Kissing numbers, sphere packings, and some unexpected proofs

FLORIAN PFENDER* and GÜNTER M. ZIEGLER**
TU Berlin, MA 6-2, D-10623 Berlin, Germany
{fpfender,ziegler}@math.tu-berlin.de

April 19, 2004

The "kissing number problem" asks for the maximal number of white spheres that can touch a black sphere of the same size in n-dimensional space. The answers in dimensions one, two and three are classical, but the answers in dimensions eight and twenty-four were a big surprise in 1979, based on an extremely elegant method initiated by Philippe Delsarte in the early seventies, which concerns inequalities for the distance distributions of kissing configurations.

Delsarte's approach led to especially striking results in cases where there are exceptionally symmetric, dense and unique configurations of spheres: In dimensions eight and twenty-four these are given by the shortest vectors in two remarkable lattices, known as the E_8 and the Leech lattice.

However, despite the fact that in dimension four there is a special configuration which is conjectured to be optimal and unique—the shortest vectors in the D_4 lattice, which are also the vertices of a regular 24-cell—it was *proved* that the bounds given by Delsarte's method aren't good enough to solve the problem in dimension four. This may explain the astonishment even to experts when in the fall of 2003, Oleg Musin announced a solution of the problem, based on a clever modification of Delsarte's method [21, 22].

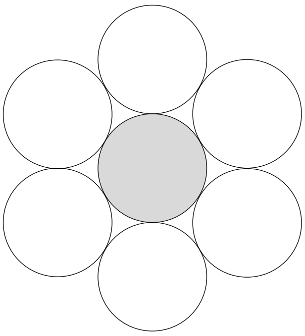
Independently, Delsarte's by now classical approach has recently also been adapted by Henry Cohn and Noam Elkies [5] to deal with optimal sphere packings more directly and more effectively than had been possible before. Based on this, Henry Cohn and Abhinav Kumar [6] have now proved that the sphere packings in dimensions eight and twenty-four given by the E_8 and Leech lattices are optimal lattice packings (for dimension eight this had been shown before) and that they are optimal sphere packings, up to an error of not more than 10^{-28} percent.

Here we try to sketch the setting, to explain some of the ideas, and to tell the story. For this we start with a brief review of the sphere packing and kissing number problems. Then we look at the remarkable kissing configurations in dimensions four, eight and twenty-four. We give a sketch of Delsarte's method, and how it was applied for the kissing number problem in dimensions eight and twenty-four. Then Musin's ideas kick in, which leads us to look at some non-linear optimization problems, as they occur as subproblems in his approach. Finally we sketch an elegant construction of the Leech lattice in dimension twenty-four, which starts from the graph of the icosahedron and very simple linear algebra. This is the lattice which Cohn and Kumar have now proved to be optimal in dimension twenty-four, by another extremely elegant and puzzling adaption of Delsarte's method. A sketch for this will end our tour.

*Supported by the DFG Research Center "Mathematics in the Key Technologies" (FZT86)
**Partially supported by Deutsche Forschungs-Gemeinschaft (DFG), via the DFG Research Center "Mathematics in the Key Technologies" (FZT86), the Research Group "Algorithms, Structure, Randomness" (Project ZI 475/3), and a Leibniz grant

Three classical problems

The "**kissing number problem**" is a basic geometric problem that got its name from billiards: Two balls "kiss" if they touch. The kissing number problem asks how many white balls can touch one given ball at the same time, if all the balls have the same size. If you arrange the balls on a pool table, it is easy to see that the answer is exactly six: Six balls just perfectly surround a given ball.

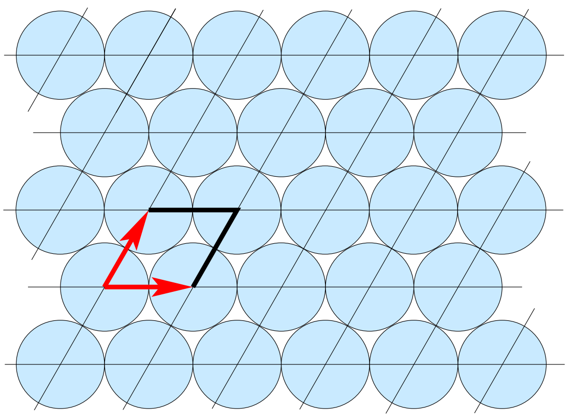


The perfect kissing arrangement for $n = 2$

The **sphere packing problem** is to determine the maximal density of a packing of balls (all of them of the same size) in Euclidean n-space.

One class of packings to consider are *lattice* packings, which are invariant under any translation that takes one ball of the packing to the other.

It is a simple exercise (recommended) to prove, for dimension two, that the "obvious" hexagonal packing of equal-sized disks (two-dimensional balls) in the plane—a lattice packing in which each disk touches $\kappa(2) = 6$ others—is the optimal lattice packing, and to compute its density.



The hexagonal lattice packing in the plane

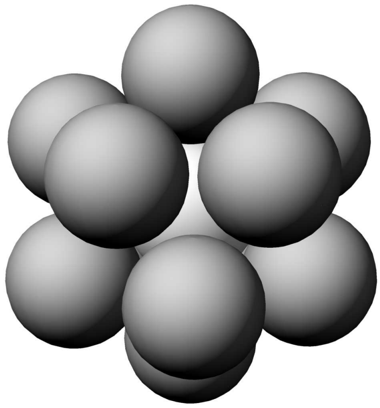
It is not so easy to prove that the hexagonal packing is indeed an optimal sphere packing for dimension two. (Experts disagree whether the first proof for this, given by Thue 1892/1910, was indeed complete; if there was a gap, it was closed by Mahler and by Segre in 1940. See e. g. [13] for a proof.)

Thus the hexagonal planar lattice packing yields optimal solutions for the two-dimensional cases of the kissing number problem, the lattice packing problem, and the sphere packing problem. However, there are various indications that solutions of these three problems in higher dimensions are not so simple, they are not just given by "one perfect lattice packing," and things are much more complicated than in dimension two. This starts to show already in dimension three.

Geometry is difficult . . .

. . . as soon as you reach dimension three.

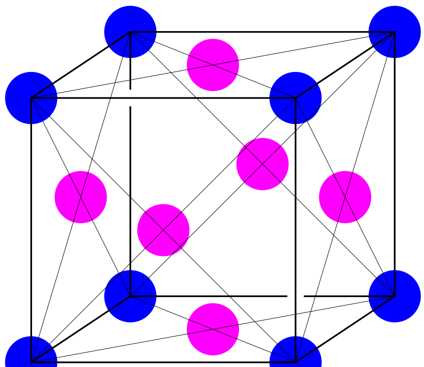
The **kissing number problem** in dimension three asks "How many balls can touch a given ball at the same time." This problem is indeed very interesting, and surprisingly hard. Isaac Newton and David Gregory had a famous controversy about it in 1694: Newton said that 12 should be the correct answer, while Gregory thought that 13 balls could fit. The regular icosahedron yields a configuration of 12 touching balls that has great beauty and symmetry, and leaves considerable gaps between the balls, which are clearly visible in our figure.



