How Many Soccer Balls Can You Fit in a High-dimensional Warehouse?

Bernardo Subercaseaux (bersub@cmu.edu)

Consider a warehouse of 1000 square feet area, 9 feet tall, and an infinite supply of soccer balls with a 5-inch radius. How many balls can we fit in the warehouse?

We may start by computing volumes; the warehouse has 9000 ft^3, and each ball is $\frac{4}{3}\pi(\frac{5}{12})^3 \approx 0.303$ ft^3. While this gives us an upper bound of $\lceil \frac{9000}{0.303} \rceil = 29702$ many balls, it certainly does not mean we can actually fit as many. No matter how the balls are packed, a certain fraction of the volume will be wasted, with empty space filling the gaps between balls. The fraction of the warehouse volume that is actually being used by the soccer balls will thus never be 100%, and it should be intuitively evident that filling up more than 50% of the space is possible. The natural question is therefore what the optimal density is.

Figure 1: Illustration of an optimal packing of spheres in 3 dimensions.

It turns out that in 1611, German polymath Johannes Kepler conjectured the answer to be around 74.05%, or more precisely, $\frac{\pi}{3\sqrt{2}}$. The way to achieve this conjectured fraction of the volume does not require any fancy techniques; see Figure 1, it's the standard way you see a pile of oranges in the grocery store! However, proving the optimality of this *3-dimensional sphere-packing density*, $\theta(3) = \frac{\pi}{3\sqrt{2}}$, is a different matter. It took nearly 4 centuries until Thomas Hales finally put the problem to rest in 2005 [7] through a tremendously complex proof that took several years to be reviewed, and was only finally formalized in 2017 [8].

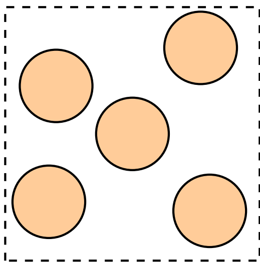
Arguably, what makes sphere-packing fascinating is how on the one hand it is an extremely natural problem with immediate industrial applications, and yet on the other hand, it is a profoundly complex mathematical problem despite its deceivingly simple appeareance. Even its significantly easier cousin, the problem of computing the optimal density of a packing of circles in the plane, escaped Lagrange (who only managed to prove a restricted form of optimality), and was only proved rigorously in 1942 to be $\theta(2) := \frac{\pi}{2\sqrt{3}} \approx 0.9069$ [6].

Lovers of generalization, mathematicians couldn't help but wonder about $\theta(d)$, the optimal density of sphere packings in $\mathbb{R}^d$. To formalize what this means, a d-dimensional sphere packing $P_d \subset \mathbb{R}^d$ is a set of d-dimensional points where every pair of points in P_d are at distance at least 2 in the natural ℓ_2 *Euclidean* metric.[1] We think of P_d as a packing of d-dimensional spheres of radius 1, with the distance $\geqslant 2$ condition enforcing that the spheres don't intersect. To properly define $\theta(d)$ we will need a couple of pieces of notation. We let $\vec{O}_d$ be the origin of $\mathbb{R}^d$, and use $B_d(c, r) := \{x \in \mathbb{R}^d \mid \|x - c\|_d \leqslant r\}$ for a ball of radius r centered around point c. We can then define the density of the sphere-packing P_d as

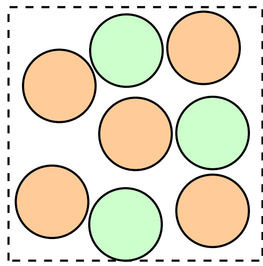
$$\delta_d(P_d) = \limsup_{r \to \infty} \frac{\mathrm{vol}\left(B_d(\vec{O}_d, r) \cap \left(\bigcup_{p \in P_d} B_d(p, 1)\right)\right)}{\mathrm{vol}(B_d(\vec{O}_d, r))},$$

[1] One can also study the problem with respect to the distance ℓ_p for general p, but let's stick to one generalization at a time.

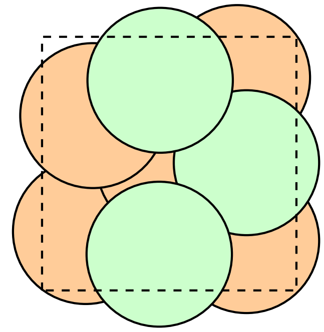
where vol(S) corresponds to the Lebesgue measure of S in $\mathbb{R}^d$. Consequently, $\theta(d) = \sup\{\delta_d(P_d) \mid P_d \subset \mathbb{R}^d$ is a sphere packing$\}$. Let us now prove a folklore lower-bound on $\theta(d)$ that will illustrate a standard argument in sphere packing. Let us say that a sphere packing P_d is *"saturated"* if no further spheres can be added to P_d. In other words, there's no sphere packing P'_d with $P_d \subsetneq P'_d$. Examples are illustrated in Figures 2a and 2b.



