Non-randomness of S-unit lattices

Daniel J. Bernstein[1,2] and Tanja Lange[3]

[1] Department of Computer Science, University of Illinois at Chicago, USA
[2] Horst Görtz Institute for IT Security, Ruhr University Bochum, Germany
djb@cr.yp.to
[3] Eindhoven University of Technology, Eindhoven, The Netherlands
tanja@hyperelliptic.org

Abstract. Spherical models of lattices are standard tools in the study of lattice-based cryptography, except for variations in terminology and minor details. Spherical models are used to predict the lengths of short vectors in lattices and the effectiveness of reduction modulo those short vectors. These predictions are consistent with an asymptotic theorem by Gauss, theorems on short vectors in almost all lattices from the invariant distribution, and a variety of experiments in the literature.

S-unit attacks are a rapidly developing line of attacks against structured lattice problems. These include the quantum polynomial-time attacks that broke the cyclotomic case of Gentry's original STOC 2009 FHE system under minor assumptions, and newer attacks that have broken through various barriers previously claimed for this line of work.

S-unit attacks take advantage of auxiliary lattices, standard number-theoretic lattices called S-unit lattices. Spherical models have recently been applied to these auxiliary lattices to deduce core limits on the power of S-unit attacks.

This paper shows that these models underestimate the power of S-unit attacks: S-unit lattices, like the lattice $\mathbb{Z}^d$, have much shorter vectors and reduce much more effectively than predicted by these models. The attacker can freely choose S to make the gap as large as desired, breaking through the core limits previously asserted for S-unit attacks.

Keywords: post-quantum cryptography, lattice-based cryptography, Ideal-SVP, S-unit attacks, Gaussian heuristic, algorithm analysis

---

Author list in alphabetical order; see https://www.ams.org/profession/leaders/culture/CultureStatement04.pdf. This work was funded in part by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2092 CASA—390781972 "Cyber Security in the Age of Large-Scale Adversaries", by the U.S. National Science Foundation under grant 1913167, by the Taiwan's Executive Yuan Data Safety and Talent Cultivation Project (AS-KPQ-109-DSTCP), and by the Netherlands Organisation for Scientific Research (NWO) under grants 628.001.028 (FASOR), 651.002.004 (CHIST-ERA USEIT), and 613.009.144 (Quantum Cryptanalysis of Post-Quantum Cryptography). "Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation" (or other funding agencies). Permanent ID of this document: a135c315e8271d05357fb3f82964060dc905ad05. Date: 2021.10.23.

1 Introduction

Power-of-2-cyclotomic Ideal-SVP is the following problem. There is a parameter, an integer $n \in \{1, 2, 4, 8, 16, \ldots\}$. The input is a nonzero ideal I of the ring $R = \mathbb{Z}[x]/(x^n + 1)$, given as $(v_1, v_2, \ldots, v_n) \in R^n$ with $I = \mathbb{Z}v_1 + \mathbb{Z}v_2 + \cdots + \mathbb{Z}v_n$. The problem is to find a short nonzero element of I. Readers not familiar with the foundational importance of this problem in the security analysis of lattice-based cryptography should see Appendix A.

Fully defining this problem requires giving a quantitative definition of "short". It is important to realize that the definitions of "short" relevant to cryptography are *not* asking specifically for a *shortest* nonzero vector in I. For each Ideal-SVP attack algorithm, one asks how short the output is—and, obviously, how long the algorithm takes. Sometimes Ideal-SVP is called "Approx-Ideal-SVP" to emphasize that "short" is only an approximation to "shortest".

The standard framework for solving this problem considers only the additive structure of I. One extends the list of known lattice vectors by taking linear combinations of those vectors (e.g., $v_1 - v_2$), using various strategies designed to move as efficiently as possible towards shorter nonzero vectors. Well-known examples of algorithms within this framework include enumeration, sieving, LLL, and BKZ. Parameters in lattice-based cryptosystems are typically chosen so that estimates of the cost of the latest variants of BKZ, using the latest variants of enumeration or sieving as subroutines, are slightly above the target security level. See, e.g., [1, Section 4.2] and [3, Section 5].

1.1. *S*-unit attacks. "Unit attacks", together with their generalization to "S-unit attacks", use multiplicative techniques to shorten the vectors found by additive attacks. See Section 5 for a review of these techniques. The history is complicated; see Appendix D for full credits.

This line of work efficiently solves many cases of cyclotomic Ideal-SVP that are not believed to be efficiently solvable by additive attacks. An early success of unit attacks was quantum polynomial-time key recovery (assuming "$h^+ = 1$"; see Appendix C) for the cyclotomic case of well-known cryptosystems introduced by Gentry [48], Smart–Vercauteren [86], Gentry–Halevi [49], and Garg–Gentry–Halevi [45], although for [45] various other security problems are known. The subsequent literature on S-unit attacks is a fascinating story of

- dividing lines between the broken and unbroken cases of Ideal-SVP being claimed to be important (e.g., "barriers"—see Appendix A); and
- these dividing lines being crossed by more advanced S-unit attacks.

The obvious technical question is whether there is a real limit to the power of S-unit attacks—a principled, well-studied barrier making clear that these attacks will not break worst-case Ideal-SVP.

1.2. Spherical models of lattices. Gauss [47] studied the number of points in $\mathbb{Z}^2$ inside the circle of radius r. Gauss showed a $\Theta(r)$ bound on the gap between this number and the area πr^2 of the circle, and concluded that the ratio converges to 1 as the circle radius increases; he noted how obvious this

conclusion was, while at the same time noting the value of proof.[1] Gauss's proof readily generalizes to prove the well-known fact, repeated as Theorem 3.2 below, that $\lim_{r\to\infty} \#(L \cap rB)/\operatorname{Vol} rB = 1/\det L$ for any d-dimensional lattice L in $\mathbb{R}^d$, where rB is the ball of radius r around 0 in $\mathbb{R}^d$.

This theorem inspired an analysis methodology used systematically within the literature on lattice-based cryptography to predict the lengths of short vectors in lattices and the effectiveness of reduction modulo those short vectors. The most common name for this methodology is the "Gaussian heuristic", but this paper uses the terminology "spherical models" for reasons explained in Section 1.6.

Write $\lambda_1(L)$ for the minimum length of nonzero vectors in a lattice L. Spherical models make a prediction r for $\lambda_1(L)$ where $\operatorname{Vol} rB$ is on the same scale as $\det L$. As motivation, notice that if r is slightly larger than this, and the lattice dimension d is large, then $\operatorname{Vol} rB$ becomes much larger than $\det L$ (since $\operatorname{Vol} rB$ increases as the dth power of r), so if $\#(L \cap rB)/\operatorname{Vol} rB$ is close to $1/\det L$ then $\#(L \cap rB)$ also has to be large. Similarly, if r is slightly smaller then the same calculation predicts that $\#(L \cap rB)$ will be close to 0; one can quibble that in fact $\#(L \cap rB) \geq 1$ since $0 \in L \cap rB$, but a modified version of Theorem 3.2 using $\#(L \cap rB) - 1$ eliminates this quibble.

This motivation is not a proof. Perhaps $\#(L \cap rB)/\operatorname{Vol} rB$ is not so close to $1/\det L$, even if it converges to $1/\det L$ as $r \to \infty$. However, the mathematical theory of random lattices—specifically, d-dimensional fixed-determinant lattices chosen from the "invariant distribution"—includes a theorem stating that, as $d \to \infty$, with probability $1 - o(1)$, such lattices L have $\lambda_1(L)$ within a factor $1 + o(1)$ of what spherical models predict; see Section 2. This still does not prove that it is safe to apply the prediction to *every* lattice that appears, but one can see the accuracy of spherical-model predictions for many lattices directly demonstrated by experiments from Chen–Nguyen [25, Figure 1].

Spherical models predict more than $\lambda_1(L)$. For example, Micciancio–Walter [70, Section 2.4] described the "Gaussian Heuristic" as saying "that for a given set S and a lattice Λ, we have $|S \cap \Lambda| \approx \operatorname{vol}(S)/\det(\Lambda)$" assuming S is "nice"; and Ducas–Laarhoven–van Woerden [38, Heuristic 2] stated as a "consequence of GH" that (1) there are $\alpha^{d+o(d)}$ lattice points in a ball of radius $\alpha\lambda_1(L)$ and (2) these lattice points are treated "as being uniformly distributed over the ball".

This distribution is important when one considers not just the lengths of short vectors but also simple reduction using a list of short vectors: replace the input t with a shorter vector $t - u$ for some u in the list, if possible, and then repeat. For example, Laarhoven [57] considered the close-vector problem for any lattice L, and used spherical models to analyze the effect of precomputing all short vectors in L and then repeatedly subtracting off one of those short vectors from the target vector. Doulgerakis–Laarhoven–de Weger [36] reported experiments with randomly generated lattices supporting Laarhoven's analysis.

1.3. Using spherical models to limit *S*-unit attacks. Internally, an S-unit attack constructs a target point and searches for vectors close to that point

[1] "Quod quamquam iam per se evidens esse videatur, tamen demonstratione rigorosa munire non aspernabimur." See [47, pages 277–278].

in an auxiliary lattice. This auxiliary lattice is a standard number-theoretic lattice called the "S-unit lattice". Given how well established spherical models are in the literature, it is natural to apply these models to S-unit lattices.

This was done in 2019 by Pellet-Mary–Hanrot–Stehlé [78], who presented and analyzed an S-unit attack by applying Laarhoven's algorithm [57] and analysis to S-unit lattices. The analysis of [78] says that S-unit attacks can reach much shorter vectors than various earlier papers had reached, but also says that this has an inherent cost: to reach approximation factor $\exp(n^{o(1)})$, one needs to spend time $\exp(n^{1+o(1)})$ building and using a database of $\exp(n^{1+o(1)})$ short lattice vectors. This relies on a spherical-model calculation in [57] concluding that each short lattice vector has probability $\exp(-n^{1+o(1)})$ of successful reduction.

Evidently this is, finally, the desired barrier to S-unit attacks: the result of taking a central, well-studied analysis technique in lattice-based cryptography and applying it to S-unit lattices.

Very recent events have illustrated the power of this barrier. In late August 2021, a talk given by Bernstein [14] reported experiments with, and conjectured subexponential scalability for, an Ideal-SVP attack. This attack is an S-unit attack, like [78], although with some differences such as a larger choice of S and a new method of precomputing short S-units. Four days later, Ducas and Pellet-Mary [40] responded "Extraordinary claims require extraordinary evidence" and concisely explained why the spherical-model analysis used in [78] already implied that the probability of success of such a large S "would be *ridiculously* small", certainly no better than [78].

1.4. Contributions of this paper. This paper shows that spherical models underestimate the power of S-unit attacks: S-unit lattices have much shorter vectors and reduce much more effectively than predicted by these models. The attacker can freely choose S to make the gap as large as desired. This resolves the dispute between [14] and [40], and shows that S-unit attacks are not subject to the limits deduced in [78].

There are many choices of the pair (n, S), and this paper does not cover them all, but this paper covers every direction in (n, S) space (see Table 1.5):

- Section 6 covers the minimal n, namely $n = 1$. This paper presents (easy) calculations regarding short S-units and reduction effectiveness, and (with more work) quantifies how inaccurate the spherical-model predictions are.
- Section 7 covers what happens for each n as S increases. It is easy to see from the structure of S-units that, as S becomes larger and larger, the shortest S-units do not grow. Furthermore, the effectiveness of reduction does not degrade—on the contrary, whatever the target vector is, it *will* be found by a sufficiently large S-unit attack. Spherical models instead make the absurd predictions that the shortest S-units become longer and longer and that reduction modulo short S-units becomes less and less likely to succeed. This is also noted in Section 6 for $n = 1$, but it is a much broader phenomenon.
- Section 8 covers what happens when S is minimal. This paper proves bounds on the length of a short S-unit; presents the results of reduction experiments

§	lattice	spherical model	reality
4	$\mathbb{Z}^d$ (warmup, already known)	$(1+o(1))(d/2\pi e)^{1/2}$	1
6	S-units for $n=1$, any S	$(1+o(1))(d/2\pi e)^{1/2}\log d$	$O(1)$
7	S-units for each n as S increases	$(1+o(1))(d/2\pi e)^{1/2}\log d$	$O(1)$
8	S-units for minimal S, any n	$d^{1+o(1)}$	$d^{1/2+o(1)}$

Table 1.5. Asymptotic overview (as $d\to\infty$) of this paper's analyses of the inaccuracy of spherical models in predicting vector lengths in various d-dimensional lattices. The set S is assumed to be $\infty\cup\{P:\#(R/P)\le y\}$. In the last row, $y=1$ and $d=n/2-1$; in the previous two rows, $d\in(1+o(1))y/\log y$. The last row assumes small h^+. In the previous row, o and O are as $d\to\infty$ for fixed n; constants can depend on n. This paper also analyzes reduction effectiveness for each lattice.

for various n; and (under the aforementioned assumption "$h^+=1$"; see Appendix C) quantifies the inaccuracy of the spherical-model predictions.

This does not rule out the possibility of the predictions being accurate for some intermediate pairs (n,S). However, even if this were to occur, it would do nothing to stop attacks: S is chosen by the attacker and in particular can be taken arbitrarily large, as reflected by the large choice of S in the talk [14].

The fact that the "Gaussian heuristic" is very wrong for some lattices L, including important lattices such as $L=\mathbb{Z}^d$, is not a new observation. The $\mathbb{Z}^d$ example contradicts not just the $\lambda_1(L)$ predictions but also the reduction predictions; see Section 4. Structurally, these counterexamples show that merely looking at the lattice dimension and lattice determinant, as in spherical models, does *not* provide enough information to predict the distribution of short vectors and the effectiveness of reduction modulo those vectors. The fact that there are some counterexamples does not say that S-unit lattices are also counterexamples, but the evidence amassed in this paper says that S-unit lattices look much more like $\mathbb{Z}^d$ than like random lattices from the invariant distribution.

Each failure that this paper describes in the shortest-nonzero-vector-length predictions is accompanied by failure of the reduction-effectiveness predictions. An open question is how general this association is: how reliably one can predict higher reduction effectiveness simply from seeing shorter nonzero vectors. In any event, given the previously known $\mathbb{Z}^d$ counterexample, the use of spherical models to analyze specific types of lattices should have been treated with much more caution; given the results of this paper, the use of spherical models of S-unit lattices is indefensible.

So how should one quantify the shortness of vectors found by S-unit attacks? Available theorems do not quantify this for most pairs (n,S) of interest. The standard scientific answer is to carry out experiments, formulate hypotheses allowing extrapolation beyond the experiments, carry out further experiments challenging the hypotheses, etc. Interestingly, some experiments were reported in [78]; however, those experiments were designed to double-check that the slow attacks in [78] worked, rather than to challenge the accuracy of spherical models. If S-unit attacks had somehow reduced *less* effectively than spherical models

predict, then this might have been detected by the experiments, but this is only one-sided evidence. The notion that S-unit attacks reduce *no more effectively* than predicted was never justified; this paper shows that the notion is incorrect.

1.6. Clarity and falsifiability. [70, Heuristic 1] states that if L is a lattice of dimension d then $\lambda_1(L) = ((d/2)!(\det L))^{1/d}/\sqrt{\pi}$. Try 100 random lattices, and notice that the statement is false for all of them. Surely the intention was instead to say that this is a *model* of the length, and that the actual length is *close* to the model—but how close? How far away would an example have to be to qualify as a counterexample? The same question arises when the word "approximately" is made explicit, as in [38, Heuristic 1].

Another clarity problem appears when statements are first made for random lattices (see, e.g., [53, page 273])—presumably meaning random lattices from the invariant distribution, presumably allowing some percentage of these lattices to be counterexamples—but are then applied to whatever lattices appear. It seems that the actual analysis methodology is assuming a statement regarding *all* lattices, not just random lattices, but this is not made explicit.

There are more of these definitional difficulties in the literature. This poses a problem for a paper whose goal is to show that the same methodology does not work for S-unit lattices. To address this issue, this paper gives a precise mathematical definition of a spherical model of a lattice, factoring the analysis of S-unit counterexamples into

- an analysis of what this definition says regarding the examples and
- a comparison of this definition to the literature (see Appendix E).

Claims that the literature meant something else do not affect the validity of the first analysis. It seems unlikely that such claims could close the large gap between the predictions and the facts; this paper includes some analysis of robustness to variations in the definitions. If someday another model is defined that accurately captures what is happening for $\mathbb{Z}^d$ and for S-unit lattices, then that model has been separated from the model in this paper; subsequent analyses can specify which of the two models they are using.

The "spherical model" terminology appears to be new in this context. There are five reasons for this paper's use of this terminology. First, there are enough variations in what the literature says regarding the "Gaussian heuristic" that new terminology will help avoid confusion. Second, the details of the definition here are handled differently; the ideas are standard, and this paper checks that the results are consistent with the literature, but the change in terminology serves as a warning that, at least formally, there is a change in details. Third, a "model" is conventionally required to be mathematically defined, capturing one of the advantages of the work here, whereas a "heuristic" can have undefined words such as "approximately". Fourth, "spherical" is more descriptive of the content than "Gaussian". Fifth, the "Gaussian heuristic" does not appear to have been introduced—or endorsed—by Gauss. One is forced to wonder whether the presence of Gauss's name here has contributed to this heuristic having been given less skepticism than it deserves.

1.7. Priority dates. The talk mentioned above included our announcement of the results of this paper: the talk [14, video, minute 20] described the S-unit lattice as an "amazingly special lattice", in particular regarding its "algebraic and analytic features". These features are a prerequisite for the success of the experiments presented in that talk. This paper provides full details of the analytic features, quantifying how far S-unit lattices are from most lattices.

1.8. Organization of this paper. Section 2 reviews theorems regarding the lengths of vectors in *most* lattices. Section 3 defines spherical models, and proves theorems regarding short vectors in spherical models and the effectiveness of reduction modulo those vectors. Section 4 quantifies how badly spherical models fail for the lattice $\mathbb{Z}^d$; this is a warmup for the subsequent analysis of S-unit lattices. Section 5 reviews S-units, the S-unit lattice, and S-unit attacks. Finally, Sections 6, 7, and 8 quantify how badly spherical models fail for S-unit lattices with, respectively, $n = 1$; each n as S increases; and each n when S is minimal.

1.9. Acknowledgments. This work began at the Simons Institute for the Theory of Computing at the University of California at Berkeley, USA, which held concurrent research programs in 2020 on "Lattices: algorithms, complexity, and cryptography" and "The quantum wave in computing"; continued during a visit to the Research Center for Information Technology Innovation at Academia Sinica, Taiwan; and performed computations on the Fast Crypto Lab cluster at the Institute of Information Science at Academia Sinica. The authors are grateful to the Simons Institute, CITI, and IIS for their hospitality.

2 Random lattices

There is a standard mathematical definition of what it means to pick a fixed-dimension fixed-determinant lattice at random. This section briefly reviews (1) the theory of random lattices and (2) theorems regarding the distribution of vector lengths in random lattices.

2.1. Invariant measure. By definition $\mathrm{SL}_d(\mathbb{R})$ is the group of determinant-1 matrices in $\mathbb{R}^{d\times d}$. One can measure subsets of $\mathrm{SL}_d(\mathbb{R})$ using (d^2-1)-dimensional Hausdorff measure, but it is more useful to work with an "invariant measure" for this group, meaning that $\int_M f(M)\,dM = \int_M f(MN)\,dM$ for each $N \in \mathrm{SL}_d(\mathbb{R})$. Siegel [85] gave an explicit construction of an invariant measure on $\mathrm{SL}_d(\mathbb{R})$, analogous to explicit constructions by Hurwitz [54] of invariant measures for some other groups.

Siegel also defined "a fundamental region F with respect to Γ_1", where Γ_1 is $\mathrm{SL}_d(\mathbb{Z})$, the group of determinant-1 matrices in $\mathbb{Z}^{d\times d}$. Siegel proved "as an immediate consequence of Minkowski's reduction theory" that "the volume of F is finite", where "volume" refers to the invariant measure. One can thus scale the measure so that the volume of F is 1, obtaining a probability measure on F.

If $M \in F$ then the lattice L generated by the rows of M is a determinant-1 dimension-d lattice. This map $M \mapsto L$ is a bijection between F and the set of determinant-1 dimension-d lattices in $\mathbb{R}^d$: any determinant-1 dimension-d lattice

has a basis matrix in $\mathrm{SL}_d(\mathbb{R})$, and modulo Γ_1 this basis matrix has a unique representative in F, the unique matrix in F whose rows generate L.

The probability measure on F mentioned above, invariant measure scaled so that F has total measure 1, now induces a probability measure on the set of determinant-1 dimension-d lattices, again called the "invariant measure" on this set. Here "invariant" means that $\int_L f(L)\,dL = \int_L f(LN)\,dL$ for each $N \in \mathrm{SL}_d(\mathbb{R})$, where now the integrals are over lattices L. The theory of random lattices considers lattices distributed according to the invariant measure.

2.2. Short vectors. Rogers [82] showed that "the number of pairs of points $\pm x$ of a lattice Λ in S has a distribution, which is asymptotic, as n becomes large, to the Poisson distribution with mean $\frac{1}{2}V$". The "n" in [82] is this paper's d; V is a fixed real number as $d \to \infty$; S is a set spherically symmetric around 0 with Borel measure V; Λ is a random determinant-1 dimension-d lattice with the invariant distribution; and $0 \in \Lambda$ is not counted (as a pair ± 0, or as a single entry), so the number of pairs of points counted can be 0.

In particular, consider a d-dimensional ball S, centered at the origin, of volume V, where V is one of the following:

- $V = 200$: A sample from a Poisson distribution with mean 100 has probability $\exp(-100) \approx 0$ of being 0.
- $V = 2$: A sample from a Poisson distribution with mean 1 has probability $\exp(-1) \approx 0.36$ of being 0.
- $V = 0.02$: A sample from a Poisson distribution with mean 0.01 has probability $\exp(-0.01) \approx 0.99$ of being 0.

One thus expects the shortest nonzero vectors in Λ to be inside the ball of volume 200, perhaps inside the ball of volume 2, and very likely not inside the ball of volume 0.02, once d is large enough for Rogers's asymptotic to apply. Ball volume grows proportionally to the dth power of radius, so these three balls have almost the same radii for large d. This forces the shortest nonzero vectors in *most* Λ, but not all, to have length almost exactly matching these radii.

Strömbergsson and Södergren [88, Remark 1.8] state a precise theorem along these lines. As an example of this theorem, the 10^6 shortest nonzero vectors in a random determinant-1 dimension-d lattice with the invariant distribution all have length $(1 + O((\log d)/d))(d/2\pi e)^{1/2}$ with probability $1 - o(1)$ as $d \to \infty$.

3 Spherical model of a lattice

This section formally defines a spherical model M of a given lattice; analyzes the minimum length of nonzero vectors in M; and analyzes the effectiveness of reduction modulo short vectors in M.

As a starting point, consider the following two well-known theorems.

Theorem 3.1. *Let d be a positive integer. Define $B = \{x \in \mathbb{R}^d : ||x||_2 \leq 1\}$. Then $\mathrm{Vol}_d\, rB = r^d \pi^{d/2}/(d/2)!$ for each nonnegative real number r.*

Here Vol_d means d-dimensional volume, and $(d/2)!$ means $\Gamma(d/2+1)$. Two examples of the theorem: $\mathrm{Vol}_2\, rB = r^2\pi/(2/2)! = \pi r^2$ for $d=2$; and $\mathrm{Vol}_3\, rB = r^3\pi^{3/2}/(3/2)! = (4\pi/3)r^3$ for $d=3$, since $(3/2)! = (3/2)(1/2)\pi^{1/2}$. Often volume subscripts are clear from context and are omitted, but $(d-1)$-volume Vol_{d-1} on $\mathbb{R}^d$, meaning $(d-1)$-dimensional normalized Hausdorff measure, is used below.

Proof. This is a standard integration exercise. For one proof see [60]; further proofs are cited in [60]. □

Theorem 3.2. *Let d be a positive integer. Define $B = \{x \in \mathbb{R}^d : ||x||_2 \le 1\}$. Let L be a d-dimensional lattice in $\mathbb{R}^d$. Then $\lim_{r\to\infty} \#(L \cap rB)/\mathrm{Vol}_d\, rB = 1/\det L$.*

Proof. Choose a basis $b_1, b_2, \ldots, b_d$ for L. Define a bijection $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ by $\varphi(c_1, c_2, \ldots, c_d) = c_1b_1 + c_2b_2 + \cdots + c_db_d$. Write $X = \varphi^{-1}(B)$. Then $\varphi^{-1}(L) = \mathbb{Z}^d$; $\#(\mathbb{Z}^d \cap rX) = \#(L \cap rB)$; and $\mathrm{Vol}\, rX = \mathrm{Vol}\, \varphi^{-1}(rB) = (\mathrm{Vol}\, rB)/\det L$.

The conditions of Davenport's theorem [35, Section 2, Theorem] are satisfied with $h=1$, $n=d$, $R=rX$: any line parallel to a coordinate axis intersects rX in at most one interval, since rX is convex; the same is true for any region obtained from rX by projecting onto any selection of m coordinates, for any $m \in \{1, \ldots, d-1\}$; and rX is closed and bounded.

