \omega_k^{\alpha,\beta}. \tag{39}$$

Let us partition the segment $[t_1,1]$ into m equal segments U_i and denote their endpoints by $u_0 = 0, u_1, \ldots, u_{m-1}, u_m = 1$. Then since $\mu(x)$ is a falling function,

$$\sum_{i=0}^{m-1}\frac{1}{m}|P_k^{\alpha,\beta}(u_i)|^2\mu(u_{i+1}) < I < \sum_{i=0}^{m}\frac{1}{m}|P_k^{\alpha,\beta}(u_{i+1})|^2\mu(u_i),$$

and when $m \to \infty$, the sums asymptotically are identical. Assume that m grows linearly in k. So if there is a point $x \in (t_1,1]$ at which (38) is violated, the lower sum will also exponentially exceed $\omega_k^{\alpha,\beta}$. This is in contradiction with (39).

For $-t_1 \le x \le t_1$ the estimate is proved by an argument similar to the proof of Theorem 4. Namely, by Theorem 3, we can bound the function $u(x) = (1-x)^{(\alpha+1)/2}(1+x)^{(\beta+1)/2}P_k^{\alpha,\beta}(x)$ from both sides by sine functions. Their maxima are known and coincide in the main term with the maxima of $|u(x)|$. This implies the required inequality. ∎

Remarkably, the bound in this theorem coincides with (29) for $x \in [-t_1,t_1]$ (we do not prove this). However, (38) is a limiting envelope of $\ln|P_k^{\alpha,\beta}(x)|$. This implies that (29) is another expression for the limiting envelope of Jacobi polynomials in the segment considered. Recall that outside this segment (29) is tight. Thus, in effect, (29) expresses the exponent of the Jacobi polynomials of growing degree for all $x \in [-1,1]$.

Since in the following section we deal with codes on $S^n(\mathbf{R})$, the case of $\alpha = \beta$ is of special interest to us. Estimates of Theorems 4–6 in this case are collected in the following theorem.

Theorem 7 *Let $P_k^{\alpha,\alpha}$ be the Jacobi polynomial and suppose that $k \to \infty$, $\alpha/k \to a$. Then up to $o(1)$ terms,*
(a) $x \in [-1,1]$

$$\frac{1}{k}\ln|P_k^{\alpha,\alpha}(x)| = 2(a+1)H\Big(\frac{1}{2(a+1)}\Big) - \ln 2 + \frac{1}{\pi}\int_0^{t_1}\frac{\ln|x^2-z^2|\sqrt{2a+1-(a+1)z^2}}{1-z^2}dz \tag{40}$$

(b) $t_1 \le |x| \le 1$

$$\frac{1}{k}\ln|P_k^{\alpha,\alpha}(x)| = (1+a)H\big(\frac{a}{1+a}\big) \mp \int_x^1 \frac{(1+2a)dz}{az \pm \sqrt{a^2z^2-(1-z^2)(1+2a)}}, \tag{41}$$

or, neglecting the vanishing terms,

$$\frac{y''}{y} \approx \left(\frac{y'}{y}\right)^2.$$

Now let us rewrite (26) as

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

Let $u = \frac{y'}{y}$. Then this equation, up to $o(1)$ terms, can be rewritten as

$$(1-x^2)u^2 + (\beta - \alpha - (\alpha+\beta+2)x)u + k(k+\alpha+\beta+1) = 0.$$

Let us prove that the solution for u must be taken with the '$-$' sign. Up to small terms, the function $u(x)$ has a constant derivative in the segment considered; see (36). So up to small perturbations, $u(x)$ is a linear function in $[t_1, 1]$. Thus, in this segment the choice of the sign in the solution is uniform. To find out which sign to take, let us use (24) and (25) 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

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

we compute $u(1) = k(k+\alpha+\beta+1)/(2(\alpha+1))$, which justifies our claim.

Now observe that

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

whence we get, for $t_{1,k} \le x \le 1$,

$$\begin{aligned}\int_x^1 d\ln P_k(z) \overset{k}{\sim} \int_x^1 \Bigg[&\frac{(\alpha+(\alpha+\beta+2)z-\beta)}{2(1-z^2)} \\ &- \frac{\sqrt{(\alpha+(\alpha+\beta+2)z-\beta)^2 - 4k(1-z^2)(k+\alpha+\beta+1)}}{2(1-z^2)}\Bigg] dz. \quad (37)\end{aligned}$$

The final answer is obtained by invoking the boundary condition (25), dividing k out and collecting the $o(1)$ terms. ∎

From the proof of Theorem 4 we see that for $t_{1,k} \le |x| \le 1$ Eq. (29) also gives the exact main term of the exponent. So for such x expressions (29) and (34) represent one and the same function (a direct proof of this would not seem immediate).