The icosahedron configuration
(Graphics: Detlev Stalling, ZIB Berlin)

So perhaps if you move all of them to one side, would a 13th ball possibly fit in? It is a close call, but the answer is no, 12 is the correct answer. To *prove* this is a hard problem, which was finally solved by Schütte and van der Waerden [26] in 1953. A short sketch of an elegant proof was given by Leech [17] in 1956: But it is a substantial challenge to derive a complete proof from this.

The **lattice packing problem** for dimension three was solved by Gauß in 1831, in an "Anzeige" (what today we'd call a book review) of a book by Ludwig August Seeber. Indeed, Gauß proved a result about ternary quadratic forms which he even interpreted geometrically, and which easily implies that the so-called "face-centered cubic (fcc)" packing is the unique densest lattice sphere packing for dimension three.



The fcc sphere packing

The centers for this sphere packing are all the integral points in $\mathbb{Z}^3$ with exactly one or exactly three even coordinates. Again it's a nice exercise to prove that this does, indeed, give a lattice packing, that we can pack spheres of radius $\frac{1}{2}\sqrt{2}$ with their centers in the lattice points, to compute the density of the resulting sphere packing, and to recognize that in this packing, each sphere is "kissed" by exactly 12 other spheres—whose touching points do *not* give a regular icosahedron.

Just recall that the general **sphere packing** problem for dimension 3, known as the "Kepler conjecture," was only recently solved, by Thomas C. Hales. The controversial story about that case has been told elsewhere (see for instance [13, 14] and [27]), and may even continue after the publication of Hales' papers (which are expected to appear in the *Annals* and *Discrete & Comput. Geometry*).

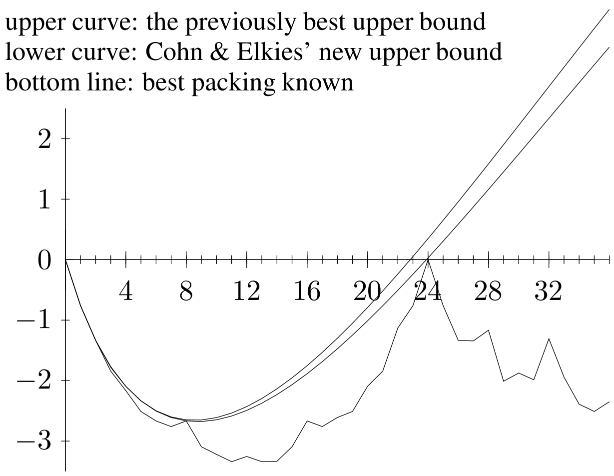
So the **lattice packing problem** is different from the general sphere packing problem, and it seems to be considerably simpler. This starts with the fact that lattice packings are easy to describe (by a basis matrix). The density of a lattice packing is easily derived from the length of a shortest nonzero lattice vector and the determinant of a basis matrix. Also the subtleties in the definition of "density" of a sphere packing disappear in the lattice case.

. . . and in high dimensions

It is likely that for most dimensions the optimal kissing arrangement is not unique and not rigid, the optimal sphere packing is not a lattice packing, and thus the methods discussed in this paper will not be able to give optimal results—but they do give the best *known* results in virtually the whole range of dimensions, from $n = 1$ to very large.

Here are three indications that the final answers in high dimensions won't be extremely simple:

- The optimal lattice packings E_7, E_8 and Λ_9 (conjectured) in dimensions 7, 8 and 9 have approximate densities 0.29530, 0.25367 and 0.14577, respectively—so there is a "sudden drop" beyond $n = 8$, it seems. A similar effect happens at $n = 24$. (See the figure below, taken from Cohn & Elkies [5] with their kind permission). This non-monotone behaviour indicates that there are "special effects" happening in special dimensions.



Plot of $\log_2 \delta + n(24 - n)/96$ vs. dimension n, where δ is the "center density"

(from Cohn & Elkies [5])

- In dimension $n = 9$ the non-lattice packing known as "P9a" contains spheres which kiss 306 others, while it is *known* that in each lattice packing the kissing number (which is the same for all spheres) cannot exceed 272. So it seems that in general, the optimal kissing configuration is not given by a lattice.
- In dimension $n = 10$ the packing "P10c" has a greater density than the best known lattice packing, "Λ_{10}."

In most dimensions, there is not even a plausible conjecture for a best sphere packing; also every dimension seems to have its own characteristics, with remarkable phenomena occuring in dimensions 4, 8 and 24—which is, however, not reflected in the upper bounds we have.

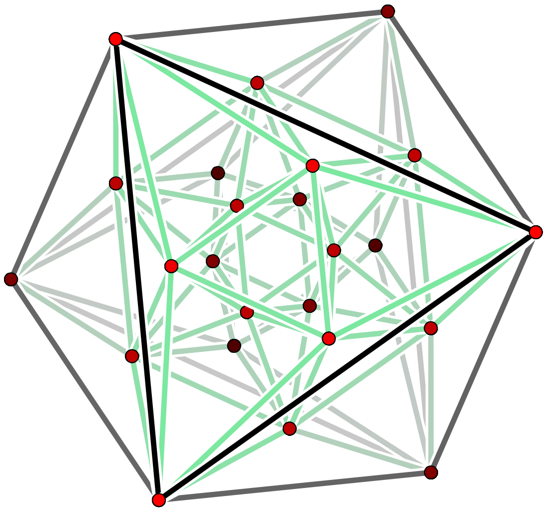
Three kissing configurations

The theory of lattices and sphere packings features some of the most beautiful objects in mathematics, including some remarkable kissing configurations in special dimensions. In the following, we describe optimal kissing configurations of spheres in dimensions 4, 8 and 24. In each of them, the vectors are the shortest vectors of a lattice of high symmetry, and there are special binary codes, large simple groups, and a lot of other miracles attached to them. Thompson's little book [28] is a nice historical account of the discoveries, Conway & Sloane's book [7] is the classical technical account, which includes a number of the key research papers in the subject, and Elkies' prize-winning *Notices* papers [11] explain a lot of the connections to other mathematical fields such as theta functions and modular forms.