Now $|\#(\mathbb{Z}^d \cap rX) - \mathrm{Vol}\, rX| \le \sum_{0\le m\le d-1} V_m$ by Davenport's theorem, where V_m is the sum of m-dimensional volumes of all $\binom{d}{m}$ regions obtained by projecting rX onto m coordinates. Each of these regions is, like rX, obtained by scaling an r-independent region by r, so V_m is proportional to r^m, so $\sum_{0\le m\le d-1} V_m \in O(r^{d-1})$. Finally

$$\left|\frac{\#(L\cap rB)}{\mathrm{Vol}\, rB} - \frac{1}{\det L}\right| = \frac{\left|\#(\mathbb{Z}^d \cap rX) - \mathrm{Vol}\, rX\right|}{\mathrm{Vol}\, rB} \le \frac{\sum_{0\le m\le d-1} V_m}{\mathrm{Vol}\, rB}.$$

The numerator is $O(r^{d-1})$, and the denominator $\mathrm{Vol}\, rB$ is proportional to r^d, so $|\#(L\cap rB)/\mathrm{Vol}\, rB - 1/\det L| \in O(1/r)$. Hence the limit as $r\to\infty$ is 0. □

For example, for $d=1$ and $L=\mathbb{Z}^1$, one has $\#(L\cap rB) = 1$ if $r<1$; $\#(L\cap rB) = 3$ if $1\le r<2$; $\#(L\cap rB) = 5$ if $2\le r<3$; etc. The general pattern for $d=1$ is that $\#(L\cap rB) = 1 + 2\lfloor(\mathrm{Vol}\, rB)/(2\det L)\rfloor$, which implies $\lim_{r\to\infty} \#(L\cap rB)/\mathrm{Vol}\, rB = 1/\det L$.

A spherical model imagines that one can extrapolate the rule $\#(L\cap rB) = 1 + 2\lfloor(\mathrm{Vol}\, rB)/(2\det L)\rfloor$ from $d=1$ to any value of d, and that lattice points are independently and uniformly distributed except for what this rule says about their lengths. This model disregards the additive structure of L, and all other information about L aside from the dimension and determinant of L. For a picture with $d=2$ see Figure 3.4.

Definition 3.3 (spherical model of a lattice). *Let d be a positive integer. Let L be a d-dimensional lattice. Let $L_{\mathbb{R}}$ be the $\mathbb{R}$-vector space generated by L. Let μ_j be a uniform random element of $\{x \in L_{\mathbb{R}} : ||x||_2^d = 2j\pi^{-d/2}(d/2)!\det L\}$ for each integer $j \ge 1$. Assume that $\mu_1, \mu_2, \ldots$ are statistically independent. Then $\{0, \mu_1, -\mu_1, \mu_2, -\mu_2, \ldots\}$ is a* **spherical model of L.**

The word "random" here has its measure-theoretic definition: there is a fixed measure space Pr of total measure 1, and a random element of a measurable space X means a measurable function from Pr to X.

The definition of a spherical model does not require L to be specifically in $\mathbb{R}^d$. Formally, the extra generality is useful for S-unit lattices: the usual definitions of a d-dimensional S-unit lattice either define the lattice embedded in $\mathbb{R}^{d+1}$, or define it embedded in a larger-dimensional vector space (the extra coordinates of lattice vectors being 0), or define it abstractly without reference to an embedding (see, e.g., [62, Section 3]). However, one can always map $L_{\mathbb{R}}$ isometrically to $\mathbb{R}^d$; then L maps to a d-dimensional lattice in $\mathbb{R}^d$ with the same determinant, and the set of elements of $L_{\mathbb{R}}$ of length at most r maps to $rB = \{x \in \mathbb{R}^d : ||x||_2 \le r\}$. This paper thus focuses on the case $L_{\mathbb{R}} = \mathbb{R}^d$ in, e.g., Theorem 3.6 below.

3.5. Using a spherical model to predict vector lengths. The number of limited-length points in a spherical model of L follows from the definition of a spherical model; see Theorem 3.6 below. In particular, the minimum nonzero length is $(2\pi^{-d/2}(d/2)! \det L)^{1/d}$.

Theorem 3.6. *Let d be a positive integer. Define $B = \{x \in \mathbb{R}^d : ||x||_2 \le 1\}$. Let L be a d-dimensional lattice in $\mathbb{R}^d$. Let M be a spherical model of L. Then*

- $\min\{||x||_2 : x \in M - \{0\}\} = (2\pi^{-d/2}(d/2)! \det L)^{1/d}$;
- *each $r \in \mathbb{R}$ with $r \ge 0$ has $\#(M \cap rB) = 1 + 2\lfloor (\mathrm{Vol}_d\, rB)/(2 \det L) \rfloor$; and*
- $\lim_{r\to\infty} \#(M \cap rB)/\mathrm{Vol}_d\, rB = 1/\det L$.

Proof. By definition M has the form $\{0, \mu_1, -\mu_1, \mu_2, -\mu_2, \ldots\}$, with $||\mu_j||_2^d = 2j\pi^{-d/2}(d/2)! \det L$. The shortest nonzero elements of M are thus $\pm\mu_1$, with length $(2\pi^{-d/2}(d/2)! \det L)^{1/d}$.

The elements of $M \cap rB$ are 0 and $\pm\mu_j$ for each positive integer j with $||\mu_j||_2 \le r$, i.e., with $2j\pi^{-d/2}(d/2)! \det L \le r^d$. The largest positive integer J with $2J\pi^{-d/2}(d/2)! \det L \le r^d$ is the largest positive integer J with $J \le r^d\pi^{d/2}/(2(d/2)! \det L)$. By Theorem 3.1, this is the largest positive integer J with $J \le (\mathrm{Vol}\, rB)/(2 \det L)$; i.e., $J = \lfloor (\mathrm{Vol}\, rB)/(2 \det L) \rfloor$. The elements of $M \cap rB$ are then 0 and $\pm\mu_j$ for each positive integer $j \le J$. These elements are distinct, so $\#(M \cap rB) = 1 + 2J = 1 + 2\lfloor (\mathrm{Vol}\, rB)/(2 \det L) \rfloor$.

Now $\#(M \cap rB)$ is between $-1 + (\mathrm{Vol}\, rB)/\det L$ and $1 + (\mathrm{Vol}\, rB)/\det L$, so $\#(M \cap rB)/\mathrm{Vol}\, rB$ is between $-1/\mathrm{Vol}\, rB + 1/\det L$ and $1/\mathrm{Vol}\, rB + 1/\det L$. Finally, $\mathrm{Vol}\, rB \to \infty$ as $r \to \infty$ by Theorem 3.1, so $1/\mathrm{Vol}\, rB \to 0$. □

3.7. Asymptotic interlude. Theorem 3.1, Definition 3.3, and Theorem 3.6 involve the quantity $(d/2)! = \Gamma(d/2+1)$. Theorem 3.8 pins down the asymptotics of $(d/2)!$ as $d \to \infty$.

Theorem 3.8. *Let d be a positive integer. Write $e = \exp 1$. Then $(d/2)! = (\pi d)^{1/2}(d/2e)^{d/2}e^{\theta/6d}$ for some $\theta \in \mathbb{R}$ with $0 < \theta < 2$.*

Proof. Robbins [81, formulas (1) and (2)] showed, for each integer $n \ge 1$, that $n! = (2\pi n)^{1/2}(n/e)^n e^{r_n}$ for some $r_n \in \mathbb{R}$ with $1/(12n+1) < r_n < 1/12n$.

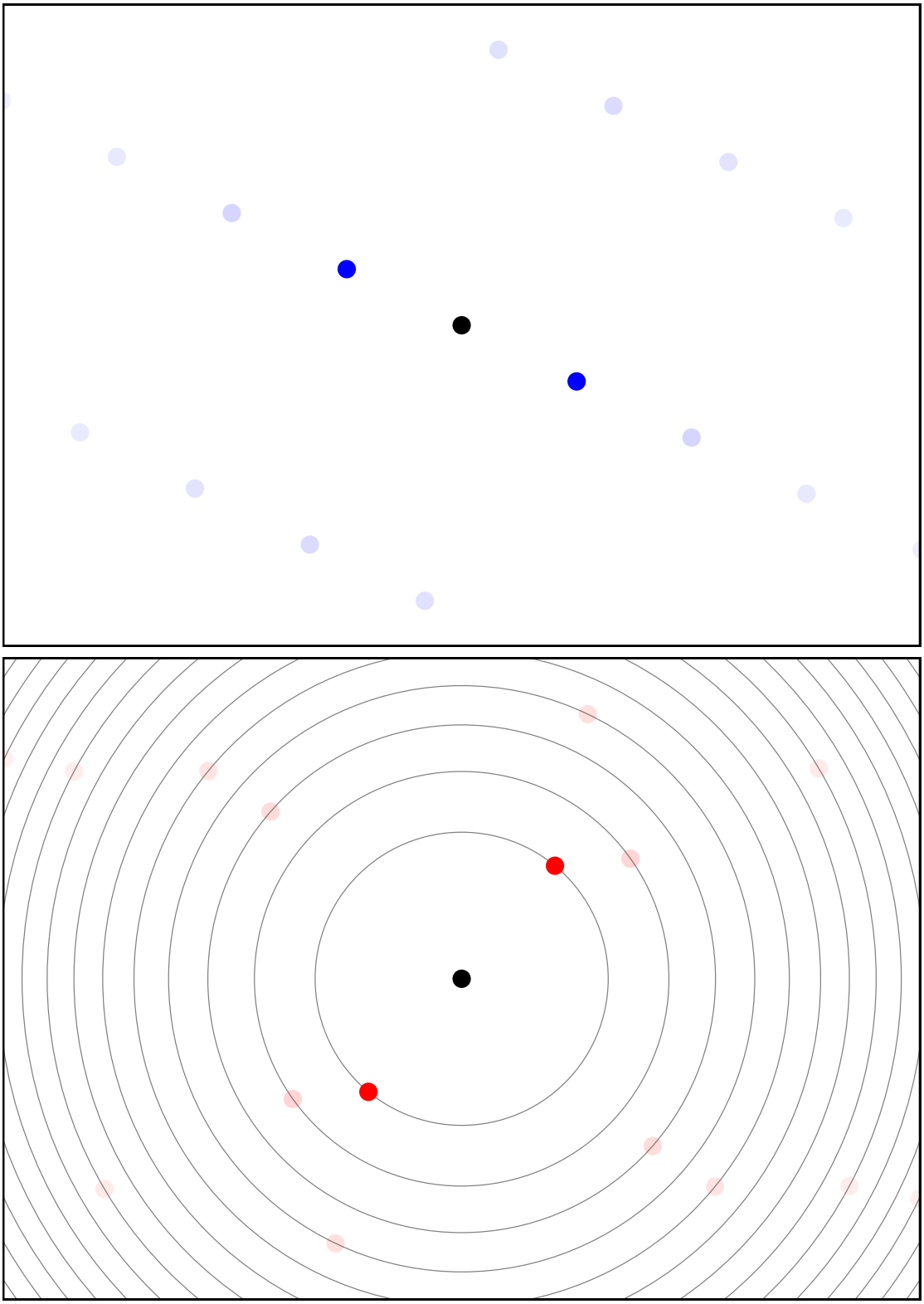

Fig. 3.4. Dots in top picture: a lattice L in dimension $d = 2$. Shortest nonzero vectors are highlighted. Dots in bottom picture: sample from a spherical model M of L. Circles in bottom picture: distribution of M, independent uniform random circle points and their negatives. Circle j has radius proportional to $j^{1/2}$, with constant of proportionality set so that $\lim_{r\to\infty} \#(M \cap rB)/\#(L \cap rB) = 1$.

The easy case of the theorem is that d is even, say $d = 2n$. Write $\theta = 6dr_{d/2}$; then $0 < \theta < 1$, and $(d/2)! = (\pi d)^{1/2}(d/2e)^{d/2}e^{\theta/6d}$ as claimed.

The rest of the proof assumes that d is odd, say $d = 2n - 1$. Then $(d/2)! = (n - 1/2)(n - 3/2)\cdots(1/2)\pi^{1/2} = (2n-1)(2n-3)\cdots(1)\pi^{1/2}/2^n$. Multiply by $n! = (2n)(2n-2)\cdots(2)/2^n$ to see that $(d/2)!n! = (2n)!\pi^{1/2}/4^n$ and thus

$$(d/2)!(2\pi n)^{1/2}(n/e)^n e^{r_n} = (4\pi n)^{1/2}(2n/e)^{2n}e^{r_{2n}}\pi^{1/2}/4^n.$$

Cancel factors and rearrange to obtain

$$(d/2)! = (2\pi)^{1/2}(n/e)^n e^{r_{2n}-r_n} = (\pi d)^{1/2}(d/2e)^{d/2}(1+1/d)^{(d+1)/2}e^{-1/2}e^{r_{2n}-r_n}.$$

In other words, $(d/2)! = (\pi d)^{1/2}(d/2e)^{d/2}e^{\theta/6d}$ where

$$\theta = 3d((d+1)\log(1+1/d) - 1) + 6d(r_{2n} - r_n).$$

All that remains is to show $0 < \theta < 2$.

The power series $(1+x)\log(1+x) - x$ is $\sum_{j\geq 2}(-1)^j x^j/j(j-1)$, which is an alternating decreasing series for $0 < x \leq 1$, so it is between $x^2/2 - x^3/6$ and $x^2/2$. Substitute $x = 1/d$ to see that

$$1/2d^2 - 1/6d^3 \leq (1+1/d)\log(1+1/d) - 1/d \leq 1/2d^2,$$

i.e., $1/2d - 1/6d^2 \leq (d+1)\log(1+1/d) - 1 \leq 1/2d$. Hence $3d((d+1)\log(1+1/d)-1)$ is between $3/2 - 1/2d$ and $3/2$.

Meanwhile $1/(24n+1) < r_{2n} < 1/24n$, and $1/(12n+1) < r_n < 1/12n$, so $1/(24n+1) - 1/12n < r_{2n} - r_n < 1/24n - 1/(12n+1)$. Hence θ is between $3/2 - 1/2d + 6d(1/(24n+1) - 1/12n) = 1 - (d^2+12d+13)/(24d^3+50d^2+26d) > 0$ and $3/2 + 6d(1/24n - 1/(12n+1)) = 1 + (8d+7)/(12d^2+26d+14) < 2$. □

Theorem 3.9. *Let d be a positive integer. Let L be a d-dimensional lattice. Let M be a spherical model of L. Write $e = \exp 1$. Then $\min\{||x||_2 : x \in M - \{0\}\} \in (1+o(1))(d/2\pi e)^{1/2}(\det L)^{1/d}$ as $d \to \infty$.*

For comparison, the minimum nonzero length in a random determinant-1 dimension-d lattice with the invariant distribution is $(1+O((\log d)/d))(d/2\pi e)^{1/2}$ with probability $1 - o(1)$ as $d \to \infty$; see Section 2. Figure 3.10 compares many identical-determinant lattices for $d = 2$ to many samples from a spherical model.

Proof. Define $L_{\mathbb{R}}$ as the $\mathbb{R}$-vector space generated by L. Map $L_{\mathbb{R}}$ isometrically to $\mathbb{R}^d$; this maps L to a d-dimensional lattice in $\mathbb{R}^d$ with determinant $\det L$, and maps M to a spherical model of that lattice. It thus suffices to consider the case $L \subseteq \mathbb{R}^d$.

Abbreviate $\min\{||x||_2 : x \in M - \{0\}\}$ as $\lambda_1(M)$. By Theorem 3.6, $\lambda_1(M) = (2\pi^{-d/2}(d/2)!\det L)^{1/d}$. By Theorem 3.8, $(d/2)! = (\pi d)^{1/2}(d/2e)^{d/2}e^{\theta/6d}$ for some $\theta \in \mathbb{R}$ with $0 < \theta < 2$. Hence $\lambda_1(M) = e^{\theta/6d^2}(4\pi d)^{1/2d}(d/2\pi e)^{1/2}(\det L)^{1/d}$. The first factor $e^{\theta/6d^2}$ converges to 1 as $d \to \infty$ since $0 < \theta < 2$; the second factor $(4\pi d)^{1/2d}$ also converges to 1; so $\lambda_1(M) \in (1+o(1))(d/2\pi e)^{1/2}(\det L)^{1/d}$. □

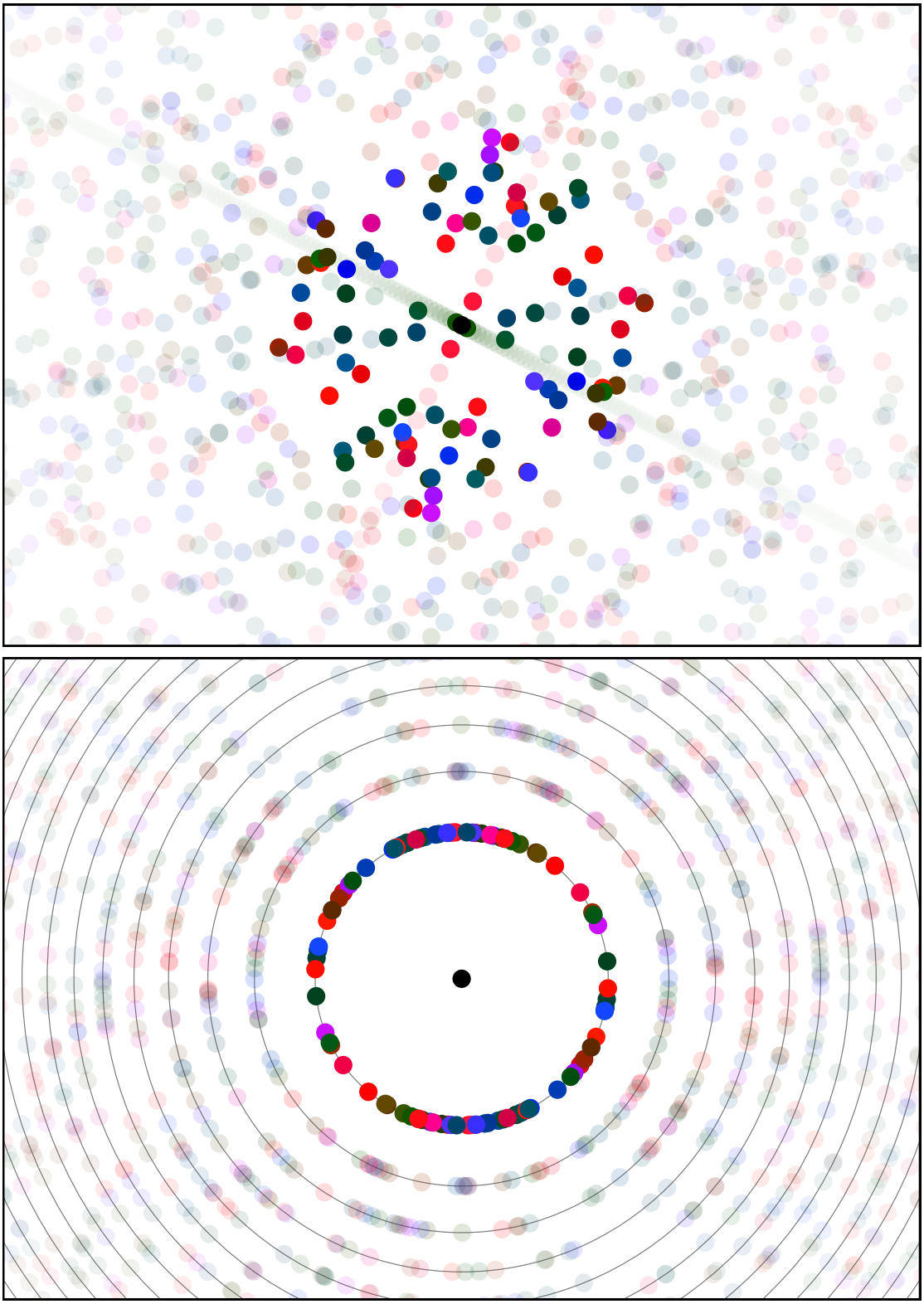

Fig. 3.10. Dots in top picture: 50 lattices L in dimension $d = 2$. Shortest nonzero vectors are highlighted. Dots in bottom picture: samples from a spherical model M of each L. Circles in bottom picture: distribution of M, independent uniform random circle points and their negatives. Circle j has radius proportional to $j^{1/2}$, with constant of proportionality set so that $\lim_{r\to\infty} \#(M \cap rB)/\#(L \cap rB) = 1$.

3.11. Using a spherical model to predict the effectiveness of reduction. Consider any nonzero vector $\nu \in \mathbb{R}^d$. Theorem 3.13 describes, for $\alpha > 0$, the probability that a uniform random vector μ of length $2\alpha||\nu||_2$ reduces ν, i.e., that $||\nu - \mu||_2 < ||\nu||_2$. The probability is the ratio between the $(d-1)$-volume of a spherical cap (see Definition 3.12) and the volume of the sphere. This ratio is easy to analyze by standard techniques; see Section 3.14.

In particular, say $\pm\mu$ are the shortest nonzero vectors in a spherical model of L. Theorem 3.6 says what the length r of μ is, and then comparing r to $||\nu||_2$ says what α is, and then Theorem 3.13 expresses the reduction probabilities for $\pm\mu$ in terms of spherical-cap volumes. Similar comments apply to all short nonzero vectors in a spherical model, not just the shortest.

Definition 3.12. *Let d be a positive integer. Let α be a real number. The* **α-cap of the unit $(d-1)$-sphere**, *denoted* $\operatorname{Cap}_\alpha^{d-1}$, *is* $\{x \in \mathbb{R}^d : ||x||_2 = 1, x_1 > \alpha\}$ *where x_1 is the first coordinate of x.*

For example, the cap is empty for $\alpha \geq 1$; the entire unit sphere for $\alpha < -1$; the "right half" of the sphere for $\alpha = 0$; and the set of $(x_1, \ldots, x_d) \in \mathbb{R}^d$ such that $x_1^2 + \cdots + x_d^2 = 1$ and $x_1 > 1/2$ for $\alpha = 1/2$. (Evidently the cap-wearer is lying flat on the sofa.)

Theorem 3.13. *Let d be a positive integer. Let r be a positive real number. Let μ be a uniform random element of $\{x \in \mathbb{R}^d : ||x||_2 = r\}$. Let α be a positive real number. Let ν be an element of $\mathbb{R}^d$ with $||\nu||_2 = r/2\alpha$. Then the probability that $||\nu - \mu||_2 < ||\nu||_2$ is* $(\operatorname{Vol}_{d-1} \operatorname{Cap}_\alpha^{d-1})/(2 \operatorname{Vol}_{d-1} \operatorname{Cap}_0^{d-1})$.

Proof. Define $x = \mu/r$. Then x is a uniform random point on the unit $(d-1)$-sphere.

Define $y = \nu/r$. Then $||y||_2 = 1/2\alpha$. Also $||\nu - \mu||_2 < ||\nu||_2$ exactly when $||y - x||_2 < ||y||_2$.

We have $||y||_2^2 = y \cdot y$ and $||y - x||_2^2 = (y - x) \cdot (y - x) = y \cdot y - 2y \cdot x + 1$, so $||y - x||_2^2 < ||y||_2^2$ exactly when $2y \cdot x > 1$.

Define $z = 2\alpha y$. Then $||z||_2 = 1$, and $2y \cdot x > 1$ exactly when $z \cdot x > \alpha$.

The desired probability is thus the probability that a uniform random point x on the unit $(d-1)$-sphere satisfies $z \cdot x > \alpha$, where z is another point on the sphere. This probability is invariant under rotations of z, so assume without loss of generality that $z = (1, 0, \ldots, 0)$. Then $z \cdot x$ is the first coordinate of x, so $z \cdot x > \alpha$ exactly when $x \in \operatorname{Cap}_\alpha^{d-1}$. The probability that this occurs is the volume ratio between $\operatorname{Cap}_\alpha^{d-1}$ and the $(d-1)$-sphere: i.e., $(\operatorname{Vol}_{d-1} \operatorname{Cap}_\alpha^{d-1})/(2 \operatorname{Vol}_{d-1} \operatorname{Cap}_0^{d-1})$. □

3.14. Spherical-cap ratios: computations and asymptotics. In general, spherical-cap volumes do not have formulas as simple as the ball volume in Theorem 3.1, but they are nevertheless easy to compute to any desired precision for any specified α and d. One approach is to compute the volumes using standard integration software, as suggested in the lattice context in [20, Section 3.2] and [92, page 66], but it is simpler and faster to use a standard equation relating the

volume ratio to the "beta distribution", as suggested in the lattice context in [15, page 50]. Theorem 3.17(2) reviews this equation. The same equation makes it easy to see, within a $\Theta(d)$ factor, asymptotics of the ratio for fixed α as $d \to \infty$; see Theorem 3.18.