Finally, let us derive a more explicit 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 $k \to \infty, \alpha/k \to a > 1, \beta/k \to b$, and $t_1 \le |x| \le 1$. Then*

$$\frac{1}{k}\ln|P_k^{\alpha,\beta}(x)| = (1+a)H\Big(\frac{a}{1+a}\Big) \mp \int_x^1 \frac{(a+(a+b)z-b) \mp \sqrt{(a+(a+b)z-b)^2 - 4(1-z^2)(1+a+b)}}{2(1-z^2)}dz + o(1), \quad (34)$$

where the $-$ sign corresponds to $x \in [t_1, 1]$ and the $+$ to $x \in [-1, -t_1]$.

Proof. We consider only the case $x \in [t_1, 1]$ since the symmetric case is almost identical. Let $x_0 = t_1, x_i = t_1 + i\delta, i = 1, 2, \ldots$, where δ is a small positive number. Divide the interval $[t_1, 1]$ into segments $[x_0, x_1], [x_1, x_2], \ldots$. In the computations below we have suppressed superscripts α, β. By (22) we have

$$\frac{P_k(x_{i+1})P_k(x_{i-1})}{P_k(x_i)^2} = \prod_{j=1}^{k} \frac{(x_i+\delta-t_j)(x_i-\delta-t_j)}{(x_i-t_j)^2} = \prod_{j=1}^{k}\Big(1 - \frac{\delta^2}{(x_i-t_j)^2}\Big).$$

Let $\delta = o(k)$. Then we can observe that

$$\frac{\delta^2}{(x_i-t_j)^2} = o\Big(\frac{1}{k}\Big).$$

From this we get

$$\ln\frac{P_k(x_{i+1})P_k(x_{i-1})}{P_k(x_i)^2} = \sum_{j=1}^{k}\ln\Big(1 - \frac{\delta^2}{(x_i-t_j)^2}\Big) = -\sum\Big[\delta^2 O\Big(\frac{1}{k^2}\Big) + \delta^4 O\Big(\frac{1}{k^4}\Big)\Big] = -\delta^2 O\Big(\frac{1}{k}\Big). \quad (35)$$

Next we have

$$\ln\frac{P_k(x_{i+1})P_k(x_{i-1})}{P_k(x_i)^2} = \Delta^2 \ln P_k(x_{i-1}),$$

where Δ^2 is the second divided difference operator. For a function y two times differentiable at a point ξ, we have

$$\Delta^2 y(\xi) = y''|_{x=\xi}\delta^2 + o_\delta(1)\delta^2,$$

where δ is a small increment. Letting $y = \ln P_k(x)$, we get

$$\Delta^2 y(x_{i-1}) = \delta^2 \frac{d}{dx}\Big(\frac{y'(x_{i-1})}{y(x_{i-1})}\Big) + o_\delta(1)\delta^2 = \delta^2\frac{y''y - y'^2}{y^2} + o_\delta(1)\delta^2.$$

Taking this together with (35), we obtain

$$\frac{y''y - y'^2}{y^2} = -O\Big(\frac{1}{k}\Big) - o_\delta(1) \quad (36)$$

where

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

We are interested in estimating the distance $t_i - t_{i+1}$ between consecutive zeros of $P_k^{\alpha,\beta}(x)$ or, which is the same, of $u(x)$. This is done using a method of Sturm–Liouville and relies on Theorem 3. Since $k \to \infty$, this distance is $o(1)$; so we also have $N(t_i) = N(t_{i+1})(1 \pm O(1/n))$. The idea is to replace equation (30) in the segment $[t_i, t_{i+1}]$ by an equation with a constant coefficient $\tilde{N}$; its solution $\sin\sqrt{\tilde{N}}x$ will approximate the required distance.

More specifically, consider the equation

$$U'' + N(t)U = 0, \tag{31}$$

where $t = t_i$ or t_{i+1} depending on whether $N(x)$ is a growing or a falling function in $[t_i, t_{i+1}]$ (note that $N'(x)$ has at most a constant number of sign changes). Then Equations (30) and (31) satisfy the conditions of Theorem 3. Hence, the function $\sin\sqrt{N(t_{i+1})}x$ bounds below the function $u(x)$ in the segment between to consecutive zeros of the sine function. We conclude that up to the main term,

$$t_i - t_{i+1} \geq \pi/\sqrt{N(x)},$$

where N is computed at any point $x \in (t_{i+1}, t_i)$.