(a) Unsaturated packing



(b) Saturated packing



(c) Amplified saturated packing

Figure 2: Illustration of the notion of saturation, key for Proposition 1. In this case $d = 2$, so the radii are doubled.

Proposition 1 (Folklore)**.** *Any saturated sphere packing $P_d \subset \mathbb{R}^d$ has density $\delta_d(P_d) \geqslant 2^{-d}$, and thus, $\theta(d) \geqslant 2^{-d}$.*

Proof. Consider an arbitrary saturated packing P_d. It is well-known that vol$(B_d(\cdot, r)) \propto r^d$, and therefore if we double the radius of each sphere in P_d, the volume of their union increases at most by 2^d. We now claim that

$$\bigcup_{p \in P_d} B_d(p, 2) = \mathbb{R}^d,$$

for if it was not true, there would exist a point $x \in \mathbb{R}^d \setminus \left(\bigcup_{p \in P_d} B_d(p, 2)\right)$, which would imply that $B_d(x, 1)$ did not intersect any of the $B_d(p, 1)_{p \in P_d}$ balls, and thus x could have been added to P_d contradicting its saturation. See Figure 2c. We have thus shown that if we amplify the density of P_d by a factor of at most 2^d, its density becomes 1, hence $\delta_d(P_d) \geqslant 2^{-d}$. □

What was known about $\theta(d)$. We have already mentioned that $\theta(2) = \frac{\pi}{2\sqrt{3}}$ and $\theta(3) = \frac{\pi}{3\sqrt{2}}$; we may as well include $\theta(1) = 1$ as a trivial case. Perhaps surprisingly, there are two more known values of $\theta(d)$:

$$\theta(8) = \frac{\pi^4}{384} \approx 0.253 \text{ (Viazovska [17])}, \quad \text{and} \quad \theta(24) = \frac{\pi^{12}}{12!} \approx 0.0019 \text{ (Cohn et al. [3])}.$$

Once again, the structure of the packings is simple: they correspond to the *exceptional lattice* E_8 and the *Leech lattice* Λ_{24}, respectively. The proofs, nonetheless, are considered "groundbreaking" and represent a significant advance in low-dimensional sphere packing [10].

In asymptotically high dimensions, the "baseline" lower bound is Proposition 1: $\theta(d) \geqslant 2^{-d}$. This was first improved by Minkowski, who proved $\theta(d) \geqslant (2 + o(1))2^{-d}$ in 1905 [12], and then furthered by Rogers to $\theta(d) \geqslant (1 + o(1))cd2^{-d}$ in 1947, with $c = 2/e$ [13]. The value of c has been improved several times since then [4, 1, 15, 16]. The best upper bound, on the other hand, is $\theta(d) \leqslant \kappa 2^{-(0.599\cdots + o_d(1))d}$, where the constant κ has recently been improved by Sardari and Zargar [14].

The breakthrough. Since 1947, no asymptotically growing improvements on the lower bound of $\theta(d)$ had been made for all sufficiently high dimensions. The breakthrough came in 2023, when Campos, Jenssen, Michelen, and Sahasrabudhe proved the following result:

Theorem 1 (Campos et al. [2])**.**

$$\theta(d) \geqslant (1 - o(1))\frac{d \log d}{2^{d+1}}, \quad \text{as } d \to \infty.$$

Albeit still exponentially far from the upper bound of Sardari and Zargar, Theorem 1 represents a major asymptotic improvement over the previous lower bounds, and has been featured by several scientific outlets including Quanta Magazine [9].

The rest of this article is devoted to sketching the proof of Theorem 1.

1 A proof sketch of Theorem 1, the new lower bound

Let $\Omega = B_d(\vec{0}_d, R)$ be a large ball (i.e., R is some large number) centered at the origin of $\mathbb{R}^d$ and $X \subseteq \Omega$ an appropriate subset. Consider a graph $G(X)$ whose vertex set is X and has edges between points $x, y \in X$ when $\|x - y\|_d \leqslant 2r_d$, with r_d the radius of a ball of volume 1 in $\mathbb{R}^d$. Note now that an independent set in G corresponds to a sphere packing in $\mathbb{R}^d$, and thus finding a large independent set in G is roughly equivalent to finding a sphere packing that is dense in Ω. Why take an appropriate subset X and not just Ω? The reason is that, in order to argue that a large independent set exists in G, we want G to be *sparse*. Let us show an illustrative example, using $\Delta(G)$ as notation for maximum degree of any vertex in G.