$n = 4$: There are 24 vectors with two zero components and two components equal to ± 1; they all have length $\sqrt{2}$, and a minimum distance of $\sqrt{2}$. Properly rescaled (that is, multiplied by $\sqrt{2}$) they yield the centers for a kissing configuration of unit spheres, and imply that $\kappa(4) \geq 24$. The convex hull of the 24 points yields a famous 4-dimensional polytope, the "24-cell" discovered in 1852 by Ludwig Schäfli. Its facets are 24 regular octahedra.

$n = 8$: Again we present a configuration with simple integer coordinates which then can be rescaled. Our configuration includes the $\binom{8}{2}4 = 112$ vectors of type "$(0^6, \pm 2^2)$," that is, with two nonzero coordinates, which are ± 2, as well as the $2^7 = 128$ vectors of type "$(\pm 1^8)$" with an even number of negative components. All the $112 + 128 = 240$ vectors have length $\sqrt{8} = 2\sqrt{2}$, which is also the minimum distance between the points.

At the same time, the vectors above are the shortest nonzero vectors of the exceptional root lattice E_8, which appears, for example, in the classification of simple Lie algebras. It consists of all integral vectors in $\mathbb{R}^8$ whose coordinates are all odd or all even, and for which the sum of all coordinates is divisible by 4.



A "Schlegel diagram" of the 24-cell
(Graphics: Michael Joswig/`polymake` [12])

$n = 24$: The configuration consists of the shortest (nonzero) vectors in a remarkable lattice, the Leech lattice, for which we will outline a simple construction further below.

The vectors have three different types: The vectors of type "$(0^{16}, \pm 2^8)$" have 16 zero coordinates, and eight coordinates that are ± 2, with an even number of minus signs. The Leech lattice contains $759 \cdot 2^7 = 97152$ of them, all of them of length $\sqrt{32} = 4\sqrt{2}$. The second type of vectors are "$(0^{22}, \pm 4^2)$," with two non-zero components, ± 4, of arbitrary sign. There are $\binom{24}{2}4 = 1104$ of them, again of length $\sqrt{32}$, and we take them all. The third type is vectors of the form "$(\pm 1^{23}, \pm(-3))$," obtained from a vector with one entry -3 and all entries $+1$ by reversing the sign on a number of coordinates which is divisible by 4. Exactly $3 \cdot 2^{15} =$

98304 of these are contained in the Leech lattice, again of length $\sqrt{32}$. Miraculously, all the resulting $97152 + 1104 + 98304 = 196560$ vectors have the same length, and the minimum distance between them is again $\sqrt{32}$: and this minimum distance is achieved *very often*.

The Delsarte method

Philippe Delsarte (Phillips Research Labs) started in the early seventies [8] to develop an approach that via linear programming yields upper bounds for cardinalities of binary codes where Krawtchouk polynomials appear at the core of the method. (See Best [4] for a beautiful exposition.)

However, Delsarte's approach was much more general, yielding cardinality bounds for "association schemes" [9]. An important case is the situation for spherical codes in the Delsarte–Goethals–Seidel method [10], where Gegenbauer polynomials play the decisive role.

Here is our sketch: If N unit spheres kiss the unit sphere in $\mathbb{R}^n$, then the set of kissing points is a rather special configuration of unit vectors, namely N vectors $x_1, \ldots, x_N \in \mathbb{R}^n$ that satisfy $\langle x_i, x_j \rangle \leq \frac{1}{2}$ for $i \neq j$, while $\langle x_i, x_i \rangle = 1$ for all i. If we write the x_i as the columns of a matrix $X \in \mathbb{R}^{n \times N}$, then the special properties amount to a matrix

$$\big(x_{ij}\big) \;:=\; X^T X \;\in\; \mathbb{R}^{N \times N}$$

with the following properties:

(i) it has ones on the diagonal,
(ii) all off-diagonal entries are at most $\frac{1}{2}$,
(iii) it has rank (at most) n, and
(iv) it is positive semidefinite.

Now we use a result that may be traced to a paper by Schoenberg [25]. He characterized the functions f one may apply to the entries of matrices with properties (i), (iii), and (iv) such that the resulting matrix

$$\big(f(x_{ij})\big)$$

is guaranteed to be still positive semidefinite. If we restrict f to be a polynomial of degree at most d, then Schoenberg's answer is that f can be an arbitrary non-negative combination of the *Gegenbauer polynomials* $G_k^{(n)}$ of degree $k \leq d$. These polynomials (also known as the *spherical* or the *ultraspherical* polynomials) may be defined in a variety of ways. One compact description is that for any $n \geq 2$ and $k \geq 0$, $G_k^{(n)}(t)$ is a polynomial of degree k, normalized such that $G_k^{(n)}(1) = 1$, and such that $G_0^{(n)}(t) = 1$, $G_1^{(n)}(t) = t$, $G_2^{(n)}(t) = \frac{nt^2-1}{n-1}$, ... are orthogonal with respect to the scalar product

$$\langle\, g(t), h(t)\,\rangle \;:=\; \int_{-1}^{+1} g(t)h(t)(1-t^2)^{\frac{n-3}{2}}\,dt$$

on the vector space $\mathbb{R}[t]$ of polynomials, which arises naturally in integration over S^{n-1}. This is just one of many possible descriptions and definitions of these remarkable polynomials. For example, the readers are invited to derive a recursion from this description by applying Gram–Schmidt orthogonalization. For $n = 3$ one obtains the Legendre polynomials, for $n = 4$ the Chebychev polynomials of the second kind (but with a different normalization than usual). Perhaps one more useful fact to know about Gegenbauer polynomials is that computer algebra systems such as `Maple` and `Mathematica` "know them".

The key property of the Gegenbauer polynomials that we need, Schoenberg's lemma, is a simple consequence of the classical addition theorem for spherical harmonics—beautifully explained and derived in the book by Andrews, Askey & Roy [2, Chap. 9], who credit Müller [20], who in turn says that this goes back to Gustav Herglotz (1881–1925).

Lemma 1 (Addition Theorem [2, Thm. 9.6.3]). *The Gegenbauer polynomial $G_k^{(n)}(t)$ can be written as*

$$G_k^{(n)}(\langle x, y\rangle) \;=\; \frac{\omega_n}{m}\sum_{\ell=1}^{m} S_{k,\ell}(x)S_{k,\ell}(y),$$

where ω_n is the $(n-1)$-dimensional area of S^{n-1}, and the functions $S_{k,1}, S_{k,2}, \ldots, S_{k,m}$ form an orthonormal basis for the space of "spherical harmonics of degree k," which has dimension $m = m(k,n) = \binom{k+n-2}{k} + \binom{k+n-3}{k-1}$.