Now let us rewrite (22) using the symmetry of zeros t_i. For every term $x - t_i$ in the product except for $t_i = 0$ (for k odd) we also have the term $x + t_i$. Moreover, since the number of factors grows, the crude logarithmic asymptotics will not change if we take the factor corresponding to $t_i = 0$ two times instead of one. Taking this all together, we can rewrite (22), neglecting the vanishing terms, as follows:

$$\frac{1}{k}\ln|P_k^{\alpha,\beta}(x)| = \frac{1}{k}\ln L_k + \frac{1}{k}\sum_{i=\lfloor k/2\rfloor}^{k} \ln|x^2 - t_i^2|, \quad x \neq t_i, \tag{32}$$

Again since $t_{i+1} = t_i + o(1)$, this expression is nothing else than the Riemann sum for the integral

$$\int_0^{t_1} \frac{\ln|t_i^2 - z^2|}{\Delta z} dz, \tag{33}$$

where Δz is the distance function between two consecutive zeros of $P_k^{\alpha,\beta}(x)$ at the point $x = z$ different from t_i. However, we already know that $\Delta z \geq \pi/\sqrt{N(z)}$. It remains to substitute this together with (33) in (32). ∎

Note that for $x \in [0, t_1]$ the integral in (29) has a singularity at the point $z = x$. However, its convergence is easily checked (we omit the details).

The result of Theorem 4 shows that the behavior of polynomials of continuous variable orthogonal on a finite segment is completely determined by the distribution of zeros and location of the extremal zero.

Zeros of $P_k^{\alpha,\beta}(x)$ for $k \to \infty$ fill the segment $[-t_{1,k}, t_{1,k}]$ densely in the sense that $t_{i,k} - t_{i+1,k} = o(1)$.

Let us proceed to bounds on $P_k^{\alpha,\beta}$. We present 3 types of results, each obtained by a different technique. The first result is obtained by transforming the differential equation (26) 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 $[-1, 1]$. The second theorem extends the method of [20] and [16] and gives an exponentially tight estimate in the interval $t_{1,k} \le |x| \le 1$. In this interval 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 $[-t_{1,k}, t_{1,k}]$ and coincides in this interval with the first estimate. All the three estimates are equal at $\pm t_{1,k}$.

The following important theorem is used to prove two of the three mentioned estimates.

Theorem 3 (a "comparison" theorem [28], p.20.) *Let $f(x)$ and $F(x)$ be functions continuous in $x_0 < x \le X_0$ with $f(x) \le F(x)$. Let the functions $y(x)$, and $Y(x)$, both not identically zero, satisfy the differential equations*

$$y'' + f(x)y = 0, \quad Y'' + F(x)Y = 0.$$

Let x' and x'', $x' < x''$, be two consecutive zeros of $y(x)$ and ξ be the first zero of $Y(x)$ to the right of x', $x' < \xi < x''$. Assuming that $y(x) > 0$; $Y(x) > 0$ in $x' < x < \xi$, and

$$\lim_{x \to x'+0} \frac{y(x)}{Y(x)} \ge 1,$$

we have $y(x) > Y(x)$ in $x' < x < \xi$.

From now till the end of this section we denote zeros of P_k by t_i, omitting the degree.

Theorem 4 *Let $P_k^{\alpha,\beta}(x)$ be the Jacobi polynomial with zeros $t_1, t_2, \ldots, t_k = -t_1$. Suppose that $k \to \infty, \alpha/k \to a > 1, \beta/k \to b$, and $x \in [-1, 1]$. Then*

$$\frac{1}{k} \ln |P_k^{\alpha,\beta}(x)| \lesssim \frac{1}{k} \ln L_k + \frac{1}{\pi} \int_0^{t_1} \ln |x^2 - z^2| \sqrt{\frac{1 + a + b + ab/2}{1 - z^2} - \frac{a^2}{4(1-z)^2} - \frac{b^2}{4(1+z)^2}}\, dz \quad (29)$$

Proof. We begin with the transformation of (26) discussed in [28], 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:

$$u'' + N(x)u = 0. \quad (30)$$

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 $t_{i,k} = t_{i,k}^{\alpha,\beta}$, $t_{k,k} < t_{k-1,k} < \cdots < t_{1,k}$. Further

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

so, in particular, zeros of $P_k^{\alpha,\beta}(x)$ are symmetric with respect to 0, and $t_{k,k} = -t_{1,k}$. Thus, we have