Proposition 2. *Any graph G on n vertices has an independent set of size at least $\left\lceil \frac{n}{\Delta(G)+1} \right\rceil$.*

Proof. We can iteratively construct an independent set I in G by repeeating the following steps: (i) pick any vertex $v \in V(G)$, (ii) add v to I, and (iii) remove v and its neighbors from $V(G)$. To see correctness, note that each step adds 1 vertex to I and removes at most $\Delta(G) + 1$ vertices from $V(G)$. Note that this implicitly uses the fact that removing vertices from $V(G)$ does not increase the maximum degree, thus preserving the invariant throughout the process. □

The result of Campos et al. [2] is based on a clever random process that constructs a reasonably large X with $G(X)$ being sparse in a precise sense. Then, a result syntactically similar to the trivial Proposition 2, but much more powerful, is used to argue that a large independent set exists in $G(X)$, which in turn implies a dense sphere packing in Ω.

In order to state the appropriate sparseness condition we need an additional piece of notation. For two distinct vertices $u, v \in V(G)$ let $d_2(u, v)$ be the number of common neighbors of u and v in G, and then define $\Delta_2(G) = \max_{u,v \in V(G)} d_2(u, v)$. Then, the key lemma relating sparseness to the size of independent sets is the following:

Lemma 1. *Let G be a graph on n vertices with maximum degree Δ and $\Delta_2(G) \leqslant 2^{-7}\Delta(\log \Delta)^{-7}$. Then, G has an independent set of size at least*

$$(1 - o(1))n\frac{\log \Delta}{\Delta},$$

where the $o(1)$ term goes to 0 as $\Delta \to \infty$.

Note immediately that the extra sparseness condition $\Delta_2(G) \leqslant 2^{-7}\Delta(\log \Delta)^{-7}$, on the common degree of vertices, is necessary since Proposition 2 is tight when only the maximum degree is considered; a disjoint union of k cliques of size $\Delta+1$ has maximum degree Δ, $|V(G)| = k(\Delta+1)$, and no independent set of size larger than $k = \lceil |V(G)|/(\Delta+1) \rceil$.

In terms of the existence of an appropriate set X, for which Lemma 1 can be applied to $G(X)$, the key lemma is the following.

Lemma 2. *Let $\Omega \subseteq \mathbb{R}^d$ be bounded and measurable. For all $d \geqslant 1000$, there exists $X \subset \Omega$ such that*

$$|X| \geqslant (1 - 1/d)\frac{\Delta}{2^d}\operatorname{vol}(\Omega), \quad \textit{where} \quad \Delta = \left(\frac{\sqrt{d}}{4\log d}\right)^d,$$

and if $G = G(X, r_d)$ then we have

$$\Delta(G) \leqslant \Delta(1 + \Delta^{-1/3}) \quad \textit{and} \quad \Delta_2(G) \leqslant \Delta \cdot e^{-(\log d)^2/8}.$$

Note that the first part of the lemma is stating that X is a large subset of Ω, in a way that uses its volume, so that one can carry forward the density argument. The second part of the lemma is the sparseness condition that allows us to apply Lemma 1 to $G(X)$. Combining Lemma 2 and Lemma 1 to obtain Theorem 1 is a straightforward calculation, and can be found in the original paper [2]. Instead, let us try to provide some intuition for the existence of the set X (Lemma 2), and then for how to leverage the more refined sparseness condition to obtain a large independent set (Lemma 1).