This easily yields Schoenberg's result:

Lemma 2 (Schoenberg [25]).
If $(x_{i,j}) \in \mathbb{R}^{N\times N}$ is a positive semidefinite matrix of rank at most n with ones on the diagonal, then the matrix $\big(G_k^{(n)}(x_{i,j})\big)$ is positive semidefinite as well. In particular, the sum of all its entries is non-negative.

Proof. We can write the matrix $(x_{i,j})$ as $X^T X$, that is, $x_{i,j} = \langle x_i, x_j\rangle$ for vectors $x_i, x_j \in S^{n-1}$. Here we prove only that the sum of all entries of $\big(G_k^{(n)}(x_{i,j})\big)$ is non-negative: For this we plainly compute

$$\sum_{i,j=1}^{N} G_k^{(n)}(\langle x_i, x_j\rangle) \;=\;$$

$$= \frac{\omega_n}{m}\sum_{i,j=1}^{N}\sum_{\ell=1}^{m} S_{k,\ell}(x_i)S_{k,\ell}(x_j)$$

$$= \frac{\omega_n}{m}\sum_{\ell=1}^{m}\Big(\sum_{i=1}^{N} S_{k,\ell}(x_i)\Big)\Big(\sum_{j=1}^{N} S_{k,\ell}(x_j)\Big)$$

$$= \frac{\omega_n}{m}\sum_{\ell=1}^{m}\Big(\sum_{i=1}^{N} S_{k,\ell}(x_i)\Big)^2 \;\geq\; 0. \qquad \square$$

To get a feel for "what this means," let (x_{ij}) be a positive semidefinite matrix of rank $n \geq 2$ with ones on the diagonal, and let's look at the polynomials $f(t)$ such that $\big(f(x_{ij})\big)$ has a non-negative sum of entries. Clearly $f(t) = 1$ has this property, and $f(t) = t$ as well. It starts to be interesting if we apply $f(t) = t^2+\alpha$, since then the set of admissible αs depends on the rank n. The claim of Schoenberg's lemma is that we can take any $\alpha \geq -\frac{1}{n}$, since $t^2+\alpha = \frac{n-1}{n}G_2^{(n)}(t) + (\frac{1}{n}+\alpha)$.

Theorem 3 (Delsarte, Goethals & Seidel [10]).
If

$$f(t) \;=\; \sum_{k=0}^{d} c_k G_k^{(n)}(t)$$

is a nonnegative combination of Gegenbauer polynomials, with $c_0 > 0$ and $c_k \geq 0$ otherwise, and if $f(t) \leq 0$ holds for all $t \in [-1, \frac{1}{2}]$, then the kissing number for $\mathbb{R}^n$ is bounded by

$$\kappa(n) \;\leq\; \frac{f(1)}{c_0}.$$

Proof. We estimate the sum of all entries of the matrix $\big(f(x_{ij})\big)$ in two ways.

The first one is the simple computation

$$\sum_{i,j=1}^{N} f(x_{ij}) \;=\; \sum_{k=0}^{d} c_k \sum_{i,j=1}^{N} G_k^{(n)}(x_{ij})$$

$$\geq\; c_0 \sum_{i,j=1}^{N} G_0^{(n)}(x_{ij}) \;=\; c_0 N^2,$$

which rests on the fact that by Schoenberg's lemma the sum of all entries of the matrix $\big(G_k^{(n)}(x_{ij})\big)$ is nonnegative.

The second, equally simple computation

$$\sum_{i,j=1}^{N} f(x_{ij}) \;=\; N\,f(1) + \sum_{i\neq j} f(x_{ij}) \;\leq\; N\,f(1) \tag{1}$$

depends on the fact that all the off-diagonal entries of the matrix $\big(f(x_{ij})\big)$ are nonpositive, due to our assumption on the function $f(t)$ in the range where the scalar products $x_{ij} = \langle x_i, x_j\rangle$ lie.

Now the two estimates yield $c_0 N^2 \leq N\,f(1)$. $\square$

$n = 8$ and $n = 24$

The kissing number problems in dimensions eight and twenty-four were solved in the late seventies, by matching the Delsarte–Goethals–Seidel bound with the very special E_8 and Leech configurations: Andrew Odlyzko and Neil Sloane (at AT&T Bell Labs) and independently Vladimir I. Levenšteĭn in Russia proved that the correct, exact maximal numbers for the kissing number problem are $\kappa(8) = 240$ and $\kappa(24) = 196560$.

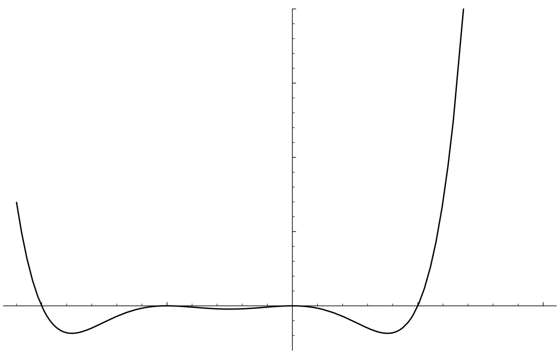
In dimensions with a candidate for a unique optimal configuration for the kissing number problem, one has a quite straightforward guess for the polynomial to be used in Delsarte's method. Namely, for the estimate (1) to be tight, we must have $f(x_{ij}) = 0$ for all scalar products x_{ij} that actually occur for $i \neq j$ in our candidate solution.

Thus, in dimension $n = 8$ the configuration given by the roots of the E_8 lattice seems so nice and dense and rigid that it might be unique. It is also very symmetric, and the only scalar products that occur (if the roots are normalized to length 1) are

± 1, $\pm\frac{1}{2}$, and 0. Thus the "obvious" function to write down is

$$f_8(t) \;:=\; (t-\tfrac{1}{2})t^2(t+\tfrac{1}{2})^2(t+1).$$

You (or your computer algebra system) have to expand this polynomial in terms of Gegenbauer polynomials, check that all coefficients in the expansion are nonnegative, compute $f_8(1)/c_0$, and get 240; this is what Odlyzko & Sloane [23] [7, Chap. 13] did in 1979, and Levenšteĭn [18] did independently at the same time.



$f_8(t)$

One can proceed similarly for $n = 24$ and the shortest roots of the Leech lattice, which have the additional scalar products $\pm\frac{1}{4}$. Thus one uses

$$f_{24}(t) \;:=\; (t-\tfrac{1}{2})(t+\tfrac{1}{4})^2t^2(t+\tfrac{1}{4})^2(t+\tfrac{1}{2})^2(t+1).$$

It works!