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

where $L_k = 2^{-k}\sum_{\nu=0}^{k}\binom{k+\alpha}{k-\nu}\binom{k+\beta}{\nu}$. Zeros of the two adjacent polynomials form two interlacing systems: $t_{i,k} < t_{i-1,k-1} < t_{i-1,k}$, and so forth. The explicit expression for $P_k^{\alpha,\beta}(x)$ has the form

$$P_k^{\alpha,\beta}(x) = 2^{-k}\sum_{i=0}^{k}\binom{k+\alpha}{i}\binom{k+\beta}{k-i}(x+1)^i(x-1)^{k-i}. \tag{23}$$

This implies the following useful relation:

$$\frac{d}{dx}P_k^{\alpha,\beta}(x) = \frac{1}{2}(k+\alpha+\beta+1)P_{k-1}^{\alpha+1,\beta+1}(x) \tag{24}$$

Finally, from (23) one obtains

$$P_k^{\alpha,\beta}(1) = \binom{k+\alpha}{\alpha}. \tag{25}$$

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

$$(1-x^2)y'' + (\beta - \alpha - (\alpha+\beta+2)x)y' + k(k+\alpha+\beta+1)y = 0 \tag{26}$$

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

$$\frac{\omega_k^{\alpha,\beta}L_{k+1}}{L_k}\sum_{i=0}^{k}\frac{1}{\omega_i^{\alpha,\beta}}P_i^{\alpha,\beta}(x)P_i^{\alpha,\beta}(y) = \frac{P_{k+1}^{\alpha,\beta}(x)P_k^{\alpha,\beta}(y) - P_k^{\alpha,\beta}(x)P_{k+1}^{\alpha,\beta}(y)}{x-y}. \tag{27}$$

For k sufficiently large by [15] we have,

$$t_{1,k} \to \frac{4\sqrt{(a+b+1)(a+1)(b+1)} - a^2 + b^2}{(a+b+2)^2}$$

In particular, for $\alpha = \beta = ak$ this implies

$$t_{1,k} \to \frac{\sqrt{1+2a}}{1+a}. \tag{28}$$

Sometimes it is possible to apply the results of this section to spaces that are not two-point homogeneous, by isometrically embedding them into one of the spaces (a)-(e). Though from the geometric point of view this is a straightforward task, this may present a separate interest for coding theory; see [8].

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

3 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 $k \to \infty$, $\alpha/k \to \text{const} > 1$. 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 general case of independent α, β especially since it is not much more difficult than the particular case mentioned.

By the previous section, we are interested in the asymptotics $k \to \infty$ and $\alpha = ak$, $\beta = bk$ (in fact, for the results to be valid it suffices to assume that at least one of the upper indices grows with k). The general analytic situation that we treat in this section stems from the derivation of the so-called "linear programming (LP) bound" [20] and its extensions to other Q-polynomial spaces [15], [17], [1]. 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 [20], [15], [18].

A more refined situation encountered in a number of problems that involve LP bounds [19], [2], [6] 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 $||P_k||_2^2 = m_k|W|$, where P_k is the Q-polynomial of degree k of the association scheme, m_k the kth eigenvalue of the scheme, and $||P_k||_2 = \langle P_k, P_k \rangle_2$ is the corresponding L_2-norm (cf. our Theorem 6 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 [20] was employed in [16] to derive an exact expression for the main term of the exponent in the interval considered. The proof in [16] 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.

Note that the estimates that we discuss differ from the classical theorems collected in ch. viii of [28]. Namely, we are concerned with the case of growing α, β, while in [28] these indices are *constant*.

Let $P_k^{\alpha,\beta}(x)$ be the Jacobi polynomial. From a large collection of properties of $P_k^{\alpha,\beta}(x)$ [28] we need the following facts. The polynomials $P_k^{\alpha,\beta}(x)$ are orthogonal on $\mathcal{I} = [-1, 1]$ with weight $\mu(x) = (1-x)^\alpha (1+x)^\beta$:

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

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):

$$\int_{\mathcal{I}} \sum_{k=0}^{l} f_k \Phi_k(x) dB(x) = \int_{\mathcal{I}} f(x) dB(x) \leq \sum_{j=1}^{m} \int_{U_j} f(x) dB(x) + f(1)$$

(we have used the fact that $t(0) = 1$). Now let us apply the mean value theorem in each segment of integration. We have, for some points $s_j \in U_j$,

$$f_0 |C| \leq f(1) + \sum_{j=0}^{m} f(s_j) b(u_i, u_{i+1}) \quad (u_{m+1} := 0).$$

Hence there exists a number $j \in [1, i]$ such that the summation term satisfies the claimed inequality. ∎

Theorem 2 *mutatis mutandis* is valid for all spaces covered by the KL theory (two-point homogeneous spaces with massive stationary subgroup). Indeed, the discrete case is covered in [19]; the continuous one by the present paper.