1.1 Sketch of the proof of Lemma 2

Consider a Poisson point process over $\mathbf{X} \subset \Omega$ with intensity $\lambda = 2^{-d}\Delta = 2^{-d}\left(\frac{\sqrt{d}}{4\log d}\right)^d$. Then, we have

$$\mathbb{E}[|\mathbf{X}|] = \lambda \operatorname{vol}(\Omega) = \Delta/2^d \operatorname{vol}(\Omega).$$

Thus, $\mathbf{X}$ can be expected to exceed the size requirement of Lemma 2 by $\frac{\Delta}{d2^d}\operatorname{vol}(\Omega)$. We will use this extra slack to ensure the sparseness condition holds by simply deleting points from $\mathbf{X}$ that violate either of the sparseness conditions. Let $\mathrm{Bad}_1 \subseteq \mathbf{X}$ correspond to the points that would have degree at least $\Delta + \Delta^{2/3}$ in $G(\mathbf{X})$, and $\mathrm{Bad}_2 \subseteq \mathbf{X}$ the points that would have common degree (with respect to any other point) at least $\Delta e^{-(\log d)^2/8}$ in $G(\mathbf{X})$. Notice now that it suffices to show that

$$\mathbb{E}[|\mathrm{Bad}_1|] \leqslant \frac{|\mathbf{X}|}{2d}, \quad \text{and} \quad \mathbb{E}[|\mathrm{Bad}_2|] \leqslant \frac{|\mathbf{X}|}{2d}, \tag{1}$$

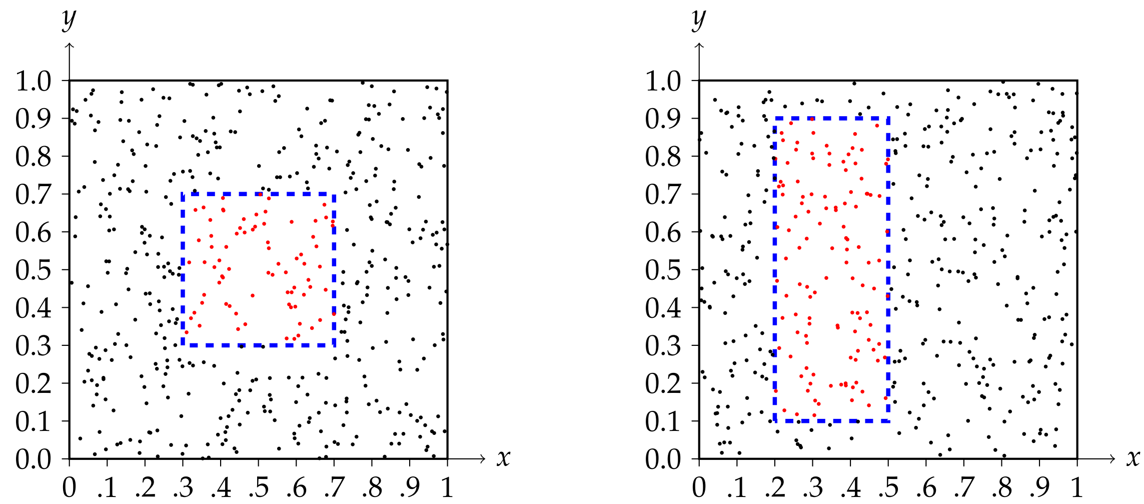
since that way

$$\mathbb{E}[|\mathbf{X} \setminus \mathrm{Bad}_1 \setminus \mathrm{Bad}_2|] \geqslant \mathbb{E}[|\mathbf{X}|] - \mathbb{E}[|\mathrm{Bad}_1|] - \mathbb{E}[|\mathrm{Bad}_2|] \geqslant (1 - 1/d)\frac{\Delta}{2^d}\operatorname{vol}(\Omega),$$

and thus the existence of an $X := \mathbf{X} \setminus \mathrm{Bad}_1 \setminus \mathrm{Bad}_2$ that satisfies all the requirements of Lemma 2 is guaranteed, as an application of the probabilistic method. Proving Equation (1) is rather technical, especially the second part, and thus we will not cover the details.

However, it is worth pointing out that a key tool is Mecke's equation [11], which justifies the use of a Poisson process. Indeed, a Poisson point process is characterized by the fact that the number of points in any region follows a Poisson distribution with mean proportional to the volume of the region (and the proportionality constant is the intensity of the process). See Figure 3 for a simple illustration. Concretely, for a bounded and measurable set $\Lambda \subset \mathbb{R}^d$ and events $(A_x)_{x \in \Lambda}$, Mecke's equation states that

$$\mathbb{E}\left[\left|\{x \in \mathbf{X} \cap \Lambda : A_x \text{ holds for } \mathbf{X}\}\right|\right] = \lambda \int_\Lambda \mathbb{P}(A_x \text{ holds for } \mathbf{X} \cup \{x\})\, dx.$$



(a) (b)

Figure 3: Illustration of Poisson point processes in $[0,1]^2$ with intensity $\lambda = 500$. Figure 3a depicts 76 out of 481 points inside the blue rectangle $[0.3, 0.7]^2$, while Figure 3b shows 119 out of 461 points inside the blue rectangle $[0.2, 0.5] \times [0.1, 0.9]$. Indeed, the number of points inside the rectangles is approximately proportional to their area: $\frac{76}{481} \approx 0.158$, $\frac{0.4^2}{1} = 0.16$ and $\frac{119}{461} \approx 0.258$, $\frac{0.3\times 0.8}{1} = 0.24$.