Definition 3.15. *Let x, a, b be real numbers with $a > 0$, $b > 0$, and $0 \le x \le 1$. Then the* **incomplete beta function at x with parameters a, b**, *denoted $\mathcal{B}(x; a, b)$, is $\int_0^x t^{a-1}(1-t)^{b-1}\,dt$.*

Theorem 3.16. *Let x, a, b be real numbers with $a > 0$, $0 < b \le 1$, and $0 \le x \le 1$. Then $x^a/a \le \mathcal{B}(x; a, b) \le x^a \mathcal{B}(1; a, b)$.*

Proof. If $x = 0$ then $x^a = 0$ and $\mathcal{B}(x; a, b) = 0$. For $x > 0$, substitute $t = xu$ into $\mathcal{B}(x; a, b) = \int_0^x t^{a-1}(1-t)^{b-1}\,dt$ to see that

$$\mathcal{B}(x; a, b) = \int_0^1 (xu)^{a-1}(1-xu)^{b-1} x\,du = x^a \int_0^1 u^{a-1}(1-xu)^{b-1}\,du.$$

If $0 < u < 1$ then $1 > 1 - xu \ge 1 - u > 0$ so $1 \le (1-xu)^{b-1} \le (1-u)^{b-1}$ since $b - 1 \le 0$. This gives the lower bound $\mathcal{B}(x; a, b) \ge x^a \int_0^1 u^{a-1}\,du = x^a/a$ and the upper bound $\mathcal{B}(x; a, b) \le x^a \int_0^1 u^{a-1}(1-u)^{b-1}\,du = x^a \mathcal{B}(1; a, b)$. □

Theorem 3.17. *Let d be an integer with $d \ge 2$. Let α be a real number with $0 \le \alpha \le 1$. Then*

$$\operatorname{Vol}_{d-1} \operatorname{Cap}_\alpha^{d-1} = \frac{\pi^{(d-1)/2}}{\Gamma((d-1)/2)} \mathcal{B}(1-\alpha^2; (d-1)/2, 1/2) \tag{1}$$

and

$$\frac{\operatorname{Vol}_{d-1} \operatorname{Cap}_\alpha^{d-1}}{\operatorname{Vol}_{d-1} \operatorname{Cap}_0^{d-1}} = \frac{\mathcal{B}(1-\alpha^2; (d-1)/2, 1/2)}{\mathcal{B}(1; (d-1)/2, 1/2)}. \tag{2}$$

Proof. See, e.g., [18, "application to integration"]: the integral of $g(\nu_1)$ over points ν on the unit $(d-1)$-sphere, where ν_1 is the first coordinate of ν, is $C \int_0^\pi (\sin\varphi)^{d-2} g(\cos\varphi)\,d\varphi$, where $C = 2\pi^{(d-1)/2}/\Gamma((d-1)/2)$. In particular, take $g(c)$ as 1 for $c > \alpha$ and 0 otherwise, and then substitute $t = (\sin\varphi)^2$:

$$\begin{aligned}
\operatorname{Vol}_{d-1} \operatorname{Cap}_\alpha^{d-1} &= C \int_0^{\arccos\alpha} (\sin\varphi)^{d-2}\,d\varphi \\
&= \frac{C}{2} \int_0^{1-\alpha^2} t^{(d-3)/2}(1-t)^{-1/2}\,dt \\
&= \frac{C}{2} \mathcal{B}(1-\alpha^2; (d-1)/2, 1/2).
\end{aligned}$$

This is the first equation; the second equation follows immediately. □

Alternative proof. [64, formula (1)] immediately gives the second equation, and in combination with $\operatorname{Vol}_{d-1} \operatorname{Cap}_0^{d-1} = \pi^{d/2}/\Gamma(d/2)$ from [64, p. 1], $\mathcal{B}(1; a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$, and $\Gamma(1/2) = \pi^{1/2}$ also gives the first equation. □

Theorem 3.18. *Let d be an integer with $d \geq 2$. Let α be a real number with $0 \leq \alpha \leq 1$. Then*

$$\frac{2}{\pi(d-1)}(1-\alpha^2)^{(d-1)/2} \leq \frac{\operatorname{Vol}_{d-1}\operatorname{Cap}_\alpha^{d-1}}{\operatorname{Vol}_{d-1}\operatorname{Cap}_0^{d-1}} \leq (1-\alpha^2)^{(d-1)/2}.$$

Standard techniques give better bounds for large d, but Theorem 3.18 suffices for this paper. For comparison, [8, Lemma 2.1] states that for "arbitrary" α with $0 < \alpha < 1$ the sphere-cap volume ratio is "$\operatorname{poly}(n) \cdot (\sqrt{1-\alpha^2})^n$" (where "$n$" is d) and cites [69, Lemma 4.1] for this; but [69, publisher version, Lemma 4.1; full version, Lemma A.3] states only a lower bound (with an α-dependent $(1-\alpha)/d$ where Theorem 3.18 has $2/\pi(d-1)$), not an upper bound.

Proof. Write $a = (d-1)/2$. The volume ratio is $\mathcal{B}(1-\alpha^2; a, 1/2)/\mathcal{B}(1; a, 1/2)$ by Theorem 3.17. This ratio is between $(1-\alpha^2)^a/(a\mathcal{B}(1; a, 1/2))$ and $(1-\alpha^2)^a$ by Theorem 3.16.

Substitute $t = (\sin\varphi)^2$ to see that $\mathcal{B}(1; a, 1/2) = \int_0^1 t^{a-1}(1-t)^{-1/2}\,dt = 2\int_0^{\pi/2}(\sin\varphi)^{2a-1}\,d\varphi$. The integrand $(\sin\varphi)^{2a-1}$ is at most 1 since $2a-1 = d-2 \geq 0$, so $\mathcal{B}(1; a, 1/2) \leq 2(\pi/2) = \pi$. Hence the volume ratio is between $(1-\alpha^2)^a/\pi a$ and $(1-\alpha^2)^a$ as claimed. □

3.19. Varying the model. One could instead define, e.g., a "probabilistic spherical model" as follows. Define D_j as the distribution of the length of the jth shortest vector modulo negation in a random dimension-d determinant-1 lattice from the invariant distribution. Take the length of $\mu_j/(\det L)^{1/d}$ to have distribution D_j.

Extrapolating from random dimension-d lattices is obviously not the same as extrapolating from dimension-1 lattices; Figure 3.10 shows some variation across lattices for $d = 2$. However, the theorems reviewed in Section 2 show that each D_j has probability $1-o(1)$ of being within a factor $1+o(1)$ of the spherical-model prediction as $d \to \infty$.

One could argue that a probabilistic spherical model is conceptually superior to a deterministic spherical model. On the other hand, one could object that a probabilistic spherical model would not stop μ_2 from sometimes being shorter than μ_1, so it would underestimate the length of the shortest nonzero vectors. A similar difficulty appears if one tries to formalize a ball model.[2] One could define a further model that accounts for the joint distributions: e.g., take the μ_1 length distributed as in a random lattice, then take the μ_2 length distributed as in a random lattice conditioned on the length of μ_1, etc. Perhaps calculations in such a model would be feasible.

All of these models, including the model from Definition 3.3, have the same basic data flow:

- The model is given the dimension and determinant of a lattice.

[2] If μ_j is uniformly distributed in a ball B_j around 0 of volume proportional to j, then it has probability $1/j$ of landing in B_1. The sum $\sum_j 1/j$ diverges, so one expects infinitely many points in B_1, including points arbitrarily close to 0.

- Based on this information, the model predicts vector lengths in a way that matches the behavior of most lattices from the invariant distribution, at least within $1 + o(1)$.
- Each vector μ_j is modeled as pointing in a uniform random direction. These directions are modeled as being statistically independent across j.

Models can vary in the exact choices of lengths, and in the difficulty of precisely computing those lengths. This paper uses Definition 3.3 as a concrete example because that definition makes the exact computations easy, but any model of this type will produce the same results for an analysis that disregards $1 + o(1)$ factors in lengths. This applies, in particular, to the asymptotic length gaps between spherical models and reality for the lattices covered in Table 1.5: for each lattice, the spherical-model prediction is $d^{1/2+o(1)}$ times the actual length, and $d^{1/2+o(1)}$ is the same as $(1 + o(1))d^{1/2+o(1)}$, so predictions in other models of this type are also $d^{1/2+o(1)}$ times the actual length.

4 Spherical model of the integer lattice

As a warmup for S-unit lattices, this section considers the determinant-1 lattice $L = \mathbb{Z}^d$, quantifies the inaccuracy of spherical-model predictions of the lengths of short vectors in L, and quantifies the inaccuracy of spherical-model predictions of the effectiveness of reduction modulo those vectors.

By Theorem 3.9, a shortest nonzero vector μ_1 in a spherical model of L has length $(1 + o(1))(d/2\pi e)^{1/2}$ as $d \to \infty$. As for reduction, if $\alpha \in \mathbb{R}$ and $\nu \in \mathbb{R}^d$ satisfy $0 < \alpha < 1$ and $||\nu||_2 = ||\mu_1||_2/2\alpha$, then the probability that μ_1 reduces ν is the α-cap volume ratio from Theorem 3.13, which, by Theorem 3.18, is $(1 - \alpha^2)^{(1/2+o(1))d}$ as $d \to \infty$ with fixed α. The shortest nonzero vector in a spherical model has, for example, $\exp(-\Theta(d))$ chance of reducing a vector of length $(10 + o(1))d^{1/2}$ as $d \to \infty$. To have a high chance of reduction one would need to try $\exp(\Theta(d))$ short vectors of similar length.

Do these conclusions regarding a spherical model translate into conclusions regarding L? No, not even close:

- The lattice has $2d$ vectors of length 1. This is not length $(1+o(1))(d/2\pi e)^{1/2}$ as $d \to \infty$.
- This list of $2d$ vectors achieves perfect reduction modulo L: iterating the reduction process reduces any lattice point to 0, and reduces a general vector to the box $[-1/2, 1/2]^d$, with length at most $d^{1/2}/2$.

See also [68], where Mazo and Odlyzko show that the number of elements of $\mathbb{Z}^d$ in a ball of radius (e.g.) $0.51 \cdot d^{1/2}$ varies by a factor exponential in d as the ball's center varies, contradicting the idea that the volume predicts the number of lattice points.

The contradictions highlighted in this section are not specific to this paper's definition of a spherical model. If a model

- sees only the determinant and dimension of a lattice,

d	actual	spherical	"random"
1	1.000000	1.000000	1.000000 ± 0.000000
2	1.000000	0.797885	0.704990 ± 0.230396
4	1.000000	0.797885	0.694549 ± 0.177018
8	1.000000	0.915335	0.865216 ± 0.137234
16	1.000000	1.143101	1.099844 ± 0.076381
32	1.000000	1.503494	1.474600 ± 0.059322
64	1.000000	2.039712	
128	1.000000	2.817717	
256	1.000000	3.933095	
512	1.000000	5.522266	
1024	1.000000	7.778923	

Table 4.2. Numerical examples of how inaccurate spherical models are for $\mathbb{Z}^d$. All entries after the first column are rounded to 6 digits after the decimal point. Second column, "actual": minimum length $\lambda_1(\mathbb{Z}^d)$ of nonzero vectors in $\mathbb{Z}^d$. Third column, "spherical": minimum length of nonzero vectors in a spherical model of $\mathbb{Z}^d$. Fourth column, "random", only for $d \leq 32$: average and standard deviation of $\lambda_1(L)$ for 128 "random" dimension-d determinant-1 lattices L; see text for details.

- predicts vector lengths in a way compatible with most lattices from the invariant distribution, and
- predicts the effectiveness of reduction in a way compatible with conjectures and experiments from the literature regarding "random" lattices,

then the model will necessarily produce incorrect predictions for $\mathbb{Z}^d$. Most lattices L with the same dimension and determinant have, compared to $\mathbb{Z}^d$, much larger $\lambda_1(L)$, and (conjecturally) much worse reduction effectiveness.

4.1. Numerical examples. Table 4.2 tabulates, for $d \in \{1, 2, 4, \ldots, 1024\}$, the exact length $(2\pi^{-d/2}(d/2)!)^{1/d}$ of each shortest nonzero vector in a spherical model of $\mathbb{Z}^d$. This length is below $\lambda_1(\mathbb{Z}^d) = 1$ for $d \in \{2, 4, 8\}$, but then rapidly increases.

For comparison, the "random" column in the table shows, for $d \leq 32$, statistics on $\lambda_1(L)$ for 128 lattices L generated as follows: in the Sage mathematics system, ask `sage.crypto.gen_lattice` for a pseudorandom dimension-d determinant-2^{64d} *integer* lattice, and then divide by 2^{64} to obtain a determinant-1 lattice.

The lattice $\mathbb{Z}^d$ is 8 standard deviations away from "random" experiments for $d = 32$. It is not surprising that spherical models are much closer to "random": Clozel–Oh–Ullmo [26, Corollary 1.8] and independently Goldstein–Mayer [50] showed that the distribution of a uniform random dimension-d determinant-D integer lattice, scaled by $D^{1/d}$, converges (in the sense of equidistribution) to the invariant distribution as $D \to \infty$; and the theorems reviewed in Section 2 say that $\lambda_1(L)$ for L from the invariant distribution is usually within $1 + o(1)$ of what a spherical model predicts as $d \to \infty$.

5 S-units and S-unit attacks

This section reviews S-units, the S-unit lattice, and S-unit attacks, specifically for power-of-2 cyclotomic fields. This section starts by reviewing unit attacks for these fields; unit attacks are a special case of S-unit attacks. See also Section 6.4 for an illustrative example, the case $n = 1$.

5.1. Notation. Throughout this section, n is an element of $\{1, 2, 4, 8, 16, \ldots\}$; R is the ring $\mathbb{Z}[x]/(x^n + 1)$; and K is the ring $\mathbb{Q}[x]/(x^n + 1)$. The ring K is a field, since $x^n + 1$ is irreducible (in $\mathbb{Z}[x]$, hence in $\mathbb{Q}[x]$) for these values of n. This section views R as a subring of K, automatically applying the natural injection $R \to K$, whether or not the definitions of quotient rings are set up to ensure that $R \subseteq K$. This paper reserves the letter P for nonzero prime ideals of R.

Define $\zeta_m = \exp(2\pi i/m)$ for each positive integer m. For each odd integer j, there is a unique ring morphism $\sigma_j : K \to \mathbb{C}$ that maps x to ζ_{2n}^j. The usual absolute-value function from $\mathbb{C}$ to $\mathbb{R}$ is written $z \mapsto |z|$. The set of "places" of K (defined below) is written V, and $\mathbb{R}^V$ means the set of V-indexed vectors with entries in $\mathbb{R}$.

5.2. Unit attacks. Say one is trying to find short nonzero elements of an ideal I of R, and has (perhaps via a fast additive attack) found a nonzero element $v \in I$, but v is not as short as desired. A unit attack outputs v/u for some unit u. The point is that v/u is an element of I and is often much shorter than v.

This paper analyzes unit attacks only in the case that I is the principal ideal generated by v; this is the usual assumption in the literature. Given a principal ideal I, one can find a generator v using the Biasse–Song [17] adaptation of the quantum polynomial-time algorithm of Eisenträger–Hallgren–Kitaev–Song [41]; a unit attack then outputs a generator v/u for some unit u. One can think of the generalization in the previous paragraph as applying the principal case to the principal ideal vR.

The remaining question is how exactly to reduce a given v to a shorter v/u. Here one uses a standard number-theoretic logarithm map $\operatorname{Log} : K^* \to \mathbb{R}^V$. If $n \geq 4$ then the set $\operatorname{Log} R^*$ is a lattice of rank $n/2 - 1 \geq 1$ in $\mathbb{R}^V$. Short v/u corresponds to short $\operatorname{Log}(v/u) = \operatorname{Log} v - \operatorname{Log} u$ when shortness of logs is defined appropriately, so the objective is to find a vector $\operatorname{Log} u$ in the unit lattice $\operatorname{Log} R^*$ close to the given $\operatorname{Log} v$.

This might seem circular—the starting problem was to find short nonzero vectors in a lattice, namely I; the objective now is to find close vectors in a lattice, namely $\operatorname{Log} R^*$. What makes this work is that $\operatorname{Log} R^*$ is a special lattice. One can, for example, simply write down $n/2 - 1$ linearly independent short vectors in $\operatorname{Log} R^*$, namely

$$\operatorname{Log}(1 + x + x^{-1}), \operatorname{Log}(1 + x^3 + x^{-3}), \ldots, \operatorname{Log}(1 + x^{n-3} + x^{3-n}).$$

See Section 8 for quantification of how short $\operatorname{Log}(1 + x + x^{-1})$ is, much shorter than predicted by a spherical model of the unit lattice.

Various close-vector algorithms appear in the literature at this point. For this paper, it suffices to consider simple reduction (mentioned in Section 1), where

one tries each $\operatorname{Log} u$ in a list of short lattice vectors; if $\operatorname{Log} v - \operatorname{Log} u$ for some $\operatorname{Log} u$ in the list is shorter than $\operatorname{Log} v$, one replaces v with v/u and repeats. One can take, for example, the list of $\pm\operatorname{Log}(1 + x^j + x^{-j})$ for $j \in \{1, 3, \ldots, n-3\}$, although experiments from Ducas–Plançon–Wesolowski [39] show that taking more vectors produces noticeably better results; see also Section 8.12.

5.3. *S*-unit attacks. More generally, an S-unit attack begins with a nonzero $v \in I$ and outputs v/u, but now u is allowed to range over a larger subset of K^*, specifically the group of S-units.

Here S is a finite set of places, a subset of the set V mentioned above. There are two types of places:

- The "infinite places" are labeled $1, 3, 5, \ldots, n-1$, except that for $n = 1$ there is one infinite place labeled 1. The entry at place j in $\operatorname{Log} \alpha$ is defined as $2 \log |\sigma_j(\alpha)|$, except that the 2 is omitted for $n = 1$. The set of all infinite places is denoted ∞, and is required to be a subset of S.
- For each nonzero prime ideal P of R, there is a "finite place" labeled P. The entry at place P in $\operatorname{Log} \alpha$ is defined as $-(\operatorname{ord}_P \alpha) \log \#(R/P)$, where $\operatorname{ord}_P \alpha$ is the exponent of P in the factorization of α as a product of powers of prime ideals. There are many choices of S here. This paper focuses on the following form of S: choose a parameter y, and take $P \in S$ if and only if $\#(R/P) \leq y$.

The group U_S of S-units of K is, by definition, the set of elements $u \in K^*$ such that the vector $\operatorname{Log} u$ is supported on S, i.e., is 0 at every place outside S. The S-unit lattice is the lattice $\operatorname{Log} U_S$, which has rank $\#S - 1$.

This paper analyzes S-unit attacks only in the case that v is an S-generator of I, meaning that $\operatorname{ord}_P v = \operatorname{ord}_P I$ for every P outside S. Again an S-generator can be found by the [17] adaptation of [41], assuming that I is S-principal, i.e., that an S-generator exists. The output v/u is then also an S-generator of I.

Short v/u again corresponds to short $\operatorname{Log} v - \operatorname{Log} u$, but care is required to ensure that $v/u \in I$, i.e., that $\operatorname{ord}_P(v/u) \geq \operatorname{ord}_P I$ for each finite place P; this was automatic for unit attacks but is not automatic for general S-unit attacks. One thus wants to find a vector $\operatorname{Log} u$ in the S-unit lattice $\operatorname{Log} U_S$ that is close to $\operatorname{Log} v$ in the following sense: $\operatorname{Log} u$ is close to $\operatorname{Log} v$ at the infinite places, and $\operatorname{ord}_P u$ is close to *but no greater than* $\operatorname{ord}_P v - \operatorname{ord}_P I$.

A standard number-theoretic conjecture reviewed in Appendix C provides an easy way to handle this extra requirement, since it guarantees for each P that $P\overline{P}$ is principal; here $\overline{P}$ is the prime ideal that one obtains from P by applying the automorphism of R that maps x to x^{-1}. If $\operatorname{ord}_P(v/u)$ happens to be below $\operatorname{ord}_P I$ then one can simply multiply by a generator of $P\overline{P}$, increasing by 1 the exponents of P and $\overline{P}$, and repeat as necessary. Sometimes P by itself is principal and one can skip $\overline{P}$.

As for closeness, for this paper it again suffices to consider simple reduction, where one tries each $\operatorname{Log} u$ in a list of short lattice vectors. As a preliminary step, if $\operatorname{ord}_P v < \operatorname{ord}_P I$ for some P, update v by multiplying by a generator of $P\overline{P}$ (or, if possible, of P) as explained above, and repeat this step. Then $v \in I$.

Next, if some u in the list has v/u shorter than v and $v/u \in I$, replace v with v/u, and repeat this step. Output the final v.

A full algorithm specification requires considering the possibility that multiple u in the list have v/u shorter than v. For definiteness, check all v/u and take the shortest; if there are ties, take the one that came first in the list of u. Similar comments apply to the order of P for checking $\operatorname{ord}_P v < \operatorname{ord}_P I$.

There is much more to say regarding efficient construction of short S-units, but for this paper it suffices to consider a brute-force precomputation that finds all of the short S-units by enumerating all short elements of K (e.g., elements with bounded numerator and denominator) and checking which elements are S-units. The point of this paper is to show how inaccurate spherical models are in predicting (1) the lengths of short S-units and (2) the effectiveness of reduction modulo short S-units; neither of these analyses depends on the speed of computing the short S-units.

As an extreme case, if $S = \infty$ (the smallest possible choice, not including any P), then $U_S = R^*$: the S-units of K are the units of R, the S-unit lattice is the unit lattice, and S-unit attacks are the same as unit attacks. Extending S to include more and more prime ideals P gives S-unit attacks the ability to modify more and more places in $\operatorname{Log} v$.

6 Spherical model of S-units for $\mathbb{Q}$

This section assumes $n = 1$, takes any number of finite places in S, and quantifies the inaccuracy of spherical-model predictions of (1) the shortness of vectors in the S-unit lattice and (2) the effectiveness of reduction modulo those vectors.

The ring $R = \mathbb{Z}[x]/(x+1)$ and field $K = \mathbb{Q}[x]/(x+1)$ are isomorphic to $\mathbb{Z}$ and $\mathbb{Q}$ respectively; this section automatically applies these isomorphisms. This section assumes that S has the form $\infty \cup \{p\mathbb{Z} : \text{prime } p \le y\}$, where $y \ge 2$.

For example, if y is chosen as 7, then $S = \infty \cup \{2\mathbb{Z}, 3\mathbb{Z}, 5\mathbb{Z}, 7\mathbb{Z}\}$, and the set of S-units is $\pm 2^{\mathbb{Z}} 3^{\mathbb{Z}} 5^{\mathbb{Z}} 7^{\mathbb{Z}}$. More and more primes appear in S as y increases. S-units in $\mathbb{Z}$ are also known as "y-smooth integers": e.g., 7-smooth integers are elements of $\pm 2^{\mathbb{N}} 3^{\mathbb{N}} 5^{\mathbb{N}} 7^{\mathbb{N}}$, where $\mathbb{N} = \{0, 1, 2, \ldots\}$.

6.1. The importance of considering $n = 1$. If a nonzero ideal I is provided as an element $v \in I$ such that $I = v\mathbb{Z}$, then one can simply output v as a short nonzero element, in fact a shortest nonzero element, of I. Why bother studying S-unit attacks for $\mathbb{Z}$?

There are three answers. First, the question addressed in this paper is the comparison between spherical models and reality, *not* the comparison between S-unit attacks and other attacks. Seeing that spherical models underestimate the power of S-unit attacks for $n = 1$ directly answers the $n = 1$ case of this question. The availability of simpler attacks for $n = 1$ has no relevance to this analysis.

Second, focusing on $n = 1$ simplifies the analysis, making it as easy as possible to see that there is a problem with applying spherical models to S-unit lattices.

Given how well established spherical models are, it is important to begin with the simplest possible counterexamples.

Third, various failures of spherical models for $n = 1$ reappear for larger values of n, as subsequent sections demonstrate. This analysis of $n = 1$ thus serves as a stepping-stone towards the analyses for larger n.

6.2. Dimension and determinant of the S-unit lattice. Define L as the S-unit lattice. Define d as the dimension of L; then d is $\#S - 1$, the number of primes $p \leq y$. By the prime-number theorem, $d \in (1 + o(1))y/\log y$ as $y \to \infty$.

One basis for L is the sequence of d rows in the matrix

$$\begin{pmatrix} \log 2 & -\log 2 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots \\ \log 3 & 0 & -\log 3 & 0 & 0 & \cdots & 0 & 0 & \cdots \\ \log 5 & 0 & 0 & -\log 5 & 0 & \cdots & 0 & 0 & \cdots \\ \log 7 & 0 & 0 & 0 & -\log 7 & \cdots & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & 0 & 0 & \cdots \\ \log p_d & 0 & 0 & 0 & 0 & \cdots & -\log p_d & 0 & \cdots \end{pmatrix}$$

where p_d is the largest prime $\leq y$; the first column is for the infinite place, and the remaining columns are for places $2\mathbb{Z}, 3\mathbb{Z}, 5\mathbb{Z}, 7\mathbb{Z}, \ldots$. Suppressing zero columns and rescaling produces the following $d \times (d + 1)$ matrix:

$$\begin{pmatrix} 1 & -1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & -1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & -1 & 0 & \cdots & 0 \\ 1 & 0 & 0 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ 1 & 0 & 0 & 0 & 0 & \cdots & -1 \end{pmatrix}$$

The latter matrix times its transpose is the $d \times d$ matrix with 2 along the diagonal and 1 elsewhere, which has determinant $d + 1 = \#S$, so $\det L = \#S^{1/2} \prod_{p \leq y} \log p$; i.e., $(\det L)^{1/d}$ is $\#S^{1/2d}$ times the geometric average of $\log p$ for the primes $p \leq y$. In particular, $(\det L)^{1/d} \in (1 + o(1)) \log y$ as $y \to \infty$.