Let us specialize this theorem to the context of this section, that is, to the unit sphere in $\mathbf{R}^n$ or $\mathbf{C}^n$. In the examples below we write $(\mathcal{Q}, t)$ to denote the metric space $\mathcal{Q}$ with the substitution function t.

(a) $(S^n(\mathbf{R}), \langle \mathbf{x}, \mathbf{y} \rangle)$. The zonal spherical functions were found by E. Cartan [7] (see also Chapter 9 in [29]) in the form

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

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 [28] that

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

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) $(S^n(\mathbf{R}), |\langle \mathbf{x}, \mathbf{y} \rangle|)$, i.e., the $(n-1)$-dimensional real projective space. In this case one can take $\Phi_k(x) = P_k^{(n-3)/2,-1/2}(2x^2 - 1)$.

(c) $(S^n(\mathbf{C}), |\langle \mathbf{x}, \mathbf{y} \rangle|)$, i.e., the $(n-1)$-dimensional complex projective space. In this case one can take $\Phi_k(x) = P_k^{(n-3)/2,0}(2x^2 - 1)$.

Other spaces for which zonal spherical functions are proportional to Jacobi polynomials, are (d) the quaternionic projective space ($\alpha = (n-2)/2, \beta = 1$) and (e) the Caley plane ($\alpha = (n-2)/2, \beta = 3$). It turns out that cases (a)-(e) exhaust the list of compact two-point homogeneous metric spaces with infinite isometry group (H.-C. Wang's theorem; see, e.g., [14]).

such stationary subgroups are called massive). The function $\phi(g) = \langle T(g)\mathbf{a}, \mathbf{a}\rangle$ is called a *zonal spherical function* corresponding to H. It easy to see [29] that $\phi(g) = \phi(h_1 g h_2)$, where $h_1, h_2 \in H$, i.e., zonal spherical functions are constant on the cosets G/H. For instance, for $S^n(\mathbf{R})$ we have $G = O(n)$, $H = O(n-1)$; so with every coset one can associate an $(n-1)$-dimensional sphere. In this case the zonal spherical functions are constant on spheres (hence the name). To any zonal spherical function $\phi(x)$ one can associate a function (invariant kernel) on $S^n \times S^n$ by putting

$$J_\phi(\mathbf{x}, \mathbf{y}) = \phi(g_{\mathbf{y}}^{-1} g_{\mathbf{x}}), \quad \text{where } g_{\mathbf{x}}\mathbf{a} = \mathbf{x},\ g_{\mathbf{y}}\mathbf{a} = \mathbf{y}.$$

Finally, let $L(g)$ be a quasiregular representation of G in the Hilbert space $L_2(S^n, \eta)$ (η is the normalized Lebesgue measure). Then it can be expanded in a countably infinite sum of irreducible (unitary) representations $L_k(g)$, each acting on a finite-dimensional space of continuous functions. Let ϕ_k be the zonal spherical functions of these representations. The set of functions

$$\Phi_k(s) = J_{\phi_k}(\mathbf{x}, \mathbf{y}), \quad k = 0, 1, \ldots,$$

where $s = t(\mathbf{x}, \mathbf{y})$, is called a complete system of zonal spherical functions of S^n. In particular, $\Phi_0 = 1$. The general geometric background for this setting is given in [29], or, on a more advanced level, in [14].

Let $\Phi_k(x)$ be a complete system of zonal spherical functions of S^n. A specific form of the zonal spherical functions depends both on the field and on the distance function on it. The main theorem of the KL theory [15] asserts that

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

Note that for $S^n(\mathbf{R})$ this theorem was proved in [11]. For finite metric spaces and, more generally, polynomial association schemes, (18) constitutes the Delsarte inequalities [10].

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

Theorem 2 *Let $B(x)$ be the distance distribution of a code $C \subset S^n$ and let m be an integer. Let $-1 \leq u_0 < 1$ and suppose that $u_0 < u_1 < \cdots < u_{m-1} < u_m = t(d(C)) < 1$ are defining points of a partition of $[u_0, t(d(C))]$ into m equal 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) $f_k \geq 0$ *for* $0 \leq k \leq l$,

(ii) $f(x) \leq 0$ *for* $-1 \leq x \leq u_0$.