The same approach *cannot* work for dimension $n = 3$, where the optimal configuration is far from unique, so the function f would have to be equal to zero in a whole range of possible scalar products to get a tight estimate $N \le 12$, or would have to be close to zero to get the estimate $N < 12.9999$, say, which would be good enough to prove $\kappa(3) = 12$.

Status 2004: Kissing Numbers

The only exact values of kissing numbers known:

n			lattice	regular polytope
$\kappa(1)$	$=$	2	A_1	
$\kappa(2)$	$=$	6	A_2	hexagon
$\kappa(3)$	$=$	12	H_3	icosahedron
$\kappa(4)$	$=$	24	D_4	24-cell
$\kappa(8)$	$=$	240	E_8	
$\kappa(24)$	$=$	196560	Λ_{24}	

Musin's trick

To determine the kissing number for $n = 4$ has been a challenge for quite a while now. There is a claim by Wu-Yi Hsiang that dates back to 1993, but that apparently hasn't been backed up by a detailed proof. It may be surprising that for $n = 4$ the Delsarte method doesn't work: After all, we have a conjectured unique optimal configuration of unit vectors, given by the 24-cell (the D_4 lattice), with only very few scalar products between distinct points ($\pm\frac{1}{2}$, 0, and -1). However, from the "obvious" polynomial

$$f_4(t) \;=\; (t-\tfrac{1}{2})t^2(t+\tfrac{1}{2})^2(t+1),$$

which is the same polynomial as for $n = 8$ and the E_8 lattice, we only get $\kappa(4) \le f_4(1)/c_0 = 28.8$.

Once we look a little bit harder for a suitable function, Delsarte's bound yields that $\kappa(4) \le 25$, but nothing better than that. Arestov & Babenko [3] have analyzed this case in detail, and *proved* that even with an optimally chosen Delsarte function, the bound obtained will not be smaller than 25.

So it came as a great surprise that now the Russian mathematician Oleg Musin, who lives in Los Angeles, has indeed found a method to modify Delsarte's method in a very beautiful and clever way, which yields better bounds. In particular he improves the upper bound from 25 to 24.

Oleg R. Musin
(photo: private)

In the meantime, an announcement [22] has appeared in print, the long version [21] is submitted and being refereed. So let's assume that the details and computations work out right (which are technical, and for some of which alternative routes are outlined in the preprint) and will be confirmed in the reviewing process. Then, indeed, $\kappa(4) = 24$ is the answer! However, Musin makes no claim to have proved that the special configuration of the 24-cell is unique.

Here we want to only sketch Musin's beautiful idea: He allows the function $f(t)$ to get positive "opposite to the given sphere," that is, close to $t = -1$. Here is the result in a nutshell.

Theorem 4 (Musin's theorem). *Fix a parameter t_0 in the range $-1 \leq t_0 < -\frac{1}{2}$. If*

$$f(t) \ = \ \sum_{k\geq 0} c_k G_k^{(n)}(t)$$

is a nonnegative combination of Gegenbauer polynomials ($c_k \geq 0$ for all k, with $c_0 > 0$), and if $f(t) \leq 0$ holds for all $t \in [t_0, \frac{1}{2}]$, while $f'(t) < 0$ for $t \in [-1, t_0]$, then the kissing number for $\mathbb{R}^n$ is bounded by

$$\kappa(n) \ \leq \ \tfrac{1}{c_0} \max\{h_0, h_1, \ldots, h_\mu\},$$

where h_m is the maximum of

$$f(1) + \sum_{j=1}^{m} f(\langle e_1, y_j\rangle)$$

over all configurations of $m \leq \mu$ unit vectors y_j in the spherical cap given by $\langle e_1, y_j\rangle \leq t_0$ whose pairwise scalar products are at most $\frac{1}{2}$. Here μ denotes the maximal number of points that fit into the spherical cap.

Proof. We argue just as in the proof of Delsarte's theorem—except that in (1) we cannot drop all non-diagonal terms: We now only get that

$$\sum_{i,j=1}^{N} f(x_{ij}) \ \leq \ \sum_{i=1}^{N} \Big(f(1) + \sum_{j:\langle x_i, x_j\rangle \leq t_0} f(x_{ij}) \Big). \tag{2}$$

Letting m denote the number of points that could appear in the last sum ($0 \leq m \leq \mu$) yields the estimate in Musin's theorem. □

Nonlinear optimization problems

The bad news about Musin's approach is that it forces him to compute, at least approximately, the numbers $h_0, \ldots, h_\mu$, and this leads to non-convex optimization problems.

Almost everything can be written as a non-linear, non-convex, constrained optimization problem. For example, the question whether 25 spheres can kiss a given sphere is immediately answered if we solve the problem

$$\min_{x_1,\ldots,x_{25}\in S^3} \ \max_{1\leq i<j\leq 25} \langle x_i, x_j\rangle. \tag{3}$$

Indeed, if the answer is $\frac{1}{2}$ or smaller, then the 25 points x_i that achieve the minimum give a kissing configuration. If the answer is larger than $\frac{1}{2}$, then a kissing configuration with 25 spheres doesn't exist. *However*, high-dimensional, non-linear, non-convex, constrained optimization problems are extremely hard to solve. We may interpret (3) as a problem in 100 variables, the coordinates of $x_1, \ldots, x_{25} \in \mathbb{R}^4$, constrained by the restrictions $x_i \in S^3$. Or we may eliminate the constraints, say by introducing polar coordinates, and thus have an unconstrained problem in 75 variables. Eliminating the symmetry of the 3-sphere will reduce the number of variables by 10, but not significantly simplify the problem. This problem is non-convex, since it has lots of minima: Indeed, any asymmetric optimal configuration will yield 25! minima, and we may assume that there are lots of "combinatorially different" optimal solutions. Numerical methods for non-linear optimization, such as local descent methods, might find some feasible point, and they should even find a stationary point, say a local maximum or minimum of the function to be optimized. Such methods exist and are widely used, the most popular ([24], `Matlab`) and the most questionable ([16], [19]) one probably being the Nelder–Mead simplex method. However, a local improvement method can't guarantee to find a global optimum.

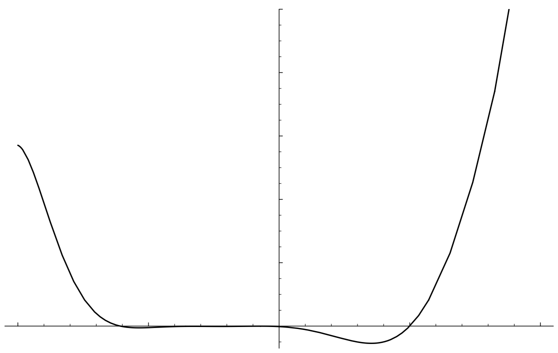
Dimension reduction

However, *the good news* for Musin's approach is that if parameters are chosen carefully, if t_0 is small enough (that is, close to -1), and if the monotonicity assumption is exploited carefully, then one gets low-dimensional problems of a type that can be treated numerically.