6.3. Length of shortest nonzero vectors. The minimum nonzero length in a spherical model of the S-unit lattice is $(1/2\pi e + o(1))^{1/2}(y \log y)^{1/2}$ as $y \to \infty$ by Theorem 3.9. Here are two ways to see that this is an artifact of the model:

- Qualitative: If it were true that $\lambda_1(L) \in (1/2\pi e + o(1))^{1/2}(y \log y)^{1/2}$ for the S-unit lattice L then a sufficient increase in y would be forced to make the shortest S-unit *larger*. This makes no sense: increasing y does not remove preexisting S-units.
- Quantitative: $\pm(\log 2, -\log 2, 0, 0, 0, \ldots)$ are the shortest nonzero vectors in the S-unit lattice. These have constant length as $y \to \infty$, far below length $(1/2\pi e + o(1))^{1/2}(y \log y)^{1/2}$.

Note that there are many further S-units in $\mathbb{Z}$ (never mind $\mathbb{Q}$) having lengths far below $(1/2\pi e + o(1))^{1/2}(y \log y)^{1/2}$ as $y \to \infty$: one can take further primes

beyond 2 and further products of primes, appearing with a density guaranteed by, e.g., the Canfield–Erdős–Pomerance theorem [23, Theorem 3.1].

6.4. Specializing S-unit attacks to $n = 1$. All ideals of $\mathbb{Z}$ are principal. Define g as the unique positive generator of I. Here are the steps of an S-unit attack from Section 5, with simple reduction, along with the specialization of these steps to $n = 1$ and $S = \infty \cup \{p\mathbb{Z} : \text{prime } p \le y\}$:

- Begin with an S-generator v of I. Specialization: Begin with a nonzero integer v such that $\operatorname{ord}_p v = \operatorname{ord}_p I$ for all primes $p > y$. Subsequent adjustments of v will preserve this property; the factorization of $v/g \in \mathbb{Q}$ always involves only primes $p \le y$.
- If $v \notin I$, then find some P with $\operatorname{ord}_P v < \operatorname{ord}_P I$, multiply by some generator of P or $P\overline{P}$, and repeat this step. Specialization: If v/g has a denominator, then find a prime number p in the denominator, replace v with vp, and repeat this step.
- Now $v \in I$. If v/u is shorter than v and $v/u \in I$ for some u in a list of short S-units, replace v with v/u, and repeat this step.
- Output v.

Regarding the list of short S-units, a brute-force computation of all elements of the S-unit lattice having (say) 1-norm at most $2 \log y$ will find, in particular, the d basis vectors listed above, $\operatorname{Log} p$ for each prime number $p \le y$. These d vectors are all that matters for the analysis of effectiveness below. Note that one can view the handling of denominators as using the negatives of these vectors.

If $y \ge 3$ then the same brute-force computation finds more S-units. Any $u = \prod_{p \le y} p^{e_p}$ where $||\operatorname{Log} u||_1 \le 2 \log y$ must have $\sum_{p \le y} |e_p \log p| \le 2 \log y$, limiting the numerator and denominator of u to y^2 and thus limiting u to $y^{O(1)}$ possibilities. One can optimize the attack to take only the basis vectors and eliminate further S-units, but the comparison to a spherical model will simply use the fact that the list size is $y^{O(1)}$.

Regarding effectiveness, it is easy to see that the attack is perfectly effective: the output must be $\pm g$, a shortest nonzero element of I. Indeed, (1) an algorithm invariant is that no primes $p > y$ appear in the factorization of v/g; (2) no prime $p \le y$ can appear in the final denominator of v/g, since if it did then v would have been replaced with vp; and (3) no prime $p \le y$ can appear in the final numerator of v/g, since if it did then v would have been replaced with v/p. Hence v/g is a unit, either 1 or -1.

6.5. Spherical model of reduction effectiveness. The spherical model necessarily draws the same conclusions for the S-unit lattice that it drew for $\mathbb{Z}^d$ in Section 4, with lengths scaled according to $\det L$. For example, a shortest nonzero vector in a spherical model has chance $\exp(-\Theta(d)) = \exp(-\Theta(y/\log y))$ of reducing a vector of length $(10 + o(1))(y \log y)^{1/2}$ as $y \to \infty$.

Again this has the astonishing qualitative feature of claiming that increasing y, using more S-units, makes S-unit attacks less effective. The claim is wrong. Quantitatively, the analysis of reduction effectiveness in Section 6.4 shows that

taking $y^{O(1)} = \exp(O(\log y))$ short vectors suffices for perfect reduction of all lattice points to 0. The effectiveness of these vectors at reducing non-lattice points does not arise in the algorithm analysis for $n = 1$: all ideals are principal.

Note that the spherical-model calculations in this paper, as in the literature, are calculations regarding 2-norms in log space. Shortness is measured differently inside S-unit attacks. Write v/g as $\pm \prod_p p^{e_p}$, and assume all $e_p \geq 0$; closeness of v to I is measured by $|v/g|$, or equivalently by $\sum_p e_p \log p$, which is a 1-norm in log space rather than a 2-norm. However, an attack modified to reduce $||\mathrm{Log}(v/g)||_2$ would still be perfectly effective: if $e_p > 0$ then $||\mathrm{Log}(v/gp)||_2 < ||\mathrm{Log}(v/g)||_2$. In the opposite direction, perhaps a spherical-model calculation regarding 1-norm reduction would produce noticeably different results from 2-norm reduction, but the point of this paper is to study the application to the S-unit lattice of the spherical-model methodology from the literature. This methodology consistently uses 2-norms.

7 An S-reversal phenomenon for every field

The analysis of S-unit attacks for $n = 1$ in Section 6 explicitly relies on the fact that all ideals are principal. The comparison to a spherical model of reduction effectiveness also relies on this fact.

Large cyclotomic fields (and, more generally, large "CM fields") have large class numbers: ideals are rarely principal. (Readers not familiar with class groups should begin with Appendix B.) For example, standard calculations show that the class number of $\mathbb{Q}[x]/(x^n + 1)$ is a multiple of 359057 for $n = 64$, a multiple of 10449592865393414737 for $n = 128$, and a multiple of

$$626250398449093235874572148252892284197821938997560 5329$$

for $n = 256$. More difficult calculations surveyed in Appendix C have proven, assuming the generalized Riemann hypothesis (GRH) in the case $n = 256$, that these are the exact class numbers, but in any case it is clear that one cannot expect ideals to be principal when n is large.

It is natural to wonder whether this increase in class numbers could rescue the applicability of a spherical model to S-unit attacks. What this section shows is that the answer is no.

Fix $n \in \{1, 2, 4, 8, 16, \ldots\}$. Define $R = \mathbb{Z}[x]/(x^n + 1)$ and $K = \mathbb{Q}[x]/(x^n + 1)$. Take $S = \infty \cup \{P : \#(R/P) \leq y\}$. This section shows that a spherical model produces two absurd conclusions regarding S-unit attacks:

- The shortest nonzero S-unit becomes longer and longer as $y \to \infty$.
- The success probability of reduction modulo short S-units converges to 0 as $y \to \infty$.

These statements already appeared in Section 6 for $n = 1$; this section generalizes to every n, in particular showing that the statements are not limited to small class groups.

This section is not specific to power-of-2 cyclotomics. Number-theorist readers should feel free to select any number field K for this section, with $R = \mathcal{O}_K$.

7.1. Length of the shortest nonzero vectors. Consider a spherical model M of the S-unit lattice L. By Landau's prime-ideal theorem [58, §5], the number of prime ideals P with $\#(R/P) \le y$ is $(1+o(1))y/\log y$ as $y \to \infty$, so the dimension d of L is $(1+o(1))y/\log y$. The geometric average of $\log \#(R/P)$ is $(1+o(1))\log y$, so $(\det L)^{1/d} \in (1+o(1))\log y$.

The shortest nonzero vectors in M have length $(1+o(1))(d/2\pi e)^{1/2}(\det L)^{1/d}$ by Theorem 3.9, i.e., $(1/2\pi e + o(1))^{1/2}(y\log y)^{1/2}$, generalizing what Section 6 said for $K = \mathbb{Q}$. In particular, this length converges to ∞ as $y \to \infty$.

The calculation here is exactly as in Section 6. Aside from $1+o(1)$ factors, the dimension and root determinant are the same functions of y for every number field; a spherical model sees no other information regarding the lattice; so it is unsurprising that these conclusions are identical across number fields K.

For comparison, as S increases, the S-unit lattice expands to include logs of more and more elements of K^*. Pick any $u \in K^*$ with $\operatorname{Log} u \neq 0$; e.g., $u = 2$. Take any y large enough to have S include all P with $\operatorname{ord}_P u \neq 0$. Then u is an S-unit, so the shortest nonzero vectors in the S-unit lattice have length at most the length of $\operatorname{Log} u$. It is therefore not true that the minimum nonzero length of vectors in the S-unit lattice converges to ∞ as $y \to \infty$.

7.2. Effectiveness of reduction. Take T as ∞ and a finite set of nonzero prime ideals generating the class group. One can bound $\#T$ and the ideal norms as functions of n using, e.g., Bach's theorem [4, Theorem 4] under GRH, or Zimmert's theorem [93] without GRH, but the exact size of T does not matter for this section.

Find a T-generator v of the input ideal I. Take $y \ge \max\{\#(R/P) : P \in T\}$. Consider an S-unit attack that starts from v, takes $S = \infty \cup \{P : \#(R/P) \le y\}$, and reduces v modulo the y shortest vectors in the S-unit lattice. Shortness here can be defined as, e.g., 1-norm or 2-norm; the choice is irrelevant to the following analysis. It also does not matter how one breaks ties in deciding which y vectors are shortest.

Consider the shortest nonzero elements of I. Does the S-unit attack find one of these elements, say α? For small y, perhaps not, but for all sufficiently large y the answer is yes, even if the number of reduction steps is limited to 1: the ratio v/α is an S-unit once y is sufficiently large, and is one of the y shortest S-units once y is sufficiently large, so the reduction tries dividing v by this S-unit v/α and sees α as desired.

Now consider a spherical model of the S-unit lattice.[3] A shortest nonzero vector in a spherical model has, exactly as in Section 6, chance $\exp(-\Theta(d)) = \exp(-\Theta(y/\log y))$ of reducing a vector of length, e.g., $(10+o(1))(y\log y)^{1/2}$. Trying y short vectors improves the chance by a factor at most y, producing overall success probability $\exp(-\Theta(y/\log y))$ of reducing a vector of this length. A longer starting vector would have higher probability of being reduced, but

[3] The calculation in this paragraph is consistent with the very recent calculation in [40]. The big difference is that [40] assumes that the calculation says something about S-unit attacks, whereas this section—as the result of analyzing how S-unit attacks actually behave—shows the opposite.

there is no hope of reducing it to length $o((y \log y)^{1/2})$, or, more extreme, to the length of α, which is constant as $y \to \infty$.

The spherical model thus predicts that increasing y—allowing more and more places in S, and allowing reduction modulo more and more S-units—somehow makes S-unit attacks less and less effective: the output length grows with y, and the chance of finding a shortest nonzero vector in the ideal converges to 0 as $y \to \infty$. In fact, the success probability is 1 for all sufficiently large y.

The same comments apply if one expands the range of targets $\alpha \in I$ to allow any short nonzero vector, rather than considering the extreme case of shortest nonzero vectors. A short nonzero vector—with any quantification of "short" for which such a vector exists—will be found by S-unit attacks for all sufficiently large y. If "short" is not very large then a spherical model incorrectly predicts that the success probability converges to 0 as $y \to \infty$.

7.3. A reason for the reversal. Say one takes a random point in a large box inside the $\mathbb{R}$-vector space $L_{\mathbb{R}}$ generated by the S-unit lattice L, and asks how effectively this point can be reduced modulo *all* S-units, i.e., how close this point is to the S-unit lattice, without regard to the cost of finding a closest lattice vector. Here are two different interpretations of this question, using two different probability spaces for the random point:

- The coordinate at each place is a uniform random real number in the specified interval. Coordinates are statistically independent across places.
- The coordinate at each infinite place is a uniform random real number in the specified interval, and the coordinate at each finite place P is a uniform random multiple of $\log \#(R/P)$ within the specified interval. Coordinates are statistically independent across places.

It is easy to see that these interpretations of the question have very different answers:

- With the first interpretation, reducing this point modulo the S-unit lattice at best reduces place P to the range $[-0.5, 0.5] \log \#(R/P)$. This is on average $(0.25 + o(1)) \log \#(R/P)$ in absolute value, assuming $\log \#(R/P) \in o(W)$ where W is the box width. As y increases, there are more and more—and larger and larger—contributions to the length of the reduced vector.
- With the second interpretation, take T to generate the class group as in Section 7.2, and assume that $T \subseteq S$. One can always reduce modulo S-units to clear all places outside T. One thus obtains a y-independent bound on the output length—never mind the possibility of reducing even more as y increases.

This difference is important for the analysis of S-unit attacks. One starts with an S-generator v of the input ideal. The entry at place P in $\operatorname{Log} v$ is a multiple of $\log \#(R/P)$, as in the second interpretation. If one disregards this information and instead models $\operatorname{Log} v$ as a point *somewhere* in log space, as in the first interpretation, then the point will almost always be much farther from the S-unit lattice. This gap is already visible for small S, and was noted for one choice

of small S by Ducas–Plançon–Wesolowski [39, Section 6.1, "adjusting"], but merely looking at one choice of S does not make clear how broken the model is; it is important to understand that the gap grows without bound as S increases.

For comparison, a spherical model is invariant under rotation, so the point being reduced is equivalent to a uniformly distributed point ν on a sphere. The distribution of ν is not exactly the box distribution from the first interpretation above, but it is close enough to produce similar conclusions. Indeed, write ν as $(\nu_1, \ldots, \nu_d)$ with $\nu_1^2 + \cdots + \nu_d^2 = r^2$. Aside from scaling, the set of ν having $\nu_1 > \alpha$ is the spherical cap from Definition 3.12, and the cumulative distribution function of ν_1 is the $\mathcal{B}$ ratio in Theorem 3.17(2). What matters here is simply that ν_1 has a smooth enough distribution that, as $r \to \infty$, the distribution of ν_1 modulo 1 converges to the uniform distribution. Similar comments apply to a coordinate at place P modulo $\log \#(R/P)$. This leads to the incorrect conclusion that almost all reduced points are farther and farther from the S-unit lattice as $y \to \infty$. This conclusion relies critically on the inaccurate model of $\operatorname{Log} v$ as having essentially arbitrary real coordinates, a model that fails to account for the number-theoretic structure of $\operatorname{Log} K^*$.

8 Spherical model of units for power-of-2 cyclotomics

The fact that spherical models of S-unit lattices become less and less accurate as y increases is sufficient reason to discard spherical models in the context of S-unit attacks that take y as large as necessary to optimize performance. However, the literature also considers S-unit attacks with small y. It is thus natural to ask whether spherical models are accurate for small y.

This section takes the smallest possible y: namely, $y = 1$. Then $S = \infty$, and the S-unit lattice is just the unit lattice. This section computes—assuming a standard number-theoretic conjecture; see below—what a spherical model of the unit lattice says regarding (1) short vectors and (2) reduction effectiveness. This section shows that these conclusions are disproven for various n by (1) calculations of the log of a cyclotomic unit and (2) statistical experiments with reduction.

The vector-length calculations in this section, unlike Section 7, are specific to power-of-2 cyclotomics. Many standard results regarding cyclotomic fields are used below. To help the reader check the material here against the cited references, this section expresses results in terms of $\mathbb{Q}(\zeta_m)$ and $\mathbb{Z}[\zeta_m]$ for $m = 2n$:

- Start with $m \in \{8, 16, 32, \ldots\}$. (The unit lattice has rank 0 for smaller m.)
- Consider the number field $K = \mathbb{Q}(\zeta_m)$. This is isomorphic to $\mathbb{Q}[x]/(x^n + 1)$ where $n = m/2$.
- The ring $\mathcal{O}_K$ is $R = \mathbb{Z}[\zeta_m]$. This is isomorphic to $\mathbb{Z}[x]/(x^n + 1)$.

Presumably one would be able to cover more general cyclotomics with some extra work, but many numbers would change depending on the choice of field.

m	n	$\mathrm{Reg}_K^+/(n/4)^{n/4}$	spherical model	actual length	ratio
8	4	0.881374	2.492901	2.492901	1.000000
16	8	0.610449	2.652102	3.766835	0.704066
32	16	0.480772	4.081293	5.673348	0.719380
64	32	0.384226	6.967780	8.189221	0.850848
128	64	0.393293	12.735518	11.719983	1.086650
256	128	0.286233	23.862591	16.663464	1.432031
512	256	0.200698	45.953088	23.631207	1.944593
1024	512	0.202244	90.089629	33.464774	2.692073
2048	1024	0.192272	178.014429	47.358628	3.758860
4096	2048	0.199056	353.577209	66.997907	5.277437

Table 8.3. Numerical examples of how inaccurate spherical models are for unit lattices for power-of-2 cyclotomic fields, assuming $h^+ = 1$. All entries after the first two columns are rounded to 6 digits after the decimal point. First column: conductor m of field $K = \mathbb{Q}(\zeta_m)$. Second column: $n = m/2$, the degree of K. Third column: the regulator Reg_K^+ of K^+ divided by $(n/4)^{n/4}$. Fourth column, "spherical model": minimum nonzero length in spherical model of the unit lattice of K. Fifth column, "actual length": length of a nonzero vector in the unit lattice of K, namely $\mathrm{Log}(1+\zeta_m+1/\zeta_m)$. Sixth column, "ratio": fourth column divided by fifth column. Compare Table 4.2.

8.1. A short nonzero vector in the unit lattice. Let $m \geq 8$ be a power of 2. Write $u = 1+\zeta_m+\zeta_m^{-1}$. To see that $u \in \mathbb{Z}[\zeta_m]^*$, multiply by $1+\zeta_m^3+\zeta_m^6+\cdots+\zeta_m^{2(m^2-1)}$ and use $\zeta_m^{m/2} = -1$. Now $\mathrm{Log}\, u$ is a nonzero vector in the unit lattice; Theorem 8.2 calculates the length of this vector.

Theorem 8.2. *Let $m \geq 8$ be a power of 2. Define $K = \mathbb{Q}(\zeta_m)$ and $u = 1+\zeta_m+\zeta_m^{-1}$. Then*

$$||\mathrm{Log}\, u||_2 = \left(\sum_{j \in \{1,3,\ldots,m/2-1\}} (2\log|1+2\cos(2\pi j/m)|)^2\right)^{1/2}.$$

Proof. Write $n = m/2$. Then $\mathrm{Log}\, u$ is 0 at the finite places of K, and

$$\left(2\log\left|1+\zeta_m+\zeta_m^{-1}\right|, 2\log\left|1+\zeta_m^3+\zeta_m^{-3}\right|, \ldots, 2\log\left|1+\zeta_m^{n-1}+\zeta_m^{1-n}\right|\right)$$

at the infinite places of K. By definition $\zeta_m = \exp(2\pi i/m)$ so $\zeta_m^j + \zeta_m^{-j} = \exp(2\pi ij/m) + \exp(-2\pi ij/m) = 2\cos(2\pi j/m)$. □

The "actual length" column in Table 8.3 tabulates $||\mathrm{Log}\, u||_2$ for $m \leq 4096$, using Theorem 8.2. As a spot-check, for $m = 16$, $\mathrm{Log}(1+x+x^{-1})$ is, at the infinite places,

$$(2\log\left|1+\zeta_m+\zeta_m^{-1}\right|, 2\log\left|1+\zeta_m^3+\zeta_m^{-3}\right|, 2\log\left|1+\zeta_m^5+\zeta_m^{-5}\right|, 2\log\left|1+\zeta_m^7+\zeta_m^{-7}\right|),$$

around $(2.093065, 1.136717, -2.899464, -0.330318)$. The sum of the squares of these entries is $\approx 3.766835^2$, matching the $m = 16$ row in the table.

The table entries are approximately

$$0.881, 0.942, 1.003, 1.024, 1.036, 1.041, 1.044, 1.046, 1.046, 1.047$$

times $m^{1/2}$. For comparison, $\int_0^{1/2} (2\log|1+2\cos 2\pi t|)^2\,dt \approx 2(1.047)^2$. To see the connection, approximate the integral by the centered Riemann sum

$$\frac{2}{m} \sum_{t\in\{1/m,3/m,\ldots,(m/2-1)/m\}} (2\log|1+2\cos 2\pi t|)^2$$
$$= \frac{2}{m} \sum_{j\in\{1,3,\ldots,m/2-1\}} (2\log|1+2\cos(2\pi j/m)|)^2 .$$

The following theorems formalize this connection precisely enough to show that $||\mathrm{Log}\, u||_2 \in m^{1/2+o(1)}$. The easy part is the lower bound, Theorem 8.4. The upper bound, Theorem 8.5, requires some care, since the integrand is ∞ at $t = 1/3$. Splitting the range of summation in each theorem into more pieces would produce tighter bounds.

Theorem 8.4. *Let $m \geq 16$ be a power of 2. Define $K = \mathbb{Q}(\zeta_m)$ and $u = 1+\zeta_m+\zeta_m^{-1}$. Then $||\mathrm{Log}\, u||_2 \geq m^{1/2}/3$.*

Proof. The quantity $\cos 2\pi t$ decreases from 1 to $\cos(\pi/4) = 1/2^{1/2}$ as t increases from 0 to 1/8, so $1+2\cos 2\pi t$ decreases from 3 to $1+2^{1/2}$, so $2\log|1+2\cos 2\pi t|$ decreases from $2\log 3$ to $2\log(1+2^{1/2})$.

There are $m/16$ choices of $j \in \{1,3,\ldots,m/8-1\}$. Write $t = j/m$; then $0 < t < 1/8$, so $2\log|1+2\cos(2\pi j/m)| = 2\log|1+2\cos 2\pi t| > 2\log(1+2^{1/2}) > 4/3$. Hence

$$\sum_{j\in\{1,3,\ldots,m/8-1\}} (2\log|1+2\cos(2\pi j/m)|)^2 \geq \frac{m}{16}\left(\frac{4}{3}\right)^2 = \frac{m}{9}.$$

Use Theorem 8.2: $||\mathrm{Log}\, u||_2^2 = \sum_{j\in\{1,3,\ldots,m/2-1\}} (2\log|1+2\cos(2\pi j/m)|)^2 \geq \sum_{j\in\{1,3,\ldots,m/8-1\}} (2\log|1+2\cos(2\pi j/m)|)^2 \geq m/9$. □

Theorem 8.5. *Let $m \geq 16$ be a power of 2. Define $K = \mathbb{Q}(\zeta_m)$ and $u = 1+\zeta_m+\zeta_m^{-1}$. Then $||\mathrm{Log}\, u||_2 \leq m^{1/2}\log m$.*

Proof. Abbreviate j/m as t. This proof partitions the range of j in the sum in Theorem 8.2, namely the set $\{1,3,5,\ldots,m/2-1\}$, into three parts, and shows for each part that $\log|1+2\cos(2\pi j/m)| = \log|1+2\cos 2\pi t|$ is between $-\log m$ and $\log m$. The sum of $(2\log|1+2\cos 2\pi t|)^2$ over all $m/4$ values of j is thus at most $m(\log m)^2$ as claimed.

The parts are defined as follows. Write X for the largest odd integer $\leq m/3$, and Y for the smallest odd integer $> m/3$. If m is a power of 4 then $m \in 4+6\mathbb{Z}$, $X = (m-1)/3$, and $Y = X+2 = (m+5)/3$; otherwise $m \in 2+6\mathbb{Z}$, $X = (m-5)/3$, and $Y = X+2 = (m+1)/3$. In both cases $X \leq (m-1)/3$ and $Y \geq (m+1)/3$.

The first part is $\{1,3,5,\ldots,m/4-1\}$; the second part is $\{m/4+1,\ldots,X\}$; the third part is $\{Y,\ldots,m/2-1\}$.

Part 1: $j \in \{1,3,5,\ldots,m/4-1\}$. Here $0 < 2\pi t < \pi/2$, so $0 < \cos 2\pi t < 1$, so $1 < 1+2\cos 2\pi t < 3$, so $0 < \log|1+2\cos 2\pi t| < \log 3 < \log m$ as desired.

Part 2: $j \in \{m/4+1, m/4+3,\ldots,X\}$. Then $1/4 < t \le X/m \le (1-1/m)/3$ so $\pi/2 < 2\pi t \le 2\pi(1-1/m)/3 < 2\pi/3$.