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

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

PROOF. We have

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

argument and write simply $f(n) \sim g(n)$. 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 $\lim_{x\to\infty}(f(x)/g(x)) \geq 1$.

2 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 the KL paper [15] 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. We begin with a (very brief) review, and the main theorem of the KL theory in this case, keeping in mind our main example, the Euclidean unit sphere.

Consider the unit sphere in $\mathbf{R}^n$ or $\mathbf{C}^n$ with the metric specified below. The sphere will be denoted by $S^n(\cdot)$ with the ground field as the argument. A code, again, is a *finite* subset of S^n. Let $\mathsf{dist}\,(\mathbf{x},\mathbf{y})$ be a certain metric and $\langle \mathbf{x},\mathbf{y}\rangle$ be the usual (Hermitian) inner product on S^n. Let $\mathcal{D}$ be the set of all possible distances on S^n with respect to $\mathsf{dist}\,(\mathbf{x},\mathbf{y})$. For instance, for $S^n(\mathbf{R})$ with the Euclidean metric we have $\mathcal{D} = [0,2]$. Let $t(\mathbf{x},\mathbf{y}) = t(d) : \mathcal{D} \to \mathcal{I} = [-1,1]$ be a function that depends only on the distance between $\mathbf{x}$ and $\mathbf{y}$, such that $t(0) = 1$, $t(\max_{d\in\mathcal{D}} d) = -1$. This substitution enables one to present results in a uniform way while not changing their analytic nature. For instance, for $S^n(\mathbf{R})$ with the Euclidean metric we have $\mathcal{D} = [0,2]$ and $t(\mathbf{x},\mathbf{y}) = \langle \mathbf{x},\mathbf{y}\rangle = 1 - \frac{1}{2}\mathsf{dist}^{\,2}(\mathbf{x},\mathbf{y})$. Thus $t(0) = 1$ ($\mathbf{x} = \mathbf{y}$) and $t(2) = -1$ (a pair of antipodal points).

Let $C \subset S^n$ be a code. Define the functions

$$b_{\mathbf{c}}(s,t) = |\{\mathbf{c}' \in C : s \leq t(\mathbf{c},\mathbf{c}') \leq t\}|, \tag{15}$$

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

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, $b(s) = b\big(s, s+\Theta(\frac{1}{n})\big)$. Observe that $b(s)$ can be thought of as the distance *density* of C. Further, let $B_{\mathbf{c}}(s) = b_{\mathbf{c}}(s,1)$, $B(s) = b(s,1)$ 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, for every interval $I \in \mathcal{I}$ let $\mu_B(I) = \sum_{x\in I}[B(x) - B(x+0)]$, where the sum is finite. Then

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

Let G be the isometry group of S^n and $T(g)$ its irreducible unitary representation. A subgroup H is called stationary if there exists an invariant vector $\mathbf{a} \in S^n$, i.e., a vector such that $T(h)\mathbf{a} = \mathbf{a}$ for all $h \in H$. It is clear that $\mathbf{a}$ is a *unique* invariant direction for H (generally,

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

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

$$F(x,\rho) = R - (1+\rho)H\left(\frac{\rho}{1+\rho}\right) + \int_x^1 \frac{4\rho(1+\rho)dz}{z+\sqrt{z^2-4(1-z^2)\rho(1+\rho)}}.$$

Together with a segment of a 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 $R \in (0, R'']$ (see Fig. 1).

The proof combines geometric, analytic, and combinatorial arguments, and is accomplished in several steps (Sections 2 to 5), each producing a result of its own. First we show that for sufficiently large n any code of large size has a rapidly growing component in the distance distribution. This group of results is proved by a new application of Delsarte's polynomial method in coding theory, discovered recently by the authors in [2],[19]. As suggested by the Kabatiansky-Levenshtein (KL) approach [15], we prove these estimates simultaneously for codes in a number of metric spaces including $S^n(\mathbf{R})$. We believe that these estimates will find further uses in coding theory as it happened with analogous results in the Hamming space [3],[5],[4]. Section 2 is devoted to the proof, in a geometric context, of general estimates on the distance distribution of codes together with a brief review of the KL theory. 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 at least one of the upper indices grows linearly with k. This is the subject of a fairly technical Section 3. 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. Though this section is expanded, we believe that it is worthwhile to present these results in view of a prominent role that Jacobi polynomials play in coding theory.

Section 4 consists of two parts. In the first part we use the estimates of Jacobi polynomials to derive 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 5, is geometrically much more intuitive. It is accomplished by an inclusion-exclusion argument first employed for the Hamming case in [19].