So, it is already a remarkable achievement that Musin's improvement of the Delsarte method yields a clean and simple proof for the Newton–Gregory problem, $\kappa(3) < 13$. Indeed, choosing $t_0 = -0.5907$ and a suitable polynomial $f_3(t)$ of degree 9, Musin gets $\mu = 4$, and the parameters $h_0 = f_3(1) = 10.11$, $h_1 = f_3(1) + f_3(-1) = 12.88$, while all other h_i's are smaller.

A sketch of Musin's proof for $\kappa(4) < 25$. Musin produced a polynomial of degree 9 satisfying the assumptions of his theorem with $t_0 \approx -0.608$:

$$\begin{aligned} f(t) &= G_0^{(4)}(t) + 2G_1^{(4)}(t) + 6.12G_2^{(4)}(t) \\ &\quad + 3.484G_3^{(4)}(t) + 5.12G_4^{(4)}(t) + 1.05G_5^{(4)}(t) \\ &= 53.76t^9 - 107.52t^7 + 70.56t^5 + 16.384t^4 \\ &\quad - 9.832t^3 - 4.128t^2 - 0.434t - 0.016. \end{aligned}$$



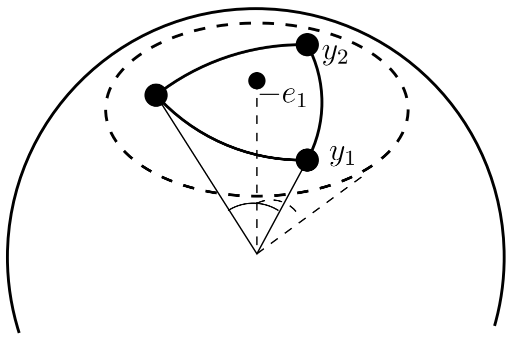
$f(t)$

This was found via discretization and linear programming—such methods had been employed already by Odlyzko & Sloane for the same purpose.

To evaluate h_m, we have to consider arrangements of m points $y_1, \dots, y_m$ in the spherical cap $C_0 := \{y \in S^3 : \langle e_1, y\rangle \le t_0\}$. The points have a minimum distance of $60° = \pi/3 = \arccos(\frac{1}{2})$ given by $\langle y_i, y_j\rangle \le \frac{1}{2}$, and this distance is larger than the radius $\arccos(-t_0)$ of the spherical cap. We know that in an *optimal* arrangement for a given m we cannot move one or several of the points towards the center of the cap while maintaining the "minimum distance" requirement, because of the monotonicity assumption on $f(t)$. From this Musin [21, Lemma 1] derives strong conditions on the combinatorics of optimal configurations. For example, for $m \ge 1$ the center $-e_1$ of the spherical cap is contained in the (spherical) convex hull of the m points y_i; for $m \ge 2$ each point has at least one other point at distance exactly $\pi/3$ (for $m \ge 2$), etc. This already yields that $h_0 = f(1) = 18.774$ and $h_1 = f(1) + f(-1) = 24.48$. Also,

$$h_2 = \max_{\varphi \le \pi/3} f(1) + f(-\cos(\varphi)) + f(-\cos(\tfrac{\pi}{3} - \varphi)),$$

which yields $h_2 \approx 24.8644$. The computations of h_m for $m = 3, 4, 5$ amount to rather well-behaved optimization problems in $m-1$ variables that can be solved numerically, and yield $h_m < h_2$; for the case $m = 6$ Musin shows that $h_6 < h_2$ by a separate argument. □



An h_3-configuration (not to scale)

Clearly there is potential for other applications of Musin's insight—in the whole range of sphere packing and coding theory problems where Delsarte's method was used in the last thirty years, with tremendous success.

Sphere packings

Surprisingly, Musin's breakthrough is not the only remarkable recent piece of progress related to the packing of spheres in high-dimensional space. Namely, by again extending and improving upon Delsarte's method, Henry Cohn (Microsoft Research) in joint papers with Noam Elkies (Harvard University) and with Abhinav Kumar (a mathematics graduate student at Harvard) has obtained new upper bounds on the density of sphere packings in n-space.

Once you know bounds for spherical codes (the kissing number is a special case of this, and these more general bounds can be derived in a similar fashion), you can bound the density of a sphere packing in this dimension. This is the classical way of using Delsarte's method in the context of sphere packings, by Kabatjanskiĭ & Levenšteĭn [15], which up to now gave the best upper bounds for the density of high-dimensional sphere packings.

Cohn and Elkies found a more direct approach to the problem. Instead of using functions defined only on the interval $[-1,1]$, they use functions defined on all of $\mathbb{R}^n$ to get good bounds, as follows.

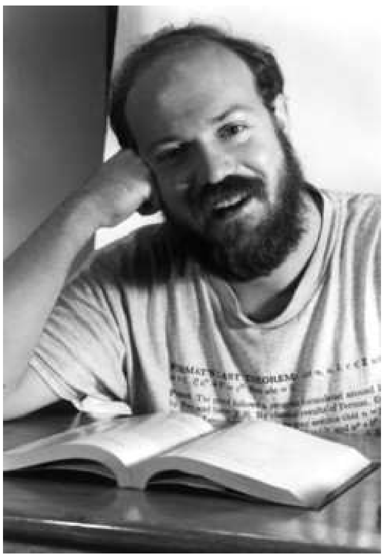
Let $f : \mathbb{R}^n \to \mathbb{R}$ be an L^1 function and let its Fourier transform $\widehat{f}$ be defined as

$$\widehat{f}(t) := \int_{\mathbb{R}^n} f(x) e^{2\pi i \langle x,t\rangle} dx.$$

The function $f(t)$ is called *admissible* if there is a constant $\varepsilon > 0$ such that $|f(x)|$ and $|\widehat{f}(x)|$ are bounded above by a constant times $(1+|x|)^{-n-\varepsilon}$. One crucial property of these functions is the Poisson summation formula

$$\sum_{x\in\Lambda} f(x+v) = \frac{1}{|\Lambda|} \sum_{t\in\Lambda^*} e^{-2\pi i\langle v,t\rangle} \widehat{f}(t),$$

for every vector $v \in \mathbb{R}^n$ and every lattice $\Lambda \subset \mathbb{R}^n$, where $\Lambda^* = \{y \in \mathbb{R}^n : \langle x,y\rangle \in \mathbb{Z} \text{ for all } x \in \Lambda\}$ is its dual lattice.