For instance, applying Mecke's equation in order to obtain the first part of Equation (1) is relatively straightforward: for each $s > 0$

$$\mathbb{E}\left[\left|\left\{x \in \mathbf{X} : \left|\mathbf{X} \cap B_d(x, 2r_d)\right| \geqslant s\right\}\right|\right] = \lambda \int_{\Omega} \mathbb{P}\left[\left|\mathbf{X} \cap B_x(2r_d)\right| \geqslant s - 1\right] \mathrm{d}x,$$

which can be combined with the following concentration result

$$\forall x \in \Omega, \quad \Pr\left[\left|\mathbf{X} \cap B_d(x, 2r_d)\right| \geqslant \Delta + \Delta^{2/3} - 1\right] \leqslant \exp(-\Delta^{1/3}/4) \leqslant 1/2d$$

to deduce $\mathbb{E}[|\mathrm{Bad}_1|] \leqslant {}^{|\mathbf{X}|}/_{2d}$. The second part of Equation (1) is more involved, and requires geometric arguments on top of Mecke's equation to bound the expected common degree.

1.2 Sketch of the proof of Lemma 1

Recall that the trivial proof for Proposition 2 worked by iteratively deleting vertices of the graph, together with their neighborhoods, and building the independent set I by adding the deleted vertices. The sparseness condition (in that case simply over the maximum degree) controls that each step deletes a small fraction of the graph, and hence guarantees the process to last sufficiently long. The proof of Lemma 1 can be seen a sophisticated version of this, where sets of vertices are randomly sampled to be deleted together with their neighborhoods. This general strategy is often referred to as *"Rödl's Nibble Method"* (see e.g., [5]).

Let $A \subseteq V(G)$ be a random subset obtained by taking each vertex independently with probability ${}^{\gamma}/_{\Delta}$, where $\gamma \leqslant {}^1/_2$ is a constant. Each "nibble" step corresponds then to deleting A and its neighborhood $N(A)$ from G. Now, roughly speaking, the idea is to show that the degrees and codegrees of the surviving vertices don't decrease too quickly. This is summarized by the following lemma:

Lemma 3 (Informal)**.** *Let $v \in (V(G) \setminus A \setminus N(A))$ be any vertex surviving a nibble step in a graph G that holds some technical conditions including having all degrees be close to the maximum degree Δ. Then, let $d(v)$ denote the degree of v before the nibble step, and $d'(v)$ its degree after the nibble step. Similarly, for a pair of vertices u, v that survive the nibble step, their common degree before the nibble is $d_2(u, v)$, whereas their common degree after the nibble is denoted $d_2'(u, v)$. We then have, for every $\alpha \in [2\gamma^2, \gamma]$, that*

$$\Pr\left[d'(v) \geqslant (1 - \gamma + \alpha)d(v)\right] \leqslant \exp\left(-\frac{\alpha^2}{4\gamma^3}\right),$$

and

$$\Pr\left[d_2'(u, v) \geqslant (1 - \gamma + \alpha)\Delta_2(u, v)\right] \leqslant \exp\left(-\frac{\alpha^2}{4\gamma^3}\right).$$

Proving Lemma 3 requires several concentration results, especially around an appropriate martingale where the bounded common degree condition is used to argue that the neighborhood of any vertex v decreases slowly on expectation. Namely, if we think of the set $T(v)$ of vertices at distance 2 from a fixed vertex v, and $N(v)$ the set of vertices at distance 1 from v, then the nibbling steps removing from $T(v)$ will decrease the size of $N(v)$, but not too quickly since the common degree sparseness ensures that no vertex $u \in T(v)$ has too many neighbors in $N(v)$, as otherwise $d_2(u, v)$ would be large.

2 Concluding remarks

The result of Campos et al. [2] is a significant advance in the field of sphere packing, and it shows that while the sphere packings that have been proved optimal (i.e., $\theta(d)$ for $d \in \{1, 2, 3, 8, 24\}$) are highly structured lattices, it is possible to obtain improved lower bounds on $\theta(d)$ by analyzing carefully constructed random structures. The proof techniques are a very nice mixture of different classic ideas in probabilistic combinatorics, and it is likely that Lemma 1 will be useful independently.