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 $\Gamma(a) \overset{a}{\sim} a(\ln a - 1)$. If both f and g depend on n, we omit the

with a Plotkin-type bound on the size of the code. Independently and earlier Plotkin-type bounds on $d(R)$ were proved by Rankin [23]. 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 [25] for discrete channels and [18] for the Gaussian channel). This bound has the form

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

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

Finally, as shown in [25], the reliability function of discrete channels is bounded above by the straight line connecting any point of the minimum-distance bound with any point of the sphere-packing bound. Sheverdyaev [26] extended this result to the Gaussian channel, showing 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:

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

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 [27], [24], we assume that code vectors can be any points on S^n. 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 [21] 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., [22].

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*

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

where R is a value of the code rate,

$$0 \le d \le \frac{\sqrt{2}\left(\sqrt{1+\rho_{kl}} - \sqrt{\rho}_{kl}\right)}{\sqrt{1+2\rho_{kl}}}, \quad d \le w \le \frac{\sqrt{2}\left(\sqrt{1+\rho} - \sqrt{\rho}\right)}{\sqrt{1+2\rho}},$$

code vectors since their probability is 0.) Under this decoding, the error probability $P_e(W)$ equals

$$P_e(W) = \frac{1}{|W|} \sum_{\mathbf{x} \in W} \Pr\{\mathbf{z} \in \cup_{\substack{\mathbf{y} \in W \\ \mathbf{y} \neq \mathbf{x}}} D(\mathbf{y}, W)\}, \tag{8}$$

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

$$\begin{aligned} P_e(R, A, n) &= \min_{W: R(W) \geq R} P_e(W) \\ E(R, A, n) &= -\frac{1}{n} \log P_e(R, A, n) \\ \overline{E}(R, A) &= \limsup_{n \to \infty} E(R, A, n) \\ \underline{E}(R, A) &= \liminf_{n \to \infty} E(R, A, n). \end{aligned}$$

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* of the channel, denoted $E(R, A)$. By abuse of notation, below we speak of upper and lower bounds on $E(R, A)$.

Shannon [24] 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 [24]

$$E(R, A) \leq E_{sp}(\theta(R), A), \tag{9}$$

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 [24]:

$$E(R, A) \geq \begin{cases} \frac{A}{4}(1 - \cos\theta(R)), & 0 \leq R \leq R_1, \quad \text{(I)} \\ \frac{A}{4}(1 - \cos\theta(R_1)) + R_1 - R_0, & R_1 \leq R \leq R_2, \quad \text{(II)} \\ E_{sp}(\theta(R), A), & R_2 \leq R \leq \mathcal{C}, \quad \text{(III)} \end{cases} \tag{10}$$

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 [24] also proved the inequality

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

which implies that bound (10.I) is tight for $R = 0$. The proof of (11) in [24] used the minimum-distance argument (the probability of mixing two code vectors at a minimum distance) together

The best known lower bound on $R(\theta)$ is Shannon's sphere-packing bound [24]:

$$R(\theta) \geq -\ln \sin\theta \quad (0 \leq \theta \leq \pi/2), \tag{1}$$

$$d^2(R) \geq 2(1-\sqrt{1-e^{-2R}}) \quad (0 \leq R < \infty), \tag{2}$$

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 [15]. One of the main results of [15] states that

$$R(\theta) \leq \frac{1+\sin\theta}{2\sin\theta} H\Big(\frac{1-\sin\theta}{1+\sin\theta}\Big) \quad (0 \leq \theta \leq \pi/2), \tag{3}$$

where $H(x) = -x\ln x - (1-x)\ln(1-x)$ is the entropy function. This bound admits a small asymptotic improvement for $0 < \theta < 63°$ [15]:

$$R(\theta) \leq -(1/2)\ln(1-\cos\theta) - 0.099$$

(the real numbers here and below are approximate). There 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

$$R = (1+\rho)H\Big(\frac{\rho}{1+\rho}\Big), \tag{4}$$

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

$$d^2(R) \leq d^2_{kl}(R) := \frac{2(\sqrt{1+\rho_{kl}} - \sqrt{\rho_{kl}})^2}{1+2\rho_{kl}}, \tag{5}$$

$$d^2(R) \leq 2e^{-2R-0.137}. \tag{6}$$

For $0.234 < R \leq \infty$ 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 [24]. Let W be a code on the sphere of radius $\sigma\sqrt{An}$. We assume that code vectors for transmission are chosen with equal probability. Then