Henry Cohn
(photo courtesy of Valerie Samn)

Theorem 5 (Cohn & Elkies [5]). *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is an admissible function, not identically zero, which satisfies the following two conditions:*

(1) $f(x) \le 0$ *for* $|x| \ge 1$, *and*
(2) $\widehat{f}(t) \ge 0$ *for all* t.

Then the density of n-dimensional sphere packings is bounded above by

$$\Delta_n \le \frac{\omega_n}{n2^n} \frac{f(0)}{\widehat{f}(0)}.$$

Proof. A *periodic packing* is given by vectors $v_1, \dots, v_N$ and a lattice Λ such that the packing consists of all spheres centered at translates of $v_1, \dots, v_N$ by elements of Λ, and such that $v_i - v_j \in \Lambda$ implies that $i = j$. It is easy to see that periodic packings come arbitrarily close in density to the densest possible sphere packing, so it is sufficient to consider packings of this type. Rescale the packing so that all spheres have radius $1/2$. The density of such a packing is $\frac{\omega_n}{n2^n}\frac{N}{|\Lambda|}$, since $\frac{1}{n}\omega_n$ is the volume of a unit n-ball.

Now we bound the quantity

$$\sum_{j,k=1}^{N} \sum_{x\in\Lambda} f(x+v_j-v_k).$$

From below, the Poisson summation formula yields a lower estimate of $N^2\widehat{f}(0)/|\Lambda|$. To get an upper bound, observe that $x+v_j$ and v_k are two centers of the packing. Thus, $|x+v_j-v_k| < 1$ if and only if $x+v_j = v_k$, i.e. $x = 0$ and $j = k$. We have $f(x+v_j-v_k) \le 0$ whenever $|x+v_j-v_k| \ge 1$, and thus, $Nf(0)$ is an upper bound for the sum. This yields

$$Nf(0) \ge \frac{N^2\widehat{f}(0)}{|\Lambda|},$$

and thus the theorem. □

The Golay code and the Leech lattice

Since the most spectacular applications of the set-up by Cohn and Elkies concern the Leech lattice, we describe it here. For this we start with a construction of the (extended binary) Golay code, a remarkable binary linear code of length 24, dimension 12 and minimal distance 8, that is, a linear

12-dimensional subspace of $(\mathbb{Z}_2)^{24}$, consisting of $2^{12} = 4096$ code words (vectors), all of which except for the zero vector have weight (number of ones) at least 8. There are myriads of ways in the literature to describe this code, the most compact one being just a list of 12 basis vectors. Here, we choose not just any basis but one based on the adjacency matrix of the graph X of the icosahedron to make use of some symmetries later, following lecture notes by Aigner [1].

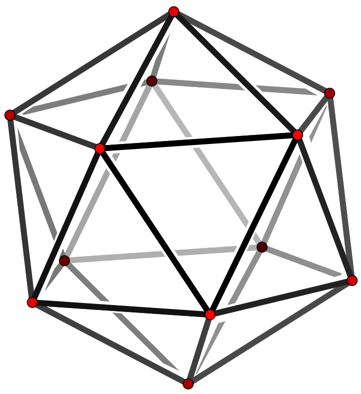
Consider the binary 12×24 matrix

$$G := (I \,|\, B) \in \mathbb{Z}^{12\times 24},$$

where I is the identity matrix of order 12, and $B = J - A$ is the all-one matrix of that order minus the adjacency matrix A of the icosahedron. Thus B is a symmetric $0/1$-matrix with seven ones in each row and each column, corresponding to the seven non-neighbors for each vertex of the icosahedron (counting the vertex itself). We will see in a minute that the code

$$C := \operatorname{rowspan}(G) \subset (\mathbb{Z}_2)^{24}$$

is a $(24, 12, 8)$-code. It has been proved that there is a unique such code, the famous Golay code.



The icosahedron graph
(Graphics: Michael Joswig/polymake [12])

Consider B^2. We have

$$(B^2)_{ij} = 12 - |N(v_i) \cup N(v_j)|$$

for every pair of vertices $v_i, v_j \in V(X)$, where we write $N(v)$ for the set of vertices adjacent to a vertex v. Therefore,

$$(B^2)_{ij} = \begin{cases} 7 & \text{if } \operatorname{dist}(v_i, v_j) = 0, \\ 4 & \text{if } \operatorname{dist}(v_i, v_j) = 1, \\ 4 & \text{if } \operatorname{dist}(v_i, v_j) = 2, \\ 2 & \text{if } \operatorname{dist}(v_i, v_j) = 3. \end{cases}$$

Thus all entries of $I - B^2$ are even, which we write as $I - B^2 \equiv O \pmod 2$, and it follows that

$$GG^T = I + B^2 \equiv O \pmod 2.$$

This implies that $C \subseteq C^\perp$, and thus $C = C^\perp$, since $\dim(C) + \dim(C^\perp) = 24$. From this, together with the fact that all rows of G have Hamming weight 8, we conclude that for each $c \in C$ the Hamming weight is divisible by 4.

All that is left to show is that there are no code words of weight 4. Note that

$$BG = B(I \,|\, B) = (B \,|\, B^2) \equiv (B \,|\, I) \pmod 2.$$

This implies that for each code word $c = (c_L \,|\, c_R)$, $(c_R \,|\, c_L)$ is a code word, too. If c has weight 4, then one of c_L and c_R has weight at most 2. All one has to do to exclude code words of weight 4 is therefore to check sums of up to two rows of G, and we have already seen that a sum of two different rows has weight at least $16 - 8 = 8$.

In order to construct the Leech lattice from the Golay code, consider the lattice

$$\Gamma_{24} = \{x \in \mathbb{Z}^{24} : x \bmod 2 \in C\}.$$

Then $\Gamma_{24} = \Gamma_1 \cup \Gamma_2$, where

$$\Gamma_1 = \left\{x \in \Gamma_{24} : \textstyle\sum x_i \equiv 0 \ (\mathrm{mod} 4)\right\},$$
$$\Gamma_2 = \left\{x \in \Gamma_{24} : \textstyle\sum x_i \equiv 2 \ (\mathrm{mod} 4)\right\}.$$

Finally, let $\Lambda_{24} = 2\Gamma_1 \cup (\mathbf{1} + 2\Gamma_2)$. One can show that this is a lattice with minimum distance $\sqrt{32}$, the Leech lattice.