While we haven't answered the question that titles this article, we can now confidently say that in a 1000-dimensional warehouse, 10 feet long in each of its d dimensions, we can fit at least

$$\frac{10^{1000}}{\left(\frac{2\pi e}{1000}\left(\frac{5}{12}\right)^2\right)^{500}} \cdot \frac{999}{1000} \cdot \frac{1000 \log 1000}{2^{1001}} \approx 3.03 \times 10^{1966}$$

many 1000-dimensional soccer balls of radius 5 inches, where we have used the upper bound $\mathrm{vol}(B_d(\vec{\cdot}, r)) \leqslant \left(\frac{2\pi e r^2}{d}\right)^{d/2}$. The obtained amount, 3.03×10^{1966}, is roughly 4000 times larger than the naive bound of Proposition 1 would give.

References

[1] K. Ball. A lower bound for the optimal density of lattice packings. *International Mathematics Research Notices*, 1992(10):217–221, Jan. 1992.

[2] M. Campos, M. Jenssen, M. Michelen, and J. Sahasrabudhe. A new lower bound for sphere packing, 2023.

[3] H. Cohn, A. Kumar, S. Miller, D. Radchenko, and M. Viazovska. The sphere packing problem in dimension 24. *Annals of Mathematics*, 185(3), May 2017.

[4] H. Davenport and C. A. Rogers. Hlawka's theorem in the geometry of numbers. *Duke Mathematical Journal*, 14(2), June 1947.

[5] D. Dubhashi, D. A. Grable, and A. Panconesi. Near-optimal, distributed edge colouring via the nibble method. *Theoretical Computer Science*, 203(2):225–251, Aug. 1998.

[6] L. Fejes. Über die dichteste kugellagerung. *Mathematische Zeitschrift*, 48(1):676–684, Dec. 1942.

[7] T. Hales. A proof of the Kepler conjecture. *Annals of Mathematics*, 162(3):1065–1185, Nov. 2005.

[8] Thomas. Hales, Mark. Adams, Gertrud. Bauer, T. D. Dang, John. Harrison, L. T. Hoang, Cezary. Kaliszyk, Victor. Magron, Sean. McLaughlin, T. T. Nguyen, Q. T. Nguyen, Tobias. Nipkow, Steven. Obua, Joseph. Pleso, Jason. Rute, Alexey. Solovyev, T. H. A. Ta, N. T. Tran, T. D. Trieu, Josef. Urban, Ky. Vu, and Roland. Zumkeller. A formal proof of the Kepler conjecture. *Forum of Mathematics, Pi*, 5:e2, 2017.

[9] K. Houston-Edwards. To Pack Spheres Tightly, Mathematicians Throw Them at Random — Quanta Magazine — quantamagazine.org. https://www.quantamagazine.org/to-pack-spheres-tightly-mathematicians-throw-them-at-random-20240430/, 2024. [Accessed 19-11-2024].

[10] E. Klarreich. Sphere Packing Solved in Higher Dimensions — Quanta Magazine — quantamagazine.org. https://www.quantamagazine.org/sphere-packing-solved-in-higher-dimensions-20160330/, 2016. [Accessed 19-11-2024].

[11] G. Last and M. Penrose. *Lectures on the Poisson Process*. Institute of Mathematical Statistics Textbooks. Cambridge University Press, Cambridge, 2017.

[12] H. Minkowski. Diskontinuitätsbereich für arithmetische Äquivalenz. *Journal für die reine und angewandte Mathematik*, 129:220–274, 1905.

[13] C. A. Rogers. Existence Theorems in the Geometry of Numbers. *Annals of Mathematics*, 48(4):994–1002, 1947.

[14] N. T. Sardari and M. Zargar. New upper bounds for spherical codes and packings. *Mathematische Annalen*, 389(4):3653–3703, Aug. 2024.

[15] S. Vance. Improved sphere packing lower bounds from Hurwitz lattices. *Advances in Mathematics*, 227(5):2144–2156, Aug. 2011.

[16] A. Venkatesh. A Note on Sphere Packings in High Dimension. *International Mathematics Research Notices*, 2013(7):1628–1642, Jan. 2013.

[17] M. Viazovska. The sphere packing problem in dimension 8. *Annals of Mathematics*, 185(3), May 2017.