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

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

$$D(\mathbf{x}, W) = \{\mathbf{z} \in S^n : ||\mathbf{z}-\mathbf{x}|| < ||\mathbf{z}-\mathbf{x}'|| \text{ for all } \mathbf{x}' \in W, \mathbf{x} \neq \mathbf{x}\} \tag{7}$$

be the Voronoi domain of $\mathbf{x}$ 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 $\mathbf{z}$ a code vector $\mathbf{x}$ such that $\mathbf{z} \in D(\mathbf{x}, W)$. (The definition ignores vectors $\mathbf{z}$ at the same distance from 2 or more

1 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 $\mathbf{x}$ is sent over the channel, the received signal is represented by a vector $\mathbf{z} = \mathbf{x} + \mathbf{y}$, where $\mathbf{y}$ is a Gaussian 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 [24] 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 n-dimensional sphere in $\mathbf{R}^n$. 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, we study only codes on the *unit* sphere in $\mathbf{R}^n$, denoted by $S^n = S^n(\mathbf{R})$ below. A *code* is a finite subset of $S^n(\mathbf{R})$. To distinguish between codes on S^n 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 $S^n(\mathbf{R})$ with spherical caps. This and closely related problems of finding the best possible fillings of $\mathbf{R}^n$ with identical spheres and the kissing number were studied long before the emergence of coding theory (see a survey in [9]). Spherical codes in information theory were introduced by Slepian [27] (paper based on a 1951 Bell Labs report) and Shannon [24]. However, studies in geometry and coding theory developed independently of each other until the second half of the 70-ies when important unifying steps were taken by Delsarte, Goethals, and Seidel [11],[12], and Kabatiansky and Levenshtein [15].

A. Parameters of Spherical Codes. Let $C \in S^n$ be a code. The distance $\mathsf{dist}(\mathbf{x},\mathbf{y}) = ||\mathbf{x}-\mathbf{y}||^{1/2}$ between two points in S^n can be also measured by the inner product $\langle \mathbf{x},\mathbf{y}\rangle = 1 - \frac{1}{2}\mathsf{dist}^2(\mathbf{x},\mathbf{y})$ or by the angle $\theta = \arccos\langle \mathbf{x},\mathbf{y}\rangle$. Each of these measures is convenient in some coding-theoretic problems. Accordingly, let

$$d(C) := \min_{\substack{\mathbf{x},\mathbf{y}\in C \\ \mathbf{x}\neq\mathbf{y}}} \mathsf{dist}(\mathbf{x},\mathbf{y}), \quad (0 \le d(C) \le 2),$$

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

Let $R(C) = \frac{1}{n}\ln C$ be the rate of the code, $0 \le R(C) < \infty$. Further, let

$$R(\theta) = \limsup_{n\to\infty} R(C)$$

be the asymptotic value of the maximum possible rate of a code (sequence of codes) of angular distance $\ge \theta$. Finally, define $d(R) = \limsup_{n\to\infty} d(C)$, where the limit is computed over all sequences of codes with rate $R(C) \ge R$. By [23], $R(\theta) = 0$ for $\pi/2 \le \theta \le \pi$.

A new upper bound on the reliability function of the Gaussian channel

A.Ashikhmin * A. Barg † S. Litsyn ‡

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 Euclidean n-dimensional sphere. First we prove that the distance distribution of codes of large size necessarily contains a large component. A general theorem establishing this estimate is proved simultaneously for codes on the Euclidean sphere and in real and complex projective spaces.

To derive specific estimates of the distance distribution, we study the asymptotic behavior of Jacobi polynomials $P_k^{\alpha,\beta}$ as $k \to \infty$ and at least one of the upper indices grows linearly in k. This group of results provides the exact behavior of the exponent of Jacobi polynomials in the entire orthogonality segment.

Since on the average there are many code vectors in the vicinity of the transmitted vector $\mathbf{x}$, one can show that the probability of confusing $\mathbf{x}$ 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.

Keywords: Spherical codes, distance distribution, Jacobi polynomials, error probability of decoding.

*Los Alamos National Laboratory, Group CIC-3, Mail Stop P990, Los Alamos, NM 87545, alexei@c3serve.c3.lanl.gov.

†Bell Laboratories, Lucent Technologies, 600 Mountain Avenue 2C-375, Murray Hill, NJ 07974, abarg@research.bell-labs.com.

‡Research done while visiting DIMACS Center, Rutgers University, Piscataway, NJ 08854. Permanent address: Department of Electrical Engineering–Systems, Tel Aviv University, Ramat Aviv 69978, Israel, litsyn@eng.tau.ac.il.