The derivative of $\cos x$ is $-\sin x$, which increases from -1 to $-3^{1/2}/2$ as x increases from $\pi/2$ to $2\pi/3$. This implies that $\cos x$ is decreasing for x between $\pi/2$ and $2\pi/3$, so $0 > \cos 2\pi t \ge \cos(2\pi(1-1/m)/3)$. This also implies that $\cos(2\pi(1-1/m)/3) \ge \cos(2\pi/3) + (3^{1/2}/2)2\pi/3m = -1/2 + \pi/3^{1/2}m$. Hence $1 > 1+2\cos 2\pi t \ge 2\pi/3^{1/2}m$, so $0 > \log|1+2\cos 2\pi t| \ge \log(2\pi/3^{1/2}m)$. This implies $\big|\log|1+2\cos 2\pi t|\big| \le \log(3^{1/2}m/2\pi) < \log m$ as desired.

Part 3: $j \in \{Y,\ldots,m/2-1\}$. Then $(1+1/m)/3 \le Y/m \le t < 1/2$ so $2\pi/3 < 2\pi(1+1/m)/3 \le 2\pi t < \pi$. Again cos is decreasing, so $\cos(2\pi(1+1/m)/3) \ge \cos 2\pi t > -1$.

We have $2\pi/3 < 2\pi(1+1/m)/3 \le 2\pi(17/16)/3$ since $m \ge 16$. On this interval, $-\sin x$ increases up to $-\sin(2\pi(17/16)/3) < -3/4$, so $\cos(2\pi(1+1/m)/3) \le \cos(2\pi/3) - (3/4)(2\pi/3m) = -1/2 - \pi/2m$, so $-1/2-\pi/2m \ge \cos 2\pi t > -1$, so $\pi/m \le |1+2\cos 2\pi t| < 1$, so $\log(\pi/m) \le \log|1+2\cos 2\pi t| < 0$. Hence $\big|\log|1+2\cos 2\pi t|\big| \le \log(m/\pi) < \log m$ as desired. □

8.6. A basis for the unit lattice. Again assume that $m \ge 8$ is a power of 2, and write $n = m/2$. The standard number-theoretic conjecture mentioned above says that $h_m^+ = 1$. Here h_m^+ is by definition the class number of the field $K^+ = \mathbb{Q}(\zeta_m + \zeta_m^{-1}) = \mathbb{R} \cap \mathbb{Q}(\zeta_m)$. Appendix C summarizes the existing evidence for this conjecture—including proofs for all $m \le 256$, a proof under GRH for $m = 512$, constraints on any primes dividing h_m^+, and class-group heuristics suggesting that h_m^+ should be small.

The unit lattice $L = \operatorname{Log} R^*$ has rank $d = n/2 - 1$. Theorem 8.7 reviews a textbook basis for the logs of the "cyclotomic units"; if $h_m^+ = 1$ then this is a basis for L.

Theorem 8.7. *Let $m \ge 8$ be a power of 2. Define $K = \mathbb{Q}(\zeta_m)$; $R = \mathbb{Z}[\zeta_m]$; and $L = \operatorname{Log} R^*$. Then*

$$\operatorname{Log}(1+\zeta_m+\zeta_m^{-1}), \operatorname{Log}(1+\zeta_m^3+\zeta_m^{-3}), \ldots, \operatorname{Log}(1+\zeta_m^{m/2-3}+\zeta_m^{3-m/2})$$

are linearly independent vectors in L, and are a basis for L if $h_m^+ = 1$.

Proof. Write $n = m/2$, $K^+ = \mathbb{R} \cap K$, and $U^+ = \mathcal{O}_{K^+}^*$. In this proof, Log is always the K logarithm map, not the K^+ logarithm map.

Define $C_m^+ = U^+ \cap (-1)^{\mathbb{Z}} \zeta_m^{\mathbb{Z}} (1-\zeta_m)^{\mathbb{Z}} (1-\zeta_m^2)^{\mathbb{Z}} \cdots (1-\zeta_m^{m-1})^{\mathbb{Z}}$. This is the group of "cyclotomic units" of K^+. See, e.g., [89, Section 8.1].

Since m is a power of 2, the group C_m^+ is generated by -1 and the conjugates of $1+\zeta_m+\zeta_m^{-1}$. To see this, take $g = -3$ in [89, Remark after Proposition 8.11].

These conjugates are $1+\zeta_m^j+\zeta_m^{-j}$ for $j \in \{1,3,5,\ldots,n-1\}$. Hence $\operatorname{Log} C_m^+$ is generated by $b_1, b_3, b_5, \ldots, b_{n-1}$, where $b_j = \operatorname{Log}(1+\zeta_m^j+\zeta_m^{-j})$. The sum

$b_1 + b_3 + b_5 + \cdots + b_{n-1}$ is 0 (since the product of conjugates is ± 1), so $\operatorname{Log} C_m^+$ is generated by $b_1, b_3, b_5, \ldots, b_{n-3}$, omitting b_{n-1}.

Since m is a prime power, C_m^+ has index h_m^+ inside U^+ by a theorem of Kummer. See, e.g., [89, Theorem 8.2]. The lattice $\operatorname{Log} C_m^+$ thus has finite index inside $\operatorname{Log} U^+$, and the index is 1 if $h_m^+ = 1$.

Since m is a prime power, $R^* = (-1)^{\mathbb{Z}} \zeta_m^{\mathbb{Z}} U^+$. See, e.g., [89, Corollary 4.13], with the notation of [89, Theorem 4.12]. Consequently $L = \operatorname{Log} R^* = \operatorname{Log} U^+$.

By Dirichlet's unit theorem, L is a lattice of rank $n/2-1$. Hence $\operatorname{Log} C_m^+$, as a sublattice of finite index, also has rank $n/2-1$. This sublattice is generated by b_j for the $n/2-1$ values $j \in \{1, 3, 5, \ldots, n-3\}$, so these b_j are a basis for $\operatorname{Log} C_m^+$. In particular, these b_j are linearly independent as claimed; and if $h_m^+ = 1$ then $\operatorname{Log} C_m^+ = L$ so these b_j are a basis for L as claimed. □

8.8. Computing the determinant of the unit lattice. Assume $h_m^+ = 1$. From the above description of a basis, one can easily calculate $\det L$ to any desired precision for any given m.

Write $d = n/2-1$. Note that L is not a full-rank lattice, even if one suppresses all the zeros at finite places: there are $d+1 = n/2$ infinite places, and L has rank only d, so one cannot simply compute a determinant of the basis matrix B. One can, however, compute $\det L$ as the square root of the determinant of the $d \times d$ product of B and its transpose.

More traditional is to compute the "regulator" Reg_K—by definition this is the absolute determinant of a $d \times d$ matrix obtained from B by removing some coordinate of $\mathbb{R}^{d+1}$—at which point $\det L = (n/2)^{1/2} \operatorname{Reg}_K$.

A faster formula for prime-power cyclotomics says that the regulator Reg_K^+ of the group C_m^+ in the proof of Theorem 8.7 is $\pm \prod_\chi \sum_a \chi(a) \log|1 - \zeta_m^a|$, where χ runs through nontrivial characters of $(\mathbb{Z}/m)^*$ and a runs through integers between 1 and $m/2$ coprime to m. See, e.g., [89, page 145]. One then has $\operatorname{Reg}_K = 2^{n/2-1} \operatorname{Reg}_K^+$ by [89, Lemma 4.16] and [89, Corollary 4.13] since m is a power of 2, and again $\det L = (n/2)^{1/2} \operatorname{Reg}_K$. The regulator examples in Table 8.3 were computed in this way. The examples for $m \leq 128$ were double-checked in another way using bounds [51, Corollary 2.4] from Grenié–Molteni, assuming GRH.

8.9. Minimum length of nonzero vectors in a spherical model of the unit lattice. Again assume $h_m^+ = 1$. The following calculations quantify the minimum length of nonzero vectors in a spherical model M of the unit lattice $L = \operatorname{Log} R^*$. The calculations begin with exact calculations for various small values of m, and conclude with Theorem 8.10, which shows that the minimum length of nonzero vectors in M grows as $m^{1+o(1)}$, much larger than the $m^{1/2+o(1)}$ length of $\operatorname{Log} u \in L$ calculated in Section 8.1.

The shortest nonzero elements of M have length $(2\pi^{-d/2}(d/2)! \det L)^{1/d}$ by Theorem 3.6. The "spherical model" column in Table 8.3 tabulates these lengths. The whole table took just a few minutes of computation on one laptop core, including high-precision interval arithmetic to protect against rounding errors. Appendix F presents some spot-checks.

Within the table, the spherical-model predictions are below $||\operatorname{Log} u||_2$ for $m \in \{16, 32, 64\}$, but above $||\operatorname{Log} u||_2$ for $m \geq 128$, with a ratio growing with m.

Beyond the table, the following back-of-the-envelope calculations suggest that the spherical-model length grows as $\Theta(m)$: if $\operatorname{Reg}_K^+ \approx (n/4)^{n/4}$ then $\operatorname{Reg}_K \approx n^{n/4}$, so $\det L \approx n^{n/4}$, so $(\det L)^{1/d} \approx n^{1/2}$, so the length is approximately $(n/4\pi e)^{1/2} n^{1/2} = n/(4\pi e)^{1/2}$.

The following theorem makes the same calculations sufficiently precise to show that the spherical-model length grows as $m^{1+o(1)}$, and thus that the ratio grows as $m^{1/2+o(1)}$, under the assumption $h_m^+ \in m^{o(m)}$, which of course follows from the assumption $h_m^+ = 1$. One can use standard techniques (see generally [89, Chapter 11]) to prove tighter bounds than $m^{1+o(1)}$, but this theorem suffices for demonstrating the asymptotic inaccuracy of spherical models of the unit lattice, on top of Table 8.3 demonstrating (with higher precision) the inaccuracy of spherical models for concrete sizes.

Theorem 8.10. *Let $m \geq 8$ be a power of 2. Define $K = \mathbb{Q}(\zeta_m)$; $R = \mathbb{Z}[\zeta_m]$; and $L = \operatorname{Log} R^*$. Let M be a spherical model of L. Assume that $h_m^+ \in m^{o(m)}$ as $m \to \infty$. Then $\min\{||x||_2 : x \in M - \{0\}\} \in m^{1+o(1)}$.*

Proof. Define $K^+ = \mathbb{R} \cap K$. Write Reg_K as the regulator of K, and Reg_K^+ as the regulator of K^+. Also write $n = m/2$ and $d = n/2 - 1$.

Write Δ for the absolute value of the discriminant of K. Then $\Delta = (m/2)^{m/2}$ by [89, Proposition 2.1].

Write Δ^+ for the absolute value of the discriminant of K^+. Then $\Delta^+ = (m/2)^{m/4}/2$ by [61, Theorem 3.8].

By the Brauer–Siegel theorem, $\log h_m^+ \operatorname{Reg}_K^+ \in (1/2 + o(1)) \log \Delta^+$. See, e.g., [89, page 44, bottom paragraph]. The notation for Δ^+ in [89] is "d_n^+" where "n" in context is m; also, [89] says $o(\log \Delta)$ instead of $o(\log \Delta^+)$, but this is equivalent.

Note that $\log \Delta^+ = (m/4) \log(m/2) - \log 2 \in (1/4 + o(1)) m \log m$, and by assumption $\log h_m^+ \in o(m \log m)$, so $\log \operatorname{Reg}_K^+ \in (1/8 + o(1)) m \log m$.

Next $\operatorname{Reg}_K = 2^{n/2-1} \operatorname{Reg}_K^+$ (by [89, Lemma 4.16] and [89, Corollary 4.13], as mentioned above), so $\log \operatorname{Reg}_K = (n/2-1) \log 2 + \log \operatorname{Reg}_K^+ \in (1/8 + o(1)) m \log m$. Similarly, $\det L = (n/2)^{1/2} \operatorname{Reg}_K$, so $\log \det L \in (1/8 + o(1)) m \log m$. Also $d \in (1/4 + o(1)) m$, so $(1/d) \log \det L \in (1/2 + o(1)) \log m$. Note that $\log(d/2\pi e) \in (1 + o(1)) \log m$.

The shortest nonzero vectors in M have length $(1 + o(1))(d/2\pi e)^{1/2} (\det L)^{1/d}$ by Theorem 3.9. The log of the length is $\log(1 + o(1)) + (1/2) \log(d/2\pi e) + (1/d) \log \det L \in (1 + o(1)) \log m$ as claimed. □

8.11. Effectiveness of reduction modulo short units: context. There is one remaining task: quantifying how inaccurately spherical models predict the effectiveness of reduction inside unit attacks. This paper's quantification consists of the statistical experiments for various m presented in Section 8.12.

Unit-attack experiments are not new. See, e.g., the software from Schanck [83]; for more context see Appendix D. However, the existing experiments were not designed to support comparisons to spherical models, and trying to base such comparisons upon the previous literature seems harder than carrying out new experiments directly on point.

Qualitatively questioning heuristics for the effectiveness of reduction modulo the unit lattice is also not new: Ducas–Plançon–Wesolowski [39, Section 4.1] wrote "those heuristics are most likely invalid for the lattices at hand which are somewhat close to orthogonal". However, this was not quantified in [39], and did not stop [40] (which shares an author with [39]) from applying such heuristics to S-unit attacks. The word "somewhat" in [39] suggests a perception of a medium-sized issue, large enough that measuring it would have been visible in the concrete numbers presented in [39], but not large enough to influence the "$\exp(-\text{subexp}(n))$" statement in [40] when $\#S$ is "$\text{subexp}(n)$".

Recall from Section 7 that the notion of reduction effectiveness converging to 0 as S increases cannot be correct—raising the question of what went wrong in the steps that led to this notion. The "most likely ... somewhat" guess from [39] was certainly not a clear, quantified statement of how wrong spherical models are for the case of unit lattices. This gap in the literature is addressed by the experiments in Section 8.12.

8.12. Effectiveness of reduction modulo short units: analysis. Let ν be a nonzero vector in $L_{\mathbb{R}}$, the $\mathbb{R}$-vector space generated by $L = \operatorname{Log} R^*$. Let α be a real number with $0 < \alpha < 1$. Let μ be a uniform random vector of length $2\alpha||\nu||_2$ in $L_{\mathbb{R}}$. By Theorem 3.13, the probability that μ reduces ν is the volume of the α-cap of the $(d-1)$-sphere divided by the volume of the $(d-1)$-sphere. The probability that one of $\pm\mu$ reduces ν is twice as large, since μ and $-\mu$ cannot both reduce ν. (One could refer to $\pm S$ for a spherical cap S as a pair of "earmuffs".)

By definition the nonzero vectors in a spherical model of L are $\pm\mu_1, \pm\mu_2, \ldots$ where $||\mu_j||_2^d = 2j\pi^{-d/2}(d/2)! \det L$. Above $\det L$ was computed for concrete m, assuming $h_m^+ = 1$, and asymptotically as $m \to \infty$, under a weaker assumption. Given $||\nu||_2$ one can then calculate, for each j, the probability that both μ_j and $-\mu_j$ fail to reduce ν. Multiplying over all small j then gives, by statistical independence of $\mu_1, \mu_2, \ldots$, the probability that all short vectors in the model fail to reduce ν.

It does not seem to be as straightforward to compute the analogous probability for the vectors $\pm u_1, \pm u_2, \ldots$ in L, even assuming $h_m^+ = 1$. One can enumerate all short vectors in L (and state formulas for the vector lengths, generalizing Theorem 8.2), but these are not statistically independent vectors, and one would not expect the reduction probabilities to be statistically independent. As an analogy, if u is one of the standard basis vectors of the lattice $\mathbb{Z}^d$, then $\pm u$ have a nonzero chance of failing to reduce an input vector of length $2d^{1/2}$, but taking all d standard basis vectors decreases the failure chance to 0.

The obvious fallback is to carry out experiments. Begin with a nonempty sequence U of short nonzero vectors in the unit lattice. Statistically sample points ν on a sphere, and check how often those points are successfully reduced by $\pm U$. There are many sphere radii of interest here, but it is easy to perform these statistical calculations for all radii simultaneously:

- Choose N independent uniform random points on the unit sphere.

- For each point ν, output $\min\{u \cdot u/(2|u \cdot \nu|) : u \in U\}$. The denominator can be 0, and then the quotient is defined as ∞; this occurs with probability 0.

For each $\lambda \geq 0$, a vector $\lambda\nu$ is reduced by $\pm u$ for some $u \in U$ if and only if $\lambda > \min\{u \cdot u/(2|\nu \cdot u|) : u \in U\}$; this is the same calculation as in Theorem 3.13. Consequently, out of the N points $\lambda\nu$ on the sphere of radius λ, the number of points reduced is the number of outputs below λ. Dividing by N gives the desired statistical estimate for the chance that a uniform random vector of length λ will be reduced.

The points that are not reduced by $\pm U$ are, by definition, the "approximate Voronoi cell" defined by $\pm U$. One can view the above computation as tracing, for each ν, the ray $\{\lambda\nu : \lambda \geq 0\}$ outwards from 0 and seeing when it reaches the boundary of the approximate Voronoi cell. The corresponding face of the cell is defined by the (unique with probability 1) vector $u \in U$ that minimizes $u \cdot u/(2|u \cdot \nu|)$. It would be interesting to demonstrate an association between how short u is and how often the u face appears, but collecting overall length statistics is more efficient and simplifies comparisons to a spherical model.

Figure 8.13 shows, for each $m \in \{16, 32, 64, 128, 256, 512\}$ and for three choices of sequences U, the distribution of outputs observed in $N = 1000$ experiments. These sets U are defined as follows. The blue curve for each m takes

$$U = \left\{ \operatorname{Log} \frac{\zeta_m^j - 1}{\zeta_m^k - 1} : j \in \{3, 5, \ldots, n-1\}, k \in \{1, 3, 5, \ldots, j-2\} \right\};$$

then $\pm U$ is the same set selected in [39, Section 4.1]. The green curve takes the same $\#U$ but uses `fpylll.Enumeration` to systematically enumerate short lattice vectors,[4] which *almost* means that the experiment is defined purely by the lattice and the specified number of vectors (not by the initial choice of generating set), except that there might be a dependence on how ties in vector length are broken; one could eliminate this dependence by considering only $\#U$ for which tie-breaking is not required. The red curve again has the same $\#U$ but takes U as all short vectors in a sample from a spherical model of L. Appendix G presents some spot-checks of the figure.

A closer look at the vectors found by enumeration shows that the set U used by the blue curve missed, e.g., $\operatorname{Log}(1 + \zeta_m^2 + \zeta_m^{-2})$ and other short vectors easily calculated from subfields, so it is not surprising that the green curve indicates noticeably better reduction than the blue curve. Compared to the red curve, the green and blue curves both show more and more advantage in reduction as m increases from 128 to 256 and then to 512. For $m = 512$, a vector of length 50 was reduced in 0/1000 experiments for the red curve, and was reduced in 999/1000 experiments for the green curve; this is overwhelming statistical evidence that this lattice reduces such vectors much more effectively than a spherical model of the same lattice.

[4] Enumeration is generally perceived as scaling poorly, but was not a problem here. Perhaps the speed of enumeration—or, to remove dependence on the input basis, the speed of verifying enumeration output—could serve as another useful metric for lattice non-randomness.

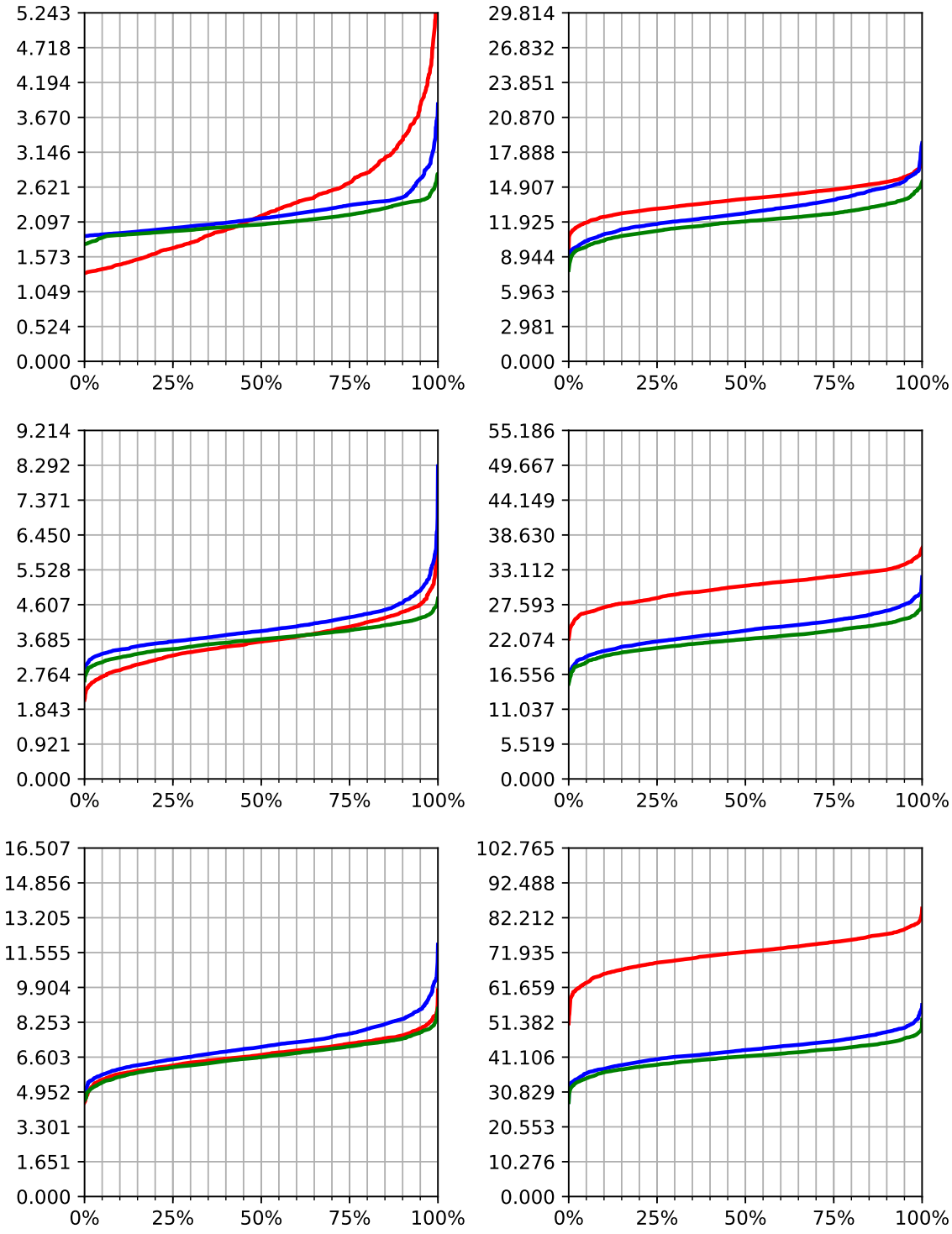

Fig. 8.13. For each $m \in \{16, 32, 64\}$ (left, top to bottom) and each $m \in \{128, 256, 512\}$ (right, top to bottom): Distribution, across 1000 samples of points ν on the unit sphere, of the maximum λ for which $\lambda\nu$ is not reduced by $\pm U$. Vertical axis is λ. The sequence U always has $(m/8)(m/4-1)$ elements. Blue curve uses $\mathrm{Log}((\zeta_m^j - 1)/(\zeta_m^k - 1))$ with odd j, k. Green curve uses U obtained by enumerating all short vectors in the unit lattice. Red curve uses U obtained by enumerating all short vectors in a sample from a spherical model of the unit lattice. Vertical axis is scaled to 2.5 times green average.