Optimality of the Leech lattice

Cohn and Elkies have conducted a systematic computer search for suitable admissible functions. In this context, the role of the Gegenbauer polynomials is played by Bessel functions (times some power of $|x|$). These are the functions whose Fourier transform is a delta function on a sphere centered at the origin, so a function with nonnegative Fourier transform is like a nonnegative linear combination of them (but with an integral instead of a sum). The best currently known function was found by Cohn and Kumar [6], a radial function that consists of a polynomial of degree 803 (evaluated at $|x|^2$, so technically it is a polynomial of degree 1606) multiplied by $e^{-\pi|x|^2}$. It yields a density bound that is above the density of the Leech lattice by a factor of less than $1+10^{-29}$. Similarly, they found a function that provides a density bound that is above the density of the E_8-lattice by a factor of less than $1+10^{-14}$. Functions which show that these lattices are in fact densest possible sphere packings are yet to be found.

But what about the easier quest of finding the densest *lattice* packing? In dimension 8 the question was settled, while in dimension 24 it was still open. This case was now answered by Cohn and Kumar [6]. Since it is known that the Leech lattice is a local optimum for the density of lattice packings, it is enough to show that every denser lattice has to be very close to the Leech lattice. This required finding the above mentioned admissible function, and then quite a bit of work involving linear programming and association scheme theory.

The reader may have arrived at this point and say "what a shame they don't explain this in more detail." Well, we feel the same—and refer you to the original papers/preprints, which are fascinating mathematics, and a pleasure to read!

Acknowledgements

We are grateful to Eiichi Bannai, Henry Cohn, Noam Elkies, and Oleg Musin for discussions, helpful remarks and many insightful comments related to this paper.

References

[1] M. AIGNER, *Gitter, Codes und Graphen.* Lecture Notes, FU Berlin, 1994/95.

[2] G. E. ANDREWS, R. ASKEY, AND R. ROY, *Special Functions*, vol. 71 of Encyclopedia of Math. and its Appl., Cambridge University Press, Cambridge, 1999.

[3] V. V. ARESTOV AND A. G. BABENKO, *Estimates for the maximal value of the angular code distance for 24 and 25 points on the unit sphere in $\mathbb{R}^4$*, Math. Notes, 68 (2000), pp. 419–435.

[4] M. R. BEST, *Optimal codes*, in Packing and Covering in Combinatorics, A. Schrijver, ed., vol. 106 of Mathematical Centre Tracts, CWI, Amsterdam, 1979, pp. 119–140.

[5] H. COHN AND N. ELKIES, *New upper bounds on sphere packings. I*, Annals of Math., 157 (2003), pp. 689–714.

[6] H. COHN AND A. KUMAR, *Optimality and uniqueness of the Leech lattice among lattices.* Preprint, March 2004, 39 pages; `arXiv:math.MG/0403263`.

[7] J. H. CONWAY AND N. J. A. SLOANE, *Sphere Packings, Lattices and Groups*, vol. 290 of Grundlehren der mathematischen Wissenschaften, Springer-Verlag, New York, third ed., 1993.

[8] P. DELSARTE, *Bounds for unrestricted codes, by linear programming*, Philips Res. Rep., 27 (1972), pp. 272–289.

[9] ——, *An algebraic approach to the association schemes of coding theory*, Philips Res. Rep. Suppl., (1973), pp. vi+97.

[10] P. DELSARTE, J. M. GOETHALS, AND J. J. SEIDEL, *Spherical codes and designs*, Geometriae Dedicata, 6 (1977), pp. 363–388.

[11] N. D. ELKIES, *Lattices, linear codes, and invariants. I/II*, Notices Amer. Math. Soc., 47 (2000), pp. 1238–1245 and 1382–1391.

[12] E. GAWRILOW AND M. JOSWIG, `Polymake`: *A software package for analyzing convex polytopes.* `http://www.math.tu-berlin.de/diskregeom/polymake/`.

[13] T. C. HALES, *Cannonballs and honeycombs*, Notices Amer. Math. Soc., 47 (2000), pp. 440–449.

[14] ——, *A computer verification of the Kepler conjecture*, in Proceedings of the International Congress of Mathematicians, Vol. III (Beijing, 2002), Beijing, 2002, Higher Ed. Press, pp. 795–804. `math.MG/0305012`.

[15] G. A. KABATJANSKIĬ AND V. I. LEVENŠTEĬN, *Bounds for packings on the sphere and in space*, Problemy Peredači Informacii, 14 (1978), pp. 3–25.

[16] J. C. LAGARIAS, J. A. REEDS, M. H. WRIGHT, AND P. E. WRIGHT, *Convergence properties of the Nelder-Mead simplex method in low dimensions*, SIAM J. Optimization, 9 (1999), pp. 112–147.

[17] J. LEECH, *The problem of the thirteen spheres*, The Mathematical Gazette, 40 (1956), pp. 22–23.

[18] V. I. LEVENŠTEĬN, *On bounds for packings in n-dimensional euclidean space*, Soviet Math. Doklady, 20 (1979), pp. 417–421.

[19] K. I. M. MCKINNON, *Convergence of the Nelder-Mead simplex algorithm to a non-stationary point*, SIAM J. Optimization, 9 (1999), pp. 148–158.

[20] C. MÜLLER, *Spherical Harmonics*, vol. 17 of Lecture Notes in Math., Springer-Verlag, Berlin Heidelberg, 1966.

[21] O. R. MUSIN, *The kissing number in four dimensions*. Preprint, September 2003, 22 pages; `math.MG/0309430`.

[22] ——, *The problem of the twenty-five spheres*, Russian Math. Surveys, 58 (2003), pp. 794–795.

[23] A. M. ODLYZKO AND N. J. A. SLOANE, *New bounds on the number of unit spheres that can touch a unit sphere in n dimensions*, J. Combinatorial Theory, Ser. A, 26 (1979), pp. 210–214.

[24] W. H. PRESS, S. A. TEUKOLSKY, W. T. VETTERLING, AND B. P. FLANNERY, *Numerical recipes in C. The Art of Scientific Computing*, Cambridge University Press, Cambridge, second ed., 1992.

[25] I. J. SCHOENBERG, *Positive definite functions on spheres*, Duke Math. J., 9 (1942), pp. 96–107.

[26] K. SCHÜTTE AND B. L. VAN DER WAERDEN, *Das Problem der dreizehn Kugeln*, Math. Annalen, 53 (1953), pp. 325–334.

[27] G. G. SZPIRO, *Kepler's Conjecture. How some of the greatest minds in history helped solve one of the oldest math problems in the world*, Wiley, 2003.

[28] T. M. THOMPSON, *From Error-correcting Codes through Sphere Packings to Simple Groups*, vol. 21 of Carus Mathematical Monographs, Mathematical Association of America, Washington, DC, 1983.