References

1. Erdem Alkim, Roberto Avanzi, Joppe Bos, Léo Ducas, Antonio de la Piedra, Thomas Pöppelmann, Peter Schwabe, and Douglas Stebila. NewHope: algorithm specifications and supporting documentation, 2017. https://csrc.nist.gov/projects/post-quantum-cryptography/round-1-submissions. 2, 43, 43, 44
2. Daniel Apon. Re: S-unit attacks, 2021. Email dated 28 Aug 2021 17:10:52 -0700; https://groups.google.com/a/list.nist.gov/g/pqc-forum/c/3mVeyEfYnUY/m/_xYg6HRcDQAJ. 44
3. Roberto Avanzi, Joppe Bos, Léo Ducas, Eike Kiltz, Tancrède Lepoint, Vadim Lyubashevsky, John M. Schanck, Peter Schwabe, Gregor Seiler, and Damien Stehlé. CRYSTALS-Kyber: algorithm specifications and supporting documentation (version 3.02), 2021. https://pq-crystals.org/kyber/data/kyber-specification-round3-20210804.pdf. 2
4. Eric Bach. Explicit bounds for primality testing and related problems. *Mathematics of Computation*, 55:355–380, 1990. https://www.ams.org/journals/mcom/1990-55-191/S0025-5718-1990-1023756-8/. 25
5. Alex Bartel and Hendrik W. Lenstra, Jr. On class groups of random number fields. *Proceedings of the London Mathematical Society*, 121:927–953, 2020. https://core.ac.uk/display/333542654. 47
6. Jens Bauch, Daniel J. Bernstein, Henry de Valence, Tanja Lange, and Christine van Vredendaal. Short generators without quantum computers: The case of multiquadratics. In Coron and Nielsen [31], pages 27–59. https://eprint.iacr.org/2017/404. 51
7. Helmut Bauer. Numerische Bestimmung von Klassenzahlen reeller zyklischer Zahlkörper. *Journal of Number Theory*, 1:161–162, 1969. https://www.sciencedirect.com/science/article/pii/0022314X69900341/pdf. 47
8. Anja Becker, Léo Ducas, Nicolas Gama, and Thijs Laarhoven. New directions in nearest neighbor searching with applications to lattice sieving. In Krauthgamer [56], pages 10–24. https://eprint.iacr.org/2015/1128. 16
9. Olivier Bernard and Adeline Roux-Langlois. Twisted-PHS: Using the product formula to solve Approx-SVP in ideal lattices. In Shiho Moriai and Huaxiong Wang, editors, *Advances in Cryptology—ASIACRYPT 2020—26th International Conference on the Theory and Application of Cryptology and Information Security, Daejeon, South Korea, December 7–11, 2020, Proceedings, Part II*, volume 12492 of *Lecture Notes in Computer Science*, pages 349–380. Springer, 2020. https://eprint.iacr.org/2020/1081. 53, 53
10. Daniel J. Bernstein. A subfield-logarithm attack against ideal lattices, 2014. https://blog.cr.yp.to/20140213-ideal.html. 49, 49, 49, 49, 50, 51, 52
11. Daniel J. Bernstein. Re: Soliloquy, 2015. Email dated 21 Feb 2015 16:53:40 -0000; https://groups.google.com/g/cryptanalytic-algorithms/c/GdVfp5Kbdb8/m/A1fwcggpJ_8J. 51, 52
12. Daniel J. Bernstein. Re: Soliloquy, 2015. Email dated 5 Apr 2015 18:05:18 -0000; https://groups.google.com/g/cryptanalytic-algorithms/c/GdVfp5Kbdb8/m/jvjfHzFGObEJ. 52, 52
13. Daniel J. Bernstein. S-unit attacks, 2016. Email dated 3 Aug 2016 09:32:23 -0000; https://groups.google.com/g/cryptanalytic-algorithms/c/mCMdsFemzQk/m/3cewE8Q5BwAJ. 52, 52, 52, 53
14. Daniel J. Bernstein. S-unit attacks, 2021. https://cr.yp.to/talks.html#2021.08.20. 4, 4, 5, 7

15. Daniel J. Bernstein, Chitchanok Chuengsatiansup, Tanja Lange, and Christine van Vredendaal. NTRU Prime: round 2, 2019. https://ntruprime.cr.yp.to/nist/ntruprime-20190330.pdf. 15, 55
16. Jean-François Biasse, Claus Fieker, Tommy Hofmann, and Aurel Page. Norm relations and computational problems in number fields, 2020. https://arxiv.org/abs/2002.12332. 51
17. Jean-François Biasse and Fang Song. Efficient quantum algorithms for computing class groups and solving the principal ideal problem in arbitrary degree number fields. In Krauthgamer [56], pages 893–902. https://epubs.siam.org/doi/pdf/10.1137/1.9781611974331.ch64. 19, 20, 44, 49
18. L. E. Blumenson. A derivation of n-dimensional spherical coordinates. *American Mathematical Monthly*, 67:63–66, 1960. 15
19. Joppe W. Bos, Craig Costello, Michael Naehrig, and Douglas Stebila. Post-quantum key exchange for the TLS protocol from the ring learning with errors problem. In *2015 IEEE Symposium on Security and Privacy, SP 2015, San Jose, CA, USA, May 17–21, 2015*, pages 553–570. IEEE Computer Society, 2015. https://eprint.iacr.org/2014/599. 43, 43, 44
20. Johannes Buchmann, Florian Göpfert, Rachel Player, and Thomas Wunderer. On the hardness of LWE with binary error: revisiting the hybrid lattice-reduction and meet-in-the-middle attack. In David Pointcheval, Abderrahmane Nitaj, and Tajjeeddine Rachidi, editors, *Progress in cryptology—AFRICACRYPT 2016—8th international conference on cryptology in Africa, Fes, Morocco, April 13–15, 2016, proceedings*, volume 96464 of *Lecture Notes in Computer Science*, pages 24–43. Springer, 2016. https://eprint.iacr.org/2016/089. 14
21. Joe Buhler, Carl Pomerance, and Leanne Robertson. Heuristics for class numbers of prime-power real cyclotomic fields. In Alf van der Poorten, editor, *High Primes and Misdemeanours: Lectures in Honour of the Sixtieth Birthday of Hugh Cowie Williams*, Fields Institute Communications, pages 149–157, 2004. https://math.dartmouth.edu/~carlp/PDF/joeleanne91603.pdf. 47, 47
22. Peter Campbell, Michael Groves, and Dan Shepherd. Soliloquy: A cautionary tale, 2014. https://docbox.etsi.org/Workshop/2014/201410_CRYPTO/S07_Systems_and_Attacks/S07_Groves_Annex.pdf. 44, 50, 51, 51
23. E. Rodney Canfield, Paul Erdös, and Carl Pomerance. On a problem of Oppenheim concerning "factorisatio numerorum". *Journal of Number Theory*, 17(1):1–28, 1983. https://math.dartmouth.edu/~carlp/PDF/paper39.pdf. 23
24. Sanjit Chatterjee, Neal Koblitz, Alfred Menezes, and Palash Sarkar. Another look at tightness II: practical issues in cryptography. In Raphael C.-W. Phan and Moti Yung, editors, *Paradigms in Cryptology—Mycrypt 2016. Malicious and Exploratory Cryptology—Second International Conference, Mycrypt 2016, Kuala Lumpur, Malaysia, December 1–2, 2016, Revised Selected Papers*, volume 10311 of *Lecture Notes in Computer Science*, pages 21–55. Springer, 2016. https://eprint.iacr.org/2016/360. 44
25. Yuanmi Chen and Phong Q. Nguyen. BKZ 2.0: Better lattice security estimates. In Dong Hoon Lee and Xiaoyun Wang, editors, *Advances in Cryptology—ASIACRYPT 2011—17th International Conference on the Theory and Application of Cryptology and Information Security, Seoul, South Korea, December 4–8, 2011. Proceedings*, volume 7073 of *Lecture Notes in Computer Science*, pages 1–20. Springer, 2011. https://www.iacr.org/archive/asiacrypt2011/70730001/70730001.pdf. 3

26. Laurent Clozel, Hee Oh, and Emmanuel Ullmo. Hecke operators and equidistribution of Hecke points. *Inventiones Mathematicae*, 144:327–351, 2001. https://gauss.math.yale.edu/~ho2/doc/hecke.pdf. 18
27. Henri Cohen. *A course in computational algebraic number theory*, volume 138 of *Graduate Texts in Mathematics*. Springer, 1993. 49
28. Henri Cohen. *Advanced topics in computational number theory*, volume 193 of *Graduate Texts in Mathematics*. Springer, 2000. 49, 49, 50, 50, 50, 50, 50, 50
29. Henri Cohen and Hendrik W. Lenstra, Jr. Heuristics on class groups of number fields. In Hendrik Jager, editor, *Number Theory Noordwijkerhout: Proceedings of the Journées Arithmétiques held at Noordwijkerhout, The Netherlands, July 11–15, 1983*, pages 33–62, 1983. https://link.springer.com/chapter/10.1007/BFb0099440. 47
30. Harvey Cohn. A numerical study of Weber's real class number calculation: part I. *Numerische Mathematik*, 2:347–362, 1960. https://link.springer.com/article/10.1007/BF01386236. 47, 47, 48, 48, 48, 48, 48, 49
31. Jean-Sébastien Coron and Jesper Buus Nielsen, editors. *Advances in Cryptology—EUROCRYPT 2017—36th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Paris, France, April 30–May 4, 2017, Proceedings, Part I*, volume 10210 of *Lecture Notes in Computer Science*, 2017. 36, 38
32. Ronald Cramer, Léo Ducas, Chris Peikert, and Oded Regev. Recovering short generators of principal ideals in cyclotomic rings. In Fischlin and Coron [42], pages 559–585. https://eprint.iacr.org/2015/313. 44, 51, 51
33. Ronald Cramer, Léo Ducas, and Benjamin Wesolowski. Short Stickelberger class relations and application to Ideal-SVP. In Coron and Nielsen [31], pages 324–348. https://eprint.iacr.org/2016/885. 52, 52, 52, 52, 52
34. Ronald Cramer, Léo Ducas, and Benjamin Wesolowski. Mildly short vectors in cyclotomic ideal lattices in quantum polynomial time. *Journal of the ACM*, 68, Article 8:1–26, 2021. https://ir.cwi.nl/pub/30736. 52
35. Harold Davenport. On a principle of Lipschitz. *Journal of the London Mathematical Society*, 26:179–183, 1951. 9
36. Emmanouil Doulgerakis, Thijs Laarhoven, and Benne de Weger. Finding closest lattice vectors using approximate Voronoi cells. In Jintai Ding and Rainer Steinwandt, editors, *Post-Quantum Cryptography—10th International Conference, PQCrypto 2019, Chongqing, China, May 8–10, 2019, Revised Selected Papers*, volume 11505 of *Lecture Notes in Computer Science*, pages 3–22. Springer, 2019. https://eprint.iacr.org/2016/888. 3, 50, 50
37. Emmanouil Doulgerakis, Thijs Laarhoven, and Benne de Weger. Sieve, enumerate, slice, and lift: Hybrid lattice algorithms for SVP via CVPP. In Abderrahmane Nitaj and Amr M. Youssef, editors, *Progress in Cryptology—AFRICACRYPT 2020—12th International Conference on Cryptology in Africa, Cairo, Egypt, July 20–22, 2020, Proceedings*, volume 12174 of *Lecture Notes in Computer Science*, pages 301–320. Springer, 2020. https://eprint.iacr.org/2020/487. 54
38. Léo Ducas, Thijs Laarhoven, and Wessel P. J. van Woerden. The randomized slicer for CVPP: sharper, faster, smaller, batchier. In Aggelos Kiayias, Markulf Kohlweiss, Petros Wallden, and Vassilis Zikas, editors, *Public-Key Cryptography—PKC 2020—23rd IACR International Conference on Practice and Theory of Public-Key Cryptography, Edinburgh, UK, May 4–7, 2020, Proceedings, Part II*, volume 12111 of *Lecture Notes in Computer Science*, pages 3–36. Springer, 2020. https://eprint.iacr.org/2020/120. 3, 6, 53, 53, 54

39. Léo Ducas, Maxime Plançon, and Benjamin Wesolowski. On the shortness of vectors to be found by the Ideal-SVP quantum algorithm. In *Advances in Cryptology—CRYPTO 2019—39th annual international Cryptology Conference, Santa Barbara, CA, USA, August 18–22, 2019, Proceedings, Part I*, pages 322–351, 2019. https://eprint.iacr.org/2019/234. 20, 27, 33, 33, 33, 33, 33, 33, 34, 50, 50, 52, 52, 52
40. Léo Ducas and Alice Pellet-Mary. Re: S-unit attacks, 2021. Email dated 24 Aug 2021 07:45:56 -0700; https://groups.google.com/a/list.nist.gov/g/pqc-forum/c/3mVeyEfYnUY/m/qgmqYdLVAQAJ. 4, 4, 25, 25, 33, 33
41. Kirsten Eisenträger, Sean Hallgren, Alexei Y. Kitaev, and Fang Song. A quantum algorithm for computing the unit group of an arbitrary degree number field. In David B. Shmoys, editor, *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31–June 03, 2014*, pages 293–302. ACM, 2014. https://www.personal.psu.edu/kxe8/unitgroup.pdf. 19, 20, 44, 49
42. Marc Fischlin and Jean-Sébastien Coron, editors. *Advances in Cryptology—EUROCRYPT 2016—35th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Vienna, Austria, May 8–12, 2016, Proceedings, Part I*, volume 9665 of *Lecture Notes in Computer Science*. Springer, 2016. 38, 41
43. Takashi Fukuda and Keiichi Komatsu. Weber's class number problem in the cyclotomic $\mathbb{Z}_2$-extension of $\mathbb{Q}$, II. *Journal de Théorie des Nombres de Bordeaux*, 22:359–368, 2011. https://jtnb.centre-mersenne.org/item/JTNB_2010__22_2_359_0/. 47, 48
44. Takashi Fukuda and Keiichi Komatsu. Weber's class number problem in the cyclotomic $\mathbb{Z}_2$-extension of $\mathbb{Q}$, III. *International Journal of Number Theory*, 7:1627–1635, 2011. 47, 47
45. Sanjam Garg, Craig Gentry, and Shai Halevi. Candidate multilinear maps from ideal lattices. In Thomas Johansson and Phong Q. Nguyen, editors, *Advances in Cryptology—EUROCRYPT 2013, 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Athens, Greece, May 26–30, 2013. Proceedings*, volume 7881 of *Lecture Notes in Computer Science*, pages 1–17. Springer, 2013. https://eprint.iacr.org/2012/610. 2, 2, 51
46. Fletcher Gates. Reduction-respecting parameters for lattice-based cryptosystems, 2018. https://hdl.handle.net/11375/24466. 44
47. Carl Friedrich Gauss. De nexu inter multitudinem classium, in quas formae binariae secundi gradus distribuuntur, earumque determinantem. In *Werke, Band II*, pages 269–291. Königlichen Gesellschaft der Wissenschaften, Göttingen, 1876. https://ia902808.us.archive.org/21/items/117771763_002/117771763_002.pdf; paper says "Commentatio prior societati regiae exhibita 1834". 2, 3
48. Craig Gentry. Fully homomorphic encryption using ideal lattices. In Michael Mitzenmacher, editor, *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31–June 2, 2009*, pages 169–178. ACM, 2009. https://dl.acm.org/doi/abs/10.1145/1536414.1536440. 2, 44, 51
49. Craig Gentry and Shai Halevi. Implementing Gentry's fully-homomorphic encryption scheme. In Kenneth G. Paterson, editor, *Advances in Cryptology—EUROCRYPT 2011—30th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tallinn, Estonia, May 15–19, 2011. Proceedings*, volume 6632 of *Lecture Notes in Computer Science*, pages 129–148. Springer, 2011. https://eprint.iacr.org/2010/520. 2, 51
50. Daniel Goldstein and Andrew Mayer. On the equidistribution of Hecke points. *Forum Mathematicum*, 15:165–189, 2003. 18

51. Loïc Grenié and Giuseppe Molteni. Explicit smoothed prime ideals theorems under GRH. *Mathematics of Computation*, 85:1875–1899, 2016. https://arxiv.org/abs/1312.4465. 31
52. Sean Hallgren. Fast quantum algorithms for computing the unit group and class group of a number field. In Harold N. Gabow and Ronald Fagin, editors, *Proceedings of the 37th Annual ACM Symposium on Theory of Computing, Baltimore, MD, USA, May 22–24, 2005*, pages 468–474. ACM, 2005. http://www.cse.psu.edu/~sjh26/unitgroup.pdf. 44, 49
53. Jeffrey Hoffstein, Jill Pipher, and Joseph H. Silverman. NTRU: A ring-based public key cryptosystem. In Joe Buhler, editor, *Algorithmic Number Theory, Third International Symposium, ANTS-III, Portland, Oregon, USA, June 21–25, 1998, Proceedings*, volume 1423 of *Lecture Notes in Computer Science*, pages 267–288. Springer, 1998. https://ntru.org/f/hps98.pdf. 6, 53, 53
54. Adolf Hurwitz. Über die Erzeugung der Invarianten durch Integration. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-Physikalische Klasse*, 1897:71–90, 1897. 7
55. Paul Kirchner, Thomas Espitau, and Pierre-Alain Fouque. Fast reduction of algebraic lattices over cyclotomic fields. In Daniele Micciancio and Thomas Ristenpart, editors, *Advances in Cryptology—CRYPTO 2020—40th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 17–21, 2020, Proceedings, Part II*, volume 12171 of *Lecture Notes in Computer Science*, pages 155–185. Springer, 2020. 51
56. Robert Krauthgamer, editor. *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10–12, 2016*. SIAM, 2016. 36, 37
57. Thijs Laarhoven. Sieving for closest lattice vectors (with preprocessing). In Roberto Avanzi and Howard M. Heys, editors, *Selected Areas in Cryptography—SAC 2016—23rd International Conference, St. John's, NL, Canada, August 10–12, 2016, Revised Selected Papers*, volume 10532 of *Lecture Notes in Computer Science*, pages 523–542. Springer, 2016. https://arxiv.org/abs/1607.04789v1. 3, 4, 4, 50, 50, 50, 50, 50, 51, 53, 53, 54, 54
58. Edmund Landau. Neuer Beweis des Primzahlsatzes und Beweis des Primidealsatzes. *Mathematische Annalen*, 56:645–670, 1903. 25
59. Adeline Langlois and Damien Stehlé. Worst-case to average-case reductions for module lattices. *Designs Codes Cryptography*, 75(3):565–599, 2015. https://eprint.iacr.org/2012/090. 45, 45
60. Jean B. Lasserre. A quick proof for the volume of n-balls. *American Mathematical Monthly*, 108(8):768–769, 2001. 9, 9
61. Derrick H. Lehmer. An extended theory of Lucas' functions. *Annals of Mathematics. Second Series*, 31:419–448, 1930. 32
62. Hendrik W. Lenstra, Jr. Lattices. In *Algorithmic number theory: Lattices, Number Fields, Curves and Cryptography*, pages 127–181. Cambridge University Press, 2008. https://www.math.leidenuniv.nl/~psh/ANTproc/06hwl.pdf. 10
63. Andrea Lesavourey, Thomas Plantard, and Willy Susilo. Short principal ideal problem in multicubic fields. *J. Math. Cryptol.*, 14(1):359–392, 2020. 51
64. Shengqiao Li. Concise formulas for the area and volume of a hyperspherical cap. *Asian Journal of Mathematics and Statistics*, 4(1):66–70, 2011. https://docsdrive.com/pdfs/ansinet/ajms/2011/66-70.pdf. 15, 15
65. Franciscus Jozef van der Linden. Class number computations of real abelian number fields. *Mathematics of Computation*, 39:693–707, 1982. https://www.ams.org/journals/mcom/1982-39-160/S0025-5718-1982-0669662-5/. 47

66. Vadim Lyubashevsky, Chris Peikert, and Oded Regev. On ideal lattices and learning with errors over rings. In Henri Gilbert, editor, *Advances in Cryptology—EUROCRYPT 2010, 29th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Monaco / French Riviera, May 30–June 3, 2010, proceedings*, volume 6110 of *Lecture Notes in Computer Science*, pages 1–23. Springer, 2010. https://eprint.iacr.org/2012/230. 42, 42, 43, 43, 43, 43, 44, 44, 44
67. John Myron Masley. Class numbers of real cyclic number fields with small conductor. *Compositio Mathematica*, 37:297–319, 1978. http://www.numdam.org/article/CM_1978__37_3_297_0.pdf. 47
68. James E. Mazo and Andrew M. Odlyzko. Lattice points in high-dimensional spheres. *Monatshefte für Mathematik*, 110(1):47–61, 1990. http://www.dtc.umn.edu/~odlyzko/doc/arch/high.dim.spheres.pdf. 17
69. Daniele Micciancio and Panagiotis Voulgaris. Faster exponential time algorithms for the shortest vector problem. In Moses Charikar, editor, *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010, Austin, Texas, USA, January 17–19, 2010*, pages 1468–1480. SIAM, 2010. https://cseweb.ucsd.edu/~daniele/papers/Sieve.pdf. 16, 16
70. Daniele Micciancio and Michael Walter. Practical, predictable lattice basis reduction. In Fischlin and Coron [42], pages 820–849. https://eprint.iacr.org/2015/1123. 3, 6, 53, 53, 53
71. John C. Miller. Class numbers of totally real fields and applications to the Weber class number problem. *Acta Arithmetica*, 164:381–397, 2014. https://arxiv.org/abs/1405.1094. 47, 47
72. John C. Miller. Class numbers in cyclotomic $\mathbb{Z}_p$-extensions. *Journal of Number Theory*, 150:47–73, 2015. https://arxiv.org/abs/1410.2921. 47
73. Chris Peikert. Lattice cryptography for the Internet. In Michele Mosca, editor, *Post-Quantum Cryptography—6th International Workshop, PQCrypto 2014, Waterloo, ON, Canada, October 1–3, 2014. Proceedings*, volume 8772 of *Lecture Notes in Computer Science*, pages 197–219. Springer, 2014. https://eprint.iacr.org/2014/070. 43, 43
74. Chris Peikert. Finding short generators of ideals, and implications for cryptography, 2016. https://web.archive.org/web/20210818101136/https://www.mathematik.uni-kl.de/~thofmann/ants/talks/mon/talk_peikert.pdf. 44
75. Chris Peikert. Lattice-based cryptography, 2016. https://www.youtube.com/watch?v=FVFw_qb1ZkY. 44, 44
76. Chris Peikert. Lattice-based cryptography, 2016. https://web.archive.org/web/20210628010013/https://web.eecs.umich.edu/~cpeikert/pubs/slides-qcrypt.pdf. 44
77. Chris Peikert. The state of lattice-based assumptions, 2018. https://www.youtube.com/watch?v=VUGHJipRfwA. 44
78. Alice Pellet-Mary, Guillaume Hanrot, and Damien Stehlé. Approx-SVP in ideal lattices with pre-processing. In Yuval Ishai and Vincent Rijmen, editors, *Advances in Cryptology—EUROCRYPT 2019—38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19–23, 2019, Proceedings, Part II*, volume 11477 of *Lecture Notes in Computer Science*, pages 685–716. Springer, 2019. https://eprint.iacr.org/2019/215. 4, 4, 4, 4, 4, 5, 5, 50, 50, 50, 50, 50, 52, 53, 53, 53
79. Ghaya Rekaya, Jean-Claude Belfiore, and Emanuele Viterbo. Very efficient lattice reduction tool on fast fading channels. In *Proceedings International Symposium on*

Information Theory and its Applications (ISITA), pages 714–717, 2004. https://ecse.monash.edu/staff/eviterbo/papers.html. 50, 50
80. Carl Gustav Reuschle. *Tafeln complexer Primzahlen, welche aus Wurzeln der Einheit gebildet sind*. Buchdruckerei der Königl. Akademie der Wissenschaften (G. Vogt), 1875. https://archive.org/details/tafelncomplexer00unkngoog. 47, 49
81. Herbert Robbins. A remark on Stirling's formula. *The American Mathematical Monthly*, 62(1):26–29, 1955. 10
82. C. A. Rogers. The number of lattice points in a set. *Proceedings of the London Mathematical Society. Third Series*, 6:305–320, 1956. 8, 8
83. John M. Schanck. logcvp, 2015. https://github.com/jschanck-si/logcvp. 32, 51
84. Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.*, 26(5):1484–1509, 1997. https://arxiv.org/abs/quant-ph/9508027. 44
85. Carl Ludwig Siegel. A mean value theorem in geometry of numbers. *Annals of Mathematics. Second Series*, 46:340–347, 1945. 7
86. Nigel P. Smart and Frederik Vercauteren. Fully homomorphic encryption with relatively small key and ciphertext sizes. In Phong Q. Nguyen and David Pointcheval, editors, *Public Key Cryptography—PKC 2010, 13th International Conference on Practice and Theory in Public Key Cryptography, Paris, France, May 26–28, 2010. Proceedings*, volume 6056 of *Lecture Notes in Computer Science*, pages 420–443. Springer, 2010. https://eprint.iacr.org/2009/571. 2, 51
87. Naftali Sommer, Meir Feder, and Ofir Shalvi. Finding the closest lattice point by iterative slicing. In *IEEE International Symposium on Information Theory, ISIT 2007, Nice, France, June 24–29, 2007*, pages 206–210. IEEE, 2007. 50, 50, 50
88. Andreas Strömbergsson and Anders Södergren. On the generalized circle problem for a random lattice in large dimension. *Advances in Mathematics*, 345:1042–1074, 2019. https://arxiv.org/abs/1611.06332. 8
89. Lawrence C. Washington. *Introduction to cyclotomic fields, second edition*, volume 83 of *Graduate Texts in Mathematics*. Springer, 1997. 30, 30, 31, 31, 31, 31, 31, 31, 32, 32, 32, 32, 32, 32, 32, 46, 46, 46, 51
90. Heinrich Weber. Theorie der Abel'schen Zahlkörper. *Acta Mathematica*, 8:193–263, 1886. https://link.springer.com/content/pdf/10.1007/BF02417089.pdf. 47
91. Heinrich Weber. *Lehrbuch der Algebra*, volume 2. F. Vieweg & Sohn, 1899. https://archive.org/details/lehrbuchderalge02webegoog/page/n833/mode/1up. 47, 47, 47, 48, 48, 48
92. Thomas Wunderer. *On the security of lattice-based cryptography against lattice reduction and hybrid attacks*. PhD thesis, Technische Universität Darmstadt, 2018. https://tuprints.ulb.tu-darmstadt.de/8082/. 14
93. Rainer Zimmert. Ideale kleiner Norm in Idealklassen und eine Regulatorabschätzung. *Inventiones Mathematicae*, 62:367–380, 1981. 25

Appendices

A Importance of Ideal-SVP

Lyubashevsky–Peikert–Regev [66] announced "very strong hardness guarantees" for Ring-LWE. The informal statement of [66, Main Theorem] hypothesized that

> it is hard for polynomial-time *quantum* algorithms to approximate the search version of the shortest vector problem (SVP) in the *worst case* on *ideal lattices* in R to within a fixed poly(n) factor

and concluded from this that any polynomial number of Ring-LWE samples "are pseudorandom to any polynomial-time (possibly quantum) attacker".

A.1. Impact of the Ideal-SVP guarantee. The conference and journal versions of [66] have together been cited 2000 times, according to Google Scholar. There is a clear path from that paper's "very strong hardness guarantees" to proposals of cryptosystems for standardization and deployment:

- Peikert's 2014 "Lattice cryptography for the Internet" [73, introduction, first paragraph] stated that "both ring-SIS and ring-LWE enjoy strong provable hardness guarantees: they are hard on the average as long as the Shortest Vector Problem is hard to approximate (by quantum computers, in the case of ring-LWE) on so-called *ideal* lattices in the corresponding ring, *in the worst case*" and characterized this as "good theoretical evidence that ring-SIS and ring-LWE are a solid foundation on which to design cryptosystems".
- Bos–Costello–Naehrig–Stebila's "Post-quantum key exchange for the TLS protocol from the ring learning with errors problem" [19, introduction, first paragraph] stated that the paper presents key exchange based on "*the ring learning with errors (R-LWE) problem* [LPR13a], which is related to hard lattice problems". The paper later clarified "hard lattice problems" by saying that "the R-LWE problem is related to the SVP on ideal lattices", and said that it was presenting "a reformulation of Peikert's KEM" from [73].
- Within the NIST Post-Quantum Cryptography Standardization Project, the submission of NewHope by Alkim–Avanzi–Bos–Ducas–de la Piedra–Pöppelmann–Schwabe–Stebila stated [1, "Justification of security strength", "Provable security reductions"] that the decision Ring-LWE problem "is hard under the assumption that the search version of the approximate shortest vector problem is hard (in the worst case) on ideal lattices in R, for appropriate parameters" and listed this as a reduction "underlying the security of NewHope-CCA-KEM". NewHope was developed as an improvement of [19].

A polynomial-time break of polynomial-approximation-factor Ideal-SVP would render [66, Main Theorem] vacuous, eliminating a core argument for the security of various cryptosystems. Usually these cryptosystems select cyclotomic fields, typically power-of-2 cyclotomics, as in the case of NewHope.

It is important to realize that there is a quantification issue rendering [66, Main Theorem] logically insufficient to guarantee the security of, e.g., NewHope. The NewHope parameter selection was designed starting with the assumption that lattice problems are *exponentially* hard with a particular exponent—see, e.g., [1, Table 12, "lower bound of $2^{.292b}$"]—while [66, Main Theorem] takes only *polynomial-time* attacks into account. A theorem guaranteeing a specified level of exponential security would need to hypothesize exponential security for Ideal-SVP. (It would also need to quantify various polynomial losses that were not

quantified in [66]; see [24] and [46].) A subexponential-time break of polynomial-approximation-factor Ideal-SVP would not render [66, Main Theorem] vacuous, but it would contradict the stronger "hard lattice problems" claim in [19] and the "approximate shortest vector problem is hard" hypothesis in [1].

A.2. Paths through the space of cryptanalytic targets. The quantum polynomial-time break of the cyclotomic case of the Gentry [48] cryptosystem was the culmination of several steps:

- Shor's algorithm [84] to compute discrete logarithms.
- Hallgren's algorithm [52] to find unit groups for constant-degree number fields.
- The Eisenträger–Hallgren–Kitaev–Song [41] algorithm to find unit groups for variable-degree number fields.
- The Biasse–Song [17] adaptation to find generators.
- The Campbell–Groves–Shepherd [22] algorithm reducing a generator to a short generator, using the textbook basis for the cyclotomic units.

If Ring-LWE is breakable, it would not be surprising for the public development of the break to follow even more steps, one of those steps being a break of worst-case Ideal-SVP (after all, according to [66], there cannot be a Ring-LWE attack without a worst-case Ideal-SVP attack) and earlier steps being breaks of *some* cases of Ideal-SVP (the same way that the Eisenträger–Hallgren–Kitaev–Song algorithm was preceded by Hallgren's algorithm handling some cases). The public record is consistent with being on this path: what has been happening recently is S-unit attacks breaking more and more cases of Ideal-SVP.

The same advances have prompted a contrary narrative saying that attacks against Ideal-SVP are nothing to worry about: there is no known reduction stating that an Ideal-SVP attack implies a Ring-LWE attack, and what we really care about is the security of Ring-LWE. The dividing line between Ideal-SVP and Ring-LWE was described in [75, minutes 61–62] and [2] as a "barrier" to this line of work. For comparison, the fact that pure unit attacks cannot beat approximation factor $\exp(n^{1/2+o(1)})$ for worst-case inputs was described in [76, PDF page 84] and [75, minutes 61–62] and [74, PDF page 75] and [77, minute 45] as a "barrier" (and a "natural barrier" and an "inherent barrier") to this line of work, and then this "barrier" was broken. Earlier [32, Section 1] had stated that "the above-described algorithms ... apply only to *principal* ideals" and characterized this as a "barrier"; this "barrier" was broken too. One has to ask what procedures are being used to declare these "barriers", and whether these procedures are properly controlling cryptographic risks.

From a cryptanalyst's perspective, it makes sense to begin with Ideal-SVP, and, within Ideal-SVP, to begin by breaking whichever cases can be broken. Each advance against Ideal-SVP is important to the vast literature relying on Ring-LWE not because the advance shows that Ring-LWE is *broken* but rather because the advance shows that Ring-LWE is *not sufficiently studied*. Further study is essential.

Similar comments apply to Module-SVP and Module-LWE, which, just like Ideal-SVP and Ring-LWE, are lattice problems for various types of R-modules.

Langlois–Stehlé [59, Section 1] stated that "M-LWE and M-SIS are obviously no easier than R-LWE and R-SIS" and "Mod-SIVP can trivially be shown to be no easier than Id-SIVP". There is again a quantification issue here—for efficiency, proposed cryptosystems using higher-rank modules generally use smaller rings, breaking the logic of [59]—but in any case it is natural for the cryptanalyst to begin with Ideal-SVP.

B Number fields

Some parts of this paper rely on the following definitions and facts from algebraic number theory.

B.1. Algebraic numbers. An element $\alpha \in \mathbb{C}$ is called an "algebraic number" if $f(\alpha) = 0$ for some monic polynomial $f \in \mathbb{Q}[x]$, and is called an "algebraic integer" if $f(\alpha) = 0$ for some monic polynomial $f \in \mathbb{Z}[x]$. The set of algebraic numbers in $\mathbb{C}$ is denoted $\overline{\mathbb{Q}}$; the set of algebraic integers in $\mathbb{C}$ is denoted $\overline{\mathbb{Z}}$. One can show that $\overline{\mathbb{Q}}$ is a subfield of $\mathbb{C}$ and that $\overline{\mathbb{Z}}$ is a subring of $\overline{\mathbb{Q}}$.

For example, i, the usual square root of -1 in $\mathbb{C}$, is in $\overline{\mathbb{Z}}$ since $f(i) = 0$ for $f = x^2 + 1$; and $(1 + i)/2$ is in $\overline{\mathbb{Q}}$ since $f((1 + i)/2) = 0$ for $f = x^2 - x + 1/2$.

Beware that most "algebraic integers" are not integers, i.e., not in $\mathbb{Z}$; the terminology "algebraic integer" violates the rule that adjectives are restrictive. Number theorists typically replace the terms "integer" and "algebraic integer" with "rational integer" and "integer" respectively, restoring restrictive adjectives but not staying consistent with use of the term "integer" outside number theory.

B.2. Number fields. Each subfield K of $\mathbb{C}$ is a vector space over $\mathbb{Q}$. One calls K a "number field" if the dimension of K as a $\mathbb{Q}$-vector space is finite. This dimension is called the "degree" of K.

One can show that all elements of a number field K are algebraic numbers: i.e., $K \subseteq \overline{\mathbb{Q}}$. Conversely, for every $\alpha \in \overline{\mathbb{Q}}$, the field $\mathbb{Q}(\alpha)$ is a number field. Here $\mathbb{Q}(\alpha)$ is, by definition, the smallest subfield of $\mathbb{C}$ containing α, i.e., the intersection of all subfields of $\mathbb{C}$ containing α.

In particular, define $\zeta_m = \exp(2\pi i/m)$ for each positive integer m. Then $\zeta_m^m = 1$, so ζ_m is an algebraic number, so $\mathbb{Q}(\zeta_m)$ is a number field, called the "mth cyclotomic field". One can show that the degree of K is $\#(\mathbb{Z}/m)^*$, the number of units in the ring $\mathbb{Z}/m$.

Often the literature defines a "number field" as a field containing $\mathbb{Q}$ and having finite dimension as a $\mathbb{Q}$-vector space. Any such field is isomorphic to a subfield of $\mathbb{C}$, as a consequence of the fundamental theorem of algebra, so it suffices to study number fields in $\mathbb{C}$. This appendix defines number fields to be in $\mathbb{C}$.

B.3. Rings of integers. Let K be a number field. Define $\mathcal{O}_K = K \cap \overline{\mathbb{Z}}$. This is the "ring of algebraic integers in K", or the "ring of integers of K", or the "maximal order of K".

For example, one can show that if $K = \mathbb{Q}(\zeta_m)$ then $\mathcal{O}_K = \mathbb{Z}[\zeta_m]$. Here $\mathbb{Z}[\zeta_m]$ is, by definition, the smallest subring of $\mathbb{C}$ containing ζ_m. If m is a power of 2

then there is a unique ring isomorphism $\mathbb{Z}[x]/(x^{m/2}+1) \to \mathbb{Z}[\zeta_m]$ mapping x to ζ_m, and a unique ring isomorphism $\mathbb{Q}[x]/(x^{m/2}+1) \to \mathbb{Q}(\zeta_m)$ mapping x to ζ_m.

B.4. Unique factorization of ideals. Define a binary operation $I, J \mapsto IJ$ on the set of ideals of $\mathcal{O}_K$ as follows: IJ is the smallest ideal containing $\{\alpha\beta : \alpha \in I, \beta \in J\}$. For example, the product of two principal ideals $\alpha\mathcal{O}_K$ and $\beta\mathcal{O}_K$ is simply $\alpha\beta\mathcal{O}_K$. This operation is commutative and associative, and has an identity element, namely $\mathcal{O}_K$.

One of the fundamental facts about this multiplication operation on ideals of $\mathcal{O}_K$ is that each nonzero ideal of $\mathcal{O}_K$ factors uniquely as a product of powers of prime ideals of $\mathcal{O}_K$. The exponent of P in the factorization of I is denoted $\operatorname{ord}_P I$. One has $\operatorname{ord}_P I = 0$ when P does not occur in the factorization. Unique factorization implies $\operatorname{ord}_P IJ = \operatorname{ord}_P I + \operatorname{ord}_P J$.

B.5. Unique factorization of elements into ideals. If $\alpha \in \mathcal{O}_K - \{0\}$ then $\operatorname{ord}_P \alpha$ is defined as $\operatorname{ord}_P \alpha\mathcal{O}_K$, the exponent of P in the factorization of the nonzero principal ideal $\alpha\mathcal{O}_K$. Then $\operatorname{ord}_P \alpha\beta = \operatorname{ord}_P \alpha + \operatorname{ord}_P \beta$.

There is a unique extension of ord_P to a function on all elements of K^* satisfying $\operatorname{ord}_P \alpha\beta = \operatorname{ord}_P \alpha + \operatorname{ord}_P \beta$. The point is that each nonzero element of K can be written as a ratio α/β for some nonzero $\alpha, \beta \in \mathcal{O}_K$.

B.6. Ideal classes. Define two nonzero ideals I, J of $\mathcal{O}_K$ as "equivalent" if $\alpha I = \beta J$ for some nonzero $\alpha, \beta \in \mathcal{O}_K$. For example, the ideals $15\mathbb{Z}$ and $17\mathbb{Z}$ of $\mathbb{Z}$ are equivalent, since $17(15\mathbb{Z}) = 15(17\mathbb{Z})$. This is, as the name suggests, an equivalence relation.

Define Cl_K as the set of equivalence classes of nonzero ideals of $\mathcal{O}_K$. It is easy to show that multiplication of ideals induces a multiplication operation on Cl_K, and that Cl_K is a group under this operation, the "class group of K".

A fundamental fact is that Cl_K is finite. The cardinality $\#\operatorname{Cl}_K$ is the "class number of K". For example, the class number of $\mathbb{Q}$ is 1, since all nonzero ideals of $\mathbb{Z}$ are equivalent.

B.7. Class numbers of cyclotomic fields. The class number of $K = \mathbb{Q}(\zeta_m)$ is called h_m. The class number of $\mathbb{R} \cap \mathbb{Q}(\zeta_m) = \mathbb{Q}(\zeta_m + \zeta_m^{-1})$ is called h_m^+. One can show that h_m^+ divides h_m; the quotient is called h_m^-. See, e.g., [89, Theorem 4.10].

For $m \in \{1, 2\}$ one has $K = \mathbb{Q}$ and $\mathcal{O}_K = \mathbb{Z}$, so $h_1 = h_2 = 1$. For $m = 4$ one has $K = \mathbb{Q}(\zeta_4) = \mathbb{Q}(i) = \{a_0 + a_1 i : a_0, a_1 \in \mathbb{Q}\}$ and $\mathcal{O}_K = \mathbb{Z}[\zeta_4] = \mathbb{Z}[i] = \{a_0 + a_1 i : a_0, a_1 \in \mathbb{Z}\}$, and with some study of the ideals of $\mathcal{O}_K$ one can show that $h_4 = 1$. With more work one can show that $h_8 = h_{16} = h_{32} = 1$. However, $h_{64} = h_{64}^- = 17$, and h_m^- increases rapidly after that, forcing $h_m = h_m^- h_m^+$ to increase rapidly. See [89, pages 412–420] for a table of h_m^- values, including m between powers of 2, and see [89, Theorem 4.20] for asymptotics.

C Evidence that $h_m^+ = 1$ when m is a power of 2

Recall from Appendix B that h_m^+ is the class number of $\mathbb{R} \cap \mathbb{Q}(\zeta_m)$ where $\zeta_m = \exp(2\pi i/m)$. A standard conjecture in number theory is that $h_m^+ = 1$ when m is a power of 2. What follows is a review of the evidence for this conjecture.

C.1. Prime constraints. A theorem of Weber [90, page 244, third paragraph] states that h_m is odd, i.e., that h_m^- is odd and that h_m^+ is odd. More recent work has proven much more here: for example, Fukuda–Komatsu [44] showed that any prime divisor of h_m^+ must be above 10^9 and must be in $\pm 1 + 32\mathbb{Z}$.

C.2. Computations for $m \leq 512$. A traditional class-group computation (via S-units) is feasible for the number field $\mathbb{R} \cap \mathbb{Q}(\zeta_m)$ for surprisingly large values of m. The main result of such a computation is a generator for each small prime ideal of $\overline{\mathbb{Z}} \cap \mathbb{R} \cap \mathbb{Q}(\zeta_m)$, at which point one is convinced that the class number is 1, although *proving* that the class number is 1 is more difficult.

Weber showed that $h_m^+ = 1$ for $m \leq 16$; Cohn [30] proved further constraints on h_m^+, and noted that computations published in 1875 by Reuschle [80] already sufficed to show that $h_{32}^+ = 1$; Bauer [7] briefly reported computations proving, among other things, that $h_{64}^+ = 1$; Masley [67] verified—with some corrections—Bauer's computations; van der Linden [65, page 705] proved that $h_{128}^+ = 1$, and proved that $h_{256}^+ = 1$ assuming GRH; Miller [71] proved unconditionally that $h_{256}^+ = 1$, and proved that $h_{512}^+ = 1$ assuming GRH.

C.3. Class-group heuristics. There are well-known conjectures regarding the distributions of class groups across various families of number fields, notably the Cohen–Lenstra [29] heuristics, followed by various extensions and corrections (see, e.g., [5]).

Buhler–Pomerance–Robertson [21] wrote "Long ago, Weber conjectured that $h^+(2^n) = 1$ for all n"; modeled $\mathbb{Q}(\zeta_m)^+$ as a random number field satisfying all known constraints on h_m^+; and concluded in this model that "the probability that Weber's conjecture is true is at least 99.3%".

Miller [72] built a more sophisticated model, accounting for the h_m^+ restrictions from [44] and the $m = 512$ computations from [71], and concluded in this model that "the probability that the Weber conjecture is true is at least 99.999998%".

C.4. Is "Weber's conjecture" due to Weber? Comments are required regarding the terminology "Weber's conjecture" used in, e.g., [21].

It is unclear that "Weber's conjecture" was made by Weber. If the conjecture actually began with a more recent paper from someone else, then miscrediting the conjecture to Weber would be overstating another form of evidence, namely the sheer amount of time since the conjecture was formulated.

Concretely, Weber [91, page 808] labeled the idea $h_{32} = 1$ as "zweifelhaft [dubious]". Weber knew that $h_{32}^- = 1$ (see [91, page 802]), so Weber was doubting that $h_{32}^+ = 1$. This suggests that if Weber ever conjectured that $h_m^+ = 1$ for all powers m of 2 then this conjecture would have been after [91].

Some sources—e.g., [30, page 348, fifth paragraph] and [43, page 359, bottom paragraph]—go further, claiming that Weber actually conjectured that $h_{64}^+ > 1$.

However, this claim appears to be based on a misreading. A quote analysis follows, for purposes of (1) ensuring appropriate credit and (2) understanding what weight, if any, should be assigned to the history.

What [30] claims is that "Weber casually conjectured [7; p. 808] on the basis of $k_6 = 17$ that $h_6 > 1$": i.e., conjectured on the basis of $h_{64}^- = 17$ that $h_{64}^+ > 1$. The claim in [43] is given without citation, and is immediately followed by a statement that [30] instead showed $h_{64}^+ = 1$.

What Weber actually wrote in [91, page 808, footnote] was the following:

> Da, wie wir gesehen haben, auch der erste Classenzahlfactor = 1 ist, so ist der Kreistheilungskörper Ω_{16} einclassig. Es gelten in ihm dieselben Gesetze der Primzahlen, wie im Körper der rationalen Zahlen. Zweifelhaft ist dies für den Körper Ω_{32} und sicher nicht mehr zutreffend im Körper Ω_{64}, in dem die Classenzahl ein Vielfaches von 17 ist.
>
> [Because, as we have seen, also the first class-number factor is equal to 1, the cyclotomic field Ω_{16} has a single class. The same laws of prime numbers apply in it as in the field of rational numbers. This is dubious for the field Ω_{32} and certainly no longer applicable to the field Ω_{64}, in which the class number is a multiple of 17.]

A straightforward reading says that "dies [this]" refers to the property $h_m = 1$ in the immediately preceding statement. Weber correctly says that the property $h_m = 1$ holds for $m = 16$; expresses doubt for $m = 32$ (as noted above, Weber had calculated $h_{32}^- = 1$; Weber did not know that $h_{32}^+ = 1$); and correctly says that the property does not hold for $m = 64$, since h_m is then a multiple of 17 (which Weber knew, since Weber had calculated $h_{64}^- = 17$).

Now consider the claim in [30] that Weber, on this page, conjectured on the basis of $h_{64}^- = 17$ that $h_{64}^+ > 1$. There are no other mentions of 17 in [91, page 808], so the claim must be referring to the quote above. The only mention of 64 is in the statement that "dies" is "sicher nicht mehr zutreffend im Körper Ω_{64}", so the claim must be interpreting "dies" not as the property that $h_m = 1$ but rather as the property that a factor of h_m, namely h_m^+, is 1.

The Weber quote does mention a property of a factor of h_m: "der erste Classenzahlfactor = 1 ist". Cohn [30, page 347] writes (in another notation) h_m^+ as the first factor of h_m; could have understood Weber's text "der erste Classenzahlfactor = 1 ist" as referring to the property $h_m^+ = 1$; and could have thought that "dies" was similarly referring to the property $h_m^+ = 1$. However, this property is farther back in Weber's text, and less emphasized, than the property $h_m = 1$, making it more difficult to explain as the referent of "dies". Also, Weber's "erste Classenzahlfactor" is actually h_m^-; see [91, page 802].

Another reason that it is implausible to interpret "dies" as the property $h_m^+ = 1$ is that "sicher nicht" is not a conjecture but a statement of certainty. This would be an overstatement of Weber's knowledge with the $h_m^+ = 1$ interpretation, whereas the straightforward $h_m = 1$ reading instead says that this is something that Weber had proven.

Beyond the question of what Weber wrote, Cohn wrote [30, page 361] that "We still have obtained no evidence to doubt that every" $h_m^+ = 1$, but also

wrote [30, page 349] that "we might believe that ultimately" $h_m^+ > 1$ since there is a "chain of fields". Neither of these is labeled as a conjecture.

D Attack credits

This appendix reviews the history of various aspects of unit attacks and, more generally, S-unit attacks, with the objective of ensuring appropriate credit. Many sources quoted here predate sources credited in the literature on lattice-based cryptography for the same ideas. Tracing the history of simple ideas can be very difficult, and it is likely that there are earlier sources for some of the ideas below.

D.1. Finding a generator. As mentioned in Appendix A, there is a quantum polynomial-time algorithm by Biasse–Song [17] to find a generator, if one exists, of an ideal I provided as input. This is an adaptation of a unit-group algorithm by Eisenträger, Hallgren, Kitaev, and Song [41], which was preceded by a constant-degree algorithm from Hallgren [52]. More generally, the same techniques find an S-generator, if one exists.

There are also surprisingly fast non-quantum algorithms for finding unit groups, generators, etc.—see generally [27, Section 6.5] and [28, Section 7.4]—with a much longer history (see, e.g., [80]). For purposes of this paper, it suffices to consider the quantum case.

D.2. Unit attacks: finding a short generator. The idea of a unit attack is to shorten generators by reducing modulo the unit lattice. The first description of unit attacks *as a threat to lattice cryptosystems* was by Bernstein [10] in 2014, but this does not mean unit attacks were introduced in [10]. On the contrary, [10] described "this approach to finding generators"—namely, trying to find a short generator "g" of a given ideal "gR" by first finding some generator "gu" and then searching for "elements of the lattice $\operatorname{Log} R^*$ close to $\operatorname{Log} gu$"—as "reasonably well known among computational algebraic number theorists".

Here is a quote confirming the "reasonably well known" statement from [10]. Cohen's 2000 textbook "Advanced topics in computational number theory"[28, pages 375–376] notes that "reducing the size of generators" of the S-unit group "can be done using variants of the LLL algorithm", and continues by mentioning the following method to obtain even shorter generators:

> We can reduce even more the size of the γ_i by replacing γ_i by γ_i/ε for a suitable unit ε, which still gives generators of $U_S(K)$. To do this, we multiply γ_i recursively by very small powers of a generating set of the unit group as long as the size of γ_i (measured in any reasonable way) decreases.

The context in the above quote, taking γ_i as one of the generators of an S-unit group for some S, obviously does not matter for this reduction algorithm. The algorithm takes any $g \in K^*$, and repeatedly replaces g by gu for some u in a specified list as long as this makes g smaller.

The history is not correctly reported in [78, Section 1], which cites [10] and [22] (in the opposite order) for the "algorithmic blueprint" of "first using class group computations to find a generator" and then using the "log-unit lattice to shorten the latter generator", and adds a "note" that this was "already suggested in" [79]. The paper [79] was four years after the textbook quote above.

D.3. Simple reduction. The idea of simple reduction is to reduce a vector v by repeatedly replacing v with some smaller $v - u$, where u comes from a database of short vectors. For example, simple reduction modulo the unit lattice reduces a vector $\operatorname{Log} v$ in log space by repeatedly replacing $\operatorname{Log} v$ with some smaller $\operatorname{Log}(v/u)$, where u comes from a database of short units.

Aside from notational quibbles about dividing by u versus multiplying by $1/u$, this is exactly what was quoted above from Cohen's 2000 textbook [28]: "multiply γ_i recursively by very small powers of a generating set of the unit group as long as the size of γ_i (measured in any reasonable way) decreases". A small special case would take the generating set U of the unit group to have $\operatorname{Log} U = \{0, \pm b_1, \pm b_2, \ldots, \pm b_d\}$ where $b_1, b_2, \ldots, b_d$ is a basis for the unit lattice, but "generating set" allows many more possibilities.

For comparison, [39, Section 4.1] says, in different terminology, that simple reduction was "proposed and analyzed" in much more recent papers: [57], [36], etc. Formally, the algorithm statement in [57] is more general than in [28] in that it covers arbitrary lattices, but is less general than in [28] in that it is restricted to 2-norm while [28] said "measured in any reasonable way"; in any event, the idea is the same. The unit attacks proposed in [39] ("we propose to use an approximate Voronoi-cell-based algorithm ... adapted to our specific setting, where we wish to minimize some carefully determined meaningful quantities rather than the Euclidean distance") are, aside from terminology, examples of the reduction modulo units that had already appeared in [28].

D.4. Analysis of simple reduction. Sommer–Feder–Shalvi [87] showed in 2007 that "iterative slicing" using a list of "Voronoi relevant vectors", at most $2^{d+1} - 2$ vectors for a d-dimensional lattice, ensures perfect reduction in the 2-norm. Laarhoven [57] in 2016 gave an analysis concluding, heuristically, that taking roughly $2^{d/2}$ short vectors is necessary and sufficient for the same idea to find closest vectors in the 2-norm with high probability for a random lattice.

The analyses in [87] and [57] (and [36]) are obviously much more detailed than the brief "reduce even more" comment in [28]. On the other hand, these analyses were only for the 2-norm, rather than "measured in any reasonable way"; the analysis in [87] provides only an upper bound on the list size, not a lower bound; and the important counterexample $\mathbb{Z}^d$ shows that *some* lattices reduce much more effectively than the heuristic analysis in [57] indicates.

In 2019, Pellet-Mary–Hanrot–Stehlé [78] applied Laarhoven's analysis to S-unit lattices (although with non-traditional place weighting; see below), and in [78, Section 4.2] stated reasons to believe that shortness of the output in one norm predicts shortness in another norm. It was claimed in [78, Section 1, "Impact"] that the preprocessing time is "exponential" so the "concrete impact is limited". If this claim is understood as referring specifically to the choice in [78]

to enumerate an exponential number of short vectors, then the claim is justified, except for questions regarding $\exp(\Theta(n))$ vs. $\exp(n^{1+o(1)})$. However, if the claim is understood as referring to S-unit attacks more broadly, then the claim relies implicitly on the unjustified idea that the heuristic lower bounds in [57] for *most* lattices are applicable to S-unit lattices.

D.5. Using subfields and automorphisms inside unit attacks. There is a long history, not reviewed here, of number theorists exploiting subfields and automorphisms of number fields. For the problem of finding a short generator, Bernstein [10] in 2014 proposed a "subfield-logarithm attack" that, given I, first finds short generators of relative norms of I in "proper subfields of the original number field" and then uses these generators to eliminate some dimensions from the unit lattice. Subsequent developments of subfield-logarithm attacks include Bauch–Bernstein–de Valence–Lange–van Vredendaal [6], Lesavourey–Plantard–Susilo [63], and Biasse–Fieker–Hofmann–Page [16].

D.6. Using cyclotomic structure inside unit attacks. Campbell–Groves–Shepherd [22] in 2014 wrote the following:

> Firstly, we assume that the cyclotomic units have index 1 in the entire group of units $\mathcal{O}^\times$, which is almost certainly true for the specific instance of Soliloquy that had been proposed. A simple generating set for the cyclotomic units is of course known. The image of $\mathcal{O}^\times$ under the logarithm map forms a lattice. The determinant of this lattice turns out to be much bigger than the typical log-length of a private key α, so it is easy to recover the causally short private key given *any* generator of $\alpha\mathcal{O}$ *e.g.* via the LLL lattice reduction algorithm.

The *analysis* in [22], referring merely to the determinant of the lattice and the shortness of the target, is inadequate. What makes the algorithm work well is the fact that the simple generating set that one finds in textbooks (e.g., [89, Proposition 8.11]) consists of *short* units.

As mentioned in Section 1, there are, under minor assumptions, fast key-recovery attacks against the cyclotomic case of cryptosystems from Gentry [48], Smart–Vercauteren [86], Gentry–Halevi [49], and Garg–Gentry–Halevi [45]. The attacks combine the Campbell–Groves–Shepherd generator-reduction algorithm with the fast algorithms mentioned above to find a generator in the first place.

Campbell–Groves–Shepherd reportedly carried out experiments to confirm that their algorithm works well for a wide range of m. These experiments were double-checked by Schanck, who published his software [83]. Cramer–Ducas–Peikert–Regev [32] worked out asymptotics for prime-power m.

The Campbell–Groves–Shepherd algorithm is frequently miscredited to [32]. See, e.g., [55, Section 1] ("it is known that by using Cramer et al. result in cyclotomic fields [9], one can solve it efficiently").

D.7. Close principal multiples. Bernstein [11] in 2015 proposed handling a not-necessarily-principal ideal I by finding a principal multiple IJ, with J small, and then applying a unit attack to IJ:

> This *almost* demonstrates a perfect match [to [10] searching for "a secret short nonzero element g" given "the principal ideal gR"] ... We also know that the ideal I contains [the target short element]. ... This means that there can't be much gap between [the target short element] and I: the ratio is some small ideal J, and replacing I by IJ will give us exactly the desired ideal ... the condition "IJ is principal" linearly constrains the exponents in J's factorization.

In [12], Bernstein considered specifically $\mathbb{Q}(\zeta_{512})$, and evaluated the situation that "we know the classes of (say) ideals $I, P_1, P_2, \ldots, P_{10}$, and we want to search through a range of 2^{182} possibilities for a principal ideal of the form $IP_1^{e_1}P_2^{e_2}\cdots P_{10}^{e_{10}}$ with small $(e_1, e_2, \ldots, e_{10})$".

Subsequent developments of the close-principal-multiple idea for cyclotomic fields $\mathbb{Q}(\zeta_m)$, by Cramer–Ducas–Wesolowski [33] in 2017 and by Ducas–Plançon–Wesolowski [39] in 2019, focused on the case that $\{P_1, P_2, \ldots\}$ is the set of prime ideals having norm p, where $p \in 1 + m\mathbb{Z}$ is prime, and used multiplication by $P\overline{P}$ to clear denominators. Each vector $(e_1, e_2, \ldots)$ in a standard number-theoretic lattice (the Stickelberger ideal) then has the property that $P_1^{e_1}P_2^{e_2}\cdots$ is principal; this lattice is reused in [33] and [39].

The extension of unit attacks to close-principal-multiple attacks is described in [33, eprint version, page 7, Section 2.2] as a "folklore approach" without reference to [11] or [12]. The full version [34] of [33] no longer says "folklore"; on the contrary, it states "To reduce the problem from arbitrary ideals to principal ideals, we introduce the *close principal multiple problem* (or CPM): given an arbitrary ideal $\mathfrak{a}$, find an integral ideal $\mathfrak{b}$ such that $\mathfrak{ab}$ is principal, and $N(\mathfrak{b})$ is small". This problem is also described in [39, eprint version, page 9] as having been "introduced" in [33].

D.8. *S*-unit attacks. Bernstein [13] in 2016 proposed S-unit attacks as a generalization of unit attacks:

> The idea of
> - solving a close-vector problem in the unit lattice, to recover a short g from any unit multiple ug

> generalizes straightforwardly to
> - solving a close-vector problem in the S-unit lattice, to recover a short g from any S-unit multiple ug.

Pellet-Mary–Hanrot–Stehlé [78] in 2019 analyzed the performance of S-unit attacks, as noted above, and did not cite [13].

D.9. Logarithmic weights in *S*-unit attacks. Bernstein's proposal of S-unit attacks in [13] used the traditional number-theoretic weights for the entries of $\operatorname{Log}\alpha$, where finite place P has entry $-(\operatorname{ord}_P \alpha)\log \#(R/P)$:

> One simply extends the logarithm map in a standard way, taking not just the logarithms of absolute values at "infinite places" but also the logarithms of absolute values at all of the "finite places" in S. For example, when the field is $\mathbb{Q}$, the absolute value $|a/b|_\infty$ is the usual $|a/b|$;

the absolute value $|(a/b)2^e|_2$ is $1/2^e$ for any odd a, b; the absolute value $|(a/b)3^e|_3$ is $1/3^e$ for any a, b coprime to 3; etc. This generalization of the logarithm map is concisely reviewed in, e.g., ...

Pellet-Mary–Hanrot–Stehlé [78] instead took entry $-(\operatorname{ord}_P \alpha)c$ at place P. Here c is a constant chosen independently of P.

Bernard–Roux-Langlois [9] presented experiments and rationales indicating that S-unit attacks with the traditional number-theoretic $\log \#(R/P)$ weight are better than S-unit attacks with a constant weight as in [78]. Modifying [78] to include the traditional $\log \#(R/P)$ is the "main contribution" claimed in [9, abstract], but the traditional $\log \#(R/P)$ was already in the proposal of S-unit attacks in [13].

E Comparison to existing heuristics

This paper's definition of a spherical model is intended to formalize existing heuristics saying how long lattice vectors are and saying that the vectors are pointing in statistically independent directions. This appendix cites illustrative examples of these heuristics, compares the heuristics to a spherical model, and notes various examples of how the heuristics are used.

[53, page 273] says that, by the "gaussian heuristic", the "expected size of the smallest vector in a random lattice of dimension n and determinant D lies between $D^{1/n}\sqrt{n/2\pi e}$ and $D^{1/n}\sqrt{n/\pi e}$". The smaller value stated here matches the spherical-model length, except for a factor $1 + o(1)$; see Theorem 3.9. The statement in [53] is for a "random lattice" without specification of the lattice distribution; the formulas are then applied specifically to various lattices that appear in attacking NTRU.

As noted in Section 1, [70, Section 2.4] describes the "Gaussian Heuristic" as saying "that for a given set S and a lattice Λ, we have $|S \cap \Lambda| \approx \operatorname{vol}(S)/\det(\Lambda)$" assuming S is "nice"; [70, Heuristic 1] then more narrowly defines the "Gaussian Heuristic" as saying $\lambda_1(L) = ((d/2)!(\det L))^{1/d}/\sqrt{\pi}$. This matches Theorem 3.6, except for a factor $2^{1/d} \in 1 + o(1)$. This is then used in [70] to analyze the performance of BKZ.

Similarly, [38, Section 2.3] states "expected value $\sqrt{d/(2\pi e)} \cdot (\det L)^{1/d}$" for $\lambda_1(L)$, describing this as a consequence of "the Gaussian heuristic". Another "consequence of GH" in [38, Heuristic 2], as mentioned in Section 1, states that (1) there are $\alpha^{d+o(d)}$ lattice points in a ball of radius $\alpha\lambda_1(L)$ and (2) these lattice points are treated "as being uniformly distributed over the ball". Note that almost all points in the unit ball have length $1-o(1)$ as $d \to \infty$, so selecting a uniform random point in the unit ball is equivalent to selecting a uniform random point on the sphere for any analysis that sees lengths only up to $1+o(1)$ factors, although taking minima across many points requires care; see Section 3.19.

Asymptotics for the probability that u reduces v, in terms of the lengths of u and v, are stated in [57, Lemma 2] for uniform random points on a sphere. The conclusions that [57] draws from this regarding the effectiveness of reduction for

```
proof.all(False)
m = 128
n = m/2
d = n/2-1
K.<zeta> = CyclotomicField(m)
R = K.regulator()
Rplus = K.subfield(zeta+1/zeta)[0].regulator()
print(Rplus/(n/4)^(n/4))
detL = R*sqrt(n/2)
print((2*pi^(-d/2)*gamma(d/2+1)*detL).n()^(1/d))
```

Fig. F.1. Sage script to double-check the regulator and spherical-model columns in the $m = 128$ row in Table 8.3.

a lattice L (see also [37]) rely on the heuristic stated in [57, Section 2.1] that, "when normalized, vectors in L follow the same distribution as vectors sampled uniformly at random from the unit sphere", along with a "Gaussian heuristic" for the vector lengths. Regarding terminology, the uniform-random-directions heuristic in [57] is not described as "Gaussian", whereas "uniformly distributed over the ball" in the subsequent paper [38] is presented as part of a "consequence of GH".

F Spot-checks of Table 8.3

This appendix presents two spot-checks of Table 8.3. Real numbers displayed here are rounded to limited precision without further comment.

First, for $m = 16$ and $n = 8$, the determinant of the matrix

$$\begin{pmatrix} 2.093065 & 1.136717 & -2.899464 \\ 1.136717 & -0.330318 & 2.093065 \\ -2.899464 & 2.093065 & -0.330318 \end{pmatrix}$$

is -19.534359, so the regulator Reg_K of $\mathbb{Q}(\zeta_m)$ is 19.534359. The determinant of the unit lattice is $(n/2)^{1/2} \mathrm{Reg}_K \approx 39.068729$. A 3-dimensional ball of radius $r = 2.652102$, the length shown in Table 8.3 for $m = 16$, has volume $(4/3)\pi r^3 \approx 78.137458$, which is twice the lattice determinant, exactly where a spherical model expects to see the shortest nonzero vectors.

Second, as a larger spot-check—assuming the accuracy of Sage's number-field computations, notably its `regulator` function—the script in Figure F.1 takes a minute on one laptop core specifically for $m = 128$ and prints results matching the $m = 128$ row in Table 8.3. If m is changed to 16 then the same script outputs 2.652102. Beware that the script becomes much slower if m is increased beyond 128 or if `proof.all(False)` is removed; Sage performs many computations to check $h_m^+ = 1$.

```
def rho(d,alpha):
  T = RealDistribution('beta',((d-1)/2,1/2))
  return T.cum_distribution_function(1-alpha^2)/2

d = 127
r = 45.953088

alpha = 0.28
while alpha < 0.38:
  Pr = 1-prod(1-2*rho(d,j^(1.0/d)*alpha) for j in range(1,8129))
  print('%.6f %.6f' % (r/(2*alpha),Pr))
  alpha += 0.002
```

Fig. G.3. Sage script to check the red curve for $m = 512$ in Figure 8.13. See text for description.

G Spot-checks of Figure 8.13

This appendix presents spot-checks of Figure 8.13. Specifically, this appendix checks the red and blue curves for $m = 512$. The further advantage of the green curve over the blue curve is not necessary for the main point of this paper.

G.1. Cap volumes. The checks below rely on calculating the probability $(\operatorname{Vol}_{d-1} \operatorname{Cap}_\alpha^{d-1})/(2 \operatorname{Vol}_{d-1} \operatorname{Cap}_0^{d-1})$ in Theorem 3.13: i.e., the percentage of the unit $(d-1)$-sphere contained in the α-cap of the sphere. One can statistically estimate this probability by choosing random sphere points as in Section 8.12; but computing the probability in a different way, specifically via Theorem 3.17, helps serve as a double-check in case of, e.g., failures in how Section 8.12 was generating sphere points.

Abbreviate $(\operatorname{Vol}_{d-1} \operatorname{Cap}_\alpha^{d-1})/(2 \operatorname{Vol}_{d-1} \operatorname{Cap}_0^{d-1})$ as $\rho_d(\alpha)$ for $0 \le \alpha \le 1$. For example, $\rho_d(1) = 0$; $\rho_d(0) = 1/2$; and $\rho_2(1/2) = 1/3$, since $1/3$ of the points (x_1, x_2) on the unit 1-sphere (i.e., the unit circle) have $x_1 > 1/2$.

The function `rho` inside the Sage script shown in Figure G.3 computes $\rho_d(\alpha)$, given (d, α). This function, based on [15, page 50], uses Sage's support for the beta distribution. By definition the cumulative beta distribution function, with parameters (a, b), maps x to $\mathcal{B}(x; a, b)/\mathcal{B}(1; a, b)$; dividing by 2 and substituting $(x, a, b) = (1 - \alpha^2, (d-1)/2, 1/2)$ gives $\rho_d(\alpha)$ by Theorem 3.17.

G.2. Red curve: spherical model. Let L be the unit lattice for $m = 512$, i.e., $n = 256$. This lattice has dimension $d = n/2 - 1 = 127$.

By definition a spherical model M of L has the form $\{0, \pm\mu_1, \pm\mu_2, \ldots\}$ where $\mu_1, \mu_2, \ldots$ are statistically independent uniform random variables in $L_{\mathbb{R}}$ subject to $||\mu_j||_2^d = 2j\pi^{-d/2}(d/2)! \det L$. One then has $||\mu_j||_2 = j^{1/d}||\mu_1||_2$. The length $||\mu_1||_2$ is approximately 45.953088; see Table 8.3.

Now consider a vector ν of length $||\mu_1||_2/2\alpha$ with $0 < \alpha < 1$. Theorem 3.13 says that μ_1 reduces ν with probability $\rho_d(\alpha)$, and that $-\mu_1$ reduces ν with the same probability. These events are disjoint, for total probability $2\rho_d(\alpha)$.

For example, if $\alpha = 0.319118667$, then ν has length $||\mu_1||_2/2\alpha \approx 72$, and $\rho_d(\alpha) \approx 0.000120578223$. For comparison, in Figure 8.13, the 71.935 horizontal line crosses the red curve for $m = 512$ around 50%, reporting that around half of the input vectors in a pool of 1000 vectors of length 71.935 are reduced. This is a much higher probability than 0.000120578223; but this is also reduction using many thousands of vectors in M, not just $\pm\mu_1$.

More generally, by the same theorem, μ_j reduces ν with probability $\rho_d(j^{1/d}\alpha)$. For example, if again $\alpha = 0.319118667$, then μ_2 reduces ν with probability $\rho_d(2^{1/d}\alpha) \approx \rho_d(0.320865131) \approx 0.000110944734$; and the 7% longer vector μ_{8128} reduces ν with probability $\rho_d(8128^{1/d}\alpha) \approx \rho_d(0.342562234) \approx 0.0000377783784$.

The number 8128 in the previous paragraph is $(m/8)(m/4-1) = 64 \cdot 127$. Each experiment in Figure 8.13 uses short vectors $\pm\mu_1, \ldots, \pm\mu_{8128}$.

The probability that all of $\pm\mu_1, \ldots, \pm\mu_{8128}$ fail to reduce ν is, by statistical independence, $\prod_j(1-2\rho_d(j^{1/d}\alpha))$. The full Sage script in Figure G.3 computes pairs $(||\mu_1||_2/2\alpha, 1-\prod_j(1-2\rho_d(j^{1/d}\alpha)))$ for various α, showing the probabilities of successful reduction for various lengths of ν. The range of α was chosen so that the probability drops from about 100% to about 0%. The pairs include, e.g.,

$$(68.382571, 0.241688), (71.801700, 0.492219), (75.086745, 0.759362);$$

comparing these to the quartiles in the red curve in Figure 8.13 shows no obvious discrepancies. Note that, since Figure 8.13 is reporting 1000 experiments, one expects horizontal deviations above 1%, and comparing probabilities very close to 0% or 100% is not meaningful.

G.4. Blue curve: $\mathbf{Log((\zeta_m^j - 1)/(\zeta_m^k - 1))}$. The blue curve uses vectors in L. Note that L has much shorter vectors than a spherical model M of L predicts. For example, $\mathrm{Log}((\zeta_m^3-1)/(\zeta_m-1)) = \mathrm{Log}(1+\zeta_m+\zeta_m^{-1})$ since $\mathrm{Log}\,\zeta_m = 0$; and $\mathrm{Log}(1+\zeta_m+\zeta_m^{-1})$ has length approximately 23.631207, as noted in Table 8.3, much shorter than the minimum length 45.953088 of nonzero vectors in M. On the other hand, this is not enough information to conclude that the 8128 vectors (modulo negation) used in the blue curve are shorter than those in the model M, or, more to the point, that reduction is as effective as reported in Figure 8.13.

The following spot-check recalculates each $u_{j,k} = \mathrm{Log}((\zeta_m^j-1)/(\zeta_m^k-1))$, and then takes two approaches to checking the blue curve:

- The first approach reuses the ρ_d function defined above to compute the probability that $\pm u_{j,k}$ reduces ν, assuming ν is uniformly distributed on a sphere, and then models these probabilities as independent. This approach relies only on $||u_{j,k}||_2$.
- The second approach samples vectors ν to see whether they are successfully reduced. This approach uses the $u_{j,k}$ vector, not just its length.

As noted in Section 8.12, one would not expect statistical independence of the reduction probabilities, so the first approach is skating on thin ice. The second approach has no such issues.

```
def rho(d,alpha):
  T = RealDistribution('beta',((d-1)/2,1/2))
  return T.cum_distribution_function(1-alpha^2)/2

m = 512
n = 256
zetam = CC(-1)^(1/256)
d = 127

def isometry(u):
  result = []
  partialsum = 0
  for j in range(d):
    result += [sqrt(RR(1+1/(d-j)))*(partialsum/(d+1-j)+u[j])]
    partialsum += u[j]
  return vector(result)

place = {}
norm = {}
U = []

for j in range(3,n,2):
  for h in range(1,n,2):
    place[h,j] = 2*log(abs(sum(zetam^(g*h) for g in range(j))))
    place[m-h,j] = place[h,j]
  norm[j] = sqrt(sum(place[h,j]^2 for h in range(1,n,2)))
  for k in range(1,n,2):
    u = [place[(h*k)%m,j] for h in range(1,n,2)]
    U += [isometry(u)]

print(norm[3])

def reduce(v):
  return any((v-u)*(v-u) < v*v for u in U)

vlen = 55
while vlen >= 30:
  Pr = 1-prod(1-2*rho(d,norm[j]/(2*vlen)) for j in range(3,n,2))^(m/8)
  with seed(31415):
    D = SphericalDistribution(dimension=d)
    Pr2 = sum(reduce(vlen*D.get_random_element()) for r in range(100))
    Pr2 /= 100
  print('%.6f %.6f %.6f' % (vlen,Pr,Pr2))
  vlen -= 1
```

Fig. G.5. Sage script to check the blue curve for $m = 512$ in Figure 8.13. See text for description.

For each $j \in \{3,5,\ldots,m-1\}$, one has $u_{j,1} = \operatorname{Log}((\zeta_m^j - 1)/(\zeta_m - 1)) = \operatorname{Log}(1 + \zeta_m + \zeta_m^2 + \cdots + \zeta_m^{j-1})$. Place h in $u_{j,1}$, for any $h \in \{1,3,5,\ldots,n-1\}$, is $2\log|1 + \zeta_m^h + \zeta_m^{2h} + \cdots + \zeta_m^{(j-1)h}|$, so

$$||u_{j,1}||_2^2 = \sum_h (2\log|1 + \zeta_m^h + \zeta_m^{2h} + \cdots + \zeta_m^{(j-1)h}|)^2.$$

The case $j = 3$ is equivalent to Theorem 8.2.

For each $k \in \{1,3,5,\ldots,m-1\}$, there is a unique ring morphism $K \mapsto K$ mapping ζ_m to ζ_m^k. This automorphism acts on $\operatorname{Log} K^*$ as a permutation of the entries of log vectors, so it preserves the length of log vectors. Applying this automorphism to $u_{j,1}$ gives $u_{jk,k}$, i.e., $u_{jk \bmod m,k}$.

Note that $\zeta_m^{m-j} - 1 = -\zeta_m^{-j}(\zeta_m^j - 1)$ so $\operatorname{Log}(\zeta_m^{m-j} - 1) = \operatorname{Log}(\zeta_m^j - 1)$ so $u_{m-j,1} = u_{j,1}$. Similarly $u_{jk \bmod m,k} = u_{jk \bmod m,m-k}$. Consequently one can restrict attention to indices $jk \bmod m$ and k in $\{1,3,5,\ldots,n-1\}$. This leaves $(m/4)(m/4-1)$ distinct lattice vectors, which up to negation are exactly the $(m/8)(m/4-1)$ lattice vectors used to define the blue curve. The lengths of these vectors are $m/4$ copies of the lengths $||u_{j,1}||_2$ described above for $j \in \{3,5,\ldots,n-1\}$. These lengths are enough information to compute α and thus $\rho_d(\alpha)$ for the first approach.

For the second approach, this appendix applies an isometry to embed each $u_{j,k}$ into $\mathbb{R}^d$, and then checks reducibility of 100 sample vectors separately for each length, without the merging across lengths described in Section 8.12.

The Sage script in Figure G.5 takes both approaches, and prints out lines such as `50.000000 0.999948 0.950000` saying that for input of length 50 the second approach was observed reducing 95 out of 100 vectors while the first approach predicted 99.9948%. Another output line `43.000000 0.731501 0.480000` is, for the second approach, consistent with the median of the $m = 512$ blue curve in Figure 8.13. Further output lines `46.000000 0.968990 0.780000` and `41.000000 0.458598 0.270000` are, for the second approach, consistent with the quartiles in the figure. The script also recalculates and prints out the length 23.631207 of $\operatorname{Log}(1 + \zeta_m + \zeta_m^{-1})$.

The match between the second approach and the blue curve means that the spot-check was successful. The mismatch between the first approach and the blue curve does not indicate a failure of the spot-check, since the first approach started with a questionable independence model. The mismatch also means that this spot-check does *not* serve as a double-check on the calculations in the first approach; the mismatch could be explained by inaccuracy in the independence model, but could also be explained by calculation errors. The details of the first approach are reported here in case this is useful for subsequent investigations of the accuracy of the independence model.